

В.Эбелинг • А.Энгель • Р. Файстель

Ф И З И К А ПРОЦЕССОВ ЭВОЛЮЦИИ

Перевод с немецкого
Ю.А.Данилова



УРСС
Москва • 2001

Настоящее издание осуществлено при поддержке "Inter Nationes" (Бонн)

Эбелинг Вернер, Энгель Андреас, Файстель Райнер

Физика процессов эволюции. Пер с нем Ю А Данилова — М Эдиториал УРСС, 2001
— 328 с

ISBN 5–8360–0233–9

Книга известного немецкого физика В Эбелинга и его учеников и единомышленников Р Файстеля и А Энгеля посвящена анализу процессов эволюции с позиций синергетики и динамики нелинейных систем. Значительное внимание авторы уделяют проблеме переработки информации нелинейными динамическими системами. Изложение опирается на широкий круг результатов, полученных исследователями в различных странах мира.

Группа подготовки издания

Директор — Доминго Марин Рикой

Заместители директора — Наталья Финогенова, Ирина Макеева

Администратор — Леонид Иосилевич

Компьютерный дизайн — Виктор Романов

Верстка — Наталия Бекетова, Михаил Кириллов, Ксения Пулькина

Редакция — Елена Кудряшова, Виктория Малышенко

Обработка графики — Наталья Аринчева

Техническая поддержка — Анна Чикунова

Менеджер по продажам — Алексей Петяев

Издательство «Эдиториал УРСС» 113208 г Москва, ул Чертановская, д 2/11, к п
Лицензия ИД № 03216 от 10.11.2000 г Гигиенический сертификат на выпуск книжной
продукции № 77-ФЗ-8-953 П-270-3-99 от 30.03.99 г Подписано к печати 14.02.2001 г
Формат 70×100/16 Тираж 1500 экз Печ л 20,5 Зак № 188

Отпечатано в АООТ «Политех-4» 129110, г Москва, ул Б Переяславская, 46

ISBN 5–8360–0233–9



© Оригинал В Эбелинг, А Энгель, Р Файстель, 1990
© Перевод с немецкого Ю А Данилов, 2001
© Произведение на русском языке Эдиториал УРСС, 2001
© Оформление, оригинал-макет Эдиториал УРСС, 2001

Никакая часть настоящей книги не может быть перепечатана, воспроизведена или использована в какой бы то ни было форме, как полностью, так и частично, какими бы то ни было средствами, как известными сейчас, так и изобретенными впоследствии, включая фотокопирование и запись на магнитный носитель, или в любой системе хранения и обработки данных, если на то нет письменного разрешения владельца авторских прав. Нарушение данного требования влечет за собой применение к виновной стороне ответственности, предусмотренной административным и уголовным законодательством РФ.



Эдиториал УРСС

научная и учебная литература

Тел./факс 7(095)135–44–23

Тел./факс 7(095)135–42–46

E-mail urss@urss.ru

Каталог изданий в Internet <http://urss.ru>

Предисловие

Вопрос о разработке физикой процессов эволюции возник с тех пор, как появились термодинамика и статистическая физика. Основополагающие идеи в этом направлении были высказаны еще Майером, Гельмгольцем, Клаузиусом, Больцманом, Пуанкаре и другими выдающимися физиками прошлого века. В первые три четверти XIX столетия идеи развития играли в физике лишь относительно узкую роль, но в последней четверти они оказались в самом центре всеобщего интереса. Причины наблюдаемой в наше время интенсификации работ по физике и химии процессов эволюции будут подробно рассмотрены в этой книге. Назовем самые главные предпосылки этой интенсификации:

- 1) разработка Гамовым и другими авторами гипотезы Большого Взрыва и постепенное экспериментальное обоснование вытекающих из этой гипотезы следствий;
- 2) разработка Пригожиным, Хакеном и др. теории самоорганизации (синергетики);
- 3) разработка Эйгеном, Шустер и др. физико-химической теории биогенеза.

Двое из авторов этой книги в своей более ранней работе выдвинули программу «наведения мостов» между уже хорошо разработанной физикой самоорганизации и физикой процессов эволюции (*Ebeling, Feistel, 1982*). В этой книге мы продолжаем намеченную линию, для чего нам в известной мере придется напомнить уже достигнутое состояние интересующей нас области знания. Сложность проблематики не позволяет нам изложить все относящиеся к затронутому кругу проблем основы и все наиболее важные приложения. Особое значение мы придавали изложению стохастической теории, так как единство необходимого и случайного составляет самую сущность эволюции и тем самым составляет центральный пункт теории. Затем мы обсуждаем вопросы стратегий эволюции и уже наметившиеся технические приложения. Ядро нашего изложения составляют вопросы эволюции информации. Относительно ранних фаз космической эволюции и их теоретико-полевого моделирования наша книга содержит лишь элементарные утверждения и ссылки на литературу для дальнейшего чтения.

Особое значение мы придавали подробному указателю литературы с полными названиями работ, в том числе и журнальных публикаций.

Для замысла этой книги решающее значение имели обсуждения на регулярно работающих семинарах «Необратимые процессы и самоорганизация» и «Система Лотки—Вольтерры» в Берлине, а также на московском семинаре «Синергетика». Особую благодарность мы хотели бы выразить своим коллегам М. Конраду, Г. Хакену, Б. Хессу, Ю. Л. Климонтовичу, Л. Кунерту, Г. Линде, А. С. Михайлову, М. Пешелю, И. Пригожину, Ю. М. Романовскому, Г.-П. Швэфелю, Е. Е. Селькову, П. Шустер, М. В. Волькенштейну, Г. Уинстону и Я. Б. Зельдовичу. Мы выражаем свою искреннюю признательность нашим друзьям, сотрудникам и студентам сектора 04 физического факультета университета имени Гумбольдта в Берлине, а также ответственным сотрудникам издательства *Akademie-Verlag* в Берлине за эффективное сотрудничество. В заключение мы хотели бы выразить особую благодарность Ю. А. Данилову за перевод нашей книги на русский язык и тщательную работу над текстом.

Вернер Эбелинг, Андреас Энгель, Райнер Файстель

Глава 1

Введение

Скрытая и замкнутая сущность Вселенной не обладает силой, которая могла бы позволить ей оказывать сопротивление мужеству познания.

Георг Фридрих Вильгельм Гегель

1.1. Возникновение сложного

Взглянув на окружающий нас мир, мы обнаружим неисчерпаемое многообразие вещей и явлений: солнце, ветер, дождь и многочисленные сложные погодные процессы, изобилующий формами и красками мир растений, царство животных и окружающий нас рукотворный мир человека. Ночью мы наблюдаем звездное небо и получаем возможность заглянуть в процессы, разыгрывающиеся на больших пространственных и временных расстояниях от нас. Все это — в движении, все развивается.

С самого начала существования человечества люди созерцали парциальные процессы эволюции нашего мира, и всегда находились такие личности, которые поднимались над повседневными заботами и ставили далеко идущие вопросы. Что кроется за непрестанной сменой и развитием, рождением и гибелью? Почему так много одних явлений происходит без видимых причин, тогда как другие явления столь регулярны и взаимосвязаны? От накопленной жизненной мудрости стариков, передающих свой опыт молодым, через первые «научные» записки к основам научных знаний наших дней путь был длинным и трудным. Из многообразия пережитого необходимо было извлекать существенное и абстрагироваться от несущественного, намечать упрощенную модель. Когда за внешне совершенно различными явлениями удавалось обнаружить одну и ту же сущность, например, единое физическое строение атомов всех химически столь различных элементов, это всегда становилось значительным шагом в познании мира. Наиболее радикально такая редукция к существенному происходила в физике. Ныне наиболее существенное содержание основных законов нашего мира спрессовано в немногочисленных теориях и уравнениях, и ведется напряженная работа по достижению объединения и этих теорий. Ведется поиск элементарных «кирпичиков» мироздания, и, судя по всему, построение единой теории станет возможным в обозримом будущем. Нет никаких сомнений в том, что такая единая теория мельчайших «кирпичиков» мироздания будет иметь огромное научное, теоретико-познавательное и практическое значение.

Но достаточно ли мы знаем для того, чтобы построить такую теорию? Все ли известно нам о веществе, если мы понимаем строение атомов? Достаточно ли мы сведущи в архитектуре, если основательно разбираемся в черепице? Целое часто представляет нечто большее, чем сумму отдельных частей. Именно это «большее» и представляет собой тот стержень, вокруг которого разворачивается в основном действие в нашей книге. Как получается, что при наложении физических процессов совершенно неожиданно возникают новые свойства? Как может быть, что из многих «мертвых» молекул возникает нечто живое? Существуют ли какие-нибудь законы

помимо известных элементарных законов физики? Как может возникнуть сам собой сложный современный мир? Таковы вопросы, изучением которых занимается молодой раздел физики — *физика процессов эволюции*.

С давних пор физика представляет собой научную дисциплину, которая изучает самые общие, самые универсальные и самые фундаментальные законы природы и облачает их в как можно более строгую математическую форму. То же можно сказать и о физике процессов эволюции: занимаясь ею, мы хотим установить, какие правила и законы управляют развитием сложного, и тем самым определить структуру и функцию сложных процессов и образований. Мы ищем такие законы, которые, как и «обычные» физические законы, не зависят от того, являются ли исследуемые объекты физическими, химическими, биологическими или даже социальными структурами. Иначе говоря, мы пытаемся навести мосты между физикой неживого и физикой живого. Правда, и поныне имеется немало открытых вопросов, а иногда нам неизвестна даже правильная постановка вопросов. Однако имеются и установленные утверждения, концепции и идеи, и о них-то и пойдет речь в нашей книге.

1.2. Какую цель ставит перед собой физика процессов эволюции?

Опыт повседневной жизни и целенаправленное наблюдение процессов в природе и общественных явлений учат нас, что многие сложные системы, состоящие из большого числа взаимодействующих подсистем, при определенных условиях обладают способностью к самоорганизации и эволюции. Самым замечательным тому примером могут служить живые существа и их развитие на нашей планете. Сколько очевиден этот факт, научная интерпретация его весьма сложна и требует слияния результатов исследований многих областей науки. Основная задача настоящей книги состоит в систематическом изложении возможного вклада физики в разработку этой чрезвычайно сложной проблематики.

Под самоорганизацией мы понимаем в дальнейшем необратимый процесс, приводящий в результате кооперативного действия подсистем к образованию более сложных структур всей системы. Самоорганизация — элементарный процесс эволюции, состоящей из неограниченной последовательности процессов самоорганизации. Термин «самоорганизация» используется в этой книге в смысле Пригожина для обозначения диссипативной самоорганизации, т. е. образования диссипативных структур. Наряду с диссипативной самоорганизацией существуют и другие формы самоорганизации, такие, как консервативная самоорганизация (образование структур кристаллов, биополимеров и т. д.) и дисперсионная самоорганизация (образование солитонных структур). Такие формы самоорганизации упоминаются в нашей книге лишь бегло.

Значение самоорганизации и эволюции для естественных и общественных процессов очевидно, и изучение ее закономерностей стало предметом бесчисленных исследований со времен Канта, Гегеля, Маркса и Дарвина. С другой стороны, в последнее время физика самоорганизации в открытых системах сформировалась в качестве новой области физики. Эта наука занимается изучением закономерностей нелинейных кооперативных процессов вдали от термодинамического равновесия; ее основные идеи восходят к Шрёдингеру, Тьюрингу, Бергаланффи, Пригожину, Эйгену и Хакену. В настоящее время физика самоорганизации переживает период определенного завершения, появились первые обзоры и монографии, последовательно излагающие весь круг затрагиваемых проблем (Гленсдорф и Пригожин (*Glansdorff, Prigogine*, 1971); Жаботинский, (1974); Романовский, Степанова и Чернавский, (1984); Эбелинг, 1979); Хакен, 1980; Haken, 1983; Николис и Пригожин, 1979, 1990); Пригожин, 1979); Эбелинг и Климонтович (*Ebeling, Klimontovich*, 1984); Хофбауэр

и Зигмунд (*Hofbauer, Sigmund*, 1984); *Васильев и др.*, 1987); Файстель и Эбелинг (*Feistel, Ebeling*, 1989)).

Решающее значение для создания физики самоорганизации имели развитие и разработка методологии следующих дисциплин:

- а) термодинамики необратимых процессов в открытых системах;
- б) нелинейной механики, электрофизики и физики лазеров;
- в) химической кинетики сильно неравновесных процессов;
- г) нелинейной динамики популяций и экологии;
- д) нелинейной теории регулирования, кибернетики и системного анализа.

Из приведенного выше перечня отчетливо виден междисциплинарный характер физики самоорганизации. Исходя из этого, физик Хакен предложил для новой области науки название синергетика (*Хакен*, 1980; *Haken*, 1973, 1975, 1981). Синергетика, по Хакену, должна охватывать все проблемы, связанные с образованием упорядоченных структур в сложных системах в результате кооперативного поведения подсистем. Если относительно названия новой области не удалось придти к единому мнению, то интерес ученых к постановке проблем самоорганизации ныне, как показывает резкое увеличение оригинальных работ и монографий, не вызывает никаких сомнений. Таким образом, контуры физики самоорганизации очерчены достаточно четко, чего нельзя сказать о физике процессов эволюции. Цель настоящей работы состоит в том, чтобы навести мост между физикой самоорганизации и физикой процессов эволюции.

Разумеется, мы должны прежде всего задать себе вопрос, разумно ли вообще наше намерение. Не суждено ли ему разбиться вдребезги о такой общеизвестный факт, как то, что даже отдельная клетка неизмеримо сложнее любого традиционного объекта физического исследования? Мы хотим поделиться с читателем своими взглядами по этому вопросу, сформулировав их в нескольких фразах. Законы физики остаются в силе и действуют беспрепятственно и в области процессов эволюции. Никаких исключений никто и никогда не наблюдал, и их никто не ожидает. Физика входит в самые общие основы процессов эволюции, поскольку она указывает возможности и запреты. С другой стороны, мы считаем, что физика сама по себе не в состоянии установить основные законы и сущность эволюции, поскольку в этом круге вопросов необходимо учитывать весьма сложные аспекты. Поэтому стремиться надлежит не к тому, чтобы, например, «свести Дарвина к Пригожину», а к тому, чтобы вооружить тех, кто изучает эволюцию, добытыми физиками достоверными знаниями и фактами, которые могут оказаться полезными при решении чрезвычайно сложных проблем эволюции. Позицию физиков относительно решения этих проблем мы продемонстрируем на двух подходящих к случаю примерах.

Николис и Пригожин пишут (*Николис, Пригожин*, 1979, с. 20): «В каком-то смысле мы находимся в положении пришельца с другой планеты, который, обнаружив пригородный дом, хочет разобраться в сути этого явления. Конечно, дом сам по себе не противоречит законам механики, иначе он развалился бы. Однако это не относится к существу дела, поскольку интерес представляют технология, которую использовали строители дома, потребности его обитателей и т. д. Понять все это невозможно, не имея представления о соответствующей цивилизации». Здесь, как мы отчетливо видим, позиция физика состоит в сотрудничестве с биологами и представителями других наук, занимающихся исследованием эволюции. При возведении здания эволюции физик выполняет функции специалиста по статике сооружений: он может выполнять важные совместные работы, но в решающих вопросах относительно смысла, цели и назначения здания должен присоединяться к мнению архитекторов.

Непревзойденное художественное описание того положения, в котором оказывается пришелец на чужой планете, принадлежит Станиславу Лему. Фантазия Лема породила планету Солярис, поверхность которой покрыта «живым Океаном», своеобразные формообразования которого остаются загадочными и непонятными для земных наблюдателей, придумывающих для них самые необычные названия. Особенно отличается в измышлении названий для причудливых формообразований приверженец систематики Гизе. «Столько лет тянулись непримиримые споры по поводу всего происходящего в “долгунах”, миллионы которых бороздят необъятные просторы живого Океана. Их считали какими-то органами, полагая, что в них происходит обмен веществ, процессы дыхания и пищеварения и что-то еще, о чем помнят теперь лишь пыльные библиотечные полки. Каждая их этих гипотез была опровергнута тысячами труднейших, а подчас и опасных опытов...

Антропоморфизм (или зооморфизм) вновь и вновь проглядывал в отчаянных поисках исследователей, они усматривали в различных видоизменениях живого Океана то “органы чувств”, то “конечности”; какое-то время ученые принимали за “конечности” “хребетники” и “мелькальцы”. Но эти протуберанцы живого Океана, вздымающиеся иногда на две мили в атмосферу, так же можно назвать “конечностями”, как землетрясения — “гимнастикой” земной коры.

Насчитывается около трехсот форм, повторяющихся с относительным постоянством и порождаемых живым Океаном сравнительно часто. За сутки можно обнаружить несколько десятков или сотен их на поверхности. Самые “нечеловеческие”, то есть абсолютно не похожие ни на что земное, формы, по утверждению школы Гизе, — это симметриады...

Хотя от этих исполинов могут сниться кошмары, самое страшное в симметриадах вовсе не их вид. Ужас наводит скорее то, что в границах симметриад нет ничего постоянного и определенного, там не действуют даже физические законы. Именно исследователи симметриад настойчивее всех утверждают, что живой Океан разумен¹⁾.

Но довольно о Станиславе Леме с его неповторимыми видениями чуждого людям мира, ускользающего от человеческого понимания. Земной физик, занимающийся изучением эволюционных процессов на нашей планете, находится в более благоприятной ситуации, чем исследователь Соляриса у Станислава Лема. Он может, во-первых, исходить из того, что нарушение физических законов никогда не наблюдалось и потому может быть исключено из рассмотрения, а во-вторых — рассчитывать на сотрудничество соответствующих специалистов. Обе основные позиции должны стать исходным пунктом любых последующих рассмотрений.

Соответствующие специалисты заведомо принесут с собой понимание того, что в дальнейшем язык физиков станет пригодным для описания областей знания, не относящихся к физике. К физикам в полной мере относится то, что Гёте некогда сказал о математиках: «Физики (в оригинале у Гёте — математики) — своего рода французы: стоит сказать им что-нибудь, как они переведут это на свой язык, и тогда получается нечто совсем другое».

Язык выступает здесь лишь как внешняя форма определенного метода подхода к исследованию вещей, о котором Ромпе и Тредер (*Rompe, Treder, 1979*) говорят следующее: «Культивируемая физикой методика исследования есть не что иное, как единственно возможный путь точного познания природы, ставящего своей целью овладение природой. Оно не ограничивается рамками физики и отнюдь не является монополией физики».

¹⁾ Станислав Лем. Солярис / Пер. Г. Гудимовой и В. Перельман // Станислав Лем. Избранное. М.: Прогресс, 1976, с. 125-130.

С другой стороны, исследования процессов эволюции привнесли в физику новое мышление и новый метод работы, которые до того были достоянием только биологических и общественных наук: историческое мышление и исторический метод. Эволюция по самой своей сути является историческим процессом. Результаты эволюционных процессов, как правило, могут быть поняты не из их мгновенного состояния, а лишь на основе исследования истории их возникновения.

Именно в этом смысле мы ставим также старый вопрос: что такое жизнь? Какие свойства живой системы не могут быть описаны или объяснены в рамках традиционной физики и химии? Мы убеждены в том, что исследование эволюции требует междисциплинарного и исторического подхода. Именно поэтому физика процессов эволюции не может сводиться к совокупности законов и методов, а должна включать в себя также историю физических аспектов эволюции.

История эволюции есть история возникновения все более сложных структур из более простых; ведь суть эволюции состоит как раз в интеграции более простых элементов в целостные образования более высокого уровня, т. е. в более сложные системы, характеризующиеся новыми качествами.

Наиболее важными фазами эволюции окружающего нас мира были следующие:

- космическая эволюция (Большой Взрыв, образование элементарных частиц, формирование атомов и молекул, возникновение галактик, звезд и планет, образование «фотонной мельницы»);

- химическая эволюция (образование системы химических элементов и соединений, возникновение органических соединений, полимеризация в цепи органических молекул);

- геологическая эволюция (образование структур земной коры, гор, вод и т. д.);

- эволюция протоклетки (самоорганизация биополимеров и хранение информации на молекулярном уровне, пространственная индивидуализация, возникновение молекулярного языка);

- дарвиновская эволюция (развитие видов животных и растений и их взаимодействие, возникновение экосистемы на Земле);

- эволюция человека (развитие труда, языка и мышления);

- эволюция общества (развитие распределения труда, общественная организация, техника, общественные формации и т. д.);

- эволюция информации и обмена информацией (обогащение и хранение знания, развитие связи, науки и т. д.).

Существенная задача «физики эволюции» состоит в том, чтобы разработать физические основы, необходимые для понимания этих отчасти весьма сложных процессов. Во избежание недоразумений сформулируем нашу точку зрения еще раз: для понимания сложных процессов одной лишь физики недостаточно, но весьма существенные аспекты эволюционных процессов не могут быть поняты без физики. Наконец, физик, занимающийся проблемами эволюции, должен способствовать не только углублению знаний, но и (в соответствии с традициями своей науки) их техническому использованию. Эволюционные стратегии берут начало в работах Бремермана (*Bremermann*, 1970), Рехенберга (*Rechenberg*, 1973), Холланда (*Holland*, 1975) и Швевеля (*Schwefel*, 1977). Авторы настоящей книги убеждены, что перед эволюционными методами открывается прекрасное будущее, и усматривают в их разработке основное назначение физики процессов эволюции.

Эволюционистическое мышление, т. е. понимание мира как процесса эволюции, существовало еще у древних греков. Именно оно неизменно находилось и находится в фокусе мировоззренческих споров (см., например, *Fuchs-Kittowski*, 1976; *Löther*, 1972; *Ley*, 1975; *Hörz*, 1974; *Hörz, Wessel*, 1983), поскольку, как заметил

Энгельс, всегда было «источником дальнейшего прогресса». Современное эволюционистическое мышление сложилось в XVIII и XIX веках и неразрывно связано с великими именами Канта, Гегеля, Маркса, Дарвина и Клаузиуса. Своего рода предтечей стало сочинение Канта «Всеобщая естественная история и теория неба». В нем Кант предпринял попытку объяснить происхождение мира, исходя из физических законов, и в необычайно остром полемическом стиле утверждал: «Если отрешиться от старых лишенных каких бы то ни было оснований предрассудков и дурацкой мировой мудрости, которая пытается под многозначительной миной скрыть глубокое невежество, то я надеюсь, опираясь на противоречивые основы, обосновать свое незыблемое убеждение в том, что мир от самого сотворения может быть познан, исходя из общих законов природы, как механическое развитие». То, что Кант — с немалой претензией — сделал для исследования космической эволюции, Гегель совершил для открытия наиболее общих законов диалектики. Марксу мы обязаны знанием некоторых законов общественной эволюции, а Дарвину (1938) — обоснованной теорией происхождения видов. Дарвин сформулировал принцип отбора и продемонстрировал его значение для эволюции в биологии. К сожалению, слишком мало кому известно, что лингвист Шлейхер около 1850 г. совершенно независимо от Дарвина установил аналогичный принцип для развития естественных языков и тем самым заложил основы теории эволюции языковой коммуникации (*Schleicher*, 1863; *Scharf*, 1975). Наконец, нельзя не упомянуть о великой заслуге Клаузиуса в формулировке важнейшего физического закона процессов эволюции — второго начала термодинамики. К Клаузиусу восходят и первые соображения относительно физически обоснованных моделей космического развития. Сколь ни сомнительным может казаться с современной точки зрения вывод Клаузиуса о «тепловой смерти» Вселенной, именно этот вывод послужил толчком к развитию теоретической мысли, которая в работах Эйнштейна, Фридмана и Гамова привела к ныне широко принятой релятивистско-термодинамической модели эволюции.

Разумеется, в связи с исследованием различных фаз эволюции следовало бы назвать еще много других имен; например, в связи с космической эволюцией — имена Гельмгольца, Больцмана, Хаббла, Толмена, Шрёдингера, Альфера, Зельдовича, Пензиаса и Вильсона, в связи с химической эволюцией — имена Оствальда, Опарина, Кальвина, Бернала, Ури и Миллера, в связи с геологической эволюцией — имена Лайеля, Вегенера и Вернадского, в связи с биологической эволюцией — имена Геккеля, Т. Хаксли (Гексли), Холдейна, Райта, Фишера, Четверикова, Лотки, Вольтерры, Дж. Хаксли, Тимофеева-Рессовского, Шмальгаузена, Т. Добжанского, Дельбрюка, Эйгена и многих других.

1.3. Об общей структуре процессов эволюции

Согласно принятому нами определению, эволюцию можно рассматривать как неограниченную последовательность процессов самоорганизации (рис. 1.1). Общая схема процесса эволюции сводится к следующему:

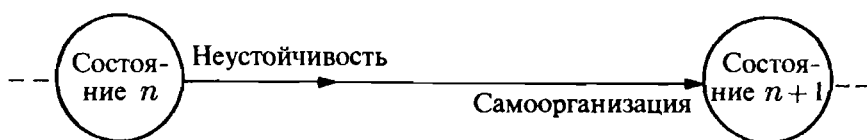


Рис. 1.1. Процесс эволюции как неограниченная последовательность процессов самоорганизации

1. Относительно стабильное n -е состояние эволюции утрачивает устойчивость. В качестве причин, вызывающих потерю устойчивости, выступают временные из-

менения внутреннего состояния или наложенных краевых условий. Особенно характерной причиной эволюционной неустойчивости является внезапное появление новой моды движения, новой разновидности молекул в химии, нового вида в биологии. Этот новый элемент в рассматриваемой динамической системе приводит к потере устойчивости состояния системы, которое до появления нового элемента было устойчивым.

2. Неустойчивость, обусловленная новым элементом в системе, запускает динамический процесс, который приводит к дальнейшей самоорганизации системы. Система порождает новые упорядоченные структуры.

3. По завершении процесса самоорганизации эволюционная система переходит в эволюционное состояние $(n + 1)$. После этого n -го эволюционного цикла начинается новый $(n + 1)$ -й эволюционный цикл.



Рис. 1.2. Спиральная структура процессов эволюции

Эволюционные циклы могут быть относительно одинаковыми, но в то же время по своему характеру качественно весьма различными. В конечном результате циклов самоорганизации заключены как малые, так и большие скачки эволюции. Характерно, что реальная эволюция никогда не заканчивается, она каким-то образом находит выход (неустойчивость) из любого тупика, и этот выход приводит к новому циклу самоорганизации. Поскольку каждый парциальный процесс поднимает систему на новую, в определенном смысле более высокую эволюционную плоскость, весь процесс в целом обладает спиральной структурой (рис. 1.2).

Как показывает приведенная выше схема, эволюционирующая система проходит последовательность состояний различного качества. При этом существенную роль играют различные факторы. Анализ этих действующих и определяющих условий, сил и механизмов имеет значение для разработки теории эволюции. Один из фундаментальных вопросов гласит: какие физические условия необходимы для процессов эволюции? Хотя окончательные ответы пока получить не удалось, имеется целый ряд важных результатов, которые представляют собой солидную основу для поиска ответов (*Шрёдингер*, 1962; *von Neumann*, 1951, 1966; *Prigogine, Nicolis, Babloyantz*, 1972; *Николис, Пригожин*, 1979, 1990; *Eigen*, 1981; *Айген*, 1973; *Эйген, Шустер*, 1982; *Волькенштейн*, 1978, 1984, 1986; *Хакен*, 1980; *Haken* 1983; *Шноль*, 1979).

С физической точки зрения особо важное значение имеют следующие факторы:

- 1) способность к экспорту энтропии путем обмена энергией и веществом с окружающей средой;
- 2) неравновесный характер «диссипативной структуры», отстоящей на закритическое расстояние от термодинамического равновесия;
- 3) нелинейность (динамика системы существенно определяется нелинейными эффектами);
- 4) кооперативный характер динамики подсистем, способность к спонтанному нарушению симметрии;
- 5) способность к самовоспроизведению, т. е. к образованию относительно точных копий исходной системы или подсистем;
- 6) конечность времени жизни системы и связанные с ней непрерывная смена поколений и процесс обновления;

7) мультистабильность, т. е. существование нескольких устойчивых состояний системы, зависимость текущего состояния от предыстории, потенциальная способность к хранению информации;

8) отбор систем и механизмов с благоприятными свойствами из большого числа возможных результатов конкурентных процессов;

9) мультистабильность вследствие случайной ошибки в процессе репродукции как источник новых структур, механизмов и информации;

10) обработка информации, т. е. способность к созданию, хранению, воспроизведению и использованию информации;

11) оптимизация и адаптация, способность приспосабливаться к изменяющимся внешним условиям, существование критериев оптимизации;

12) морфогенез, т. е. формообразование системы и ее органов;

13) образование архетипов, или эталонов, с тенденцией к увеличению многообразия и сложности;

14) ветвление, т. е. все более сильное расщепление реального и в еще большей мере потенциального пути эволюции;

15) сетевая структура с тенденцией к образованию все более сложных соотношений и зависимостей между подсистемами;

16) игровой характер процесса, единство действия необходимых и случайных факторов (при основополагающих закономерностях, определяющих существо процесса);

17) дифференциация, специализация и распределение функций подсистем;

18) интеграция путем соединения в целое все более возрастающей сложности и все большей потенции к действию;

19) иерархическое строение в пространственном, временном и функциональном отношении, разбиение на плоскости, элементы которого объемлют друг друга и вложены один в другой, существование «параметров порядка»;

20) ускорение эволюции, т. е. постоянное нарастание средней скорости эволюционного процесса вследствие механизмов обратной связи.

Разумеется, приведенный длинный перечень можно было бы легко продолжить, включив в него другие важные факторы эволюции. Однако мы должны ограничиться приведенными выше пунктами и в особенности их физическими аспектами, содержащимися, главным образом, в первых десяти пунктах. Почти каждый из этих факторов будет подробно рассмотрен в отдельной главе. Существенно меньше места отводится обсуждению пунктов 11–20, так как в них на первый план выходят нефизические аспекты. Однако и в них содержатся физические компоненты, которые мы поясним в дальнейшем. Эти факторы перекрываются. Более подробную информацию о факторах 11–20 интересующийся читатель сможет найти в литературе (Polanyi, 1968; Шмальгаузен, 1968; Calvin, 1969; Dose, Rauchfuss, 1975; Crow, Kimura, 1979; Odum, 1969, 1971; May, 1973; Деборин и др., 1975; Eigen, Winkler, 1975; Zuckerkandl, 1975, 1976; Романовский, Степанова, Чернавский, 1984; Bresch, 1977; Хакен, 1980; Шноль, 1979; Greenwood, 1985).

Существенным вкладом в физическое понимание качественных переходов при эволюционных процессах мы обязаны аналогии с фазовыми переходами в термодинамике. Как обнаружили впервые Хакен и Шлёгль, явления, известные в равновесной термодинамике под названием фазовых переходов, обладают сходством с качественными сильно неравновесными переходами — кинетическими переходами (Хакен, 1980). Мы демонстрируем такого рода аналогии на примере из экологии (рис. 1.3 и 1.4). Рассмотрим биологические виды, занимающие на определенной территории одну экологическую нишу. При внедрении нового вида, который существенно лучше использует ту же нишу, наступает фаза перехода от существования к полному

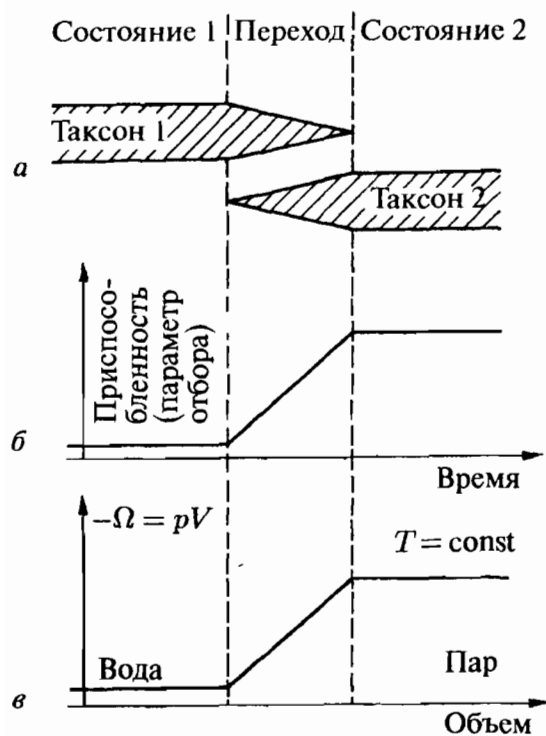


Рис. 1.3. Качественный ход типичного фазового перехода 1-го рода: (а), (б) вытеснение вида при отборе и (в) переход вода—пар

вытеснению исходных видов (рис. 1.3 а). В то время как такие переходы повышают селекционную ценность смешанной популяции от ценности старой популяции до ценности новой. Среди прочих можно отметить следующие свойства такого рода процессов:

- состояния 1 и 2 разделены переходной областью конечной величины;
- в переходной области оба состояния (вида) отличимы друг от друга;
- симметрия относительно использования видами экологической ниши не должна нарушаться;
- скачкообразный переход может быть обойден, например, с помощью медленного улучшения селекционной ценности таксона 1 до ценности таксона 2.

Поразительно похожие утверждения характеризуют термодинамический фазовый переход первого рода, например, переход вода—пар. На рис. 1.3 в для сравнения представлен термодинамический потенциал $\Omega = -pV$ как функция от объема. С такими переходами первого рода мы снова встретимся в главах 4 и 5.

Другим весьма важным для эволюции процессом является специализация, дифференциация или распределение функций (рис. 1.4 а). Например, такой процесс происходит, когда один вид использует две экологические ниши, но в ходе эволюции одна часть вида специализируется по отношению к одной, другая — к другой нише, пока не образуются два различных вида. Поскольку ресурсы обеих ниш используются при этом с разделением функций, становится возможным быстрое улучшение приспособленности (рис. 1.4 б). В подобных случаях удастся установить следующие свойства:

- переход происходит в определенной точке, а именно, с прекращением образования смешанного потомства;
- с наступлением перехода нарушается симметрия использования ниш подвидами;
- в точке перехода оба состояния совпадают.

То, что такой переход обладает аналогией с кинетическим переходом второго рода, мы покажем на примере лазера (рис. 1.4 в). Переходы подобного типа мы обсудим в гл. 5. Аналогия с равновесными фазовыми переходами весьма интересна и с другой точки зрения.

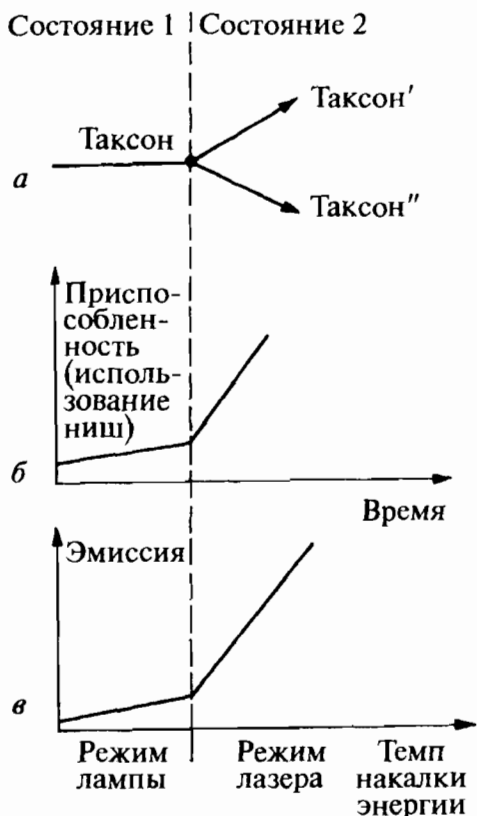


Рис. 1.4. Качественный ход типичного фазового перехода 2-го рода: (а), (б) дифференциация вида и (в) работа лазера

В результате фазового перехода второго рода кооперативное поведение большого числа подсистем приведет к возникновению микросостояния со спонтанно нарушенной симметрией. Такое нарушение симметрии можно рассматривать как элементарный процесс возникновения сложных структур из простых основных элементов. Такие свойства, как макроскопическая намагниченность, кристаллическое дальней упорядочение или сверхтекучесть, которые статистическая механика позволяет рассматривать как следствие взаимодействия высокосимметричных элементарных «кирпичиков», показывают, какие качественно новые явления могут возникать уже в относительно простых макроскопических системах (Anderson, 1972, 1984).

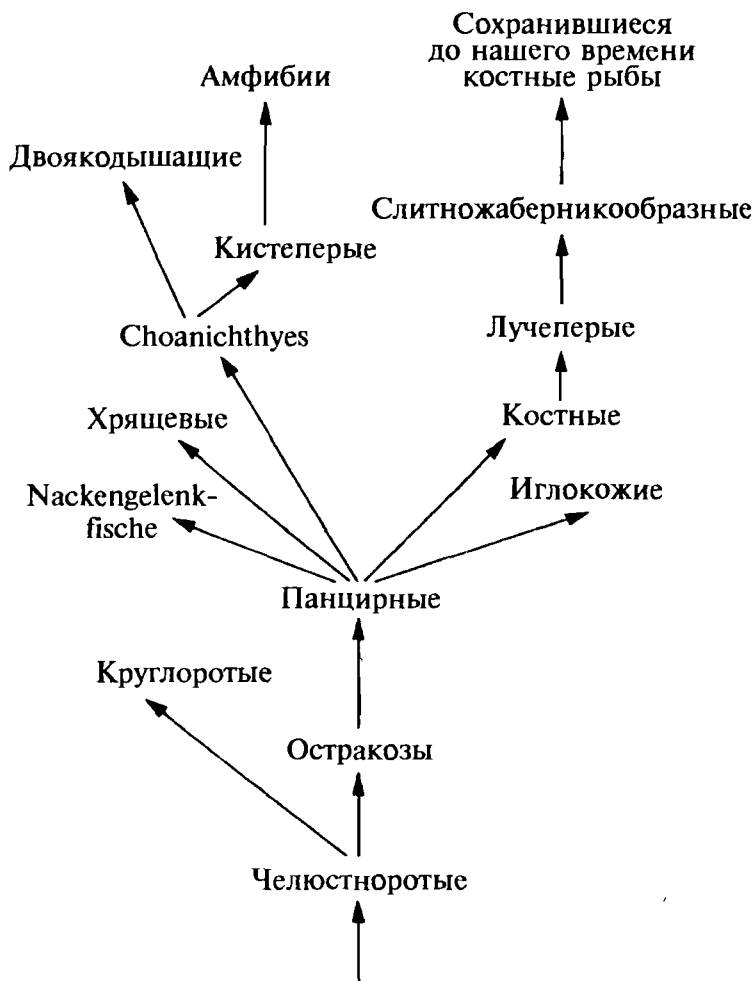


Рис. 1.5. Дерево эволюции рыб (по Крумбигелю (Krumbiegel, 1977))

На рис. 1.4 а показана еще одна существенная черта процессов эволюции — ветвление. Теоретическое весьма наглядное описание этого свойства дает система понятий теории графов или связанный с ней математический формализм. Например, если вершинам графа поставить в соответствие виды некоего рода или класса, возникающие в ходе эволюции жизни на Земле, и соединить ребрами (стрелками) виды, происходящие друг от друга, то мы получаем особого вида ориентированный граф — так называемое дерево эволюции. На рис. 1.5 вы видите дерево эволюции рыб. Оно обладает рядом специфических свойств. У такого дерева всегда ровно один корень (начало, источник) и в большинстве случаев несколько стоков (концов), соответствующих вымершим или рецетным (ныне живущим) видам; циклические последовательности «предок—потомок» наблюдаются лишь в исключительных случаях (например, у вирусов гриппа).

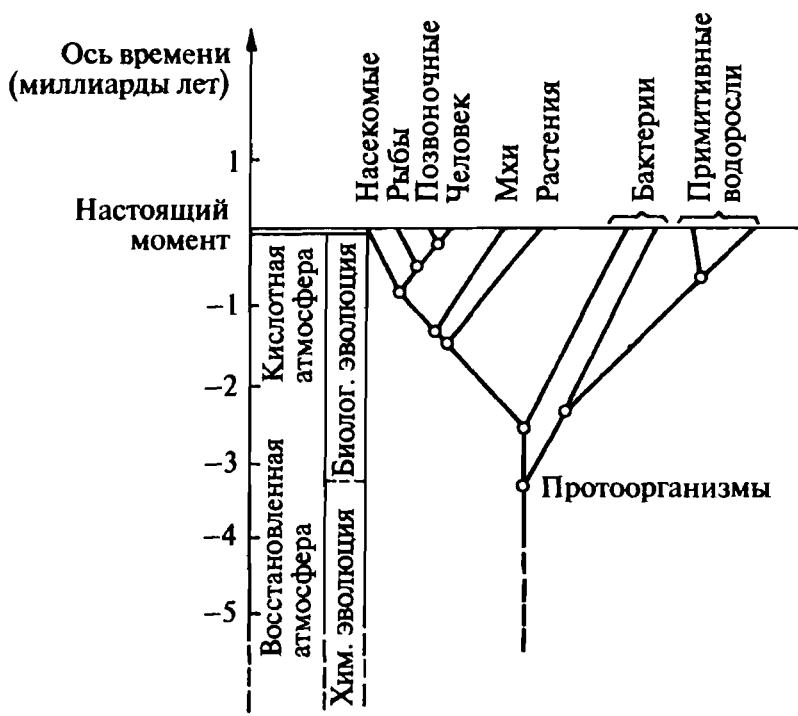


Рис. 1.6. Дерево эволюции жизни на Земле (по Дейхоффу (Dayhoff, 1969))

Важное значение имеет еще одно свойство: не существует сходящихся подграфов. Последнее связано с тем, что, по определению, исключается перенос генетической информации между различными видами одного поколения, и происхождение одного вида от другого означает восприятие генетической информации. Теми же свойствами обладает и общее дерево биологического развития (рис. 1.6). Картина изменяется, если от видов мы перейдем к расам или подвидам (рис. 1.7). Если информационный обмен возможен, то, соответственно, возникают сходящиеся подграфы. В пребиологической эволюции, до того, как было «найдено» разделение полов, могут поэтому встречаться только расходящиеся деревья эволюции. Как писал Бреш (Bresch, 1977), конвергенция всегда привносит с собой значительный прогресс в эволюции: обмен генетической информацией при половых процессах и коммуникация с помощью современных средств связи действуют аналогично. Разумеется, неслучайно, что в подобных и в аналогичных случаях возникают речевые структуры. Говоря о языке, мы понимаем этот термин в обобщенном смысле как линейную последовательность молекул, звуков или символов, построенную по определенным правилам. Передача информации с помощью языка является существенным фактором высших ступеней эволюции (гл. 8 и 11). С другой стороны, возникновение и распространение естественных языков также является хорошо изученным примером процесса эволюции, на котором может быть прослежено действие различных факторов (Scharf, 1975; Klix, 1980). На дереве эволюции естественных языков, представленном на рис. 1.8, хорошо заметно сильное ветвление.

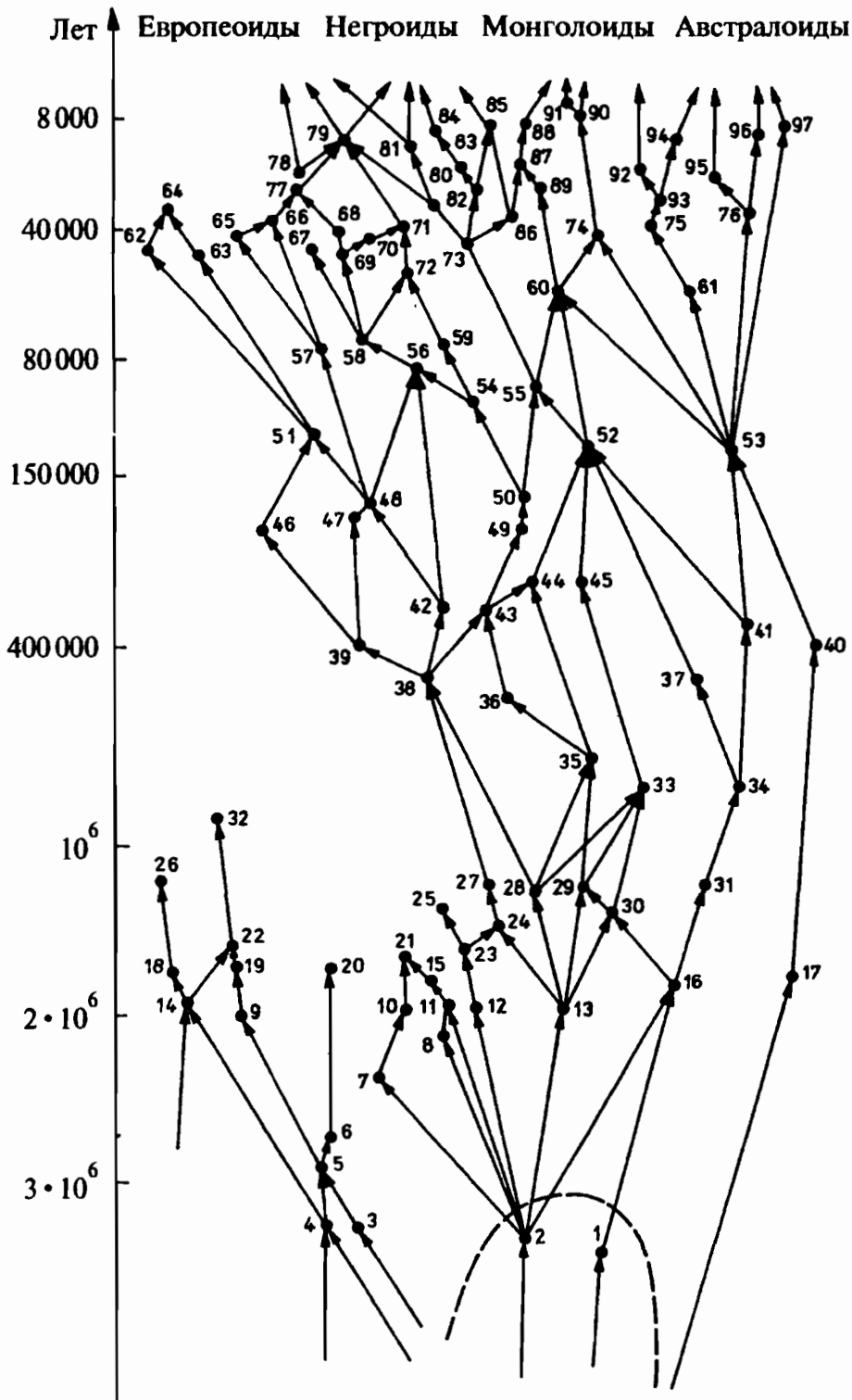


Рис. 1.7. Дерево эволюции человеческого рода (по Фойстелю (Feustel, 1978); обозначения вершин — из работы Эбелинга и Файстеля (Ebeling, Feistel, 1982))

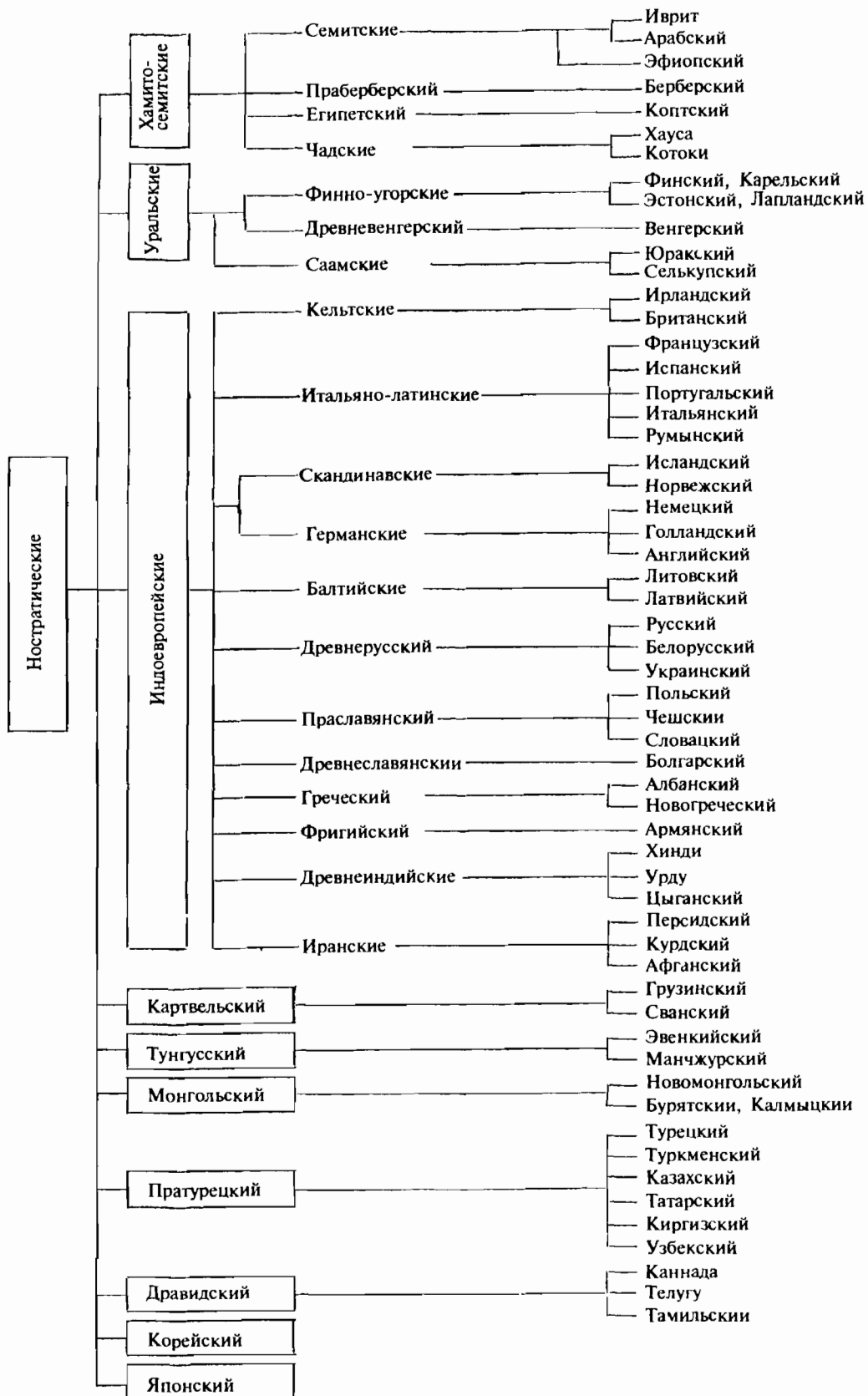


Рис. 1.8. Дерево эволюции естественных языков (по Долгопольскому, 1973)

Глава 2

Сценарий эволюции Метагалактики, Земли и жизни

Три источника обусловили успех космологии: наблюдения, физическая теория и смелость.

Я. Б. Зельдович

2.1. Большой Взрыв и расширение

Современные представления об истории возникновения нашей Метагалактики, то есть системы галактик, принадлежащих нашему Млечному Пути, основываются на пяти важных экспериментальных наблюдениях (*Blome, Priester, 1984*):

1. Исследования спектральных линий звезд показывают, что Метагалактика в среднем обладает единым химическим составом. Преобладают водород и гелий. Остальное составляют главным образом кислород и железо.

2. В спектрах элементов далеких галактик обнаруживается систематическое смещение в сторону красной части. Величина этого смещения возрастает по мере удаления галактик от наблюдателя.

3. Измерения радиоволн, приходящих из космоса в сантиметровом и миллиметровом диапазонах, указывают на то, что космическое пространство равномерно и изотропно заполнено слабым радиоизлучением. Спектральная характеристика этого так называемого фонового излучения соответствует излучению абсолютно черного тела при температуре около 2,7 градуса Кельвина, плотность составляет около 450 протонов на см^3 .

4. По астрономическим наблюдениям, крупномасштабное распределение галактик соответствует постоянной плотности массы, составляющей, по современным оценкам, по крайней мере 0,3 бариона на каждый кубический метр.

5. Анализ процессов радиоактивного распада в метеоритах показывает, что некоторые из их компонентов должны были возникнуть от 14 до 24 миллиардов лет назад.

Первые из таких наблюдений восходят к Густаву Кирхгофу и Роберту Бунзену, которые еще в 1859 г. разработали в Гейдельберге метод спектрального анализа. Этот новый метод обнаружения химических элементов сразу же оказался мощным средством исследования Солнца и других удаленных небесных тел. В результате продолжительных исследований было установлено, что Солнце и звезды не содержат других элементов, кроме тех, которые имеются на нашей Земле. Спектральные исследования все более далеких галактик позволили установить, что наша Метагалактика с точностью до систематических расхождений между звездами, планетами и межзвездной материей имеет однородный химический состав: 77% водорода, 22% гелия, 0,8% кислорода, 0,1% железа и 0,1% всех остальных элементов.

Совокупность других наблюдений подтвердила поразительную однородность и изотропию и относительно других важных физических параметров. Так как тяжелые

ядра образуются лишь существенно позже, для модели эволюции прежде всего важно распределение масс относительно суммарной массы легких ядер. Измерения спектров в оптическом и радиодиапазонах приводят к следующим значениям:

$$\text{He-2: } 0,22-0,03;$$

$$\text{He-3: } 5(+6, -5) \cdot 10^{-5};$$

$$\text{H-2: } 5(+6, -4) \cdot 10^{-5};$$

$$\text{Li-7: } 6 \cdot 10^{-10}.$$

Второе из перечисленных выше наблюдений восходит к американскому астроному Эдвину П. Хабблу, открывшему в 1929 г., что красное смещение далеких галактик пропорционально расстоянию R до них:

$$V = H \cdot R. \quad (2.1)$$

Так называемая постоянная Хаббла H имеет размерность обратного времени. Точное значение ее неизвестно, но имеющиеся наблюдения приводят к оценке

$$H^{-1} = 20(+3, -10) \cdot 10^9 \text{ лет.} \quad (2.2)$$

Относительно большая погрешность связана с тем, что измерение расстояний до соседних скоплений галактик, отстоящих от нас на 50 миллионов световых лет, сопряжено со значительными трудностями.

Третьим решающим наблюдением стало открытие в 1965 г. фонового излучения. Два американских радиоастронома Пензиас и Уилсон систематически исследовали внеземные радиоисточники с помощью рупорного отражателя, чувствительность которого им удалось довести до пределов возможного. В начале 1965 г. им после тщательной проверки всех измерений удалось придти к неопровержимому выводу о том, что существует лишенное выделенного направления слабое радиоизлучение, приходящее на Землю со всех сторон с одинаковой интенсивностью. Измерение спектрального распределения этого излучения дало температуру около 2,7 К. Исследователи пришли к выводу, что все космическое пространство заполнено фотонным газом, находящимся в тепловом равновесии при температуре около 2,7 К.

Четвертое из перечисленных выше утверждений представляет собой экстраполяцию всего накопленного астрономией опыта (коперниканский принцип). Каждое из многократных расширений нашего «радиуса наблюдений» подтверждает это основополагающее предположение. Оценка средней плотности барионной материи была получена в 1958 г. Яном Хендриком Оортом из наблюдений светимости галактик с учетом зависимости масса—светимость. Полученное значение

$$\rho \lesssim 1,7 \cdot 10^{-29} \text{ г/см}^3$$

еще содержало большую погрешность, но независимые наблюдения на гелии и дейтерии привели к оценке

$$\rho \lesssim 5 \cdot 10^{-29} \text{ г/см}^3$$

(Blome, Priest, 1984). По современным оценкам, средняя плотность нуклонов составляет

$$n_n \lesssim 0,3 \text{ м}^{-3}.$$

Наконец, пятое экспериментальное наблюдение основано на определении возраста метеоритов. В качестве космических часов служат ядра радиоактивных атомов с большими периодами полураспада: торий-232, уран-238, плутоний-244, йод-129, лутеций-176, гафний-176, рений-187 и осмий-187.

По оценке Тилемана (*Thielemann et al*, 1983), средний возраст Метагалактики составляет

$$t_0 = (17,6 \pm 4) \cdot 10^9 \text{ лет.}$$

Так как галактики образовались предположительно в первый миллиард лет, возраст нашей Метагалактики с некоторой долей уверенности можно оценить как лежащий в интервале от 14 до 24 миллиардов лет (*Blome, Priester*, 1984).

Каждая гипотеза относительно истории возникновения нашей Метагалактики должна по крайней мере давать объяснение пяти кратко перечисленным выше экспериментальным фактам. Эта проблема и поныне оживленно обсуждается специалистами по космологии. По общему мнению, теоретическая интерпретация этих фактов должна основываться на сформулированной Эйнштейном в конце 1915 г. общей теории относительности. Фундаментальная работа, опубликованная Эйнштейном в 1916 г. в журнале «*Annalen der Physik*», носит название «Основы общей теории относительности». Несколько позже Эйнштейн представил Прусской академии наук работу «Вопросы космологии и общая теория относительности», которая была опубликована в «Сообщениях» (*Sitzungsberichte*) этой академии за 1917 г.

Основные уравнения общей теории относительности, выведенные Эйнштейном, имеют следующий вид:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R_{\alpha\beta}g^{\alpha\beta} = (8\pi Gc^{-2})T_{\mu\nu}. \quad (2.3)$$

Здесь $R_{\mu\nu}$ — тензор Риччи, $g_{\mu\nu}$ — метрический тензор, G — гравитационная постоянная:

$$G = 6,6 \cdot 10^{-8} \text{ см}^3 \text{ г}^{-1} \text{ с}^{-1},$$

c — скорость света, $T_{\mu\nu}$ — тензор энергии-импульса, который, например, в случае идеальной жидкости имеет вид

$$T_{\mu\nu} = (\rho c^2 + p)u_\mu u_\nu + pg_{\mu\nu}, \quad (2.4)$$

где ρ — плотность массы, p — скалярное давление, u_μ — 4-скорость.

Как видно из уравнений (2.3), характеризуемая тензором Риччи $R_{\mu\nu}$ кривизна четырехмерного пространства времени определяется распределением материи, которое задается тензором энергии-импульса $T_{\mu\nu}$. С другой стороны, тензор $T_{\mu\nu}$ зависит от геометрических свойств пространства, поэтому, если принять дополнительно уравнение состояния для среды, заполняющей космос, мы получаем замкнутую систему уравнений для метрики $g_{\mu\nu}$. Движение континуума гидродинамически характеризуется ковариантными градиентами $u_{\mu;\nu}$ поля скоростей. В однородных изотропных моделях Вселенной для четырехмерного интервала ds между соседними мировыми точками справедливы соотношения (*Dautcourt*, 1976):

$$ds^2 = \left(1 + \frac{1}{4}kr^2\right)^{-2} R^2 dr^2 - c^2 dt^2, \quad (2.5)$$

$$r^2 = x^2 + y^2 + z^2,$$

а для градиентов скоростей —

$$u_{\mu;\nu} = (g_{\mu\nu} + u_\mu u_\nu) \frac{\dot{R}}{R}. \quad (2.6)$$

Здесь $R(t)$ — скалярный множитель, также определяющий метрику и тензор кривизны. Для пространственных компонент $l, m = 1, 2, 3$ получаем

$$\begin{aligned} g_{lm} &= R^2 \gamma_{lm}, \\ R_{lm} &= k \gamma_{lm}, \\ \dot{\gamma}_{lm} &= 0. \end{aligned} \quad (2.7)$$

Сравнение уравнений (2.5) и (2.7) показывает, что γ_{lm} — метрика трехмерного пространства постоянной кривизны k . При $k = 1$ получается эллиптическое, при $k = 0$ — евклидово и при $k = -1$ — гиперболическое пространство. Из полевых уравнений (2.3) с учетом уравнений (2.6) и (2.7) следуют уравнения Фридмана (Dautcourt, 1976):

$$\frac{\dot{R}^2}{R^2} = \frac{1}{3} \kappa \rho - \frac{k}{R^2} + \frac{1}{3} \Lambda, \quad (2.8)$$

$$\dot{R} \left(\rho + \frac{p}{c^2} \right) = -\frac{1}{3} \dot{\rho} R. \quad (2.9)$$

Здесь Λ — введенная Эйнштейном космологическая постоянная. В качестве единственного статического решения существует вселенная Эйнштейна:

$$p = 0, \quad \Lambda = R^{-2}, \quad \rho = \frac{2}{\kappa R^2}. \quad (2.10)$$

Статическое решение Эйнштейна (1917) находится в очевидном противоречии с наблюдаемым расширением галактики. Решения, найденные Фридманом (1922, 1924) в предположении $\Lambda = 0$, соответствуют расширяющейся вселенной. При $p = 0$ эти решения имеют следующий явный вид (Dautcourt, 1976):

Открытая (гиперболическая) вселенная ($k = -1$)

$$\begin{aligned} R(t) &= \frac{cq_1}{H_1(1-2q_1)^{3/2}} (\operatorname{ch} \tau - 1), \\ ct &= \frac{cq_1}{H_1(1-2q_1)^{3/2}} (\operatorname{sh} \tau - \tau). \end{aligned} \quad (2.11)$$

Плоская (евклидова) вселенная ($k = 0$)

$$R(t) = R_0 \left(\frac{t}{t_1} \right)^{3/2}. \quad (2.12)$$

Замкнутая (эллиптическая) вселенная ($k = 1$)

$$\begin{aligned} R(t) &= \frac{cq_1}{H_1(2q_1-1)^{3/2}} (1 - \cos \tau), \\ ct &= \frac{cq_1}{H_1(2q_1-1)^{3/2}} (\tau - \sin \tau). \end{aligned} \quad (2.13)$$

Величина H_1 здесь — постоянная Хаббла:

$$H_1 = \left(\frac{\dot{R}}{R} \right)_{t=t_1},$$

q_1 — значение параметра замедления

$$q_1 = - \left(\frac{\ddot{R}R}{\dot{R}^2} \right)_{t=t_1}$$

в настоящее время, т. е. при

$$t_1 \approx 20 \cdot 10^9 \text{ лет.}$$

Все решения (2.11)–(2.13) содержат при $t = 0$ космологическую сингулярность ($R = 0$).

Предположение о нулевом давлении $p = 0$ дает в настоящее время хорошее приближение, но утрачивает силу в окрестности сингулярности. В случае ультрарелятивистского газа с уравнением состояния (вселенная, заполненная излучением)

$$p = \frac{1}{3} \rho c^2 \quad (2.14)$$

мы получаем следующее решение уравнений (2.8)–(2.9):

$$R^2(t) = 2q_1^{1/2} R_1^2 H_1 t - (q_1 - 1)t^2 R_1^2 H_1^2, \quad (2.15)$$

$$\rho(t) = \rho_1 \left(\frac{R_1}{R(t)} \right)^4. \quad (2.16)$$

В дальнейшем мы покажем, что предположение о вселенной, заполненной излучением, при малых временах после начала расширения является очень хорошим приближением, поэтому на начальной стадии расширения можно принять закон квадратного корня

$$R(t) \sim \sqrt{t}, \quad (2.17)$$

$$\rho(t) \sim t^{-2}. \quad (2.18)$$

Как показывает это решение, плотность вселенной вскоре после начала расширения должна быть необычайно высокой. По законам термодинамики мы можем ожидать вместе с Альфером, Херманом и Гамовым (*Alpher, Hermann, Gamov*, 1948), что вскоре после Большого Взрыва во вселенной царят очень высокие температуры. Из этой идеи возникла модель «горячего взрыва». Оценим закон, по которому изменяется температура.

Кинетическая энергия релятивистских частиц (фотонов) в расширяющейся вселенной падает как

$$E = p \sim R^{-1} \sim t^{-1/2}.$$

Следовательно, температура как средняя кинетическая энергия также убывает по закону квадратного корня:

$$T(t) = ct^{-1/2}.$$

Так как температура фотонов фонового излучения ныне, т. е. спустя $t = 10^{10}$ с после начала расширения, составляет около 3 К, мы ожидаем, что

$$c \approx 10^{10} \text{ К} \cdot \text{с}^{1/2}.$$

Температура нерелятивистских частиц падала бы по закону

$$T(t) \sim \frac{p^2}{2m} \sim \frac{1}{t}.$$

Поскольку число фотонов в Метагалактике примерно в 10^9 раз больше числа массивных частиц, на начальных фазах развития Вселенной температура спадает по единому закону

$$T(t) \approx \frac{1}{\sqrt{t}} 10^{10} \text{ К} \cdot \text{с}^{1/2}. \quad (2.19)$$

Более точный вывод этого важного соотношения исходит из плотности энергии релятивистской материи (Dolgov, Zeldovich, 1981):

$$\rho_E = \frac{3}{32\pi G t^2}. \quad (2.20)$$

Связь с температурой получается, если предположить, что выполняется закон Стефана—Больцмана:

$$\rho_E = \frac{\pi^3}{30} N_{\text{ст.св.}} T^4, \quad (2.21)$$

где $N_{\text{ст.св.}}$ — число эффективных степеней свободы. Для фотонов $N_{\text{ст.св.}} = 2$. Приравняв соотношения (2.20) и (2.21), получаем

$$T(t) = \left(\frac{45}{16\pi^3 N_{\text{ст.св.}} G} \right)^{1/4} t^{-1/2}. \quad (2.22)$$

В рамках классической общерелятивистской космологии вопрос о геометрическом типе вселенной остается открытым. Предпочтение одному из трех возможных случаев $k = 1$, $k = 0$ и $k = -1$ может быть отдано только на основании экспериментальных данных. Замкнута ли вселенная или открыта, плоска или искривлена, зависит от средней плотности массы во вселенной. Известное в настоящее время значение $\rho \simeq 10^{-29} \text{ г} \cdot \text{см}^{-3}$ расположено относительно близко к критическому значению

$$\rho_{\text{крит}} = (0,5 - 2) \cdot 10^{-29} \text{ г} \cdot \text{см}^{-3}.$$

Многие исследования предполагают существование скрытых масс, приводящих к за- критической плотности и тем самым склоняющих чашу весов в пользу замкнутой модели мира. При определенных условиях за скрытые массы могут отвечать гипотетические массы покоя нейтрино. В рамках квантотеоретической космологической модели явное предпочтение отдается замкнутой вселенной (Фомин, 1986).

2.2. Многократное нарушение симметрии и возникновение элементов

Расширение Метагалактики оставляет для эволюции широкие рамки, которые подлежат заполнению. Следуя различным авторам, мы делим весь процесс на 12 эпох (Fritsch, 1983; Wilczek, 1983; Linde, 1984; Ebert, 1984; Feistel, Ebeling, 1989). Возникновение химических элементов охватывает первые восемь эпох:

1-я эпоха (физический вакуум):

$$t \lesssim t_{\text{Planck}} \simeq 10^{-43} \text{ с}.$$

Относительно этой самой ранней эпохи после начала расширения по существу ничего не известно, кроме более или менее смелых измышлений. По-видимому, ни пространство, ни время не имели ни сколько-нибудь определенного смысла, ни размерности. Вселенная находилась в состоянии с высокой симметрией.

2-я эпоха (выход из вакуума):

$$10^{-43} \text{ с} \lesssim t \lesssim 10^{-33} \text{ с.}$$

В начале этой эпохи Вселенная распадается на различные области (Linde, 1984). Одна из таких областей имеет пространственно-временную размерность $d = 4$ и обладает симметрией $SU(3) \times SU(2) \times U(1)$. Наша Метагалактика рождается в результате первого нарушения симметрии. Вся Метагалактика заполнена однородной газобразной смесью праматерии: кварков, электронов, нейтрино, фотонов, глюонов, X -бозонов. Как показывает оценка (2.19), такой газ праматерии в начале эпохи обладает чрезвычайно высокой температурой:

$$T \approx 10^{32} \text{ К.}$$

На основе анализа различного рода космологических трудностей многие авторы приходят к предположению о том, что начало этой эпохи связано с фазой инфляционного расширения, в ходе которой «протопузырь» быстро увеличивается в размерах в 10^{50} раз (Guth, 1983; Guth, Steinhart, 1984; Linde, 1984). Мы ограничимся грубым приближением и рассмотрим простое фридмановское расширение, в среднем связанное с охлаждением по закону квадратного корня (2.19). Каким бы ни был временной ход эволюции во 2-ю эпоху, во многом неясную и таящую много неизвестного, мы исходим из того, что в конце эпохи температура падает по закону (2.19) до

$$T \approx 10^{28} \text{ К.}$$

Такая температура соответствует энергиям около 10^{24} эВ. «Высвобождаются» X -бозоны. Вследствие этого происходит нарушение симметрии, приводящее к разделению сильного и электрослабого взаимодействий. Метагалактика вступает в более привычную нам фазу, легче поддающуюся теоретическому анализу. Относительно этой фазы в настоящее время известно значительно больше (стандартная модель).

3-я эпоха (квази-фотонный газ):

$$10^{-32} \text{ с} \lesssim t \lesssim 10^{-11} \text{ с.}$$

В начале 3-й эпохи протогаз состоит из кварков и антикварков, лептонов, фотонов и X -бозонов. После распада X -бозонов плотность кварков, хотя и на очень малую величину (около 10^{-8}), превосходит плотность антикварков. Метагалактика ведет себя в дальнейшем (с термодинамической точки зрения) как обычный адиабатически расширяющийся газ. К описанию его применима стандартная модель адиабатического расширения, изложенная в основных чертах в работе Альфера, Хермана и Гамова (Alpher, Hermann, Gamov, 1948). Рассмотрим эти основные черты термодинамической модели.

Вследствие чрезвычайно высоких температур, достигающих к началу 3-й эпохи 10^{28} К, протогаз можно считать идеальным газом. Эффекты неидеальности пренебрежимо малы. Для адиабатического процесса в идеальном газе справедливо следующее соотношение между температурой и плотностью:

$$T = \text{const} \cdot \rho^{\gamma-1}. \quad (2.23)$$

Показатель адиабаты $\gamma = C_p/C_v$ принимает для атомарного (массивного) газа значение $\gamma = 5/3$, а для релятивистского газа (с нулевой массой покоя) — значение $\gamma = 4/3$ (Ландау, Лифшиц, 1988, 1995). Так как на более поздней стадии развития Вселенной преобладают (и по числу, и по вкладу в плотность энергии) релятивистские частицы, в среднем выполняется релятивистское соотношение

$$T = \text{const} \cdot \rho^{1/3}. \quad (2.24)$$

Учитывая зависимость

$$\rho(t) \sim R(t)^{-3} \sim t^{-3/2}, \quad (2.25)$$

мы приходим к характерному закону квадратного корня

$$T(t) \sim t^{-1/2}.$$

Адиабатическое расширение устанавливает для эволюции Метагалактики симметричные по времени рамки. Временная симметрия нарушается, эволюция становится необратимой.

Нарушение симметрии

обратимость $\longrightarrow$ необратимость

вводит в игру второе начало термодинамики. Оно играет роль основного физического закона эволюции.

Как было показано выше, Метагалактика на раннем этапе своего развития уже обладает стрелой времени, но еще не наделена пространственной структурой. С точностью до малых флуктуаций материя однородно заполняется имеющееся пространство. Температуры предельно высокие, и Вселенная, «переполненная» энергией, напоминает заряженную до отказа электрическую батарею или сжатую до отказа пружину. Вследствие адиабатического охлаждения температуры к концу 3-й эпохи понижаются до

$$T \approx 10^{16} \text{ К} \approx 1000 \text{ ГэВ}.$$

Падение средней энергии частиц приводит к нарушению электрослабой симметрии, а именно: при средних энергиях ниже масс $W^\pm$ - и Z^0 -бозонов эти частицы «высвобождаются». Происходит разделение слабого взаимодействия, с одной стороны, и электромагнитного взаимодействия — с другой. Это нарушение симметрии влечет за собой переход к новой эпохе эволюции.

4-я эпоха (аннигиляция кварков):

$$10^{-11} \text{ с} \lesssim t \lesssim 10^{-6} \text{ с}.$$

В начале этой эпохи в протогазе преобладают кварки q и антикварки $\bar{q}$. Существует шесть различных сортов кварков, в том числе u -кварки и d -кварки (Lanius, 1980; Fritsch, 1981). В результате усиливающегося адиабатического расширения протогаза температура к концу 4-й эпохи убывает до

$$T \approx 10^{13} \text{ К} \approx 1 \text{ ГэВ}.$$

Тем самым протогаз оказывается в области, где резко возрастает вероятность попарной аннигиляции кварка и антикварка. Тепловое равновесие реакций аннигиляции

$$\begin{aligned} q + \bar{q} &\rightleftharpoons 2\gamma, \\ q + \bar{q} &\rightleftharpoons e + \bar{e} \end{aligned}$$

сильно смещено вправо. Такой вывод следует из закона действия масс

$$\begin{aligned} \frac{n_\gamma^2}{n_q n_{\bar{q}}} &= K_1(T) \sim \exp \left\{ -\frac{\Delta g_1}{k_B T} \right\}, \\ \frac{n_e n_{\bar{e}}}{n_q n_{\bar{q}}} &= K_2(T) \sim \exp \left\{ -\frac{\Delta g_2}{k_B T} \right\}. \end{aligned} \quad (2.26)$$

Здесь в левой части каждого равенства стоят соответствующие плотности частиц, а в правой — свободные энтальпии реакций, составляющие величины порядка

$$\Delta g_i \approx -10 \text{ ГэВ}.$$

В реакциях аннигиляции кварков энергия связи произвольно полагается равной величине порядка 10^{12} эВ. Аннигиляция термодинамически выгодна, и почти все кварки находятся в связанных состояниях, соответствующих более глубоко лежащим значениям потенциала Гиббса. Мы уже упоминали о том, что в протогазе имеется небольшой (составляющий примерно одну миллионную долю процента) избыток кварков по сравнению с антикварками. После аннигиляции остаются практически только кварки, а антикварки исчезают. Симметрия между материей и антиматерией нарушается.

Нарушение симметрии

материя—антиматерия $\rightarrow$ преобладание материи

приводит к весьма глубоким последствиям для структуры нашей современной Вселенной, в которой античастицы встречаются лишь чрезвычайно редко. Именно оно подготавливает почву для перехода к следующей эпохе, в которой образуются протоны и нейтроны.

5-я эпоха (образование нуклонов):

$$10^{-6} \text{ с} \leq t \leq 10^{-3} \text{ с}.$$

На основе притягивающих хромодинамических взаимодействий три кварка могут образовывать связанные состояния; эти связанные состояния соответствуют хорошо известным нуклонам. Один u -кварк и два d -кварка образуют нейтрон:

$$u + 2d \rightleftharpoons n,$$

а два u -кварка и один d -кварк — протон:

$$2u + d \rightleftharpoons p.$$

Так как энтальпия реакции этого процесса составляет около 100 МэВ, равновесие в ходе 5-й эпохи сдвигается далеко вправо. Свободные кварки исчезают из Метагалактики. К концу 5-й эпохи Метагалактика состоит из нуклонов с плотностью около

$$n_n \approx 10^{27} \text{ см}^{-3}.$$

Помимо нуклонов во Вселенной имеются электроны, позитроны, фотоны и нейтрино с существенно более высокими плотностями. Суммарная плотность составляет величину около 10^{36} частиц на см^3 , и плотность энергии определяется фотонами и в меньшей степени лептонами.

В ходе 5-й эпохи отчетливо проявляется нарушение симметрии относительно барионов. Барионы составляют в Метагалактике абсолютное большинство частиц по сравнению с антибарионами, которые практически не встречаются в естественных условиях.

6-я эпоха (появление свободных нейтрино):

$$10^{-3} \text{ с} \leq t \leq 1 \text{ с}.$$

На протяжении этой эпохи структура Метагалактики изменяется несущественно, но плотность и температура продолжают падать. К концу эпохи температура понижается до

$$T \approx 10^{10} \text{ К} \approx 1 \text{ МэВ}.$$

Плотность массы этой горячей плазмы составляет примерно величину

$$\rho \approx 10^8 \text{ г} \cdot \text{см}^{-3}.$$

Средняя плотность материи убывает в Метагалактике как куб расстояния. Современная Метагалактика состоит из относительно однородного фотонного газа с плотностью около 500 фотонов на см^3 , что, по существу, соответствует реликтовому излучению, а также нейтринному газу с плотностью около 300 нейтрино и 300 антинейтрино на см^3 . Впрочем, существование такого нейтринного газа пока экспериментально не подтверждено. О средней плотности материи с ненулевой массой покоя мы, к сожалению, располагаем лишь весьма неполной информацией. Если бы вся имеющаяся в виде вещества материя была распределена равномерно, то в нулевом приближении это соответствовало бы газу со средней плотностью около 1 атомного ядра на м^3 , причем 70% приходилось бы на протоны, а 30% — на ядра гелия с небольшой примесью ядер других элементов. Отношение между плотностью фотонов и ядер атомов составляет около $5 \cdot 10^9$; оно почти не изменилось с эпохи горячей протоплазмы (Новиков, 1979; Зельдович, 1983; Зельдович, Новиков, 1967, 1975).

Исходя из приведенных выше соображений, мы находим следующие законы эволюции для средних плотностей фотонов, нуклонов и нейтрино в Метагалактике:

$$\begin{aligned} n_\gamma(t) &\approx 500 \text{ см}^{-3} \left[\frac{R(t_1)}{R(t)} \right]^3, \\ n_n(t) &\approx 1 \text{ м}^{-3} \left[\frac{R(t_1)}{R(t)} \right]^3, \\ n_\nu(t) &\approx 300 \text{ см}^{-3} \left[\frac{R(t_1)}{R(t)} \right]^3. \end{aligned} \quad (2.27)$$

Здесь t_1 — эталонное время, $R(t_1)$ — эталонное расстояние; настоящее время соответствует выбору $t_1 \approx 10^{20}$.

Закон расширения Хаббла $R(t)/R(t_1) = Ht$ выполняется только для ближайшего прошлого, т. е. лишь при $(t - t_1) \ll H^{-1}$, а при малых временах $t \ll H^{-1}$ мы, решая уравнения Эйнштейна (см. разд. 2.1), приходим к закону расширения

$$R(t) = \text{const} \cdot t^{1/2}.$$

В дальнейшем мы будем рассматривать функцию расширения $R(t)$ как заданную монотонно возрастающую функцию. Исходя из такого предположения, мы исследуем термодинамику эволюции нашей газовой модели Метагалактики. Поскольку средние плотности малы, мы можем, как и выше, воспользоваться в первом приближении законами идеального квантового газа (Ландау, Лифшиц, 1995).

Для адиабатического расширения атомарного газа имеет место инвариант $n_{\text{ат}} T_{\text{ат}}^{-3/2} = \text{const}$. Отсюда для температуры следует закон эволюции

$$T_{\text{ат}}(t) = \text{const} \cdot R(t)^{-2}. \quad (2.28)$$

Аналогично мы получаем для плотности энтропии

$$s_{\text{ат}}(t) = \frac{S_{\text{ат}}(t)}{V(t)} = \text{const} \cdot R(t)^{-3}. \quad (2.29)$$

Для фотонного газа адиабатический инвариант имеет вид $n_\gamma T_\gamma^{-3} = \text{const}$, из чего следует закон эволюции для температуры

$$T_\gamma(t) = \text{const} \cdot R(t)^{-1}. \quad (2.30)$$

Наконец, для плотности энтропии мы получаем

$$s_\gamma(t) = \text{const} \cdot T_\gamma(t)^3 = \text{const} \cdot R(t)^{-3}. \quad (2.31)$$

Как показывают вычисления, средние по пространству от плотности массы, температуры и плотности энтропии в расширяющейся Вселенной монотонно убывают. Эти факты служат важной основой эволюции сложных структур в Метагалактике.

Наши оценки при $t = 1$ приводят к средним плотностям частиц

$$n_\gamma \approx 10^{30} \text{ см}^{-3}, \quad n_n \approx 10^{21} \text{ см}^{-3}, \quad n_\nu \approx 10^{30} \text{ см}^{-3}.$$

Тем самым по крайней мере для плотностей нуклонов достигаются уже вполне «земные» порядки величин, так как 10^{21} нуклонов на 1 см^3 примерно соответствуют реальным плотностям нуклонов в твердых телах и в жидкостях. Здесь в игру вступают некоторые особенности нейтрино. Эти частицы, существование которых было постулировано Паули в 1930 г., не участвуют ни в сильном, ни в электромагнитном взаимодействиях, но участвуют в слабом и в гравитационном взаимодействиях. Поэтому нейтрино могут проникать в земную толщу на километровые глубины, не испытывая сколько-нибудь существенного воздействия со стороны среды.

При этих условиях нейтрино практически превращаются в независимый газ. Полное термодинамическое равновесие системы нарушается отделением нейтринного газа. Метагалактика состоит с этого момента из двух фактически независимых подсистем — нейтринного газа и остального газа, состоящего из фотонов, лептонов и барионов. Экспериментальное доказательство существования нейтринного газа вследствие малого сечения взаимодействия нейтрино и поныне не свободно от возражений. Надежное подтверждение существования нейтринного газа принадлежит к числу наиболее насущных и важных задач современной экспериментальной физики.

Выделение нейтринного газа можно рассматривать как нарушение симметрии

$$\text{тепловое равновесие} \longrightarrow \text{распад на две независимые тепловые системы.}$$

В отличие от рассмотренных выше нарушений симметрии здесь впервые нарушается симметрия относительно чисто термодинамической переменной — температуры. При рассмотрении последующих эпох необходимость принимать во внимание нейтринный газ полностью отпадает.

7-я эпоха (нарушение нейтрон-протонной и электрон-позитронной симметрии):

$$1 \text{ с} \leq t \leq 100 \text{ с.}$$

Первоначально плотность протонов в Метагалактике равна плотности нейтронов. Однако когда температура падает ниже

$$T \approx 1,4 \cdot 10^{10} \text{ К} \approx 1,3 \text{ МэВ},$$

разница в массах нейтрона и протона становится заметной. Как известно, масса покоя нейтрона составляет 939,550 МэВ, а масса покоя протона — 938,256 МэВ. Поэтому большинство нейтронов до тех пор распадается на протоны и электроны, пока, наконец, не устанавливается соотношение 75% протонов и 25% нейтронов.

Вторым важным результатом 7-й эпохи является аннигиляция электронов и позитронов:

$$e + \bar{e} \rightarrow \gamma.$$

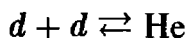
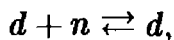
Лишь немногие электроны переживают этот процесс. Вследствие электронейтральности Метагалактики средняя плотность оставшихся электронов в точности равна плотности протонов. В результате этого процесса Вселенная в конце 7-й эпохи превращается в нейтральную плазму, состоящую из протонов, нейтронов, электронов и фотонов. На заднем плане, в качестве фона, как и прежде, существует море

обособившихся нейтрино. Температура плазмы понижается до одного миллиарда Кельвин, а плотность приближается к плотности внутри звезд.

8-я эпоха (синтез ядер):

$$10^2 \text{ с} \leq t \leq 10^4 \text{ с.}$$

Температура и плотность во Вселенной — такие же, как внутри звезд-гигантов. Идут процессы нуклеосинтеза, и, в частности, равновесие в реакциях



сдвигается все более вправо. Вследствие стабильности ядер гелия почти все нейтроны оказываются «встроенными» в ядра гелия. Наконец, устанавливается соотношение

$$\text{H} : 23\%, \quad \text{He} : 77\%.$$

Результат вычислений подтверждает приведенные в разд. 2.1 экспериментальные данные. Замечательный результат состоит в том, что гипотеза горячего взрыва может быть столь хорошо подтверждена экспериментально найденными распространённостями легких ядер. Ядра остальных более тяжелых элементов возникают лишь в результате процессов, протекающих позднее, например, взрывов сверхновых.

В заключение мы можем констатировать, что гипотеза горячего взрыва позволяет удовлетворительным образом интерпретировать все пять рассмотренных выше экспериментальных фактов. Именно поэтому современные представления о возникновении нашей Метагалактики основаны на изложенной нами модели, хотя многие вопросы все еще остаются открытыми.

2.3. Структурирование протоплазмы и образование фотонной мельницы

В результате восьми описанных в предыдущем разделе эпох, охватывающих по времени около трех часов (рис. 2.1), образовалась горячая плазма, состоящая из ядер различных элементов, а также электронов и фотонов. Плотность электронов в протоплазме в начале **9-й эпохи** составляет величину около

$$n_e \approx 10^{17} \text{ см}^{-3},$$

а температура — около

$$T \approx 10^8 \text{ К.}$$

Такие параметры соответствуют, например, плазме в термоядерной установке (существующей пока только на чертежной доске). Тем самым Вселенная из некоей экзотической релятивистской фазы переходит в фазу, приближающуюся к земным условиям, что делает ее более доступной экспериментальным исследованиям. Теорию экзотических — первых восьми — эпох мы рассмотрели лишь в самых общих чертах, не вдаваясь в детали; более подробное изложение предполагало бы у читателя обширные познания в общерелятивистской теории гравитации и современной физике элементарных частиц и теории поля. Так как мы намереваемся заняться вопросами эволюции сложных структур на выбранной модельной планете Земля, нам придется по всем вопросам, относящимся к более ранней эволюции Вселенной, отослать интересующегося читателя к более глубоким изложениям в специальной литературе (Зельдович, Новиков, 1967, 1975; Treder, 1968, 1986; Schmutzer, 1968; Peebles, 1971; Пиблс, 1983; Dautcourt, 1976; Вайнберг, 1981; Новиков, 1979, 1983; Barrow, Silk, 1980; Зельдович, 1983; Dolgov, Zeldovich, 1981; Гуревич, Чернин, 1987).

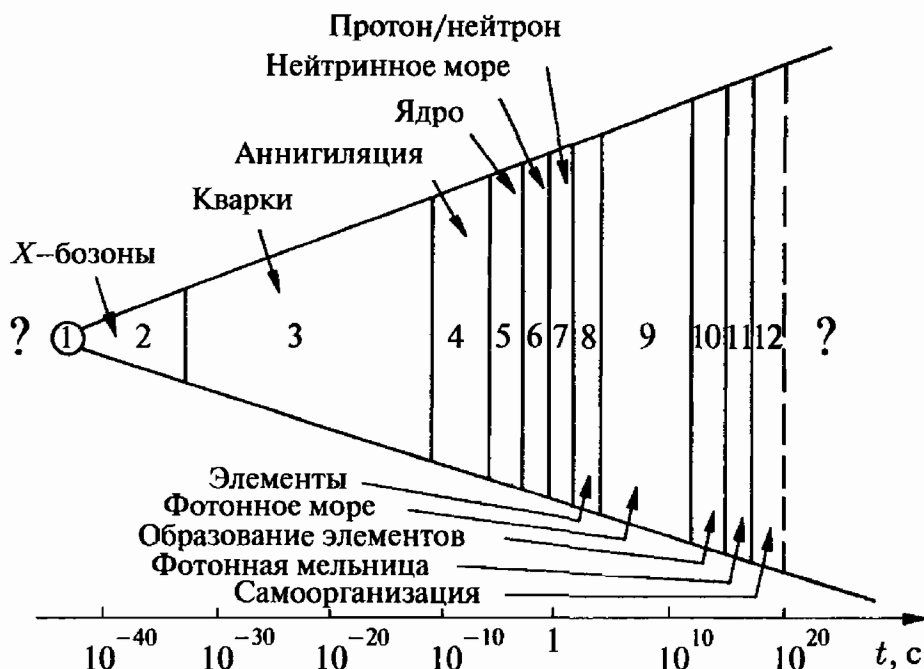


Рис. 2.1. Двенадцать эпох эволюции расширяющейся Метагалактики

Рассмотрим теперь фазу плазмы, в которой Вселенная оказалась в начале 9-й эпохи через три часа после Большого Взрыва. Относительно начала этой фазы мы пока располагаем скудной информацией, в основном, в виде гипотез (Вайнберг, 1981; Новиков, 1983). Однако мы располагаем надежным источником информации о состоянии Вселенной в тот период в форме космического фонового излучения, которое называется также реликтовым излучением. Как уже было сказано, речь идет о тепловом излучении с температурой 2,7 К, максимум которого лежит в центральной части радиодиапазона. Это тепловое излучение заполняет всю Вселенную с плотностью около $5 \cdot 10^8$ фотонов на кубический метр и приходит на Землю со всех направлений. По гипотезе Гамова, Дорошкевича, Новикова, Дикке, Пиблза и других авторов, реликтовое излучение происходит от горячей плазмы, заполнявшей Вселенную через несколько часов после Большого Взрыва.

Так как плотность фотонов в протоплазме весьма значительно превосходила плотность остальных частиц, мы оцениваем отношение числа фотонов к числу нуклонов примерно в 10^9 и говорим о фотонной плазме. При адиабатическом расширении Вселенной числовое значение этого отношения, связанного с энтропией системы, практически не изменяется. В современной Вселенной мы обнаруживаем в среднем около одного атома водорода на кубический метр, а все остальные атомы встречаются существенно реже. Что же касается средней плотности фотонов, то она составляет почти 10^9 на кубический метр. Но в отличие от современных соотношений фотоны протоплазмы к началу 9-й эпохи обладали весьма высокой средней тепловой энергией — около 10 кэВ. В ходе последующей эволюции адиабатически расширяющаяся Вселенная постоянно охлаждается, причем температура убывает примерно по закону квадратного корня:

$$T \approx 2 \cdot 10^6 \cdot t^{-1/2},$$

T — в Кельвинах, t — в годах.

В ходе расширения температура после 10^5 – 10^6 лет падает примерно до 2 000–4 000 К, и тогда внезапно происходит образование атомов водорода и гелия, а также молекул водорода. Непрозрачная фотонная плазма превращается в прозрачный нейтральный газ. Наш повседневный опыт подтверждает этот процесс. Если выключить

люминесцентную лампу, наполненную горячей непрозрачной плазмой, то газ-наполнитель внезапно становится прозрачным. Атомарные и молекулярные газы слабо взаимодействуют с излучением. Они прозрачны для света. Грубое приближение хода адиабаты на плоскости плотности свободных электронов и температуры представлено на рис. 2.2. Известно, что примерно через 300 000 лет адиабата входит в область образования атомов водорода (Эбелинг и др., 1979; Kraeft et al., 1986). Наблюдаемое в настоящее время реликтовое излучение дает нам своеобразную «фотографию» Вселенной в том виде, в каком та была в момент образования атомов. Если с протоплазмой фотоны находились в термодинамическом равновесии, то с только что образовавшимся разреженным газом фотоны взаимодействуют лишь чрезвычайно слабо и поэтому могут беспрепятственно пролетать огромные расстояния. Некоторые из образовавшихся в ту эпоху фотонов улавливаются теперь земными телескопами как реликтовое излучение и дают нам информацию о событиях, разыгрывавшихся через несколько часов после Большого Взрыва.

10-я эпоха (пространственно-временное самоструктурирование).

В состоянии относительно однородного разреженного нейтрального газа Вселенная пребывала предположительно около 10^8 – 10^9 лет, не переставая расширяться. Естественно спросить, как в такой разреженной нейтральной Вселенной могло произойти образование упорядоченных структур. В ответ мы сошлемся на три наиболее важные физические причины:

- 1) дальнейшее понижение температуры и уменьшение плотности энтропии вследствие расширения;
- 2) возникновение гравитационных неустойчивостей;
- 3) передача энтропии фоновому излучению.

Основным механизмом структурообразования в 10-й эпохе являются гравитационные неустойчивости. Еще в 1662 г. Ньютон высказал гипотезу о том, что галактики и звезды могли образоваться из первоначально равномерно распределенного газа под действием гравитации. Ныне эта гипотеза общепризнанна и подтверждена и многочисленными теоретическими работами, и модельными расчетами (Peebles, 1971; Пиблс, 1983; Гуревич, Чернин, 1987).

Рассмотрим эволюцию медленно вращающегося газового шара, который мог образоваться под действием гравитационных неустойчивостей. Гравитационное притяжение приводит к все более сильному сжатию газового шара, причем вследствие сохранения углового момента частота вращения все более возрастает. В результате сжатия возрастает число атомных столкновений, т. е. температура газового шара увеличивается. С повышением температуры, как следует из формулы Саха, возрастает степень ионизации газа (Эбелинг и др., 1979; Kraeft et al., 1986). Сильное кулоновское отталкивание сначала удерживает ядра водорода и гелия на большом расстоянии друг от друга, но при повышении температуры эти ядра все больше сближаются вплоть

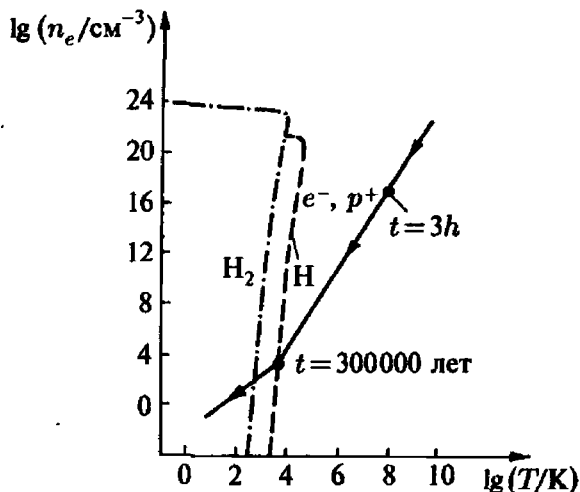


Рис. 2.2. Приближенный ход адиабатического расширения в водородной Вселенной. Излом при $t = 300\,000$ лет соответствует переходу между двумя показателями адиабаты, соответствующими расширению Вселенной с преобладанием излучения и Вселенной с преобладанием вещества ($\gamma = 5/3$) (штрихпунктирная линия — 50%-я диссоциация молекулярного водорода; штриховая линия — 50%-я диссоциация атомарного водорода)

до расстояний около 10^{-15} м. Важную роль в преодолении кулоновских барьеров играет квантовомеханический туннельный эффект (рис. 2.3). Когда многочисленные ядра водорода сближаются до расстояний 10^{-15} м, вступают в действие ядерные силы притяжения. В центрах газовых шаров инициируются реакции ядерного синтеза, и тем самым происходит рождение звезд первого поколения. Звезды развиваются по зонам по мере того, как содержащийся в них водород медленно сгорает, превращаясь в гелий. Дальнейшая история их развития варьируется в весьма широких пределах и сильно зависит от массы звезд (Шкловский, 1984). Как правило, массы звезд заключены в пределах от 10^{29} до 10^{33} кг. В ходе эволюции звезд синтезируются не только ядра гелия, но и ядра легких элементов Li, Be, B, C, N, O и т. д. вплоть до железа. Впрочем, синтез этих ядер требует весьма высоких температур внутри звезд, поэтому гелиевое горение $3\text{He}^4 \rightarrow \text{C}^{12}$ начинается лишь при температуре около 10^8 К. В случае не очень больших звезд с массами, не превосходящими 10^{31} кг, постоянное сжатие может быть приостановлено давлением вырождения электронов. Возникают небольшие компактные звезды размером с Землю, но с массой, равной примерно массе Солнца, — так называемые белые карлики.

Гораздо более сложное поведение демонстрируют более массивные звезды. Например, образуются так называемые красные гиганты, внутри которых плотность ядер гелия достигает высокого уровня. Эти ядра заключены внутри менее плотной плазмы, отдающей ядра и атомы в окружающее пространство. Широкий класс массивных звезд оказывается нестабильным. Такие звезды вырождаются в нейтронные звезды или в черные дыры, или, взрываясь, инжектируют свою материю снова в космическое пространство. Вследствие происходящего при этом интенсивного нейтронного излучения происходит образование ядер тяжелых элементов — «по ту сторону» железа. Некоторые из возникающих таким образом газовых облаков, содержащих большинство элементов периодической системы, конденсируются в газовые шары второго поколения, плотность которых все более возрастает.

Один из таких газовых шаров с массой, примерно вдвое превышающей массу Солнца, по-видимому, стал исходным пунктом образования нашей планетной системы с центральным телом — Солнцем. Для объяснения происхождения солнечной системы со времен Канта (1775) и Лапласа (1796) привлекались усилия лучших теоретиков. Число предложенных гипотез на сегодня очень велико, и ни одна из них не может считаться доказанной (Unsöld, 1975; Dorschner, 1976; Амбарцумян и др., 1986; Альфен, Аррениус 1987; Жарков, 1983).

Рассмотрим лишь один из вариантов этих гипотез: в результате коллапса межзвездного газового облака образуется газовый шар размером с солнечную систему и массой, равной примерно двум солнечным массам. Вследствие все усиливающегося сжатия это протосолнце вращается все быстрее, принимает форму эллипсоида, и на экваторе происходит отрыв массы. Истекающий ионизированный газ остается сначала еще связанным магнитным полем протосолнца и уносит часть момента импульса полной системы. Эта магнитная связь, теоретически до конца еще не объясненная, тормозит возникающее центральное Солнце и ускоряет истекающий газ, в результате чего образуется плоский вращающийся диск. Возможно, что

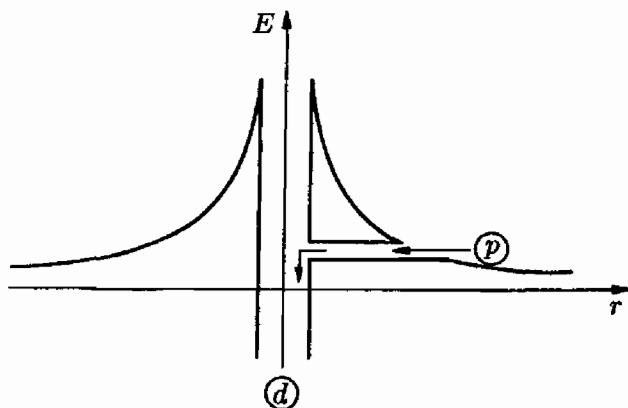


Рис. 2.3. Туннельный эффект по Гамову: протон туннелирует сквозь потенциальную стенку дейтрона и образует с ним ядро гелия

на распределении углового момента сказываются радиальные конвективные токи. В результате кластеризации твердых частиц дискообразной туманности сначала, по-видимому, образуются сгустки диаметром от десяти сантиметров до метра — так называемые планетезимали. С возрастанием массы гравитационное поле таких зародышей планет приводит к захвату все большей массы туманности и, наконец, к образованию планет. Остаток туманности солнечный ветер развеивает по окружающему пространству. Примерно так может происходить формирование планетной системы. Однако многие пункты нуждаются в пояснении. Основополагающая идея естественной эволюции солнечной системы восходит к Иммануилу Канту (1775), ставшему пионером теории эволюции нового времени. Идеи Канта имели фундаментальное значение для дальнейшего развития теории эволюции.

11-я эпоха (образование фотонной мельницы).

Решающее значение в эту эпоху обретает термодинамический механизм, своего рода двигатель дальнейшей эволюции, — так называемая фотонная мельница. Как мы уже знаем, судьба фотонов в ходе первых восьми эпох сравнительно несложна, они охлаждаются при все прогрессирующем расширении по закону (2.19) до тех пор, пока не достигают своего современного состояния. Ныне излучение образует море фотонов, соответствующих $T = 3$ К, которые однородно заполняют Вселенную с плотностью около 500 фотонов на см^3 . Более сложна судьба материи, имеющейся в виде вещества. В отличие от фотонов, частицы вещества имеют отличную от нуля массу покоя. Тем самым они в большей мере, чем фотоны, обладающие только массой движения, подвержены действию гравитационных сил. Если температура еще достаточно высока, то есть в начальной фазе расширения, то неупорядоченное тепловое движение настолько интенсивно, что никакие неоднородности, зародыши структурообразования, возникнуть не могут. С уменьшением температуры действие гравитации становится все сильнее. Гравитация усиливает случайные флуктуации плотности, образуются сгустки материи в форме вещества. Разумеется, такой процесс становится возможным лишь после того, как температуры опустятся существенно ниже тысячи градусов. Аналогичный эффект наблюдается в газах с притягивающими взаимодействиями — так называемых вандерваальсовских газах. При высоких температурах такие газы находятся в однородных состояниях. Если же их охлаждать, подвергая адиабатическому расширению, то молекулы газа «слипаются» в капли. В такого рода системах должны наблюдаться эффекты конденсации. Можно представить себе, что существующие ныне галактические структуры возникают из гигантских сгустков масс в более ранней Метагалактике. В то время как галактики образуются из огромных сгустков газа, внутри галактик также происходит структурообразование. Шарообразные неоднородности сжимаются все сильнее под действием притягивающих гравитационных сил. Вследствие адиабатического сжатия температура повышается с увеличением плотности по закону

$$T \sim n^{2/3}.$$

Достаточно большие газовые шары вновь достигают в результате гравитационного коллапса температур в несколько миллионов градусов, начинается синтез ядер, рождаются звезды. Эти процессы были и остаются поныне весьма сложными, и мы не будем обсуждать их подробно. Для нас важно лишь, что из бесструктурного однородного распределения материи в виде вещества за миллиарды лет возникли такие сложные структуры и суперструктуры, как красные гиганты, солнца, белые карлики, планеты, галактики и галактические субструктуры. Типичными представителями этих образований являются наше Солнце с температурой поверхности около 6 000 К и наша Земля с температурой поверхности около 300 К.

Тем самым возникает следующая структура нашей Вселенной: относительно плотные и вновь разогревшиеся сгустки материи в форме вещества (звезды и планеты) погружены в море излучения (фотоны с энергией, соответствующей температуре 3 К). Именно этот механизм приводит в движение мельницу самоорганизации и эволюции на планетах.

В принципе фотонная мельница действует так же, как любая обычная водяная мельница или тепловая машина. Действие водяной мельницы основано на том, что вода с более высокого уровня потенциальной энергии падает на более низкий уровень. При падении она задевает мельничное колесо, которое забирает часть разности энергий и превращает ее в полезную энергию вращательного движения. Тепловые машины, впервые исследованные Карно и Клаузиусом, работают на основе аналогичного принципа. Они состоят из теплового резервуара при более высокой температуре, например, парового котла, и теплового резервуара при более низкой температуре, например, реки, моря или окружающего воздуха. В этом случае действие тепловой машины основано на том, что часть тепла, перетекающего из верхнего резервуара в нижний, заимствуется для совершения полезной механической работы. Термодинамическая теория таких тепловых машин ныне хорошо разработана. Клаузиусу мы обязаны знанием того, что важную роль играет не поток энергии, а поток величины, получившей название энтропии. Если тело получает при температуре T количество тепла ΔQ , то, по Клаузиусу, это соответствует приращению энтропии

$$\Delta S = \frac{\Delta Q}{T}. \quad (2.32)$$

Тепловая машина забирает тепло при более высокой температуре и отдает тепло при более низкой температуре, поэтому по формуле (2.32) тепловая машина отдает энтропии больше, чем получает. Это позволяет говорить об экспорте энтропии. Бельгийский исследователь Пригожин (лауреат Нобелевской премии по химии 1977 г.) показал, что экспорт энтропии имеет основополагающее значение не только для тепловых машин, но и для любых систем, обладающих способностью к самоорганизации. Рассмотрим с этой точки зрения «фотонную мельницу Земли» (рис. 2.4). Эта система принимает поток тепла около 10^{17} Вт от Солнца, температура которого достигает 5 800 К. Эта величина соответствует эффективной температуре поверхности Солнца. Земля излучает во вне равное количество тепла, 10^{17} Вт, но при температуре всего лишь 260 К. Такова температура теплового излучения планеты Земля, которую наблюдатель мог бы измерить на большой высоте. Как показывают несложные вычисления, экспорт энергии в единицу времени составляет величину

$$\frac{\text{Поглощенное излучение}}{\text{Температура}} - \frac{\text{Испущенное излучение}}{\text{Температура}} \approx \\ \approx 10^{17} \text{ Вт} \left(\frac{1}{5\,800 \text{ К}} - \frac{1}{260 \text{ К}} \right) \approx -4 \cdot 10^{14} \text{ Вт/К}.$$

Средний экспорт энтропии на квадратный метр земной поверхности составляет величину около $1 \text{ Вт/м}^2 \text{ К}$.

Это замечательное число отражает основополагающее свойство экспорта энтропии Земли. Суть дела можно сформулировать еще проще: планета Земля представляет собой своего рода «фотонную мельницу», действующую в данном случае между горячими фотонами солнечного излучения и холодными фотонами, излучаемыми земной поверхностью. Число $1 \text{ Вт/м}^2 \text{ К}$ характеризует термодинамическую интенсивность силы, приводящей в движение фотонную мельницу, которая, в свою очередь, движет

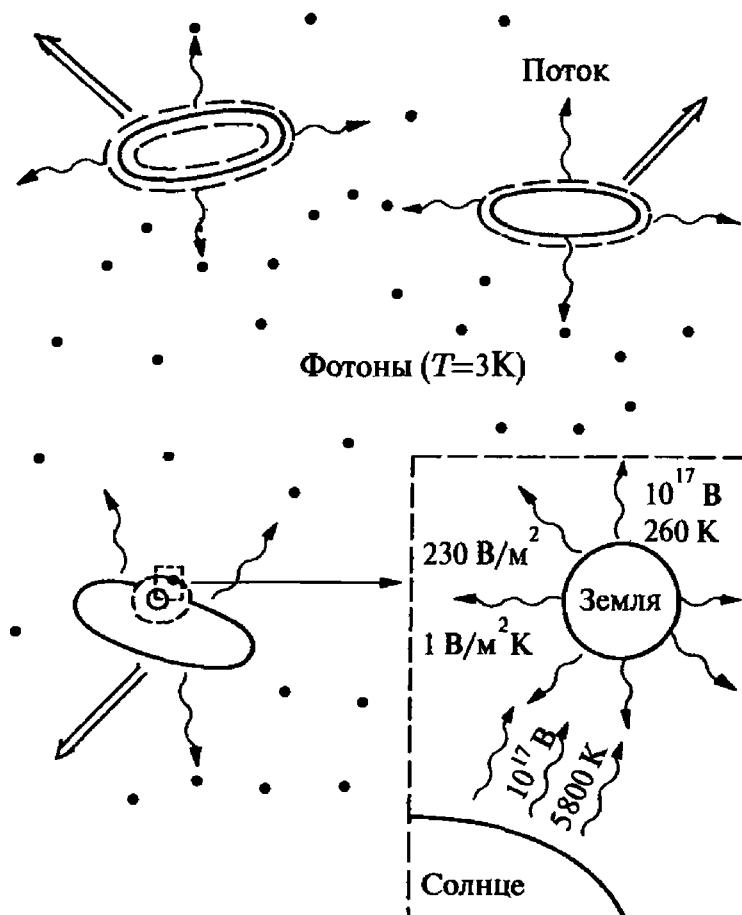


Рис. 2.4. Схематическое изображение механизма фотонной мельницы в расширяющейся Метагалактике

процессами самоорганизации и эволюции на Земле. Возникновение механизма фотонной мельницы в 11-ю эпоху эволюции служит предпосылкой появления особых планет, например, нашей Земли, в 12-ю эпоху, охватывающую биогенез, а также эволюцию экосистемы и общественной системы.

2.4. Возникновение жизни

Вплоть до первой половины XIX столетия в науке господствовала точка зрения, согласно которой жизнь представляет собой нечто особенное, противостоящее физико-химическим закономерностям. Эта точка зрения нашла свое выражение в принятии гипотезы «*vis vitalis*» — особой «жизненной силы». Революционное свершение таких выдающихся исследователей, как Майер, Гельмгольц, Больцман и Оствальд, состоит в том, что им удалось показать не только отсутствие какого-либо противостояния между сущностью жизни и физико-химическими закономерностями, но и следование этим закономерностям по всем пунктам. Разумеется, это не означает, что существование жизни сводится к физике и химии. Было бы необычайной дерзостью пытаться объяснить, например, все многообразие существующих ныне видов животных (более 1,2 миллиона), исходя из одних лишь физико-химических законов. Можно понимать жизнь по Дарвину лишь как результат длящегося несколько миллиардов лет процесса эволюции, а все науки от физики и химии до биологии и палеонтологии относить к философии. Так как к явлениям жизни на нашей Земле принято относить столь сложные явления, как общество, музыка, живопись, литература и т. д., мы можем рассматривать жизнь только как предмет

изучения биологии. Мы придерживаемся эволюционного и междисциплинарного подхода к жизни. Что же такое жизнь?

Мы хотим пояснить этот вопрос (но отнюдь не ответить на него) с помощью мысленного эксперимента. Представим себе, что мы путешествуем по Вселенной на космическом корабле и встречаем на своем пути различные формы существования материи. Какие из них можно считать живыми существами? Ясно, что не метеоры, газовые облака, камни. Но представим себе, что нам встретила информационная система, т. е. система, занимающаяся обменом и обработкой информации. Естественно, мы бы спросили, возникла ли такая система сама по себе или была создана. В первом случае мы могли бы говорить о живом существе совершенно независимо от физико-химической природы системы — так же, как говорит о «живом Океане» Станислав Лем в своем романе «Солярис». Так мы приходим к следующему рабочему определению: «Живые существа — естественные информационные системы, т. е. системы, возникающие сами собой, а не в результате построения или составленной кем-то программы, от информационной системы более высокого уровня в ходе естественной эволюции». Мы видим, что в этом рабочем определении понятия эволюции и информации играют центральную роль. Еще одно центральное понятие — понятие самовоспроизведения, или саморепродукции, — нам не хотелось бы включать в определение, поскольку, как будет показано в дальнейшем, самовоспроизведение является решающим механизмом эволюции для обработки информации.

Перейдем теперь к вопросу о том, как возникает жизнь. Среди бесчисленных солнечных систем нашей Метагалактики находится одна весьма своеобразная с семью крупными планетами, среди которых находится планета Земля, с твердой поверхностью, средней плотностью $5,51 \text{ г} \cdot \text{см}^{-3}$, отстоящая на расстоянии $1,5 \cdot 10^{11} \text{ м}$ от центральной звезды — Солнца, масса которой составляет около $2 \cdot 10^{30} \text{ кг}$. Образование этой планеты Земля произошло около 4,5 эонов (миллиардов лет) назад. Космическая эволюция шла дальше, рождались и погибали новые звезды, звездные системы и солнечные системы. Звезды, эволюционируя, превращались в красные гиганты, белые карлики, нейтронные звезды или черные дыры. Мы не будем проследивать их эволюцию более подробно, так как для нас основной интерес представляет самоорганизация и эволюция материи, происшедшие 4,5 эона назад на выделенной модельной планете Земля. Предшествующая космическая эволюция явилась для разворачивающейся теперь драмы не более, чем прологом. Мы сознательно изложили этот пролог лишь в общих чертах, чтобы его действие могло разворачиваться и так, и эдак. Неопределенные расплывчатые места по ходу действия пролога связаны главным образом с открытой проблемой гравитационных неустойчивостей турбулентных газовых облаков. Унзольд (*Unsöld*, 1975) говорит по этому поводу следующее: «Большие нерешенные проблемы в астрофизике сегодня имеют по существу гидродинамическую природу. Это связано с тем, что большинство космических течений турбулентны». Интересно отметить, что успехи астрофизики тесно связаны с решением старых проблем классической теоретической физики. В силу предистории своего развития наша планета окружена газовой оболочкой. Земные горные породы выделяют газы, водород и гелий улетучиваются в окружающее пространство, поэтому состав газовой оболочки Земли непрерывно изменяется. На протяжении двух эонов земная атмосфера состояла главным образом из метана, водяного пара, аммиака, водорода и сероводорода. Такая газовая оболочка на первый взгляд может показаться «негостеприимной». Но, как выяснилось, такая газовая смесь таила в себе большие эволюционные потенции. Газовая смесь подвергалась действию интенсивного солнечного излучения. В бурной атмосфере бушевали грозы, действующие вулканы создавали локальные очаги высоких температур и значительные градиенты. Некоторые газы, в частности водород, конденсировались и скапли-

вались в низинах, образуя зародыши протоокеана и конкурируя с потоками лавы, изливавшимися из вулканов. В этих условиях, как следует из гипотезы, впервые выдвинутой Опариным (Опарин, 1924, 1960), возникли первые, еще примитивные, органические молекулы, в частности аминокислоты и нуклеиновые кислоты. Ныне эти процессы можно наблюдать экспериментально. В 1953 г. Миллер, тогда еще студент, пропуская в лаборатории Ури день за днем электрические заряды сквозь газовую смесь, моделировавшую протоатмосферу Земли, обнаружил в образовавшейся смеси аминокислоты. Работы Миллера и Ури породили огромное число аналогичных исследований. Было установлено, что при подводе энергии (например, в форме электрических разрядов, ультрафиолетового излучения, радиоактивного излучения и тепла) к газовой смеси, содержащей углерод, водород, кислород и азот, в восстановительной среде образуются все важнейшие детали для построения биовеществ: аминокислоты, гидроокиси, сахара, пуриновые и пиримидиновые основания (Kaplan, 1978; Ponnamperuma, 1972; Fox, Dose, 1972; Miller, Orgel, 1973; Körner, 1974, 1977, 1979, Амбарцумян и др., 1986).

С инициацией химических процессов на планете Земля началась фаза химической эволюции. Она началась около 4–4,5 эонов назад и длилась около 1 эона. Основным результатом первой стадии химической эволюции стала интеграция простых атомов H, C, N, P, ... в относительно сложные органические молекулы. Следующий эволюционный процесс состоит в интеграции этих молекул в макромолекулы. Некоторые из возникших в протоатмосфере органических молекул соединяются в цепи: полипептиды и полинуклеотиды образуются в постоянной конкуренции с разрушающим цепи процессом гидролиза. Множество полимерных цепей состоит из мономеров не одного, а нескольких сортов, т. е. являются самыми настоящими гетерополимерами. Тем самым возникает определенный линейный молекулярный «узор» — эталон, или паттерн. С момента своего образования паттерн начинает играть основополагающую роль (Bresch, 1977). Химическая эволюция по существу завершается. Ее основной результат состоит в интеграции атомов химических элементов во многие сложные органические молекулы, а молекул — во многие еще более сложные цепные молекулы. Важную роль в этих превращениях играли следующие химические элементарные процессы: гомогенный и гетерогенный катализ, автокатализ, бистабильность и колебания.

Рассмотрим следующую фазу эволюции: через некоторое время возникают особые органические цепные молекулы, наделенные способностью к самовоспроизведению или саморепродуцированию. В случае полинуклеотидов это свойство основано на том, что водородные связи между нуклеотидами имеют различную силу. Молекулу рибонуклеиновой кислоты мы представляем в виде линейной цепи, состоящей из «кирпичиков» (звеньев) четырех сортов: аденина, цитозина, гуанина и урацила. Нуклеотид аденин обладает способностью образовывать относительно прочную водородную связь с нуклеотидом урацилом, а нуклеотид цитозин — с нуклеотидом гуанином. Вследствие этого исходная цепь нуклеотидов может присоединять к своим звеньям дополнительные (в указанном выше смысле) звенья. В результате образования химических связей между ассоциируемыми нуклеотидами может возникнуть дополнительная цепь (Kaplan, 1978). Таким образом, относительно короткие цепи рибонуклеиновой кислоты могут инициировать образование дополнительных цепей, а те в свою очередь — инициировать образование исходного прототипа. Тем самым за конечное число шагов происходит удвоение молекул-прототипов. Впрочем, из-за недостаточной каталитической поддержки и весьма слабых концентраций спонтанная репликация молекул происходила в водах протоокеана чрезвычайно редко. Все переплетение этих трудностей переносится и на полипептиды. Ныне экспериментально доказано, что еще первичные абиогенно возникшие полипептиды (так

называемые протеноиды) вследствие своей пространственной структуры — складок цепи — обладали разнообразными каталитическими свойствами (*Kaplan, 1978; Segal, Körner, Leiterer, 1982*). В силу этого спонтанно возникавшие в достаточных количествах протеноиды могли решающим образом ускорить протекание процессов синтеза и репликации. Вещество обрело тем самым важнейшее свойство самовоспроизведения и вступило в новую фазу эволюции — фазу самоорганизации через самовоспроизведение. При этом решающую роль стали играть образование модификаций (мутантов) исходных молекул при небольших ошибках копирования и последующий отбор благоприятных вариантов. Возможно, появились даже целые сети полинуклеотидов и полипептидов, обладавшие способностью к репликации. Эти гипотетические структуры получили название гиперциклов (*Айген, Шустер, 1982*) и циклов рибонуклеиновые кислоты—репликаза (*Ebeling, Feistel, 1979*).

Решающим процессом, приводившим к локальному повышению концентраций макромолекул, по-видимому, было образование коацерватов (*Опарин, 1963; Segal, Körner, Leiterer, 1982*). Коацерватами называются каплевидные сгустки. Они возникают в тех случаях, когда в растворах высокомолекулярных соединений сольватные оболочки («шубы») ионизированных макромолекул отчасти состоят из других заряженных молекул и сливаются. Коацервация является одной из форм процесса, противоположного перемешиванию: коацерватная фаза содержит макромолекулы в гораздо более высокой концентрации, чем остальная жидкая фаза. *Опарин (1960)* придавал огромное значение коацерватам как одной из возможных ступеней, предшествующих возникновению жизни. С коацерватами тесно связаны исследованные *Фоксом* микрошарики. Речь идет о крохотных шариках протеноидов диаметром до 3 мкм (*Fox, Dose, 1972*). В коацерватах и микрошариках могли протекать реакции, поэтому их в некотором смысле можно рассматривать как химические микрореакторы. Эти микрореакторы обладают способностью к росту и к делению, а также к обмену веществ с окружающей средой и тем самым к мутации своего состава. Наряду с репликативными сетями коацерваты выступают как новые репликативные единицы (*Ebeling, Feistel, 1982; Feistel, 1983*). Образование микрореакторов является важным шагом на пути к пространственному ограничению биологических особей. Такая индивидуализация несомненно представляется решающим скачком в биогенезе.

Другим шагом, имевшим особенно большое значение, стало образование молекулярного языка биополимеров. Что, собственно, следует понимать в данном случае, говоря о языке? Рассмотрим для простоты немецкий язык (письменную речь). Формально немецкая письменная речь состоит из последовательностей символов, насчитывающих всего 35 знаков: 26 букв, пробел и знаки препинания. Эти символы упорядочены в различные последовательности. Пробелами, запятыми, точками, отступлениями, нумерацией глав и т. д. последовательности символов делятся на слова, части предложения, предложения, абзацы, главы и т. д. Этим подпоследовательностям символов придается различный смысл. Порядок следования слов отнюдь не произволен, а подчиняется правилам грамматики. Формальному описанию грамматических структур мы обязаны выдающемуся языковеду *Хомскому (Хомский, 1962)*. Нетрудно видеть, что понятия языкознания применимы к описанию биополимеров (*Ratner, 1974; Ратнер, 1983, 1986; Eigen, Winkler, 1975; Ebeling, Jiménez-Montaño, 1980; Ebeling, Feistel, Herzel, 1987; Küppers, 1986; Николис, Пригожин, 1990*). «Элементарный язык биологической системы — это химический язык» (*Scheler, Mohr, 1979*). Этот химический язык имеет алфавит, состоящий из различных сортов нуклеотидов и аминокислот; он обладает грамматическими правилами и позволяет выстраивать последовательности символов различной длины — единицы мутации, кодирования и рекомбинации (цистроны и более длинные участки). В биологическом отношении такие подпоследовательности символов имеют строго определенное значение, например,

они кодируют определенные метки или ферменты. Элементы этого молекулярного языка должны были образоваться в протоокеане. Предпосылкой тому было существование носителей информации, наделенных способностью к репликации. В протоокеане эту роль исполняли полинуклеотиды — сначала РНК, а впоследствии ДНК.

Эффективная репликация осуществлялась в связи с сетями с помощью специальных ферментов — репликаз. Возникали все более сложные репликативные системы, конкурировавшие друг с другом. Одни репликативные системы, обеспечивавшие быстрое и надежное (с относительно малым числом ошибок) копирование прототипа, оказывались впереди, другие отставали. В качестве «сырья» служили репликативные системы всюду встречающихся простейших органических молекул, таких, как аминокислоты и «заготовки» нуклеотидов, образующиеся под действием солнечного света, ударов молний и извержений вулканов. Со временем протоокеан все более заполнялся быстро копирующими репликативными системами, и сырьевой ресурс все более истощался, обмен веществ при репликации становился все более ограниченным. Происходил жесткий естественный отбор, определенные репликативные единицы, работавшие на основе полинуклеотидов и полипептидов, усвершенствовались. Они пространственно изолировались от окружающей среды (возможно, путем стабилизации микрореакторов) и отделились от внешнего мира мембранами, через которые происходил обмен веществ с окружающей средой: произошло рождение первых живых существ.

Можно предположить, что первая микрофлора и первая микрофауна появилась 3,3–4 эона назад. Первыми микроорганизмами могли быть бактерии или примитивные водоросли. К числу первых микроорганизмов относятся также прототипы одноклеточных с ядром (эукариоты). Древнейшие ископаемые с внутриклеточным ядром обнаружены в Северной Австралии. Их возраст достигает примерно 1,5 эона. Древнейшие эукариоты размножались неполовым путем митоза (простого деления клеток). Лишь через некоторое время появились микроорганизмы, обладающие способностью к половому размножению. Это было значительным шагом в эволюции, поскольку половое размножение значительно расширило генетическую вариабельность и внесло ощутимый вклад в стабилизацию типов (Körner, 1979). С усовершенствованием репликации сырьевой ресурс жизни — образовавшийся чисто химическим путем, т. е. вне живого организма, аминокислоты и прототипы нуклеотидов, — резко сокращается. Именно в тот период возникает новый класс микроорганизмов, производящих молекулярные структурные единицы («кирпичики») из воды и молекул газа под действием солнечного света. Такие представители древней микрофлоры размножались необычайно быстро. Из неассимилирующих организмов шанс на выживание имели лишь паразиты на протофлоре. Так появился принцип гетеротрофии. Под гетеротрофом принято понимать любой организм (травоядный, плотоядный или всеядный), который питается своей добычей — другими организмами. Основанием возникшей трофической цепи служили автотрофные растения. Они медленно изменяли состав атмосферы. Аммиак превращался в азот и водород, двуокись углерода — в углерод и кислород, метан — в двуокись углерода и водород. Для построения своих тканей растениям были необходимы прежде всего молекулы углерода и водорода. Азот и кислород растения в огромных количествах выделяли в атмосферу.

Возникновение содержащей кислород атмосферы, начавшееся около 2 эонов назад, глубоко изменило условия существования жизни. Для живых существ той далекой эпохи кислород был высокотоксичным газом, который в результате процесса окисления мог привести к разрушению органических молекул. Мутация и отбор помогли преодолеть и эту смертельную угрозу: возникли живые организмы, снабженные сначала примитивными органами, а впоследствии жабрами и легкими, которые

развили высокоэффективные механизмы обмена веществ для атмосферы, содержащей кислород. Одной из важнейших предпосылок возникновения высокоразвитых форм жизни стало образование колоний клеток путем скопления клеток с ядрами (эукариотов) и распределения функций между ними. Возникновение примерно 0,6 зона назад многоклеточных эукариотов привело к взрывообразному увеличению числа высокоразвитых форм жизни (*Schopf*, 1976). Как свидетельствуют ископаемые относящиеся к той эпохе (началу так называемого кембрия), в течение сравнительно короткого геологического периода появились многие виды беспозвоночных и макроскопические водоросли. Для того чтобы произошел этот эволюционный скачок, понадобилось три шага:

- 1) развитие полового размножения;
- 2) открытие принципа гетеротрофии;
- 3) образование колоний клеток с распределением функций.

Дальнейший ход событий, если подходить к ним с точки зрения эволюции, разворачивался на протяжении относительно короткого времени. Чтобы хотя бы бегло обозреть важнейшие даты, мы воспользуемся одной идеей Сагана (*Sagan*, 1978) и рассмотрим «календарь эволюции» с сильно сжатой шкалой времени. Писатель Александр Полещук в своем утопическом романе «Тайна стеклянного конуса» повествует об экспериментах, в которых героям удалось существенно сократить пространственно-временные масштабы протекания реальных процессов. Вообразим себе, что у нас имеется «машина Полещука» и мы хотим с ее помощью исследовать более подробно в лаборатории процесс эволюции Вселенной и планеты Земля. В своем первом гипотетическом эксперименте мы выберем коэффициент сокращения временных масштабов $1 : 2 \cdot 10^{10}$. Тогда наша машина времени проектировала бы на один год отрезок времени в 20 миллиардов лет, и вся новая эра от начала уместилась бы примерно в 3 секунды. Для простоты мы совместим начала отсчета времени для реальной и гипотетической эволюции. Тогда эволюционный эксперимент начался бы в январе, феврале или марте года, предшествующего Большому Взрыву, а в первые 3 секунды нового года уложилась бы вся история с начала новой эры. Результат такого «эксперимента» представлен в табл. 2.1 и 2.2.

Первый эксперимент обнаружил в процессе эволюции множество тонкостей. Чтобы разобраться в них, проведем второй гипотетический эксперимент продолжительностью около 20 лет с коэффициентом сжатия временной шкалы $1 : 10^9$. В новом эксперименте цейтраферная машина сокращает миллиард лет до одного года. Предположим, что наш второй эксперимент начался в 1970 г. Перелистаем лабораторные записи, представленные в табл. 2.3–2.5.

Более совершенная оптика позволяет нам рассмотреть множество деталей, временное упорядочение которых сопряжено с известными неопределенностями. Сильнейшее впечатление от нашего гипотетического эксперимента несомненно связано со значительным ускорением эволюционного процесса по мере продвижения по времени. Глубокую причину такого ускорения ныне усматривают в эволюционной обратной связи (*Николис, Пригожин*, 1979), которая заключается в воздействии механизмов эволюции на породившие ее факторы (*Riedl*, 1975).

Если принять за возможную меру величину интегрированных в целостные модули (паттерны) структур, то, по Брешу (*Bresch*, 1977), эволюция следует экспоненциальному закону, в котором время от времени происходят качественные скачки. В первой пребиологической фазе, охватывающей период от Большого Взрыва до возникновения протобионтов, в такие целостные модули (макромолекулы), интегрировалось до 10^5 нуклонов. По Брешу, величина паттерна удвоилась примерно за 500 миллионов лет. В следующей — биологической — фазе величина

паттерна удваивалась примерно каждые 50 миллионов лет, и в конце этого периода в целостный модуль (в форме млекопитающих) интегрировалось уже около 10^{29} нуклонов. Для следующей — интеллектуальной — фазы со значительно более высокой способностью к интеграции — до 10^{49} нуклонов в целостном модуле (что соответствует человечеству со всеми средствами производства) — время удвоения паттерна, по оценкам Бреша, составляет около 100 000 лет. Если принять для отдаленного будущего гипотезу об установлении космических связей (космическая фаза), то экстраполяция этих оценок сразу же приводит к «физическим стоп-сигналам», устанавливаемым конечностью скорости света и массы космоса, составляющей,

Таблица 2.1. Календарь докембрийской эпохи с сжатием временной шкалы в $1 : 2 \cdot 10^{10}$ раз

Новый год	Большой (горячий) Взрыв.
Июнь	Возникновение Млечного Пути — Галактики.
Сентябрь	Возникновение солнечной системы и образование планеты Земля.
Октябрь	Первые живые организмы; древнейшие из известных остаточных пород; древнейшие из известных ископаемых форм микроорганизмов.
Ноябрь	Распространение производящей кислород микрофлоры и микрофауны; развитие полового размножения; появление фотосинтезирующих растений; первые клетки с ядром (эукариоты), способные к митозу.
Новый год	Образование атмосферы с высоким содержанием кислорода; интенсивные извержения вулканов.
Декабрь	Формирование миотического-полового размножения.
Середина декабря	Развитие гетеротрофных одноклеточных; появление первых многоклеточных организмов, начало макроскопической жизни.

Таблица 2.2. Календарь кембрия и послекембрийской эпохи с сжатием временной шкалы в $1 : 2 \cdot 10^{10}$ раз

20.12.	Возникновение беспозвоночных.
21.12.	Первый океанический планктон; кровь трилобитов.
22.12.	Ордовик (период): первые позвоночные (рыбы).
23.12.	Силур (период): риниофиты заселяют сушу.
24.12.	Девон (период): первые насекомые; животные заселяют сушу; первые амфибии; летающие насекомые.
25.12.	Карбон (период): первые папоротники; первые пресмыкающиеся.
26.12.	Пермь (период): первые динозавры.
27.12.	Триас (период): первые млекопитающие.
28.12.	Юра (период): первые птицы.
29.12.	Мел (период): первые цветковые растения; вымирание динозавров.
30.12.	Третичный период: разнообразные млекопитающие; первые приматы; первые гоминиды.
31.12.	Возникновение проконсула и рамапитека — около 14.00.00 часов. Первые люди — около 22.30.00 часов. Использование каменных орудий — около 23.00.00 часов. Изобретение земледелия — около 23.59.00 часов. Первые города — около 23.59.30 часов. Изобретение письменности — около 23.59.54 часов. Бронзовый век — около 23.59.56 часов. Железный век — около 23.59.57 часов. Евклидова геометрия, архимедова физика — около 23.59.59 часов. Начало отсчета времени — 24.00.00 часов.
1.1.	(Новый год) Введение нуля и десятичной системы счисления — около 00.00.01 часов. Эпоха Возрождения и современная наука — около 00.00.02 часов. Современная эпоха — около 00.00.03 часов.

по оценкам, 10^{79} масс протона. К сожалению, даже наша машина Полешука не позволяет нам заглянуть в более далекое будущее. Правда, с некоторой вероятностью можно предположить, что ускорение интеллектуальной фазы нашей планеты приведет к установлению на Земле некоего стационарного состояния, характеризующегося максимальной степенью интеграции и кооперации (монон-состояние в смысле Бреша). Разумеется, все рассуждения о далеком будущем нашей Вселенной носят чисто умозрительный характер. Дайсон в своем интересном исследовании (Dyson, 1979) показал, что в открытой космологической модели фридмановского типа эволюция непременно должна завершаться установлением стационарной фазы.

Таблица 2.3. Календарь земной эволюции с сжатием временной шкалы в $1 : 10^9$ (даты — по работам *Krumbiegel, 1977; Sagan, 1978; Kaplan, 1978; Монин, Шишков, 1979*)

15–20 лет назад	Большой Взрыв (точная дата неизвестна).
10 лет назад	Возникновение нашей Галактики.
6 лет назад	Образование солнечной системы.
5 лет назад	Возникновение Земли
4 года назад	Возникновение первых форм жизни.
3 1/2 года назад	Образование древнейших из известных ископаемых: бактерий, синих водорослей (фотосинтез!).
2 года 9 месяцев назад	Начало архея (архейского эона); начало эпохи водорослей; появление свободного O_3 в атмосфере.
2 года 3 месяца назад	Развитие полового размножения.
2 года назад	Начало протерозоя; образование большого материка Гондваны.
1 год 10 месяцев назад	Увеличение распространенности одноклеточных организмов.
1 год 9 месяцев назад	Расхождение эволюционных путей прокариотов и эукариотов.
1 год 8 месяцев назад	Первые многоклеточные.
6 месяцев назад	Кембрий, господствующая фауна: морские беспозвоночные; область, где расположено современное Марокко, находилась у южного полюса.
5 месяцев назад	Ордовик: первые позвоночные.
4 3/4 месяцев назад	Силур: слияние микроконтинентов в большой континент Лавразию; образование Каледонской складчатости; Гренландия и Огненная Земля расположены на экваторе; первые коралловые рифы и первые рыбы; растения завоевывают сушу.
4 1/2 месяца	Девон: глобальное понижение температуры; Рио де Жанейро, расположенный у побережья Намибии, находится на южном полюсе. В флоре риниофиты уступают господство водорослям; в фауне из рыб (кистеперых) развиваются первые сухопутные позвоночные (амфибии). Появление первых насекомых. Животные завоевывают сушу.
4 месяца назад	Карбон: дальнейшее глобальное понижение температуры; Лавразия и Гондвана сливаются в праcontinent Пангею; варискийская (герцинская) складчатость; Ботсвана на южном полюсе; Сибирь — вблизи северного полюса; каменноугольные леса; гигантские стрекозы; первые пресмыкающиеся.
3 1/2 месяца назад	Пермь: глобальный минимум температуры; Антарктика на южном полюсе; трилобиты вымирают; голосеменные становятся доминирующей флорой; первые динозавры; первые предшественники млекопитающих.
2 1/2 месяца назад	Триас: глобальное повышение температуры; распад Пангеи на 4 материка: — Евразию/Северную Америку, — Южную Америку/Африку, — Индию, — Антарктику/Австралию. Камчатка лежит у северного полюса; пресмыкающиеся нарушают господство амфибий.

Продолжение таблицы 2.3.

2 месяца назад	Юра: дальнейшее глобальное повышение температуры; период расцвета гигантских динозавров; первая птица — археоптерикс.
7 недель назад	Мел: глобальный максимум температуры; начало отделения Южной Америки от Африки и Северной Америки от Евразии; образование Кордильер; Африка, Индия и Антарктида/Австралия дрейфуют на север, Африка встречается с Евразией; образование альпийской складчатости; динозавры вымирают; покрытосеменные начинают доминировать над голосеменными.
Последние 3 недели	Третичный период: температура глобально убывает; Южная Америка отделяется от Африки; после северо-западной части северо-восточная часть Африки примыкает к евразийскому матерiku; образуется германское среднегорье; Австралия отделяется от Антарктики; начинается господство млекопитающих; первые приматы.
Вчера	Четвертичный период: дальнейшее повышение средней глобальной температуры; наступление «быстрых» колебаний температуры (ледниковых периодов); развитие человека.

Таблица 2.4. Календарь эволюции человека с сокращением временной шкалы в $1 : 10^9$ раз (даты по Фойстелю (*Feustel*, 1978))

7-11 дней назад	Расхождение путей эволюции понгид (предшественников орангутанов, горилл и шимпанзе) и гоминид (человекообразных).
5 дней назад	Появление рамапитеков — древнейших из известных гоминид; двуногое хождение; использование подходящих естественных предметов в качестве «орудий труда»; охота на небольших животных.
2 дня назад	Начало переходной фазы животное—человек; появление австралопитеков (предков человека) как прямых предшественников гоминид (человека); четко выраженное хождение на двух ногах; использование необработанных (естественных) или обработанных с помощью телесных органов предметов в качестве орудий труда; охота на небольших и крупных животных; собирательство; жизнь в стадах.
1 день назад	Завершение переходной фазы животное—человек; обработка естественных предметов (например, палок, камней) с помощью других предметов как индивидуальное, а затем группоспецифическое занятие (трудовая деятельность); начало общественного сознания; первобытное человеческое стадо.
20 часов назад	Появление габилинов (первых людей или человекообразных); древнейшие из известных каменных орудий (оббитых галек Олдована); первые человеческие гоминиды: <i>Homo habilis</i> (Человек умелый).
17 часов назад	Появление архантропов (первобытных людей); ручное зубило и простейшие орудия из оббитых камней; использование огня; первые хижины; охотничье-собирательское хозяйство; кооперация и распределение труда в первобытных группах; возникновение языка.
9 часов назад	Начало ледниковых периодов (похолодание и потепление следует каждые 30 минут); вымирание австралопитеков и габилинов; распространение гоминид; стандартизация орудий труда и оружия.
2 часа назад	Появление палеоантропов (древних людей); обработанные орудия труда из оббитых камней (наконечники, скребки, каменные ножи и т. д.); наконечники из каменных нуклеусов и отщепов; комбинированные орудия из камня и дерева; костяные орудия; добывание огня; жизнь в хижинах; человеческое стадо.
80 минут назад	Начало последнего ледникового периода; использование огня; господство более сложной техники; появление большого числа орудий разнообразного назначения; век неандертальцев.
30 минут назад	Поселения австралийцев; наконечники, гарпуны, добывание огня, погребения, культ теней, каннибализм.
8 минут назад	Окончание последнего ледникового периода.
6 минут назад	Массовое производство полуфабрикатов; многочисленные специальные инструменты; искусство; магия охоты и плодородия; появление современного человека; дифференциация рас; начало углубленной фазы распределения труда; ранний родовой строй; первые города.

Таблица 2.5. Календарь эволюции человеческого общества с сокращением временной шкалы в $1 : 10^9$ раз

5 минут назад	Неолитическая цивилизация; первые династии в Шумере, Эбле и Египте.
2 1/2 минуты назад	Возникновение астрономии; первое классовое (рабовладельческое) общество.
2 1/4 минуты назад	Аккадская империя; изобретение алфавита.
2 минуты назад	Среднее царство в Египте.
1 3/4 минуты назад	Троянская война; бронзовый век.
1 1/2 минуты назад	Закладка Карфагена; железный век.
1 минута назад	Начало новой эры.
45 секунд назад	Гибель Римской империи; введение десятичной системы счисления с нулем; расцвет феодализма.
35 секунд назад	Империя Карла Великого, короля франков.
28 секунд назад	Возникновение готической архитектуры.
15 секунд назад	Открытие Колумбом Америки.
12 секунд назад	Основание Галилеем современной физики.
10 секунд назад	Первое капиталистическое государство.
2 секунды назад	Первое социалистическое государство; взрывообразный расцвет науки и техники.
1 секунда назад	Конец второй мировой войны.

Глава 3

Термодинамические условия эволюции

Без языка математики большая часть глубоких взаимосвязей между вещами навсегда осталась бы неизвестной.

Анри Пуанкаре

3.1. Энергия и энтропия

Над термодинамическими условиями эволюции немало размышляли еще основатели термодинамики Майер, Гельмгольц, Клаузиус и Больцман. Их идеи, получившие дальнейшее развитие в трудах Оствальда, Шрёдингера, Бергаланффи и Пригожина, составляют ныне центральное ядро физики процессов эволюции. Как показывает рассмотренный выше сценарий процессов эволюции в Метагалактике и на нашей планете, основное участие в эволюции принимают тепловые потоки, потоки вещества и потоки излучения. Обратимся теперь к физическому анализу взаимодействия этих потоков, допускающего количественное описание в рамках термодинамики.

Что, собственно, мы понимаем под теплотой, энергией и энтропией?

Гуляем ли мы зимой по морозцу или нежимся под солнечными лучами на пляже, «тепло» и «холод» всегда и вполне ощутимы. Поскольку они имеют решающее значение для нашего биологического существования, природа наделила нас специальными сенсорами для восприятия тепла; тепло и холод относятся к нашим непосредственным ощущениям.

Именно поэтому тепло еще в глубокой древности вошло в число наиболее важных составных частей окружающего нас мира: вспомним элемент «огонь» у греков и «теплород», или «флогистон», у алхимиков. И хотя ныне мы точно знаем, что теплота не является веществом в собственном смысле, но даже в современной физике тепло принято рассматривать как «материю весьма особенного свойства». Поэтому прежде чем мы приступим к изложению теории эволюции, уместно предварительно рассмотреть несколько более подробно классическую теорию теплоты — термодинамику.

Поскольку с изобретением паровой машины в истории человечества началась новая — индустриальная, или промышленная, — эпоха, вопрос о природе и свойствах тепла стал одной из насущнейших практических проблем, и в прошлом столетии основания термодинамики были разработаны трудами Майера, Джоуля, Гельмгольца, Клаузиуса, Карно, Томсона, Максвелла и других. Теплоту они представляли как совокупность микроскопических неупорядоченных движений мельчайших составных частей вещества — атомов и молекул. Совместными усилиями удалось сформулировать два фундаментальных и универсальных утверждения — так называемых начал термодинамики.

Первое начало. Теплота — одна из форм энергии. Она может превращаться в другие формы энергии (энергию механического движения, электрическую энергию,

химическую энергию и т. д.), и другие формы энергии могут превращаться в нее. Энергия не может возникать или уничтожаться.

Второе начало. В замкнутой системе определенная физическая величина, характеризующая значением энергии и называемая энтропией, всегда возрастает или остается постоянной, но никогда не убывает.

Вплоть до последнего времени проводились бесчисленные эксперименты с целью доказать, что в особых условиях начала термодинамики могут нарушаться, но исход их неизменно оказывался отрицательным. Отсюда можно сделать вывод о том, что в случае начал термодинамики речь идет о физических законах, которые могут быть экспериментально лишь подтверждены. В то время как первое начало термодинамики гладко, «без швов», стыкуется со всем остальным зданием физики и утверждение о сохранении энергии может быть выведено из основных законов механики и электродинамики, второе начало термодинамики занимает особое положение. Точнее говоря, второе начало выходит даже за рамки «нормальной» физики.

Чтобы понять, почему так происходит, необходимо несколько точнее проанализировать понятие энтропии. Энтропия — не только мера, позволяющая оценивать значение энергии, но и количественная мера упорядочения тела или системы. Чем меньше энтропия, тем больше упорядочения в положении и движении составных частей тела. Рассмотрим в качестве примера автомашину. Предположим, что в одном случае она при температуре 20°C едет со скоростью 100 км/ч по шоссе, а во втором — при температуре 21°C стоит в гараже. Если отвлечься от вращения колес и движений частей двигателя, то положения атомов, образующих автомашину, в обоих случаях одинаковы. Как нетрудно показать с помощью вычислений, полная энергия в обоих случаях одинакова и, следовательно, одинакова доля энергии, приходящаяся в среднем на каждый отдельный атом автомашины. Вместе с тем значения энергии и, следовательно, значения энтропии заметно отличаются: в случае покоящейся автомашины все атомы движутся совершенно неупорядоченно (большая энтропия), в случае мчащейся автомашины наряду с нерегулярным тепловым движением имеется и изрядная доля «высококачественной» энергии направленного движения (меньшая энтропия). Если едущую автомашину затормозить до полной остановки, то в силу первого начала термодинамики полная энергия сохраняется, и автомашина в среднем нагревается на 1°C . Таким образом, состояние с меньшей энтропией переходит в состояние с более высокой энтропией, или энергия более высокого качества упорядоченного движения автомашины превращается в энергию более низкого качества теплового движения: энергия движения автомашины «диссипируется» (рассеивается).

Второе начало термодинамики позволяет утверждать, что обратный процесс физически невозможен, т. е. заторможенную до полной остановки автомашину невозможно снова разогнать до скорости 100 км/ч , просто охлаждая ее на 1°C и не расходуя ни капли бензина. Это, несомненно, наилучшим образом согласуется с нашим повседневным опытом. В чем же проблема?

Проблема возникла, когда физики попытались доказать закон возрастания (неубывания) энтропии, исходя из основных уравнений, полностью определяющих движение всех атомов, электронов и т. д. физических тел. Хотя правильность уравнений движения не вызвала ни малейших сомнений, все попытки такого рода оказывались безуспешными. Более того, Анри Пуанкаре даже удалось доказать, что невозможно вычислить энтропию, исходя из координат, скоростей, масс и др. характеристик всех участвующих в процессе частиц. Таким образом, теоретического доказательства правильности первого и второго начал термодинамики не существует и поныне, и остается по-прежнему неизвестным, существуют ли какие-то экстремальные условия, при которых начала становятся либо вообще, либо частично

неприменимыми. До сих пор такие условия удавалось воспроизвести лишь в некоторых экзотических экспериментах, например, в случае приготовления ядерных спинов, и поэтому мы должны исходить из того, что в нашей Метагалактике закон возрастания энтропии носит универсальный характер.

Сформулируем основной вывод еще раз, чтобы подчеркнуть главное: хотя все части рассматриваемых физических тел строго подчиняются известным законам механики, квантовой механики и электродинамики, существует величина, называемая энтропией, которая не сводится к свойствам составных частей тела, а описывает свойства всего тела в целом. Энтропия — один из важных примеров, подтверждающих правильность тезиса о том, что целое представляет собой нечто большее, чем сумму своих составных частей. Естественно напрашивается вопрос о том, существуют ли другие физические законы аналогичного типа, которые бы также не сводились к законам движения отдельных частиц. Наряду со столь простыми процессами, как нагревание тормозных колодок при остановке автомашины, в природе существуют гораздо более сложные процессы, например, эволюция Вселенной, образование геологических структур в земной коре, возникновение живых существ в нашей биосфере, социально-экономических структур, языков или даже искусства. Не кроются ли за всем разнообразием и сложностью этих процессов какие-то законы, достаточно универсальные и строгие для того, чтобы их можно было включить в число физических законов? Такова точная постановка вопроса, на который пытается в конечном счете найти ответ теория самоорганизации.

Но вернемся еще раз к энтропии. Имеется веское основание считать, что энтропия и энергия — физически различные величины. Действительно, все основные законы, описывающие движение частиц или электромагнитные поля, могут быть записаны в виде обратимых уравнений. Это означает, что каждое движение, каждый процесс в принципе могут протекать одинаково как в одну, так и в другую сторону — как вперед, так и назад. Например, если мы снимем на киноплёнку обращение спутника вокруг Земли и прокрутим фильм один раз «вперед», а другой раз «назад», то по тому, что мы увидим на экране, невозможно установить, какой из вариантов «правильный». Что же касается второго начала термодинамики, то оно задает направление времени: энтропия со временем не убывает, а может лишь возрастать или оставаться постоянной. Применительно к нашему примеру с фильмом правильным, или «настоящим», является то направление времени, в котором энтропия не возрастает. Это хорошо заметно хотя бы при просмотре видеозаписи футбольного матча: с первого же взгляда совершенно очевидно, в каком направлении в действительности протекала игра.

Принято говорить, что основные уравнения динамики частиц и полей «симметричны» относительно обращения времени, тогда как второе начало термодинамики несимметрично относительно обращения стрелы времени. Такое явление, как выбор одного из двух симметричных (в некотором смысле эквивалентных) вариантов и отбрасывание другого, принято называть нарушением симметрии. Мы часто будем встречаться с нарушениями симметрии в связи с самоорганизацией: нарушение симметрии по существу представляет собой основное явление в образовании структур.

3.2. Начала термодинамики и самоорганизация

Один из ключевых вопросов состоит в том, каким образом в природе вопреки царящей повсюду неизбежной диссипации энергии возникают высокоорганизованные структуры и беспрестанно могут образовываться все новые и новые структуры. Диссипация («обесценивание») энергии — фундаментальное свойство реального процесса, сформулированное во втором начале термодинамики как закон природы.

По Клаузиусу, справедливо следующее утверждение: энтропия в замкнутой системе непрерывно возрастает до тех пор, пока не достигает своего максимального значения в состоянии термодинамического равновесия.

Так как, по Больцману и Планку, между энтропией S и термодинамической вероятностью W существует соотношение

$$S = k_B \ln W \quad (3.1)$$

(k_B — постоянная Больцмана), увеличение энтропии соответствует переходу в более вероятные состояния, и последние более неупорядочены. Еще в прошлом столетии основатели термодинамики пришли к заключению, что порядок может возникнуть только при условии, если к системе подводится высокоценная энергия, а от системы отводится малоценная энергия. Майер и Больцман особенно подчеркивали роль Солнца как движущей силы структурообразования на нашей планете.

Основными законами термодинамики являются три начала термодинамики и фундаментальное уравнение Гиббса. Рассмотрим макроскопическую систему, характеризующуюся по доказанной теореме такими экстенсивными переменными, как энергия E , энтропия S , объем V и количества химических компонент $N_1, N_2, \dots, N_s$, а также такими интенсивными переменными, как температура T , давление p и т. д. В качестве единицы количества вещества мы используем в зависимости от целесообразности 1 частицу или $6,023 \cdot 10^{23}$ частиц. В общих формулах выбор единицы не играет никакой роли.

Приведем краткие формулировки трех начал термодинамики.

Первое начало. Энергия не рождается и не исчезает. Энергия — количественная величина, баланс которой может быть записан в следующем виде:

$$dE = d_e E + d_i E. \quad (3.2)$$

Иначе говоря, изменение количества энергии в системе равно сумме двух вкладов, учитывающих, соответственно, обмен энергией и производство энергии. Поскольку энергия не может ни рождаться, ни исчезать,

$$d_i E \doteq 0.$$

Что же касается члена, описывающего обмен энергией, то его можно представить в виде суммы совершенной работы и подведенного тепла:

$$d_e E = d' A + d' Q.$$

В случае квазистатического процесса обменный член можно представить в виде билинейной формы

$$d_e E = \sum l_i dL_i + \sum \mu_k d_e N_k + d' Q. \quad (3.3)$$

Здесь l_i и L_i — сопряженные переменные работы, например, $l_1 = -p$ и $L_1 = V$; μ_k — химический потенциал k -й компоненты; $d_e N_k$ — изменение числа частиц, вызванное обменом; $d' Q$ — часть потока энергии, не связанная с изменением координат работы или с обменом частицами. В реальных (необратимых) процессах необходимо учитывать еще один вклад — $d' A_{\text{необр}}$:

$$d_e E = \sum l_i dL_i + \sum \mu_k d_e N_k + d' A_{\text{необр}} + d' Q. \quad (3.4)$$

Здесь $d' A_{\text{необр}}$ — часть работы, связанная с необратимыми процессами, происходящими внутри системы, например, с трением.

Второе начало. Энтропия может только рождаться, но не уничтожаться. Энтропия — также количественная величина, баланс которой можно представить в виде

$$dS = d_e S + d_i S. \quad (3.5)$$

Так как энтропия не может уничтожаться, справедливо неравенство

$$d_i S \geq 0. \quad (3.6)$$

Обменный член, по Клаузиусу, прямо пропорционален подведенному теплу и обратно пропорционален температуре:

$$d_e S = \frac{d'Q}{T}. \quad (3.7)$$

Как показали Больцман и Планк, энтропия системы служит мерой царящего в ней беспорядка. По Оствальду, энтропию можно рассматривать и как количественную меру величины содержащейся в системе энергии.

Третье начало. Энтропия и энергия — ограниченные снизу переменные состояния, и при подходящем выборе начала отсчета можно считать, что

$$S \geq 0 \quad \text{и} \quad E \geq 0. \quad (3.8)$$

По Нернсту и Планку, энтропия системы, находящейся в равновесии относительно внутренних переменных, обращается в нуль при абсолютном нуле температуры вместе со своими производными по экстенсивным переменным:

$$S = 0 \quad \text{и} \quad \frac{\partial S}{\partial L_k} = 0 \quad \text{при} \quad T = 0. \quad (3.9)$$

Энтропия тела при конечной температуре всегда больше, чем при $T = 0$, откуда и следует, что $S > 0$ при $T > 0$. Положительность энергии гарантируется, если учесть энергию покоя $E_0 = m_0 c^2$ по Эйнштейну (m_0 — масса покоя).

Четвертым из важных для нас законов термодинамики принято считать фундаментальное уравнение Гиббса, устанавливающее взаимосвязь между экстенсивными переменными состояния E , S и L . Обоснование этого соотношения следует из первых двух начал термодинамики. Из соотношений (3.2)–(3.4) и (3.7) для квазистатического процесса следует, что

$$dE = \sum l_i dL_i + \sum \mu_k dN_k + T dS. \quad (3.10)$$

Это соотношение означает, что при квазистатическом переходе из состояния 1 в соседнее состояние 2 изменения энергии, энтропии и координат L_i связаны между собой линейным соотношением. Так как в соотношение (3.10) входят только переменные состояния, оно должно выполняться и для необратимых переходов из состояния 1 в состояние 2, поскольку значения переменных состояния не зависят от пути, по которому происходит переход. Тем самым предполагается, что фундаментальное уравнение Гиббса должно распространяться и на необратимые переходы при условии, что значения переменных состояния в начальном и в конечном состояниях вполне определены.

В дальнейшем мы наряду с энергией E будем использовать внутреннюю энергию U :

$$U = E - E_{\text{пот}} - E_{\text{кин}},$$

и часто предполагать, что $dE = dU$.

Итак, мы сформулировали в явном виде основные законы термодинамики, которые понадобятся нам в дальнейшем. Рассмотрим теперь различные частные случаи.

Начнем с (макроскопически) изолированной системы. Для нее имеем:

$$d_e S = 0, \quad dS = d_i S \geq 0. \quad (3.11)$$

Энтропия — монотонная неубывающая функция времени, асимптотически стремящаяся при больших временах к максимальному значению, соответствующему равновесным значениям термодинамических параметров (энергии, объема и т. д.). Следовательно, макроскопически изолированная система со временем приближается к состоянию с максимальным неупорядочением.

Рассмотрим теперь замкнутую систему. По определению, она не может обмениваться веществом (на макроскопическом уровне) с окружающей средой. Это позволяет нам записать соотношения:

$$dE = d_e E = \sum l_i dL_i + d' A_{\text{необр}} + d' Q, \quad (3.12)$$

$$d_e S = \frac{d' Q}{T}, \quad d_i S = \frac{d' A_{\text{необр}}}{T},$$

$$dS \geq \frac{d' Q}{T}. \quad (3.13)$$

Последнее неравенство (неравенство Клаузиуса—Карно) означает, что подвод тепла увеличивает беспорядок, в то время как увеличение порядка (уменьшение энтропии) требует отвода тепла $d' Q < 0$. Образование структур при отводе тепла (переходе к более низким температурам) — универсальное свойство материи. Каждому известно, что физические тела при низких температурах переходят в агрегатное состояние с большей степенью упорядоченности, например, в кристаллическое состояние. При низких температурах становятся реализуемыми относительно маловероятные состояния с малыми энергиями, соответствующие более высокой степени упорядочения. Это следует из распределения вероятностей Больцмана

$$p_i = C \exp \left\{ -\frac{E_i}{k_B T} \right\}, \quad (3.14)$$

делающее при низких температурах более предпочтительными состояния с малыми энергиями E_i . Другая возможность отвести энтропию из замкнутой системы состоит в том, чтобы подводить к системе тепло при более высокой температуре и отводить тепло при более низкой температуре. Именно таков механизм экспорта энтропии нашей Земли (см. разд. 3.5).

Величина $d' Q$ в соотношении (3.12) означает бесконечно малое количество тепла, подведенное к системе из окружающей среды; L_i — координаты работы экстенсивного характера, l_i — коэффициенты работы интенсивного характера. Наряду с обратимой работой система может также совершать и необратимую работу, например, при вязкости и проводимости. Например, при прохождении электрического тока силы I , индуцированного приложенным извне напряжением Φ , справедливо соотношение

$$d' A_{\text{необр}} = I \Phi dt. \quad (3.15)$$

Перейдем теперь к рассмотрению открытых систем. Определение работы задается соотношением (3.4). Хотя вычисление работы по действующим извне причинам, например, работы электрических сил, работы трения, работы расширения, работы

ускорения, может оказаться сложным, однако эта работа всегда вполне определена. Нетривиально и неоднозначно разделение между бесконечно малой работой и бесконечно малым количеством тепла (Хаазе, 1967). Рассмотрим систему с температурой T , химическими потенциалами μ_k и парциальными мольными энтропиями s_k компонент $k = 1, 2, \dots, s$. В дальнейшем мы будем предполагать, что изменение количества k -й компоненты состоит из вклада внутренних реакций и вклада обмена веществ:

$$dN_k = d_i N_k + d_e N_k. \quad (3.16)$$

Подставляя в фундаментальное уравнение Гиббса, получаем

$$\begin{aligned} d_e S + d_i S &= \frac{1}{T} d_e E - \frac{1}{T} \sum l_i dL_i - \frac{1}{T} \sum \mu_k (d_e N_k + d_i N_k) = \\ &= \frac{1}{T} d' Q + \frac{1}{T} d' A_{\text{необр}} - \frac{1}{T} \sum \mu_k d_i N_k. \end{aligned} \quad (3.17)$$

Если в правой и левой частях уравнения (3.17) сохранить только вклады внутренних (диссипативных) процессов, то с учетом соотношения (3.12) мы получаем

$$d_i S = \frac{1}{T} d' A_{\text{необр}} - \frac{1}{T} \sum \mu_k d_i N_k. \quad (3.18)$$

В химической литературе отдается предпочтение другому определению потока тепла (Prigogine, Defay, 1962):

$$dQ = d' Q + \sum \mu_k d_e N_k. \quad (3.19)$$

В этом случае получаем:

$$\begin{aligned} d_e S &= \frac{1}{T} dQ - \frac{1}{T} \sum \mu_k d_e N_k, \\ d_e E &= dQ + \sum l_i dL_i + d' A_{\text{необр}}. \end{aligned} \quad (3.20)$$

Еще одна возможность определения теплового потока связана со столь же непротиворечивым соотношением (Хаазе, 1967):

$$d^* Q = d' Q - T \sum s_k d_e N_k. \quad (3.21)$$

В этом случае получаем:

$$\begin{aligned} d_e S &= \frac{1}{T} d^* Q + \sum s_k d_e N_k, \\ d_e E &= d^* Q + \sum l_i dL_i + \sum h_k d_e N_k + d' A_{\text{необр}}, \end{aligned} \quad (3.22)$$

где парциальные мольные энтропии определяются соотношением $h_k = \mu_k + T s_k$.

В литературе поток тепла $d^* Q$, как правило, называется приведенным теплом (Prigogine, 1955, 1967; Хаазе, 1967). Это определение потока тепла в открытых системах обладает рядом принципиальных преимуществ, которые развили в своих работах Хаазе (Хаазе, 1967), а также Де Гроот и Мазур (De Groot, Mazur, 1962).

1. Представление приведенного потока тепла $d^* Q$ по формуле (3.22) показывает в явном виде, что обмен веществом сопряжен с обменом энтропией.

2. Если к однородной жидкой смеси добавить жидкость с теми же параметрами, то

$$d^* Q = 0.$$

Ни одно другое определение потока тепла не позволяет достичь этого свойства, которое представляется вполне разумным.

3. Приведенный поток тепла d^*Q инвариантен относительно сдвига нулевой точки внутренней энергии и относительно изменения относительной скорости (Хаазе, 1967).

4. Выражение для производства энтропии принимает при использовании приведенного теплового потока особенно симметричную форму (De Groot, Mazur, 1962).

Имея в виду перечисленные выше преимущества, мы будем в дальнейшем использовать приведенный тепловой поток и исходить из уравнений (3.22).

В открытых системах изменение энтропии dS не имеет определенного знака. В зависимости от теплообмена и обмена веществом изменение энтропии может быть и положительным, и отрицательным.

3.3. Глобальные условия самоорганизации

Энтропия является ключевой физической величиной при описании самоорганизации (Ebeling, Ulbricht, 1986). Она служит мерой ценности содержащейся в системе энергии и мерой беспорядка. Самоорганизация связана с ценными формами энергии и с упорядоченными состояниями системы. Энтропия системы может уменьшаться, если система экспортирует энтропию ($d_e S < 0$) и если экспорт в единицу времени превосходит соответствующее производство энтропии внутри системы, т. е.

$$dS < 0, \quad \text{если} \quad |d_e S| > d_i S \geq 0. \quad (3.23)$$

Такая ситуация мыслима только вдали от равновесия, поскольку вблизи равновесия всегда доминирует неравенство $d_i S > 0$. Для того чтобы в системе могло начаться образование структур, экспорт энтропии должен превзойти некоторое критическое значение. Самоорганизация — явление «надкритическое», или «закритическое». Это означает, что оно возможно, если параметр системы превосходит определенное критическое значение.

Экспорт энтропии, превосходящий производство энтропии, возникает не спонтанно, а требует «энтропийного насоса». Для приведения этого насоса в действие, как для приведения в действие любой машины, необходима расходуемая свободная энергия или свободная энтальпия, которую насос может черпать из внешних или внутренних источников. В конечном счете такие источники энергии всегда могут быть сведены к ядерным или химическим реакциям, протекающим на Земле или на звездах. Энтропийные насосы встречаются как внутри, так и вне структурообразующей системы. Соответственно, мы различаем активные и пассивные структурообразующие системы. Пассивные структурообразующие системы (ячейки Бенара, электрические цепи, лазеры и т. д.) непременно должны быть связаны с окружающей средой. Окружающая среда должна содержать энтропийные насосы, которые накачивают в систему электричество, тепло при более высокой температуре или коротковолновое излучение. Активные структурообразующие системы (живые организмы, двигатели Отто и т. д.) содержат энтропийные насосы внутри себя и поэтому должны, как правило, обладать высокой внутренней организацией. Кроме того, активные структурообразовательные системы должны получать из окружающей среды приток высокоэнергетических сырьевых веществ. Таким образом, энтропийный насос уводит как активные, так и пассивные системы в сторону от равновесия.

В качестве примера экспортирующего энтропию процесса рассмотрим изотермическо-изохорическую систему. Для нее справедливы следующие соотношения:

$$dF = d_i F + d_e F = dE - TdS = d_e E - Td_e S - Td_i S,$$

$$d_e F = dE - T d_e S, \quad (3.24)$$

$$d_e S = \frac{1}{T}(dE - d_e F).$$

Из условия экспорта энтропии $d_e S < 0$ и $|d_e S| > d_i S$ следует, что

$$d_e F > (dE + T d_i S). \quad (3.25)$$

Иначе говоря, чтобы гарантировать экспорт энтропии, свободную энергию следует подвести к системе в количестве, превышающем изменение энергии и ее затраты вследствие производства энтропии. Для изотермическо-изобарической системы справедливы соотношения

$$d_e S = \frac{1}{T}(dH - d_e G),$$

т. е. условие экспорта энтропии гласит:

$$d_e G > (dH + T d_i S). \quad (3.26)$$

Иначе говоря, чтобы гарантировать экспорт энтропии, изотермически-изобарическая система должна иметь подвод свободной энтальпии в количестве, покрывающем как изменение энтальпии, так и ее расход на производство энтропии.

Если система сильно отклоняется от равновесия, то переменные, вообще говоря, удовлетворяют уже не линейным уравнениям, как прежде, а более глубоким нелинейным закономерностям. Нелинейность — существенная и общая характерная особенность сильно неравновесных процессов, протекающих в природе. С другой стороны, надкритический экспорт энтропии требует особых внутренних структур системы. Это означает, что самоорганизация является свойством материи, существующим лишь при специфических внутренних и внешних условиях. Самоорганизация отнюдь не ограничена каким-либо особым классом веществ. Резюмируя, мы можем утверждать, что в природе существуют необратимые процессы двух основных типов.

1. Искажение, или возмущение, структур в окрестности равновесия как общее свойство систем при произвольных условиях.

2. Образование структур вдали от равновесия при специальных внутренних и внешних условиях; к числу этих условий принадлежат нелинейность внутренней динамики и надкритичность значения внешних параметров системы.

Пригожин предложил называть устойчивые пространственные, временные или пространственно-временные структуры, образующиеся вдали от равновесия при надкритических значениях параметров в нелинейной области, *диссипативными* структурами (Prigogine, 1967; Glansdorff, Prigogine, 1971; Николис, Пригожин, 1979). Каждый из отличительных признаков, отобранных и включенных в определение диссипативной структуры, имеет определенное значение для процесса. К числу наиболее важных признаков относятся устойчивость относительно малых возмущений и надкритическое расстояние от равновесия. Важнейший подкласс диссипативных структур образуют стационарные диссипативные структуры, которые при постоянных внешних условиях не претерпевают изменений со временем:

$$\begin{aligned} dS &= d_e S + d_i S = 0, \\ d_e S &= -d_i S < 0, \\ dE &= d_e E = 0. \end{aligned} \quad (3.27)$$

Последнее неравенство должно выполняться потому, что всем истинно неравновесным процессам соответствует неравенство $d_i S > 0$. Отсюда следует, что $d_e S < 0$,

т. е. что система должна отдавать энтропию в окружающую среду, чтобы компенсировать производство энтропии внутри системы на основе необратимых процессов. Для частного случая изотермически-изохорической системы имеем:

$$d_e F = T d_i S > 0, \quad (3.28)$$

а для изотермически-изобарической системы —

$$d_e G = T d_i S > 0, \quad (3.29)$$

т. е. стационарным неравновесным системам необходим подвод свободной энергии или свободной энтальпии.

Мы воспользуемся для обозначения устойчивого стационарного неравновесного состояния введенным Оствальдом и Берталанффи понятием текущего равновесия и введем следующее определение.

Стационарные неравновесные состояния, устойчивые относительно малых флуктуаций, называются состояниями *текущего равновесия*. Тем самым, диссипативные структуры соответствуют надкритическим состояниям текущего равновесия.

Введем теперь время и проследим, как изменяются во времени переменные состояния и производство энтропии в случае неравновесных процессов. Рассмотрим открытую систему, которая обменивается с окружающей средой произвольной экстенсивной величиной Z , например, энергией, энтропией, количеством вещества. За элементарный интервал времени dt величина Z изменяется на бесконечно малую величину

$$dZ = d_i Z + d_e Z, \quad (3.30)$$

$$\frac{dZ}{dt} = Q_Z + \Phi_Z, \quad (3.31)$$

где, по определению производства Z и потока Z , мы ввели величины

$$Q_Z = \frac{d_i Z}{dt}, \quad \Phi_Z = \frac{d_e Z}{dt}. \quad (3.32)$$

Три наиболее важных глобальных уравнения баланса для энергии E (измеренной в Дж), энтропии S (измеренной в Дж/К) и количества компоненты N_k (измеренного в молях) имеют следующий вид:

$$\frac{dE}{dt} = Q_E + \Phi_E, \quad (3.33)$$

$$\frac{dS}{dt} = P + \Phi_S, \quad (3.34)$$

$$\frac{dN_k}{dt} = Q_k + \Phi_k. \quad (3.35)$$

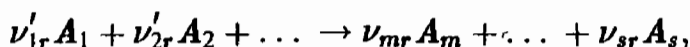
В открытых системах величины потоков Φ_E , Φ_S и Φ_k полностью определяются характером погружения и подвержены лишь малым флуктуациям их знаков и величин. С другой стороны, термы, описывающие производство соответствующих величин, ограничены законами термодинамики. В частности, справедливо равенство $Q_E = 0$, так как внутри системы вследствие первого начала термодинамики энергия не может ни рождаться, ни уничтожаться. Второе начало термодинамики определяет знак производства энтропии:

$$P = \frac{d_i S}{dt} \geq 0. \quad (3.36)$$

Наконец, законы химической стехиометрии устанавливают соотношения, в которых вещества могут появляться или исчезать вследствие реакций. Например, для реакции гремучего газа $2\text{H}_2 + \text{O}_2 \rightarrow 2\text{H}_2\text{O}$ имеют место соотношения

$$Q(\text{H}_2\text{O}) = -2Q(\text{O}_2) = -Q(\text{H}_2).$$

В общем случае в системе могут протекать R реакций, которые мы перенумеруем индексом $r = 1, 2, \dots, R$. Если в системе имеется всего s сортов различных веществ, то реакции формально можно записать следующим образом:



где ν_{kr} — стехиометрические коэффициенты r -й реакции. Перенесем теперь все члены в одну сторону и условимся считать стехиометрические коэффициенты расходуемых веществ отрицательными:

$$\nu_{1r} = -\nu'_{1r}, \quad \nu_{2r} = -\nu'_{2r}, \quad \dots$$

В результате мы получаем символическое уравнение

$$\sum_{k=1}^s \nu_{kr} A_k = 0. \quad (3.37)$$

Например, для реакции горючего газа стехиометрические коэффициенты оказываются следующими:

$$\nu(\text{H}_2\text{O}) = +2, \quad \nu(\text{O}_2) = -1, \quad \nu(\text{H}_2) = -2.$$

Далее удобно ввести в r -ю реакцию координату реакции, которая показывает, сколько «оборотов формулы» уже произошло. Например, $\xi_r = 1$ означает, что в r -й реакции образовалось ν_{kr} молей k -го вещества ($\nu_{kr} > 0$) или израсходовано ν_{kr} молей того же вещества ($\nu_{kr} < 0$); $\xi_r = 2$ означает, что образовалось (или израсходовано) $2\nu_{kr}$ молей и т. д. Введенные нами понятия позволяют представить члены, описывающие производство, в следующем виде:

$$Q_k = \frac{d_i N_k}{dt} = \sum_{r=1}^R \nu_{kr} \frac{d\xi_r}{dt}. \quad (3.38)$$

Для описания гомогенных систем выведенные выше глобальные уравнения баланса служат более чем достаточным основанием. Определим серию функций плотности, таких, как плотность производства энтропии σ и концентрация вещества k -го сорта c_k :

$$\sigma = \frac{P}{V}, \quad c_k = \frac{N_k}{V}. \quad (3.39)$$

Для скорости изменения концентрации k -й компоненты из уравнений (3.32) и (3.38) для изохорической системы следует уравнение

$$\begin{aligned} \frac{dc_k}{dt} &= \sum_{r=1}^R \nu_{kr} w_r + \frac{1}{V} \Phi_k, \\ w_k &= \frac{1}{V} \frac{d\xi_r}{dt}. \end{aligned} \quad (3.40)$$

Величина w_k называется скоростью r -й реакции. Вычислим теперь плотность производства энтропии. Используя соотношение (3.18), получаем:

$$\sigma = \frac{1}{V} \frac{d_i S}{dt} = \frac{1}{TV} \frac{d' A_{\text{необр}}}{dt} - \frac{1}{T} \sum_k \mu_k \frac{dc_k}{dt}. \quad (3.41)$$

Воспользуемся теперь соотношением (3.32) и определим химическое сродство r -й реакции по формуле

$$A_r = - \sum_{k=1}^s \nu_{kr} \mu_k. \quad (3.42)$$

Тогда

$$\sigma = \frac{1}{VT} \frac{d' A_{\text{необр}}}{dt} + \sum_{r=1}^R \frac{A_r}{T} w_r. \quad (3.43)$$

Предположим теперь, что частицы k -го сорта под действием внешней силы F_k перемещаются с плотностью тока J_k . Для работы, диссипируемой при таком процессе в системе, мы воспользуемся формулой Ома:

$$d' A_{\text{необр}} = V \sum_k J_k F_k dt. \quad (3.44)$$

В результате для плотности производства энтропии в гомогенной системе получаем

$$\sigma = \sum_{k=1}^s \frac{F_k}{T} J_k + \sum_{r=1}^R \frac{A_r}{T} w_r. \quad (3.45)$$

Формально правую часть соотношения (3.45) можно записать в виде билинейной формы с $(3s + R)$ слагаемыми:

$$\sigma = \sum_{\alpha=1}^{3s+R} X_{\alpha} J_{\alpha}. \quad (3.46)$$

Здесь мы ввели обобщенные термодинамические силы X_{α} и потоки J_{α} с помощью соотношений

$$X_{\alpha} = \frac{1}{T} F_{k\alpha}, \quad J_{\alpha} = J_{k\gamma} \quad (\gamma = 1, 2, 3, \quad k = 1, 2, \dots, s) \quad (3.47)$$

при $\alpha = 1, 2, \dots, 3s$ и

$$X_{\alpha} = \frac{1}{T} A_r, \quad J_{\alpha} = w_r \quad (r = 1, 2, \dots, R) \quad (3.48)$$

при $\alpha = 3s + 1, \dots, (3s + R)$.

Изложим кратко основы теории Онсагера для гомогенных систем, исходя из соотношения (3.46).

В состоянии термодинамического равновесия термодинамические силы и потоки одновременно обращаются в нуль: $X_{\alpha}^{(0)} = 0$, $J_{\alpha}^{(0)} = 0$. Поэтому в окрестности состояния равновесия потоки допустимо рассматривать как функции сил и разлагать по степеням последних.

Пренебрегая квадратичными членами (и членами более высокого порядка), получаем:

$$J_{\alpha} = \sum_{\beta} L_{\alpha\beta} X_{\beta}, \quad L_{\alpha\beta} = \left(\frac{\partial J_{\alpha}}{\partial X_{\beta}} \right)^{(0)}. \quad (3.49)$$

Это так называемое линейное соотношение Онсагера приводит к квадратичному выражению для производства энтропии, так как из соотношений (3.46) и (3.49) следует, что

$$\sigma = \sum_{\alpha, \beta} L_{\alpha\beta} X_{\alpha} X_{\beta}. \quad (3.50)$$

Из второго начала термодинамики $d_i S \geq 0$ получаем неравенство $\sigma_i \geq 0$, причем $\sigma = 0$ лишь при $X_\alpha = 0$. Из положительной определенности квадратичной формы следуют важные утверждения относительно коэффициентов связи. Рассмотрим сначала случай двух связанных необратимых процессов с

$$\sigma = [L_{11}X_1^2 + (L_{12} + L_{21})X_1X_2 + L_{22}X_2^2] \geq 0. \quad (3.51)$$

Алгебраические условия положительной определенности формы (3.51) гласят:

$$\begin{aligned} L_{11} > 0, \quad L_{22} > 0, \\ [4L_{11}L_{22} - (L_{12} + L_{21})^2] > 0. \end{aligned} \quad (3.52)$$

Нетрудно привести и алгебраические условия положительной определенности в общем случае n связанных процессов: матрица коэффициентов $\{L_{\alpha\beta}\}$ должна быть положительно определенной, т. е. должна иметь только собственные значения с положительными действительными частями. Это эквивалентно условию положительности всех главных миноров симметричной матрицы

$$\frac{1}{2}(L_{\alpha\beta} + L_{\beta\alpha}).$$

Эти условия для коэффициентов связи следуют из второго начала термодинамики и поэтому имеют совершенно общий характер. Дополнительные условия на матрицу связи возникают из пространственных и временных симметрий системы. Рассмотрим случай пространственно-изотропной среды. Как видно из соотношения (3.45), протекающие процессы можно разделить на два различных класса: на скалярные процессы (химические реакции) и на векторные процессы (потoki частиц). В силу изотропии системы скалярные причины не могут приводить к анизотропным следствиям (эффектам) и наоборот. Иначе говоря, причины не могут обладать более высокой симметрией, чем те следствия, к которым они приводят. Этот принцип был впервые сформулирован П. Кюри при исследовании проблем кристаллофизики, например, пьезоэлектричества, и применен Пригожиным к термодинамике необратимых процессов. В своей общей формулировке принцип Кюри—Пригожина гласит: коэффициенты Онсагера связи между потоками и силами, описываемыми тензорами различных рангов, равны нулю.

В 1931 г. Онсагер показал, что временные симметрии также оказывают определенное воздействие на коэффициенты связи. Исходя из предположения о детальном равновесии для стохастических микропроцессов, или симметрии динамических уравнений микрофизики относительно обращения времени, Онсагеру удалось вывести соотношение симметрии

$$L_{\alpha\beta} = L_{\beta\alpha} \quad (3.53)$$

(De Groot, Mazur, 1962).

Выведем для линейной области весьма общий принцип минимума, восходящий к Пригожину (Prigogine, 1967). Рассмотрим производство энтропии, потоки и силы в окрестности значений $\sigma^{(0)}$, $J_\alpha^{(0)}$, $X_\alpha^{(0)}$, соответствующих стационарному состоянию. Справедливо соотношение

$$\begin{aligned} \delta\sigma &= \sigma - \sigma^{(0)} = \\ &= \sum_{\alpha,\beta} \left\{ L_{\alpha\beta} (X_\alpha^{(0)} + \delta X_\alpha) (X_\beta^{(0)} + \delta X_\beta) - L_{\alpha\beta} X_\alpha^{(0)} X_\beta^{(0)} \right\} = \\ &= \sum_{\alpha,\beta} \left\{ L_{\alpha\beta} X_\alpha^{(0)} \delta X_\beta + L_{\alpha\beta} \delta X_\alpha X_\beta^{(0)} \right\} + \sum_{\alpha,\beta} L_{\alpha\beta} \delta X_\alpha \delta X_\beta. \end{aligned} \quad (3.54)$$

Первый член равен нулю, так как в стационарном состоянии $\frac{d\sigma}{dt} = 0$, в линейном приближении должно выполняться равенство $\delta\sigma = 0$. Следовательно, должно выполняться соотношение

$$\delta\sigma = \sum_{\alpha, \beta} L_{\alpha\beta} \delta X_{\alpha} \delta X_{\beta}. \quad (3.55)$$

Из положительной определенности матрицы $\{L_{\alpha\beta}\}$ получаем:

$$\begin{aligned} \delta\sigma &\geq 0, \quad \delta P = V\delta\sigma \geq 0, \\ \delta\sigma = \delta P = 0 &\text{ лишь при } \delta X_{\alpha} = 0. \end{aligned} \quad (3.56)$$

Отсюда следует теорема Пригожина: производство энтропии линейного необратимого процесса достигает минимума в стационарном состоянии. При приближении к стационарному состоянию изменение сил и потоков всегда происходит в таком направлении, которое соответствует уменьшению производства энтропии.

3.4. Локальные равновесия и критерии эволюции

Рассмотрим находящийся в состоянии покоя объем V , ограниченный поверхностью O . В этом случае произвольная экстенсивная величина $Z(t)$ системы может быть связана с локальной плотностью этой величины $z(r, t)$ соотношением

$$Z(t) = \int z(r, t) dV. \quad (3.57)$$

Разложение $dZ = d_i Z + d_e Z$ приводит к разложению временной вариации величины $Z(t)$ на вклад внутреннего производства и приток через поверхность по формуле (3.32).

Вводя локальную плотность производства величины Z и плотность тока j_z ,

$$Q_z = \int q_z dV, \quad \Phi_z = - \int j_z dO, \quad (3.58)$$

мы получаем по теореме Гаусса

$$\begin{aligned} \int \frac{\partial}{\partial t} z(r, t) dV &= \int q_z(r, t) dV - \int j_z dO = \\ &= \int \{q_z(r, t) - \text{div } j_z(r, t)\} dV. \end{aligned} \quad (3.59)$$

Так как соотношение (3.59) должно выполняться для произвольного объема, из него следует уравнение локального баланса величины Z :

$$\frac{\partial}{\partial t} z(r, t) + \text{div } j_z(r, t) = q_z(r, t). \quad (3.60)$$

Для всех сохраняющихся величин член q_z (источник) равен нулю. Например, если $Z = M$ — масса и $z(r, t) = \rho(r, t)$ — плотность массы, то, поскольку $q_M = 0$ (сохранение массы) и $j_M = \rho u$ (u — скорость массы, $u = \sum M_k c_k u_k / \sum M_k c_k$), мы получаем классическое уравнение неразрывности

$$\frac{\partial}{\partial t} \rho(r, t) + \text{div} (\rho(r, t) u(r, t)) = 0. \quad (3.61)$$

Если отождествить Z с числом молей l -й компоненты и ввести молярную концентрацию $C_k(r, t)$ в качестве его (числа) локальной плотности, j_k — в качестве локальной плотности потока, определить поток диффузии $J_k(r, t)$ с помощью соотношения

$$j_k(r, t) = \rho(r, t)u(r, t) + J_k(r, t), \quad (3.62)$$

а плотность производства k -й компоненты — с помощью соотношения

$$q_k(r, t) = \lim_{V \rightarrow 0} \frac{1}{V} \frac{d_i N_k}{dt} = \sum_k \nu_{kr} w_r(r, t), \quad (3.63)$$

то мы получаем уравнение баланса

$$\frac{\partial}{\partial t} c_k(r, t) = \operatorname{div} [c_k u(r, t) + J_k(r, t)] = \sum \nu_{kr} w_r(r, t). \quad (3.64)$$

Чтобы получить уравнение баланса для среднего импульса, мы исходим из гидродинамического уравнения движения:

$$\rho(r, t) \frac{d}{dt} u(r, t) = \sum c_k(r, t) F_k(r, t) - \nabla P(r, t). \quad (3.65)$$

Здесь $F_i(r, t)$ — плотность силы, а $P(r, t)$ — тензор плотности. Используя операторное равенство

$$\frac{d}{dt} = \frac{\partial}{\partial t} + u \cdot \nabla,$$

получаем после несложных преобразований

$$\frac{\partial}{\partial t} (\rho u) + \nabla (\rho u u + P) = \sum_k c_k F_k. \quad (3.66)$$

Следовательно, величины

$$j_P = \rho u u + P, \quad q_P = \sum c_k F_k \quad (3.67)$$

можно интерпретировать как плотность потока и плотность производства локального импульса.

Рассмотрим теперь уравнение баланса энергии, представимой в виде интеграла от плотности энергии $\varepsilon(r, t)$:

$$E = \int \rho(r, t) \varepsilon(r, t) dV.$$

В силу сохранения энергии $q_\varepsilon = 0$, и уравнение баланса имеет вид

$$\frac{\partial}{\partial t} (\rho \varepsilon) + \operatorname{div} j_\varepsilon = 0. \quad (3.68)$$

Уравнение баланса для внутренней энергии ($e(r, t)$ — плотность внутренней энергии)

$$U = E - E_{\text{кин}} - E_{\text{пот}} = \int \rho(r, t) e(r, t) dV$$

уже не свободно от источника и имеет вид

$$\frac{\partial}{\partial t} (\rho e) + \operatorname{div} j_e = q_e. \quad (3.69)$$

Определим с учетом соотношения (3.22) тепловой поток J_Q соотношением

$$\dot{j}_e = \rho h u + J_Q.$$

Здесь h — плотность энтальпии, определяемой из соотношения

$$H = \int \rho(r, t) h(r, t) dV. \quad (3.70)$$

Наконец, примем для плотности производства внутренней энергии q_e физически «прозрачное» выражение

$$q_e = \sum_k F_k \cdot J_k - P \cdot \cdot \nabla u. \quad (3.71)$$

(Здесь и далее: символ $\cdot \cdot$ означает тензорное произведение.) Первый член в правой части формулы (3.71) соответствует работе диффузионного потока против внешних полей, а второй — работе потока импульса против тензора давления (внутреннего трения). Более строгий вывод уравнений (3.69)–(3.71) мы получаем, умножая уравнение (3.66) скалярно на u и вычитая получившееся уравнение из уравнения (3.68) (De Groot, Mazur, 1962; Хаазе, 1967).

При выводе уравнения баланса для плотности энтропии, определяемой как

$$S = \int \rho(r, t) s(r, t) dV,$$

воспользуемся фундаментальным уравнением Гиббса (3.10), которое, по предположению, справедливо в определенных границах и для необратимых процессов:

$$T dS = dU + p dV - \sum_k \mu_k dN_k.$$

При переходе к локальной формулировке нам понадобятся соотношения

$$V = \int \rho v dV, \quad N_k = \int \rho v_k dV.$$

Здесь $v = \rho^{-1}$ — удельный объем и плотность числа молей, связанные с единицей массы. Решающее предположение термодинамики необратимых процессов (Glansdorff, Prigogine, 1971) сводится к гипотезе, что фундаментальное уравнение Гиббса выполняется также и локально, т. е.

$$T \frac{ds}{dt} = \frac{de}{dt} + p \frac{dv}{dt} - \sum_k \mu_k \frac{dv_k}{dt}. \quad (3.72)$$

Как показали Гленсдорф и Пригожин (Glansdorff, Prigogine, 1971), существенной предпосылкой, при которой выполняется это уравнение, является существование локального термодинамического равновесия.

После некоторых преобразований мы получаем

$$\begin{aligned} \frac{\partial}{\partial t}(\rho s) + \nabla \left(\rho s u + \frac{1}{T} J_Q + \sum_k s_k J_k \right) = J_Q \cdot \nabla \frac{1}{T} + \\ + \frac{1}{T} \sum_k J_k \cdot (F_k - \nabla \mu_k) - \frac{1}{T} \Pi \cdot \cdot \nabla u + \frac{1}{T} \sum_r w_r A_r. \end{aligned} \quad (3.73)$$

Здесь мы воспользовались разложением тензора давления на упругую часть $p_{\text{упр}}$ и диссипативную часть Π :

$$P = p_{\text{упр}} + \Pi, \quad p_{\text{упр}} = p \delta. \quad (3.74)$$

Так как общее уравнение баланса плотности энтропии должно иметь вид

$$\frac{\partial}{\partial t}(\rho s) + \nabla \cdot (\rho s u + J_s) = \sigma, \quad (3.75)$$

мы, сравнивая уравнение (3.75) с уравнением (3.73), получаем следующие выражения для плотности потока энтропии J_s и плотности производства энтропии σ :

$$J_s = \frac{1}{T} J_Q + \sum_k s_k J_k, \quad (3.76)$$

$$T\sigma = J_Q \cdot T \nabla \frac{1}{T} + \sum_k J_k \cdot (F_k - \nabla \mu_k) - \Pi \cdot \nabla u + \sum_r w_r A_r.$$

Как показывает сравнение с производством энтропии в случае однородной системы по формуле (3.45), в неоднородных системах возникает ряд дополнительных источников производства энтропии. К их числу относятся теплопроводность, диффузия и внутреннее трение.

Локальное производство энтропии в этом случае также имеет структуру билинейной формы:

$$\sigma(r, t) = \sum_{\alpha} J_{\alpha}(r, t) X_{\alpha}(r, t). \quad (3.77)$$

Тем самым, общие выводы, к которым мы пришли в предыдущем разделе, остаются в силе. Основные исходные допущения теории изотропных систем сводятся к следующим.

1. Линейные соотношения Онсагера:

$$J_{\alpha} = \sum_{\beta} L_{\alpha\beta} X_{\beta}. \quad (3.78)$$

2. Соотношения симметрии Онсагера:

$$L_{\alpha\beta} = L_{\beta\alpha}. \quad (3.79)$$

3. Принцип симметрии Кюри—Пригожина: связь между тензорами различного ранга невозможна.

Из них в конкретном случае следуют соотношения

$$w_r = \sum_s L'_{sr} \frac{A_s}{T},$$

$$J_k = \sum_l L_{kl} \frac{1}{T} (F_l - \nabla \mu_l) + L_{kQ} \nabla \left(\frac{1}{T} \right),$$

$$J_Q = L_{QQ} \nabla \left(\frac{1}{T} \right) + \sum_l L_{Ql} \frac{1}{T} (F_l - \nabla \mu_l),$$

$$\Pi = -L \cdot \frac{1}{T} [\nabla u]_{\text{симм}}$$

с соотношениями симметрии

$$L'_{rs} = L'_{sr}, \quad L_{kl} = L_{lk}, \quad L_{Ql} = L_{lQ}. \quad (3.81)$$

Заметим, что коэффициенты $L_{Ql} = L_{lQ}$ связаны с эффектом Соре, или термодиффузией, а величина L пропорциональна коэффициентам внутренней вязкости:

$$\Pi_{ij} = -\eta(\partial_j u_i + \partial_i u_j).$$

Вместе с уравнением (3.66) это линейное соотношение эквивалентно уравнению Навье—Стокса в гидродинамике (*Ebeling, Klimontovich, 1984*). Линейные соотношения неравновесной термодинамики обоснованы только в том случае, если потоки или силы линейно независимы. Линейную зависимость, если таковая представится, следует устранять до применения общей схемы путем исключения переменных. Рассмотрим в качестве примера диффузионный поток относительно скорости масс u :

$$J_k = c_k(u_k - u). \quad (3.82)$$

В силу соотношений (3.80) между потоками имеет место линейная зависимость

$$\sum_{k=1}^s M_k J_k = 0, \quad (3.83)$$

откуда

$$J_s = -\frac{1}{M_s} \sum_{k=1}^{s-1} M_k J_k. \quad (3.84)$$

Один из потоков линейно зависим от остальных и может быть полностью исключен. Для вклада диффузии в производство энтропии мы получаем выражение

$$\sigma_{\text{дифф}} = \sum_{k=1}^s J_k \cdot X_k = \sum_{k=1}^{s-1} J_k \cdot \left(X_k - \frac{M_k}{M_s} X_s \right), \quad (3.85)$$

$$X_k = \frac{F_k - \nabla \mu_k}{T}.$$

Вводя приведенные силы

$$X'_k = X_k - \frac{M_k}{M_s} X_s, \quad (3.86)$$

мы получаем для $\sigma_{\text{дифф}}$ билинейную форму, содержащую только линейно независимые потоки

$$\sigma_{\text{дифф}} = \sum_{k=1}^{s-1} J_k \cdot X'_k. \quad (3.87)$$

По правилам линейной теории получаем:

$$J_k = \sum_{l=1}^{s-1} L_{kl} X'_l + L_{kQ} \nabla \left(\frac{1}{T} \right), \quad (3.88)$$

$$J_Q = L_{QQ} \nabla \left(\frac{1}{T} \right) + \sum_{l=1}^{s-1} L_{Ql} X'_l.$$

Подставляя соотношения (3.88) в уравнения баланса (3.64) и (3.69), получаем, соответственно, уравнение диффузии и уравнение теплопроводности (*De Groot, Mazur, 1962*).

Аналогичным образом можно поступать во всех случаях, когда потоки или силы окажутся линейно зависимыми. В дальнейшем мы всегда будем предполагать, что предварительно произведена редукция к линейно независимым потокам и силам. Рассмотрим теперь критерий эволюции Пригожина для линейных необратимых процессов. Поскольку при выводе условий (3.56) были использованы лишь совершенно

общие соотношения, которые, как было показано, выполняются и в неоднородном случае, мы воспользуемся условиями (3.56) и получим для локального производства энтропии

$$\begin{aligned}\delta\sigma(r, t) &\geq 0, \\ \delta\sigma(r, t) &= 0 \quad \text{при} \quad X_\alpha(r, t) = X_\alpha^{(0)}(r),\end{aligned}\tag{3.89}$$

где $X_\alpha^{(0)}(r)$ — стационарные локальные силы. Аналогично, для полного производства энтропии в системе

$$P = \int \sigma(r, t) dV$$

справедливы соотношения

$$\begin{aligned}\delta P &\geq 0, \\ \delta P &= 0 \quad \text{при} \quad X_\alpha(r, t) = X_\alpha^{(0)}(r).\end{aligned}\tag{3.90}$$

В стационарном состоянии производство энтропии в линейном необратимом процессе достигает минимума. Так как необратимый процесс в линейной области всегда стремится к некоторому однозначно определенному состоянию, мы заключаем, что в линейной области процесс эволюции не может существовать, поскольку он представляет собой неограниченную последовательность процессов. Тем самым поиск процессов эволюции должен быть ориентирован на область нелинейных неравновесных состояний.

Критерий эволюции для нелинейной области был сформулирован Гленсдорфом и Пригожиным. Частный вид этого критерия эволюции следует из формулы (3.76) для химических реакций:

$$\sigma = \sum_{r=1}^k w_r \left(\frac{A_r}{T} \right).\tag{3.91}$$

Варьируя химические сродства A_r , получаем

$$d_x \sigma = \frac{1}{T} \sum w_r dA_r.$$

Подставляя вместо A_r их выражения (3.42), приходим при $\Phi_k = 0$ к соотношению

$$d_x \sigma = -\frac{1}{T} \sum_k \frac{dc_k}{dt} d\mu_k.$$

Если учесть соотношение

$$d\mu_k = \sum_l \frac{\partial \mu_k}{\partial N_l} dN_l,\tag{3.92}$$

выполняющееся для изотермических процессов, то окончательно мы получаем

$$\frac{d_x \sigma}{dt} = -\frac{1}{T} \sum_{k,l} \frac{\partial \mu_k}{\partial c_l} \left(\frac{dc_k}{dt} \right) \left(\frac{dc_l}{dt} \right).\tag{3.93}$$

Как показал еще Гиббс, для химического потенциала должно выполняться следующее фундаментальное неравенство:

$$\sum_{k,l} \left(\frac{\partial \mu_k}{\partial c_l} \right) z_k z_l \geq 0,\tag{3.94}$$

где z_k — некоторый набор действительных чисел (Prigogine, Defay, 1962). Наконец, с помощью неравенства Гиббса мы получаем неравенство

$$\frac{d_x \sigma}{dt} \leq 0, \quad (3.95)$$

известное под названием теоремы Гленсдорфа—Пригожина (Glensdorff, Prigogine, 1971). По этой теореме, изменение удельного производства энтропии, обусловленное вариацией сил, может быть либо отрицательным, либо равным нулю. Эквивалентная форма неравенства (3.95) возникает при интегрировании его по объему:

$$\frac{d_x P}{dt} \leq 0. \quad (3.96)$$

Для частного случая линейного процесса из критерия Гленсдорфа—Пригожина (3.95) следует теорема Пригожина. Действительно, в этом случае

$$\begin{aligned} \frac{d\sigma}{dt} &= \sum L_{kl} \left[\left(\frac{A_l}{T} \right) \frac{d}{dt} \left(\frac{A_k}{T} \right) + \left(\frac{A_k}{T} \right) \frac{d}{dt} \left(\frac{A_l}{T} \right) \right] = \\ &= 2 \sum L_{kl} \left(\frac{A_k}{T} \right) \frac{d}{dt} \left(\frac{A_l}{T} \right) = 2 \frac{d_x \sigma}{dt}. \end{aligned} \quad (3.97)$$

Следовательно, для линейного процесса мы получаем из неравенства (3.95) критерий Пригожина

$$\frac{d\sigma}{dt} \leq 0 \quad \text{и} \quad \frac{dP}{dt} \leq 0. \quad (3.98)$$

Из теоремы Пригожина следует также, что лежащие в линейной области стационарные состояния автоматически устойчивы, поскольку в силу тенденции величины σ к убыванию любое отклонение сразу же сходит на нет.

На рис. 3.1 показан временной ход производства энтропии для линейного необратимого процесса. Из критерия эволюции Пригожина нетрудно заключить, что неограниченная последовательность процессов невозможна в области линейных необратимых процессов. В линейной области процесс всегда стремится к состоянию с наименьшим производством энтропии и по истечении конечного времени релаксации асимптотически достигает его. Этот вывод имеет чрезвычайно важное значение для теории процессов эволюции. В области линейных процессов лежат процессы эволюции, которые состоят из неограниченной последовательности процессов, что противоречит точным основным

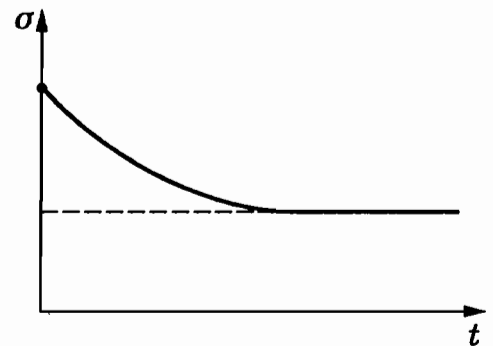


Рис. 3.1. Временной ход производства энтропии для линейной системы (в соответствии с принципом Пригожина)

законам термодинамики. Эволюционные процессы этого типа в принципе возможны только в нелинейной области при больших отклонениях от термодинамического равновесия. Вопрос о существовании общих принципов эволюции для нелинейной области не вполне ясен и поныне, и в литературе приводятся самые различные варианты (Guarati, 1970; Зотин, 1974; Стратонович, 1985; Лопушанская, 1987).

Теорема Гленсдорфа—Пригожина представляет собой обобщение теоремы Пригожина, не исключающее бесконечную последовательность шагов самоорганизации. В заключение мы рассмотрим возможности и границы применения термодинамики необратимых процессов к проблемам самоорганизации и эволюции (Fleischer, 1979).

Сведения о термодинамике необратимых процессов, полученные при рассмотрении простых физических систем, позволяют по аналогии строить заключения о сложных процессах эволюции и тем самым подкрепляют наше понимание эволюции. Высокая степень абстракции термодинамики необратимых процессов позволяет также интерпретировать некоторые аспекты механизмов эволюции; законы, которым подчиняются эти механизмы, остаются в силе без каких бы то ни было ограничений и в нелинейной области. Процессы эволюции представляют собой частные случаи нелинейных необратимых процессов; они подчиняются изложенным выше фундаментальным общим законам, но их сущность определяется не этими законами. Одна из характерных особенностей эволюции состоит, очевидно, в том, что при возникновении новых структур эволюция «не скупится» на издержки-производство энтропии, в то время как по отношению к уже установившимся структурам ограничивается лишь самым необходимым уровнем производства энтропии (Jantsch, 1979).

3.5. Термодинамические действующие силы эволюции Земли

Нашу Землю можно рассматривать как гигантскую открытую систему, в которой на протяжении последних 4,5 миллиардов лет разыгрываются процессы самоорганизации и эволюции. Планета Земля имеет массу около $6 \cdot 10^{24}$ кг и обращается вокруг Солнца на расстоянии около $1,5 \cdot 10^{11}$ м. Процессы эволюции на Земле по существу развертывались в тонком слое земной поверхности, толщина которого составляет лишь незначительную долю радиуса Земли, — слое, состоящем из скальных и осадочных пород, вод и атмосферы. Существенные этапы этой эволюции представлены в календаре эволюции (табл. 2.1–2.5). Эволюция в поверхностных слоях Земли охватывает геологические, атмосферные и биологические процессы. Энтропийными насосами, приводящими в действие эти процессы эволюции, служат солнечное излучение и запасы энергии в ядре Земли, передаваемые поверхностным слоям посредством теплопроводности и тепловой конвекции, радиоактивного излучения, извержений вулканов, энергии тектонических движений. В качестве космического источника энергии следует также упомянуть космическое излучение, хотя оно составляет не более 0,1% солнечного излучения. На рис. 3.2 представлена схема термодинамических потоков энтропии и энергии, или свободной энтальпии, через поверхность рассматриваемой открытой системы, имеющей форму сферической оболочки. Возможные процессы самоорганизации и эволюции на Земле приводятся в действие перепадом, или градиентом, температур между Солнцем или ядром Земли с одной стороны и температурой космического пространства (фоновое излучение) с другой стороны. Собственное излучение Земли оказывает лишь пренебрежимо слабое влияние на температуру Солнца или космического пространства, поэтому Землю можно рассматривать как пассивную систему, сквозь которую «прокачивается» энергия. Процесс, создающий этот поток энергии, приводит к деградации энергии, т. е. к производству энтропии.

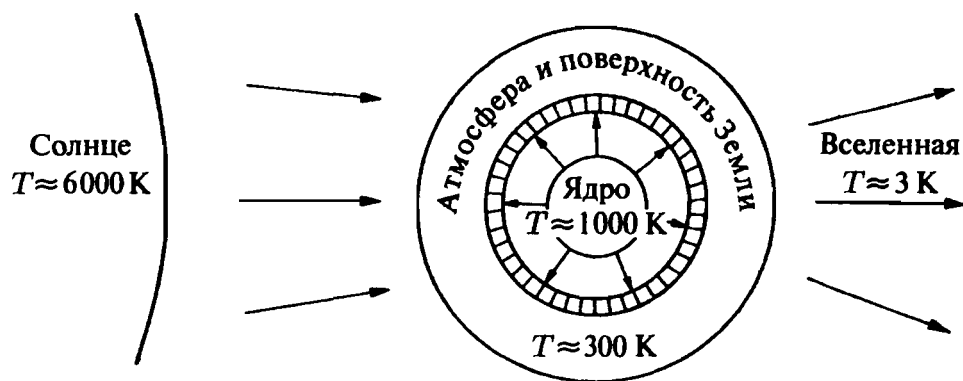


Рис. 3.2. Потоки энергии через открытую систему «земная мантия и атмосфера»

Земля поглощает энергию космических источников своей атмосферой и верхними слоями земной мантии. Поглощающий слой имеет толщину около 30 000 км. Еще на большой высоте над поверхностью Земли потоки частиц захватываются магнитным полем Земли. В области от 5 000 км до 500 км космические кванты высоких энергий, взаимодействуя с верхними слоями газовой оболочки, порождают частицы с большой энергией (около 10^8 эВ), образующие внутренний пояс излучений. В слоях выше 80 км плотность газа уже столь велика, что поглощаются почти все кванты электромагнитного излучения с длинами волн примерно до 200 мм. Поглощенные кванты взаимодействуют с отдельными частицами газа и вызывают диссоциацию и ионизацию молекул и атомов. Эта область, расположенная выше 90 км, называется гетеросферой. Слой, лежащий ниже, называется гомосферой. Плотность газа в гомосфере существенно выше, что приводит к интенсивному обмену энергией и веществом в турбулентном режиме. Потоки энергии изображены на рис. 3.3.

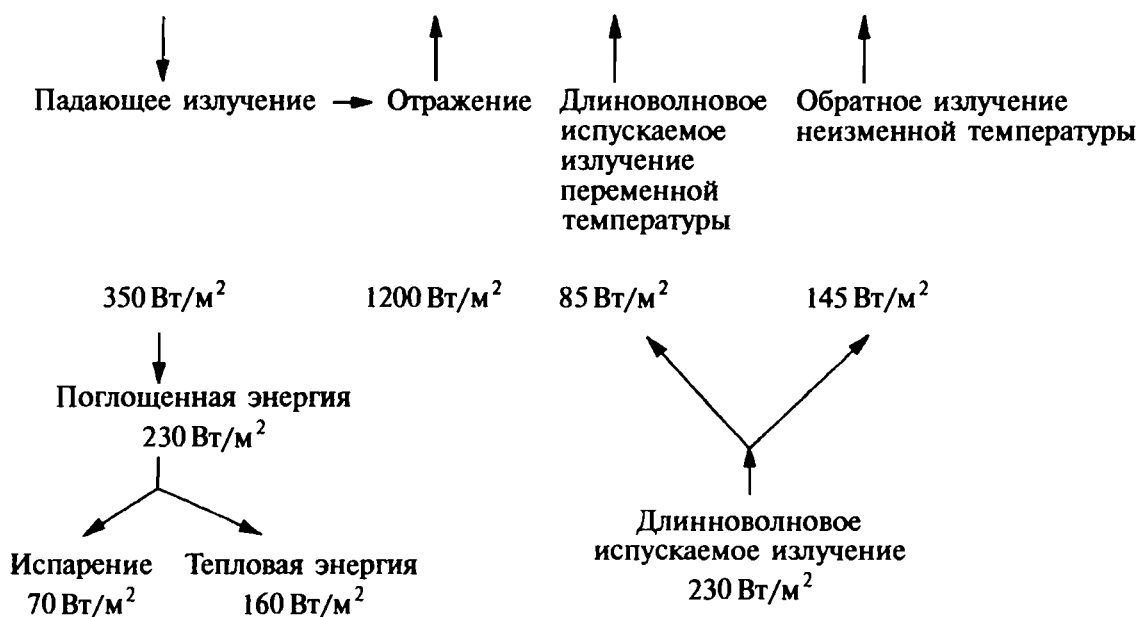


Рис. 3.3. Схема протоков энергии в атмосфере (Сватков, 1974)

Приведенные числа дают лишь весьма приблизительные оценки (Сватков, 1974). Данные различных авторов расходятся примерно на 10% (см., например, Монин, Шишков, 1979; Будыко, 1978; Монин, 1987). Из схемы, приведенной на рис. 3.3, видно, что баланс энергии уравновешен, т. е. выполняется условие стационарности $dU = 0$. Излучение, приходящее на Землю, достигает величины около 230 Вт/м² и соответствует температуре излучения около 6 000 К (температура поверхности Солнца). Измеренное спектральное распределение энергии солнечного излучения представлено на рис. 3.4. Спектральное распределение энергии излучения, испускаемого Землей, носит существенно более сложный характер. В грубом приближении это излучение можно представить как суперпозицию черного излучения с первоначальной температурой и длинноволнового излучения с температурой около 250 К. Вклад длинноволнового излучения составляет, по Сваткову, около 230 Вт/м². Спектр длинноволнового излучения представлен на рис. 3.5. Нетрудно видеть, что спектральное распределение длинноволнового излучения заключено между распределениями черного излучения с температурами 218 К и 288 К. Объясняется это тем, что земная поверхность со средней температурой 288 К испускает тепло, как черное тело, но в космическое пространство это тепло может излучаться только в тех спектральных областях, в которых атмосфера прозрачна. В остальных областях спектра из космического пространства «видна» не поверхность Земли, а поверхность стратосферы,

средняя температура которой составляет лишь 218 К (подобно тому, как с Земли в видимой области спектра мы можем наблюдать лишь наружные слои газа на Солнце).

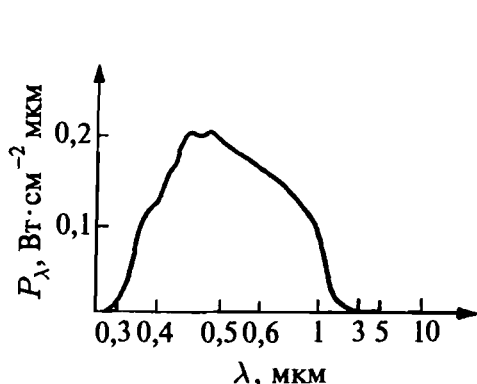


Рис. 3.4. Спектр солнечного излучения (Сватков, 1974)

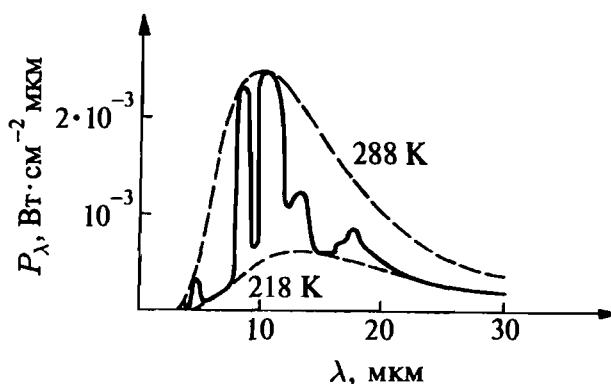


Рис. 3.5. Схема спектрального распределения земного излучения (Сватков, 1974)

Более подробное распределение потоков энергии приводит в своей работе Будыко (Будыко, 1978). По его данным, в среднем за сутки на единицу поверхности Земли падает из космического пространства 340 Вт/м^2 энергии, около 102 Вт/м^2 сразу же отражается, а из оставшейся энергии 81 Вт/м^2 поглощается атмосферой и 157 Вт/м^2 — поверхностью Земли. За счет испарения воды с поверхности Земли и конденсации в атмосфере в последнюю переносится еще 88 Вт/м^2 . К этой величине следует добавить еще 17 Вт/м^2 за счет турбулентного потока тепла. Лишь оставшиеся 52 Вт/м^2 поверхность Земли излучает в космическое пространство. Гораздо большая часть энергии — 186 Вт/м^2 — покидает Землю в виде длинноволнового излучения атмосферы.

По оценке Мони́на и Сискова (Монин, Сисков, 1979), внутренняя энергия атмосферы составляет $8,6 \cdot 10^{23} \text{ Дж}$, или $1,7 \cdot 10^9 \text{ Дж/м}^2$. Примерно половина ее — $3,6 \cdot 10^{23} \text{ Дж}$, или $0,7 \cdot 10^9 \text{ Дж/м}^2$, — приходится на долю потенциальной энергии. Для поддержания турбулентных течений, по мнению авторов, требуется около 4 Вт/м^2 .

Для оценки потока энтропии в секунду, создаваемого Солнцем, мы предполагаем, что в среднем поглощается $1,2 \cdot 10^{17} \text{ Вт}$ коротковолнового излучения с температурой $T = 5770 \text{ К}$ и испускается такое же количество излучения с температурой $T = 257 \text{ К}$. Тогда экспорт энтропии под действием излучения составляет

$$\frac{d_e S}{dt} = \frac{4}{3} \cdot 1,2 \cdot 10^{17} \text{ Вт} \cdot \left(\frac{1}{5770 \text{ К}} - \frac{1}{257 \text{ К}} \right) = -6 \cdot 10^{14} \text{ Вт/К}. \quad (3.99)$$

При этом мы использовали при вычислениях планковский формфактор $4/3$ для черного излучения (Fortak, 1979). Действительно, из известных соотношений для плотностей энергии и энтропии черного излучения (Ландау, Лифшиц, 1995) получаем

$$J_e = \sigma T^4, \quad J_s = \frac{4}{3} \sigma T^3,$$

откуда $J_s = (4/3)J_e/T$, что и оправдывает множитель $(4/3)$. Приведенному выше значению полного экспорта энтропии соответствует следующая плотность потока энтропии на поверхности Земли:

$$J_s = \frac{4}{3} \cdot 230 \text{ Вт/м}^2 \cdot \left(\frac{1}{5770 \text{ К}} - \frac{1}{257 \text{ К}} \right) \approx -1 \text{ Вт/м}^2 \text{К}. \quad (3.100)$$

Такой экспорт энтропии обеспечивает возможность компенсировать среднюю плотность производства энтропии от 10^{-3} до 10^{-4} Вт/м³К в нижних слоях атмосферы. Экспорт энтропии за счет резервов энергии ядра Земли на несколько порядков меньше. По оценке Жаркова (Жарков, 1983), поток энергии от ядра Земли к поверхности составляет около 0,2 Вт/м². Если предположить, что температура в ядре Земли находится в интервале от 10^3 до 10^4 К, а температура на поверхности Земли составляет около 300 К, то, по оценке, экспорт энтропии за счет теплового потока Земли составляет $\sim 10^{11}$ Вт/К. Таким образом, коротковолновое солнечное излучение оказывается важнейшим энтропийным насосом Земли. Именно оно является в конечном счете ответственным за многообразные процессы самоорганизации и биологической эволюции на поверхности нашей планеты. Наибольших энергетических и энтропийных затрат при этом требуют погодные и климатические процессы. Хотя остальные вклады в экспорт энтропии относительно малы, в эволюции Земли они тем не менее играют важную роль, в частности являются действующей силой тектонических процессов. Если оценить числа Рэлея для магмы, то при определенных предположениях получаются надкритические значения. Следовательно, в магме должны возникать конвективные течения, которым соответствует рассматриваемый в разд. 4.2 эффект Бенара. Модель таких конвективных магматических течений представлена на рис. 3.6. Действие сил Кориолиса приводит к тому, что эффект Бенара в магме порождает существенно более сложные структуры, чем в простых жидкостях. Аналогичные эффекты играют роль при образовании планетарной системы ветров. Было бы интересно выяснить, не полем ли скоростей магмы на поверхности Земли обусловлены сложные вращательные и трансляционные движения континентальных плит и эффекты образования коры («растекание морского дна»). В том случае, если подобные гипотезы имеют под собой какое-то реальное основание, построение теоретической модели движения магмы позволило бы проследить движения континентов и тем самым достичь более полного и глубокого понимания геологической эволюции. Подобно тому, как ныне долгопериодические колебания климата (например, продолжительности ледовых периодов) успешно моделируются математически на основе (модельных) представлений о нелинейных самовозбуждающихся колебаниях (Монин, Сисков, 1979; Сергин, Сергин, 1978), ответ на вопрос о длиннопериодических колебаниях земного магнитного поля можно надеяться получить с помощью нелинейной математической модели. Некоторые интересные результаты в этом направлении после основополагающих работ Чандraseкара и Рункорна по теории конвекции магмы были получены в последнее время (Монин, Сисков, 1979; Uyeda, 1978; Runcorn, 1962; Монин, 1987).

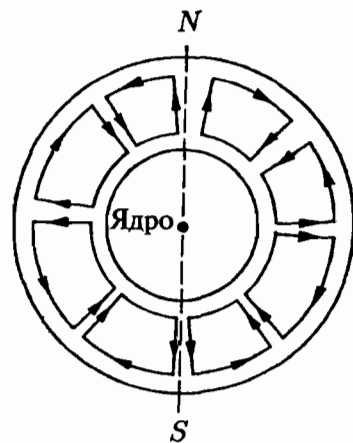


Рис. 3.6. Крупномасштабные магматические конвективные течения как возможная причина континентального дрейфа

3.6. Термодинамика самоорганизации живых организмов

Рассмотренные до сих пор примеры самоорганизации открытых систем были заимствованы из неорганического мира. В то время как таким системам часто присуще нечто необычное, в органическом мире самоорганизация и эволюция относятся к числу основных явлений. Проблеме термодинамической интерпретации самоорганизации и эволюции живых систем были посвящены основополагающие работы Оствальда (Ostwald, 1931), Берталанффи (Bertalanffy, 1942, 1956), Бауэра (Baуэр, 1935),

Шрёдингера (*Шрёдингер*, 1962), а также Пригожина и Виаме (*Prigogine, Wiame*, 1946), заложившие основы современных исследований (*Зотин*, 1974; *Blumenfeld*, 1977; *Glaser*, 1986; *Николис, Пригожин*, 1979; *Волькенштейн*, 1978; *Morowitz*, 1978).

Рассмотрим сначала живой организм в стационарном состоянии (*Penzlin*, 1976). В стационарном состоянии производная энергии и энтропии по времени равна нулю, т. е. выполняются соотношения

$$\begin{aligned}\frac{dU}{dt} &= \frac{d_i U}{dt} + \frac{d_e U}{dt} = 0, & \frac{d_e U}{dt} &= -\frac{d_i U}{dt} = 0, \\ \frac{dS}{dt} &= \frac{d_i S}{dt} + \frac{d_e S}{dt} = 0, & \frac{d_e S}{dt} &= -\frac{d_i S}{dt} < 0.\end{aligned}\quad (3.101)$$

Как правило, термодинамические процессы в живом организме протекают изотермически и изобарически, т. е. при постоянной температуре и постоянном давлении. Из соотношения (3.26) следует, что при этом

$$\frac{d_e G}{dt} = T \frac{d_i S}{dt} > 0, \quad (3.102)$$

т. е. для поддержания стационарного состояния живой организм должен постоянно получать определенное количество свободной энтальпии в единицу времени. Поскольку живой организм на 70% состоит из несжимаемых жидкостей, свободную энтальпию можно приближенно считать равной свободной энергии.

Живые организмы представляют собой открытые системы, структура и функционирование которых поддерживается путем постоянного обмена веществом и энергией (подвода свободной энтальпии) с окружающей средой. Э. С. Бауэр (*Бауэр*, 1935) так сформулировал общий закон биологии: «Живые системы никогда не находятся в равновесии и функционируют за счет своей свободной энергии, постоянно совершая работу против навязываемого внешними условиями равновесия». При этом происходит производство энтропии — выполняется неравенство $\frac{d_i S}{dt} > 0$, т. е. имеет место текущее равновесие. Функция обмена веществ в живом организме с точки зрения термодинамики состоит в том, чтобы путем обмена веществом и теплом с окружающей средой обеспечить по крайней мере такое количество энтропии, какое продуцируется в единицу времени внутри системы. Для поддержания стационарности должен сохраняться и баланс энергии, т. е. система должна поглощать ровно столько энергии, сколько она расходует: $\frac{dU}{dt} = 0$. Если обозначить через $d'Q$ тепло, подводимое к системе в течение элементарного интервала времени dt , а через $d_e N_k$ — количество k -й компоненты, измеренное в молях или в числе молекул, подводимое за время dt , то доля обмениваемой энтропии по формуле (3.22) составляет

$$d_e S = \frac{d'Q}{T} + \sum_k s_k d_e N_k. \quad (3.103)$$

Здесь T — температура системы, s_k — парциальная энтропия k -й компоненты.

Необходимый экспорт энтропии ($-d_e S$) $\geq d_i S > 0$ обеспечивается тремя процессами:

- 1) теплоотдачей;
- 2) обменом веществом с окружающей средой;
- 3) превращением веществ внутри системы.

Особенно важную роль играет теплопередача. Это понятие мы будем использовать в полном объеме, понимая под теплопередачей и теплопроводность, и тепловое излучение (рис. 3.7). Все живые организмы, как правило, постоянно отдают тепло

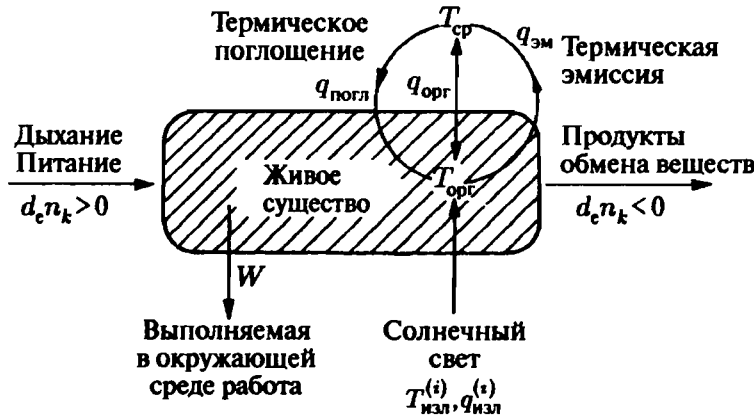


Рис. 3.7. Обмен энергией и веществом открытой системы «живого существа»

окружающей среде и тем самым экспортируют энтропию. Для этого между живым организмом и окружающей средой должна поддерживаться определенная разность температур $\Delta T > 0$. Температура многих живых организмов (растений, пойкилотермных (= холоднокровных) животных) $T_{\text{орг}}$ следует за температурой окружающей среды $T_{\text{ср}}$, т. е.

$$T_{\text{орг}} = T_{\text{ср}} + \Delta T, \quad \Delta T \approx \text{const}.$$

Другие организмы (гомойтермные (=теплокровные) животные) поддерживают постоянство своей температуры $T_{\text{орг}} \approx \text{const}$ с помощью механизмов терморегуляции. В единицу времени такой организм может отдавать в окружающую среду посредством теплопроводности количество тепла $q_{\text{орг}}$. Кроме того, в единицу времени в организм может поступать энергия излучения $q_{\text{изл}}$. Если падающее на организм излучение приближенно рассматривать как излучение черного тела с температурой $T_{\text{изл}}$, то приходящее излучение соответствует поступлению энтропии $4q_{\text{изл}}/3T_{\text{изл}}$. Коэффициент $4/3$ учитывает особенности транспорта энтропии посредством излучения (планковский формфактор). Если учесть и обмен веществ, то экспорт энтропии приближенно можно записать в виде

$$-\frac{d_e S}{dt} \approx -\frac{4q_{\text{изл}}}{3T_{\text{изл}}} + \frac{q_{\text{орг}} \Delta T}{T_{\text{орг}}^2} - \sum_k s_k \frac{d_e N_k}{dt}. \quad (3.104)$$

Более точный баланс обмена энтропии должен учитывать различные типы излучения. Пусть $q_{\text{исп}}$ — энергия излучения, испускаемая живым организмом, $q_{\text{погл}}$ — энергия излучения, поглощаемая живым организмом из окружающей среды; $T_{\text{орг}}$ и $T_{\text{ср}}$ — соответствующие температуры излучения. Некоторые трудности возникают в связи с энергией излучения $q_{\text{фот}}$, поглощаемой при фотосинтезе. В силу селективного характера поглощения излучения при фотосинтезе предположение о существовании эффективной температуры излучения представляет лишь грубое приближение. Однако описание можно улучшить, предположив существование различных тепловых компонент с температурами излучения $T_{\text{изл}}^{(i)} (i = 1, 2, \dots)$. Суммируя все вклады, получаем для экспорта энтропии в единицу времени выражение

$$\frac{d_e S}{dt} = q_{\text{орг}} \left(\frac{1}{T_{\text{ср}}} - \frac{1}{T_{\text{орг}}} \right) + \frac{4q_{\text{исп}}}{3T_{\text{орг}}} - \frac{4q_{\text{погл}}}{3T_{\text{ср}}} - \sum_i \frac{4q_{\text{фот}}^{(i)}}{3T_{\text{изл}}^{(i)}} - \sum_k s_k \frac{d_{\text{исп}} N_k}{dt}. \quad (3.105)$$

Необходимым условием функционирования живого организма является положительный баланс экспорта энтропии (см. формулу (3.23)):

$$\left\{ q_{\text{орг}} \left(\frac{1}{T_{\text{ср}}} - \frac{1}{T_{\text{орг}}} \right) + \frac{4q_{\text{исп}}}{3T_{\text{орг}}} - \right.$$

$$-\frac{4q_{\text{погл}}}{3T_{\text{ср}}} - \sum \frac{4q_{\text{фот}}^{(i)}}{3T_{\text{изл}}^{(i)}} - \sum s_k \frac{d_e N_k}{dt} \Big\} \geq \frac{d_1 S}{dt} > 0. \quad (3.106)$$

Производство энтропии при этом имеет вид

$$P = \frac{d_1 S}{dt} = \frac{1}{T_{\text{орг}}} \frac{d' A_{\text{дисс}}}{dt} - \frac{1}{T_{\text{орг}}} \sum \mu_k \frac{d_1 n_k}{dt}. \quad (3.107)$$

Условие (3.106) можно было бы назвать четвертым началом термодинамики живого. Можно утверждать, что по крайней мере на определенных стадиях индивидуального развития живого организма, например, при эмбриогенезе, в детстве, при заживлении ран и регенерации организмов должно выполняться более строгое условие структурообразования

$$\left(-\frac{d_e S}{dt} \right) > \frac{d_1 S}{dt} > 0. \quad (3.108)$$

Сравним различные структурообразующие системы с живыми организмами. Существенное различие между системой Бенара, лазерной системой и Землей, с одной стороны, и живым организмом — с другой, состоит в том, что порядок в живом организме поддерживается не внешним насосом, а некоторым внутренним механизмом. Движущая сила экспорта энтропии через организм исходит изнутри самой системы. С термодинамической точки зрения организм подобен автомашине. В двигателе автомашины поступающий бензин сгорает и с помощью хитроумного механизма превращается в полезную работу, тепло, двуокись углерода и водяной пар. В стационарном состоянии (на холостом ходу) или при езде с постоянной скоростью поступающий в двигатель бензин расходуется только на поддержание внутреннего динамического порядка, т. е. с точки зрения термодинамики на компенсацию энтропии, производимой внутри двигателя и за счет трения о покрытие шоссе. При более интенсивном поступлении бензина (если водитель «поддаст газу») двигатель производит больше полезной работы, которую можно затратить на ускорение системы или грузов.

Решающее различие между термодинамическими системами автомашины и живого организма состоит, как нетрудно видеть, в том, что в первом случае принцип функционирования придуман человеком, а во втором — является результатом естественного процесса эволюции. Общее с точки зрения термодинамики между автомашиной и организмом состоит в том, что обеспечивающие экспорт энтропии насосы «вмонтированы» внутрь системы. Для поддержания своего функционирования энтропийные насосы нуждаются в расходовании свободной энергии или свободной энтальпии, подводимых к системе в форме высокоценного горючего или пищи. Если двигатель Отто нуждается в непрерывном подводе бензина, то живому организму для поддержания своего существования требуются углеводороды, белки, жиры, витамины и микроэлементы. В смысле того, о чем говорилось в разд. 3.3, автомашины, как и живые организмы, можно было бы отнести к активным структурообразующим системам. И тем, и другим для функционирования требуется подвод из окружающей среды веществ с высоким содержанием энергии и низким содержанием энтропии, а также подвод тепла.

Для нас особый интерес представляет развитие рыб, земноводных, пресмыкающихся и птиц из яйца. Как известно, начальная стадия развития птиц до вылупления проходит внутри замкнутой известковой оболочки — скорлупы, содержащей яйцо. Птенец имеет полный запас питательных веществ (белков, жиров, витаминов) в желтке, а также в белке яйца и в воде белкового отстоя. Ядро оплодотворенной клетки образует зародышевый диск, который располагается в верхней части желтка. Если

устанавливаются условия, благоприятные для выведения птенцов, т. е. определенная температура (38–39° С для домашних кур) и влажность (60–65% для домашних кур), то начинается процесс дробления клетки. Развитие зародыша начинается в 1-й день выведения с образования так называемой первичной борозды с зачатком будущей хорды и с образования основной формы тела с головой и задней частью, а также перикарда во 2-й день. На 3-й и 4-й день образуется кровообращение и аллантоис — мешок, в котором собираются продукты обмена веществ. В последующие дни происходит дифференциация концов крыльев и ног, а также формирование системы органов, связанных с использованием желтка. Затем происходит процесс проникновения птенца в окружающую среду — птенец разрушает известковую оболочку. На протяжении всего процесса выведения птенца происходит обмен яйца с окружающей средой теплом, кислородом, двуокисью углерода и водяным паром. Следовательно, с термодинамической точки зрения яйцо представляет открытую систему (рис. 3.8).

Рассмотрим теперь возникновение упорядоченных структур при индивидуальном развитии живого организма из яйцеклетки. При этом мы ограничимся кратким изложением эмбрионального развития животных из яйцеклетки, которое представляет классический пример самоорганизации открытой системы. Как и обычные клетки, яйцеклетки наделены ядром и цитоплазмой с аппаратом синтеза. В отличие от обычных клеток яйцеклетка содержит такие сырьевые вещества, как аминокислоты, белки, гликоген, фосфолипиды, витамины и др., запасенные в качестве желтка в цитоплазме. Желток служит резервом энергии и «строительного материала» (сырьевых веществ) для образования будущего эмбриона. Эмбриональное развитие животных по времени можно разделить на четыре периода: 1) оплодотворение; 2) образование зародышевого диска; 3) возникновение основной формы; 4) дифференциация органов.

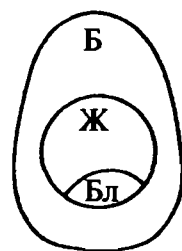


Рис. 3.8. Строение яйца птицы: Б — белок, Ж — желток, Бл — бласто-диск

Для термодинамического анализа процессов в яйце птицы обозначим полную энтропию яйца через S , энтропию зародыша через S_z и энтропию желтка и белка через $S_{\text{ж}}$. Тогда

$$\begin{aligned} S &= S_z + S_{\text{ж}}, \\ dS &= dS_z + dS_{\text{ж}}. \end{aligned} \quad (3.109)$$

С другой стороны, необходимо учитывать баланс между производством и экспортом энтропии:

$$dS = d_e S + d_i S.$$

Так как в зародышевом диске протекают существенно необратимые процессы, членом $d_i S_{\text{ж}}$ можно пренебречь. Обмен с окружающей средой ограничен тепло- и газообменом подсистемы желток—белок через скорлупу, поэтому справедливо соотношение

$$d_e S = \frac{d^* Q}{T} + dS_{\text{газ}}. \quad (3.110)$$

Член $dS_{\text{газ}}$ обусловлен непрерывным поглощением кислорода и выделением двуокиси углерода и водяного пара. В единицу времени величина его составляет

$$dS_{\text{газ}} = s_{\text{O}_2} d_e N_{\text{O}_2} + s_{\text{CO}_2} d_e N_{\text{CO}_2} + s_{\text{H}_2\text{O}} d_e N_{\text{H}_2\text{O}}. \quad (3.111)$$

Как правило, этим членом можно пренебречь по сравнению с теплоотдачей, и поэтому для изменения энтропии зародышевого диска мы получаем

$$dS_z = dS - dS_{\text{ж}} = d_i S + d_e S - dS_{\text{ж}} \approx d_i S_z + \frac{d^* Q}{T} - dS_{\text{ж}}. \quad (3.112)$$

Выражение для производства энтропии в зародышевом диске, соответственно, имеет вид

$$d_i S_3 \approx -\frac{d^* Q}{T} + dS_3 + dS_{\text{ж}}. \quad (3.113)$$

В первом приближении относительно малым членом, соответствующим вариации энтропии, можно пренебречь по сравнению с производством тепла. В результате мы получаем (Prigogine, 1967)

$$d_i S_3 \approx -\frac{d^* Q}{T}. \quad (3.114)$$

Следовательно, по производству энтропии зародыша можно приближенно судить о его теплообмене и метаболизме (обмене веществ). Имеется огромное число калориметрически полученных экспериментальных данных относительно производства тепла живой системой (Zotin, 1974). Типичная зависимость производства тепла как функции времени представлена на рис. 3.9. Это термограмма оплодотворенного яйца земноводного, построенная по измерениям Гайды. На ранней стадии эмбриогенеза наблюдается подъем, который после достижения максимума сменяется монотонным спадом. Интерпретация такого хода кривых была достигнута после ряда дискуссий (Zotin, 1974; Hiernaux, Babloyantz, 1976; Lurie, Wagensberg, 1979), однако окончательную ясность еще предстоит внести последующим исследованиям.

Если интерес представляет временная зависимость энтропии самого эмбриона, то приближение (3.114), разумеется, недопустимо. Экспериментальное определение величины dS_3 по формуле (3.112) затруднительно, поскольку в ее левую часть входит разность двух больших членов. Теоретически следует ожидать, что на изменение энтропии особенно сильное влияние оказывают процессы дифференциации (самоорганизации) и рост биомассы. Чтобы разделить эти два эффекта, исследуем изменение во времени удельной энтропии, отнесенной к единице биомассы:

$$\frac{d}{dt} \frac{S_3}{m} = \frac{1}{m} \frac{dS_3}{dt} - S_3 \frac{1}{m^2} \frac{dm}{dt}. \quad (3.115)$$

Переставляя члены, получаем

$$\frac{dS_3}{dt} = m \frac{d}{dt} \left(\frac{S_3}{m} \right) + \left(\frac{S_3}{m} \right) \frac{dm}{dt}. \quad (3.116)$$

Процесс дифференциации означает уменьшение плотности удельной энтропии (S_3/m) зародыша. Рост соответствует положительной производной dm/dt . Конкретный вид функции роста $m(t)$ рассматривал в числе других Бергаланффи (Bertalanffy, 1960). Анализ соотношения (3.116) показывает, что изменение энтропии зародыша определяется комбинацией положительного (рост) и отрицательного (дифференциация) членов. На этом основании не следует ожидать монотонной зависимости изменения энтропии во времени (Lurie, Wagensberg, 1979).

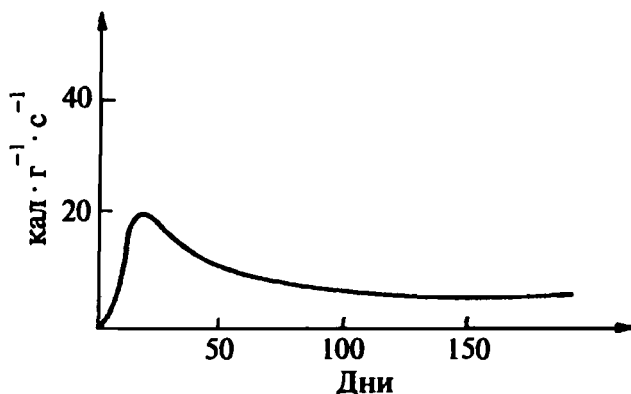


Рис. 3.9. Типичный временной ход производства тепла оплодотворенным яйцом амфибии (Lurie, Wagensberg, 1979)

3.7. Термодинамика экологических систем

Для предмета экологии — «взаимосвязи и взаимодействия живого и окружающей среды на надорганизмической ступени» (Stugren, 1978) — характерны закономерности

особого рода, не сводимые к физическим законам. Вместе с тем физика устанавливает пределы экологического подхода; на нашем современном уровне знаний ни один экологический процесс не привел «к отмене» физических законов. Поэтому при оценке экологических законов мы должны требовать их совместимости с физическими принципами, поэтому физика выступает в роли весьма ценного вспомогательного средства при установлении законов экологии. Экология может также использовать физику еще в одном отношении. Экологические процессы в природе отличаются необычайным разнообразием взаимодействий. Следовательно, при их математическом моделировании неизбежно сильное упрощение и сведение широкого спектра свойств живого организма в единый комплекс, обладающий лишь небольшим числом удобообозримых свойств. Для анализа таких упрощенных систем физика располагает большим запасом подробно разработанных математических моделей, системой понятий и накопила немалый опыт их применения. Поэтому заключения по аналогии с физическими системами, если к ним подходить с надлежащей осторожностью и требовательностью, могут стать ценным источником знаний для экологов.

Наиболее обширной, поистине всеобъемлющей, экологической системой на Земле является биосфера, охватывающая весь живой мир и все места его сосредоточения (Stugren, 1978). Понятие биосферы было введено Зюссом в 1909 г. Биосфера включает в себя гидросферу, осадочные породы земной коры и нижние слои атмосферы. Верхней границей биосферы служит лежащий на высоте около 30 км озонный слой (озоносфера), который играет роль своего рода экрана, предохраняющего жизнь на Земле от губительного ультрафиолетового излучения из космического пространства. Нижняя граница биосферы проходит по дну океана, а на суше — несколькими километрами ниже поверхности Земли. С точки зрения термодинамики биосфера представляет собой открытую систему. Обмен веществ играет в биосфере лишь весьма незначительную роль по сравнению с обменом энергией. Биосфера содержит и неограниченную компоненту — запас сырья (рис. 3.10). Относительно автономная система биосферы называется экосистемой. По Стугрену (Stugren, 1978), «экосистема представляет собой пространственно-временное образование, в котором интегрированы совокупность живых организмов (биоценоз) и биотоп. Такое единство живого и окружающей среды высокого порядка лежит в основе всех экологических систем». Примерами экосистем могут служить дубовый лес, пшеничное поле, деревенский пруд и т. д.

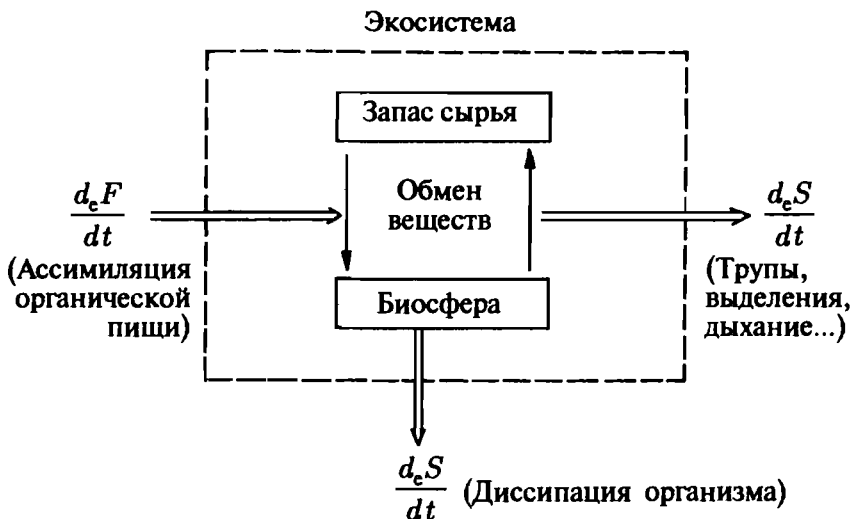


Рис. 3.10. Термодинамическая схема экосистемы

Мы покажем на примерах, как просто могут быть построены экологические модели и как при их исследовании можно использовать методы феноменологической термодинамики. Методы современной термодинамики необратимых процессов впервые, насколько нам известно, применил Мауэрсбергер (*Mauersberger*, 1981). Объектом наших моделей мы выбираем некую открытую заданную экосистему.

Экосистема черпает необходимые вещества из имеющегося запаса, поглощает свободную энергию из солнечного излучения и из органической пищи и возвращает вещества, отдавая энтропию (в основном в виде тепла) (рис. 3.10). При этом организмы и как отдельные особи, и как совокупность постоянно диссипируют энергию, чтобы поддерживать свои (диссипативные) структуры и развиваться. Поэтому для любого живого организма доступ к свободной энергии является центральным вопросом существования; борьба за выживание есть прежде всего борьба за свободную энергию.

Рассмотрим теперь две термодинамические модели такой экосистемы. Пусть наша первая (общая) модель обладает следующими свойствами.

1. Экосистема состоит из химических компонент с числом молей n_k , $k = 1, 2, \dots, s$, и особей, которые могут быть подразделены на конечное число видов $i = 1, 2, \dots, n$. Число N_i особей i -го вида в некотором фиксированном объеме (ареале) V условимся считать переменной величиной. Изменение чисел N_i подчиняется феноменологическим дифференциальным уравнениям

$$\frac{d}{dt}N_i = f_i(N_1, N_2, \dots, N_n), \quad i = 1, 2, \dots, n, \quad (3.117)$$

отражающим действие экологических закономерностей. В основе уравнений (3.117) лежат такие физические требования, как неотрицательность и конечность числа частиц N_i , а также конечность функций f_i для каждой возможной популяционной системы $\{N_1, N_2, \dots, N_n\}$. От требования целочисленности величин N_i мы отказываемся в рамках феноменологической теории. Как правило, феноменологические дифференциальные уравнения содержат также нелинейные члены. Простым примером может служить дифференциальное уравнение Ферхюльста—Перла

$$\frac{dN_i}{dt} = R_i \left(1 - \frac{N_i}{N_i^{(1)}} \right) N_i. \quad (3.118)$$

Здесь R_i — скорость роста i -го вида, $N_i^{(1)}$ — максимально возможное в заданных условиях (например, при заданном запасе пищевых ресурсов) число особей i -го вида. Зависящие от времени решения уравнения (3.118) имеют вид

$$N_i(t) = N_i(0)N_i^{(1)} [N_i(0) + (N_i^{(1)} - N_i(0)) \exp \{-R_i t\}]^{-1}. \quad (3.119)$$

По этому закону роста число особей i -го вида сходится при больших временах $t \rightarrow \infty$ к предельному значению $N_i^{(1)}$. Другим часто встречающимся примером могут служить отношения хищник—жертва между группой видов i (жертвы) и группой видов j (хищники), описываемые дифференциальными уравнениями Лотки—Вольтерры

$$\frac{dN_i}{dt} = \epsilon_i N_i - \sum_j \gamma_{ij} N_i N_j. \quad (3.120)$$

2. Экосфера в целом может удовлетворять условиям баланса энергии и энтропии, т. е. для нее должны выполняться первое и второе начала термодинамики (*Mauersberger*, 1981). Пусть изменения во времени внутренней энергии U и энтропии S

связаны между собой фундаментальным уравнением Гиббса

$$\frac{dU}{dt} = T \frac{dS}{dt} - p \frac{dV}{dt} + \sum_i \tilde{\mu}_i \frac{dN_i}{dt} + \sum_k \mu_k \frac{dn_k}{dt}. \quad (3.121)$$

Здесь μ_k — химический потенциал k -й компоненты (субстрата), $\tilde{\mu}_i$ — химический потенциал k^i -го вида, T — температура, p — давление, V — объем экосистемы. Если определить свободную энергию экосистемы

$$F = U - TS, \quad (3.122)$$

то для изотермического процесса

$$\frac{dF}{dt} = \frac{dU}{dt} - T \frac{dS}{dt}. \quad (3.123)$$

Для формулировки первого и второго начал термодинамики разложим, как обычно, изменения термодинамических потенциалов на внутреннее слагаемое (i) и обменный член (e):

$$\begin{aligned} \frac{dU}{dt} &= \frac{d_i U}{dt} + \frac{d_e U}{dt}, \\ \frac{dS}{dt} &= \frac{d_i S}{dt} + \frac{d_e S}{dt}, \\ \frac{dF}{dt} &= \frac{d_i F}{dt} + \frac{d_e F}{dt}. \end{aligned} \quad (3.124)$$

Тогда первое начало термодинамики сводится к утверждению

$$\frac{d_i U}{dt} = 0, \quad (3.125)$$

соответственно, для изотермического процесса

$$\frac{d_i F}{dt} = -T \frac{d_i S}{dt}.$$

Второе начало термодинамики утверждает, что

$$P = \frac{d_i S}{dt} > 0, \quad (3.126)$$

соответственно, для изотермического процесса

$$\frac{d_i F}{dt} = -TP < 0. \quad (3.127)$$

Знаки равенства в соотношениях (3.126) и (3.127) исключаются потому, что производство энтропии P организмов всегда отлично от нуля. Словесно второе начало термодинамики можно сформулировать следующим образом: организмы за время жизни всегда производят энтропию, т. е. всегда поглощают свободную энергию. Рассмотрим теперь снова энтропию как меру упорядоченности. Для здоровой экосистемы энтропия в среднем не должна возрастать, т. е. должно выполняться неравенство

$$\frac{dS}{dt} \leq 0.$$

Здесь среднее надлежит брать по определенному естественному жизненному периоду — году, суткам и т. д.

С учетом соотношения (3.124) мы получаем условие

$$\left(-\frac{d_e S}{dt}\right) \geq \frac{d_i S}{dt} > 0. \quad (3.128)$$

Это «четвертое начало термодинамики здоровой экосистемы» можно сформулировать следующим образом: экспорт энтропии не должен быть меньше, чем производство энтропии внутри системы. Если процесс происходит при заданной температуре и постоянном объеме или давлении, то условие (3.128) представимо в виде

$$\frac{d_e F}{dt} \geq T \frac{d_i S}{dt} > 0 \quad (3.129)$$

или

$$\frac{d_e G}{dt} \geq T \frac{d_i S}{dt} > 0.$$

Таким образом, изотермические процессы, протекающие в экосистемах, требуют подвода либо свободной энергии, либо свободной энтальпии. Для биосферы в целом единственным существенным источником необходимой свободной энергии или свободной энтальпии является солнечный свет. В этой связи следует заметить, что существующий ныне биологический механизм фотосинтеза использует лишь около 10% имеющейся свободной энергии (Лампрехт, Зотин, 1976). Современная плотность растений (около 1 г/см² поверхности Земли) теоретически может вырасти до $2 \cdot 10^4$ г/см², если в качестве ограничивающего фактора действует только производство энтропии.

В качестве второй более специальной термодинамической модели экосистемы — пруда или леса — мы рассмотрим так называемую «модель раствора», основные идеи которой восходят к теории идеальных растворов Планка. Аналогичную модель параллельно с нами разработал Мауэрсбергер (Mauersberger, 1981). Мы исходим из следующих модельных представлений (Feistel, Ebeling, 1981).

1. Живые организмы погружены в окружающую среду (растворитель), которая находится в термодинамическом равновесии и характеризуется заданной температурой T и заданным химическими потенциалами субстрата $\mu_1, \mu_2, \dots, \mu_s$ (рис. 3.11).

2. Химические компоненты допустимо рассматривать как растворитель с поверхностью, совпадающей с поверхностью рассматриваемой экосистемы с объемом $\tilde{V}$. Пусть $p + \tilde{p}$ — давление внутри экосистемы, p — давление вне $\tilde{p}$. Тем самым экологическое давление $\tilde{p}$ соответствует осмотическому давлению раствора.

3. В окружающей среде (в растворителе) живые организмы ведут себя, как идеальный раствор в смысле теории растворов Планка.

4. Временная зависимость переменных $N_i(t)$ (числа особей i -го вида) определяется из экологических уравнений движения (3.117).

Рассмотрим теперь термодинамические потенциалы нашей экосистемы. Произвольный экстенсивный термодинамический потенциал зависит от числа молей субстрата $n_1, n_2, \dots, n_s$ и числа особей каждого вида $N_1, N_2, \dots, N_n$:

$$Z = Z(n_1, \dots, n_s; N_1, \dots, N_n; p, T). \quad (3.130)$$

Разложим вклад живого организма Z на два слагаемых:

$$Z = Z_0 + \tilde{Z}, \quad (3.131)$$

$$Z_0 = Z(n_1, \dots, n_s; 0, \dots, 0; p, T).$$

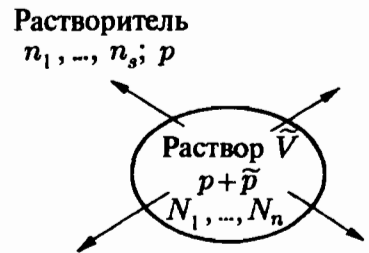


Рис. 3.11. Раствор как модель экологической системы

Если соотношения (3.121) и (3.131) применить к модели, представленной на рис. 3.11, то для подсистемы живого организма мы получаем следующее фундаментальное уравнение Гиббса:

$$\frac{d\tilde{U}}{dt} = T \frac{d\tilde{S}}{dt} - \tilde{p} \frac{d\tilde{V}}{dt} + \sum \tilde{\mu}_i \frac{dN_i}{dt}. \quad (3.132)$$

Так как $\tilde{U} = 0$ при $N_1 = N_2 = \dots = N_n = 0$, величину $\tilde{U}$ при малых плотностях можно представить как линейную функцию чисел N_i :

$$\tilde{U} = \sum_i N_i u_i(n_1, \dots, n_s; p, T). \quad (3.133)$$

Энергию особи можно разложить в сумму кинетической и потенциальной энергии, а также энергии взаимодействия особи с субстратом:

$$u_i = t_i + v_i + w_i.$$

В первом приближении организмы можно рассматривать как тела, находящиеся в тепловом равновесии и обладающие теплоемкостью c_i , т. е. $t_i = c_i T$. Для человеческого организма на долю кинетической энергии приходится величина порядка 10^8 Дж. Примерно такая же величина составляет вклад потенциальной энергии, по существу определяемой запасом химической энергии (потенциальной энергией молекулярных взаимодействий). Потенциальная энергия, зависящая от положения тела в гравитационном поле, примерно на три порядка меньше.

Приближение (3.133) соответствует теории идеальных разбавленных растворов Планка. Для энтропии в этом приближении справедливо выражение

$$\tilde{S} = \sum_i N_i \left[s_i(p, T, n_1, \dots, n_s) - k \ln \frac{N_i}{\tilde{V}} \right]. \quad (3.134)$$

Если ввести полное число живых организмов N и относительную долю видов $q_i = N_i/N$, то энтропия на одну особь принимает следующий вид:

$$\frac{\tilde{S}}{N} = \sum_i q_i s_i - k \sum_i q_i \ln q_i - k \ln \frac{N}{\tilde{V}}. \quad (3.135)$$

Формула (3.135) позволяет проанализировать зависимость энтропии от богатства видов экосистемы. Если все удельные энтропии s_i считать примерно равными ($s_i \approx s$), то максимальное значение энтропии соответствует равномерному распределению $q_i = 1/n$:

$$\left(\frac{\tilde{S}}{N} \right)_{\max} = s - k \ln \left(\frac{N}{\tilde{V}} \right) + k \ln n.$$

Из этой формулы следует также, что максимальное значение энтропии логарифмически растет с числом видов n . Своего минимального значения энтропия достигает, когда в экосистеме доминирует один вид:

$$q_1 \approx 1, \quad q_2 = q_3 = \dots = q_{n-1} = q_n \approx 0.$$

В этом случае

$$\left(\frac{\tilde{S}}{N} \right)_{\min} \approx s - k \ln \left(\frac{N}{\tilde{V}} \right).$$

Однако приведенные выше утверждения нуждаются в существенных поправках, если учесть, что величины s_i для отдельных видов могут сильно отличаться. Максимум энтропии достигается тогда не для равномерного распределения $q_i = 1/n$, а для распределения, отдающего предпочтение видам с большими s_i :

$$q_i = \frac{\exp \{s_i/k\}}{\sum_j \exp \{s_j/k\}}.$$

Разумеется, определить величины s_i для конкретной экосистемы чрезвычайно трудно. Если пренебречь различиями между s_i , то формула (3.135) сводится к выражению, предложенному для энтропии Стугреном (*Stugren*, 1978).

Для свободной энергии получаем

$$\tilde{F} = \tilde{U} - T\tilde{S} = \sum_i N_i \left[c_i T + v_i + w_i + kT \ln \left(\frac{N_i}{\tilde{V}} \right) - Ts_i \right]. \quad (3.136)$$

Наконец, для экологического давления справедлива формула

$$\tilde{p} = -\frac{\partial \tilde{F}}{\partial \tilde{V}} = \sum_i \frac{kTN_i}{\tilde{V}}. \quad (3.137)$$

Это выражение хорошо известно из физики идеальных газов и идеальных растворов. В экосистемах из-за того, что число «частиц» каждого сорта в 10^{20} раз меньше, давление $\tilde{p}$ исчезающе мало и почти не поддается измерению. Для химического потенциала i -го вида получаем

$$\begin{aligned} \tilde{\mu}_i &= \frac{\partial \tilde{F}}{\partial \tilde{N}_i} = \mu_i^{(0)} + kT \ln \left(\frac{N_i}{\tilde{V}} \right), \\ \mu_i^{(0)} &= c_i T + v_i + w_i + kT - Ts_i. \end{aligned} \quad (3.138)$$

Это позволяет записать свободную энергию в обычном виде:

$$\tilde{F} = \sum_i N_i (\tilde{\mu}_i - kT), \quad (3.139)$$

а свободную энтальпию — в виде

$$\tilde{G} = \tilde{F} + \tilde{p} \tilde{V} = \sum_i N_i \tilde{\mu}_i. \quad (3.140)$$

Если перекрыть подвод свободной энергии (энтальпии), то по второму началу термодинамики экосистема стремится к состоянию термодинамического равновесия, соответствующего минимуму свободной энергии (энтальпии):

$$\tilde{F}(T, \tilde{V}, N_1, \dots, N_n) = \min \quad (3.141)$$

или, соответственно,

$$\tilde{G}(T, \tilde{p}, N_1, \dots, N_n) = \min.$$

Как и при рассмотрении химических реакций, при нахождении минимума необходимо соблюдать дополнительные условия на число «частиц» каждого сорта (и всех вместе). В рамках заданной экологической модели все условия на вариацию числа частиц каждого сорта содержатся в экологических уравнениях движения

$$\frac{d}{dt} N_i = f_i(N_1, \dots, N_n). \quad (3.142)$$

С учетом сказанного необходимо исследовать, какие условия следуют из уравнений (3.142) в окрестности термодинамического равновесия. Если прекратить подвод свободной энергии к экосистеме, то экосистема устремится к термодинамическому равновесию, а это означает не что иное, как смерть всего живого. В состоянии термодинамического равновесия могут существовать только такие организмы, которые образуются в результате «спонтанного рождения» из субстрата. Поскольку этот случай практически исключается, мы в дальнейшем всегда предполагаем, что

$$N_i^{\text{равн}} \ll 1. \quad (3.143)$$

При подобном предположении всеми нелинейными членами в правой части уравнения (3.142) можно пренебречь, откуда следует, что в окрестности термодинамического равновесия величины N_i можно варьировать независимо с экологической точки зрения. Пусть

$$\left(\frac{\partial \tilde{F}}{\partial N_i} \right)_{T, \tilde{V}} = \mu_i^{\text{равн}}, \quad \left(\frac{\partial \tilde{G}}{\partial N_i} \right)_{T, p} = \mu_i^{\text{равн}} \quad (3.144)$$

— равновесные значения химических потенциалов. Они определяются химическим составом организмов. Далее находим:

$$\tilde{\mu}_i = \mu_i^{\text{равн}} + kT \ln \left(\frac{N_i}{N_i^{\text{равн}}} \right), \quad (3.145)$$

$$G = \sum N_i \mu_i^{\text{равн}} + kT \sum N_i \ln \left(\frac{N_i}{N_i^{\text{равн}}} \right). \quad (3.146)$$

Число представителей сорта N_i в экосистемах всегда очень велико по сравнению со значениями в состоянии термодинамического равновесия.

Как правило, экосистемы не находятся в состоянии термодинамического равновесия, а далеко выведены из него. Вместо условий термодинамического равновесия типа (3.141) система описывается экологическими уравнениями движения (3.142). Если решить эти уравнения движения, то в рамках «модели раствора» термодинамические функции можно вычислить по формулам (3.133)–(3.140).

Недостаток модели раствора состоит в том, что она не учитывает моторику живых организмов. Для учета этого аспекта можно развить модель газа, на которой мы не будем останавливаться более подробно. В рамках модели газа каждому виду на основе энергии движения ставится в соответствие его собственная экологическая температура. Следует признать, что экологические температуры оказываются достаточно высокими.

Мы видим, что для описания определенных аспектов экосистем могут быть предложены различные термодинамические модели. Но во всех этих моделях экосистемы рассматриваются как диссипативные системы, функционирующие вдали от равновесия. Этим и объясняется значение теории диссипативных структур для экосистем.

В реальных экосистемах для всех встречающихся в них видов всегда выполняется неравенство

$$N_i \gg N_i^{\text{равн}}, \quad (3.147)$$

откуда

$$\tilde{\mu}_i \gg kT, \quad \tilde{G} \gg NkT, \quad \tilde{F} \gg NkT. \quad (3.148)$$

Эти неравенства не выполняются для газов и растворов, изучаемых в физике и химии, поэтому они выражают качественное различие между физико-химическими и экологическими многочастичными системами. Так как в экологических системах,

естественно, всегда встречаются различные виды с $N_i \gg 1$, речь идет о сильно неравновесных условиях, которые могут поддерживаться лишь с помощью экспорта или импорта свободной энергии (энтальпии).

Мы ограничились изложением одной специальной модели и лишь в общих чертах коснулись той большой роли, которую термодинамика играет в современной экологии (*Morowitz, 1978*). Термодинамику часто рассматривают как своего рода логику экологических процессов, которая определяет не то, что происходит в действительности, а то, что может произойти (*Conrad, 1977*). Большой интерес представляют попытки придать с помощью термодинамики точный смысл различным довольно расплывчатым экологическим понятиям, например, понятию экологической эффективности (*Слободкин, 1960; Conrad, 1973, 1977*).

Глава 4

Неустойчивости и пространственно-временные структуры

Природе заглянуть в нутро
Лишь изредка счастливчику дано.

Вильгельм Буш

4.1. Термодинамические неустойчивости

Наличие запаса высокоценной энергии или постоянного экспорта энтропии являются термодинамическими условиями самоорганизации и эволюции. Но «пусковыми кнопками» процессов в каждом конкретном случае служат определенные неустойчивости системы. Новые макроскопические структуры образуются при усилении микроскопических флуктуаций в неустойчивых системах. Николис и Пригожин назвали этот механизм «порядком через флуктуации». Как правило, усиление флуктуаций связано с нарушением определенных симметрий системы. Во второй главе было рассказано о том, как ранние этапы эволюции Метагалактики разворачивались как целая серия нарушающих симметрию неустойчивостей. К числу их относится и нарушение симметрии между материей и антиматерией. Симметрия ранней Метагалактики относительно пространственных трансляций была нарушена лишь на гораздо более поздней стадии эволюции. Первоначально фотоны и материя в виде вещества были распределены равномерно — плотность массы была постоянной. Но со времен Ньютона известно, что однородное распределение гравитирующих масс неустойчиво относительно флуктуаций плотности. Пока температура материи составляет несколько тысяч или даже миллионов Кельвин, эти неустойчивости не работают. Как показывает оценка (2.19), температура через 10^{17} с от начала Вселенной (что соответствует примерно одному миллиарду лет после Большого Взрыва) падает ниже 100 К. В этом холодном газе, в котором кинетическая энергия частиц меньше 0,01 эВ, в игру вступают гравитационные силы притяжения и междуатомные взаимодействия. Материя в виде вещества собирается в «капли».

Процесс образования «капель» из гравитирующей материи напоминает каплеобразование в переохлажденном паре. Как показано в разд. 2.4, конденсация капель и образование обширных водоемов играет большую роль в ранней истории Земли. С высокой вероятностью можно утверждать, что жизнь зародилась в водной фазе. При биогенезе также происходит образование коацерватов или микросфер, т. е. капель с повышенным содержанием неких химических соединений, возникающих в протосупе при разделении веществ, и этот процесс играет важную роль.

Все эти процессы, имевшие огромное значение в ходе эволюции Метагалактики, Земли и жизни, основаны на термодинамических неустойчивостях и могут быть интерпретированы как фазовые переходы первого рода. Термодинамическая первопричина этих процессов состоит в том, что неоднородные пространственно структурированные конфигурации при определенных условиях термодинамически выгоднее, чем однородные конфигурации. Рассмотрим прежде всего условия термодинамической стабильности, следующие из второго начала термодинамики.

1. **Замкнутая система** ($E = \text{const}$, $V = \text{const}$). Из неравенства

$$dS = d_1 S \geq 0 \quad (4.1)$$

следует, что энтропия возрастает или в предельном случае остается постоянной:

$$S \rightarrow \max. \quad (4.2)$$

Условие равновесия имеет вид

$$(\delta S)_{E,V} = 0, \quad (4.3)$$

а условие устойчивости —

$$(\delta^2 S)_{E,V} < 0. \quad (4.4)$$

Замкнутая система устойчива, если каждое конечное изменение состояния приводит к уменьшению энтропии.

2. **Изоэнтропийно-изохорическая система.** Аналогичным образом получаем для энергии условия равновесия и устойчивости

$$E \rightarrow \min, \quad (\delta E)_{S,V} = 0, \quad (\delta^2 E)_{S,V} > 0. \quad (4.5)$$

3. **Изоэнтропийно-изобарическая система.** Для энтальпии

$$H = E + pV$$

следуют условия

$$H \rightarrow \min, \quad (\delta H)_{S,p} = 0, \quad (\delta^2 H)_{S,p} > 0. \quad (4.6)$$

4. **Изотермически-изохорическая система.** Свободная энергия

$$F = E - TS$$

удовлетворяет условиям

$$F \rightarrow \min, \quad (\delta F)_{T,V} = 0, \quad (\delta^2 F)_{T,V} > 0. \quad (4.7)$$

5. **Изотермически-изобарическая система.** Наконец, свободная энтальпия

$$G = H - TS$$

удовлетворяет условиям

$$G \rightarrow \min, \quad (\delta G)_{T,p} = 0, \quad (\delta^2 G)_{T,p} > 0. \quad (4.8)$$

Резюмируя, мы можем утверждать следующее. Если термодинамическая система обладает возможностью к оптимизации термодинамической функции состояния посредством какого-нибудь процесса, например, пространственного разделения на фазы различной плотности или концентрации, то система неустойчива. Запустить оптимизирующий процесс, который, начавшись, протекает спонтанно, могут малые флуктуации. Специальным и особенно важным процессом, который может приводить к оптимизации термодинамических функций, является образование кластеров элементов материи, существующей в виде вещества (атомов, молекул, планетезималей и т. д.). Чтобы проиллюстрировать микроскопические процессы, ведущие к агрегации материи в виде вещества, рассмотрим весьма простую модель (Schweitzer et al., 1988).

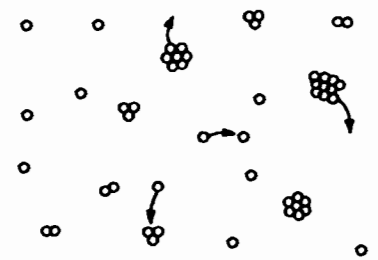
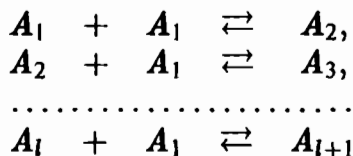


Рис. 4.1. Модель стохастического образования и распада кластеров в изохорической системе

Предположим, что дана система из N элементарных точечных масс A_1 , образующих более крупные кластеры в соответствии с уравнениями реакций



(рис. 4.1). Система поддерживается при постоянном объеме и постоянной температуре и в исходной конфигурации состоит только из мономеров. В предположении идеального перемешивания давление и свободная энергия задаются выражениями

$$p = k_B T \sum_{l=1}^N \left(\frac{N_l}{V} \right), \quad (4.9)$$

$$\begin{aligned} F(T, V, N_1, \dots, N_N) = \\ = \sum_{l=1}^N N_l \left\{ f_l(T) + k_B T \left[\ln \frac{N_l}{V} + \frac{3}{2} \ln \left(\frac{h^2}{2\pi l m k_B T} \right) - 1 \right] \right\}. \end{aligned} \quad (4.10)$$

Вполне разумным представляется следующий выбор:

$$f_1 = 0, \quad f_2 = -a, \quad f_l = -Al + Bl^{2/3} \quad \text{при } l \gg 1. \quad (4.11)$$

Здесь a — энергия связи димера, A — энергия связи, приходящаяся в кластере на один мономер, B — константа, пропорциональная поверхностной энергии. Для модели (4.11) был проведен ряд численных экспериментов (Schweitzer et al., 1988). Вероятность перехода мономера в l -мер в этих экспериментах принята равной

$$l^{2/3} N_l \left(\frac{N_1}{V} \right). \quad (4.12)$$

Вероятности распада определялись из условия, что при $t \rightarrow \infty$ должно устанавливаться химическое равновесие. На рис. 4.2 представлены результаты численного моделирования при

$$T = 290 \text{ К}, \quad N = 150, \quad A = 19,08 k_B T, \quad B = 5,32 k_B T.$$

Начальное пересыщение соответствовало значению 12,5, т. е. со временем давление (и число частиц) падало примерно в 12,5 раз (рис. 4.2). В этих численных экспериментах были обнаружены три этапа перехода от термодинамического равновесия.

1-й этап: за весьма короткое время устанавливается равновесие подкритических кластеров (мономеров, димеров, тримеров, ...).

2-й этап: существенно большее время требуется, чтобы один или несколько кластеров выросли до надкритических размеров.

3-й этап: один или несколько кластеров вступают в конкуренцию и сосредотачивают основную часть массы (оствальдовское созревание).

Свободная энергия как функция числа частиц обладает двумя минимумами. Первый минимум расположен при

$$N_1^{(1)} \approx 70, \quad N_2^{(1)} \approx 25, \quad N_3^{(1)} \approx 10,$$

второй — в окрестности

$$N_1^{(2)} \approx 20, \quad N_2^{(2)} \approx 5, \quad N_{120}^{(2)} = 1.$$

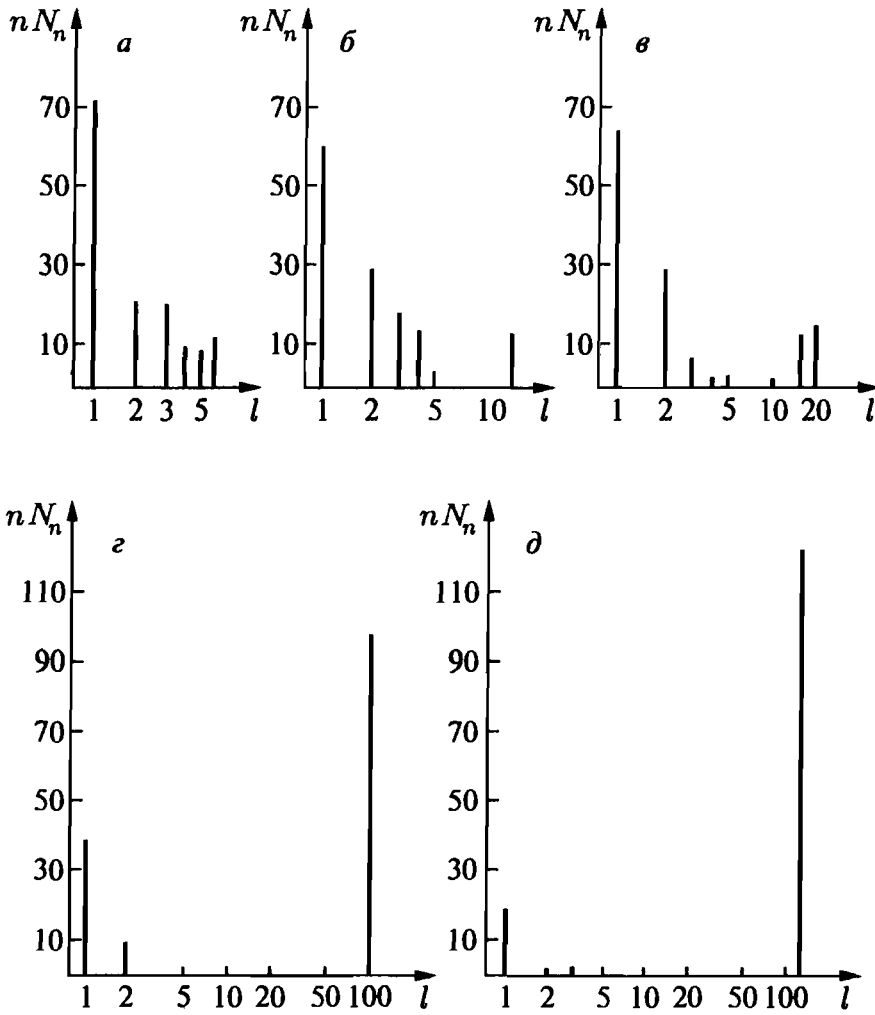


Рис. 4.2. Результаты моделирования (из работы: *Schweitzer et al.*, 1988) образования кластеров в сильно пересыщенном паре ($t = 0$, $p/p' = 12,5$):
 (а) $t = 0,36$, $p/p' = 8,35$; (б) $t = 0,79$, $p/p' = 8,15$; (в) $t = 1,47$, $p/p' = 7,30$;
 (г) $t = 6,75$, $p/p' = 3,17$; (д) $t = 27,25$, $p/p' = 2,17$

Соответственно, распределение вероятностей

$$P(N_1, \dots, N_N) = \text{const} \cdot \exp \left\{ -\frac{F(N_1, \dots, N_N)}{k_B T} \right\} \quad (4.13)$$

обладает двумя максимумами. В процессе релаксации на первом этапе достигается первый минимум свободной энергии и первый максимум распределения вероятности. На втором этапе система проходит седловую точку свободной энергии и распределения вероятности с тем, чтобы на третьем этапе перейти во второй, более глубокий, минимум (соответственно, достичь второго, более высокого, максимума); устанавливается равновесие между популяциями, соответствующими двум экстремумам.

В соответствии с тремя этапами система обладает тремя временами релаксации.

Описанная выше система, состоящая из 150 агрегирующих элементов, носит модельный характер и может в лучшем случае иллюстрировать лишь одну сторону интересующего нас процесса образования пространственно-временных структур вследствие термодинамических неустойчивостей. Альтернативное описание исходит из существования непрерывного параметра порядка

$$n(r, t),$$

зависящего от точки пространства и времени. В однокомпонентных жидких системах параметр порядка можно отождествить, например, с плотностью, а в бинарных системах — с мольной долей (Паташинский, Шумило, 1979; Allen, Cahn, 1979). Для соответствующей термодинамической функции (например, свободной энергии) получаем

$$H[n] = \int dV \left[V(n(r, t)) + \frac{1}{2} D^* (\nabla n(r, t))^2 \right]. \quad (4.14)$$

Здесь первый член описывает вклад локальных плотностей, аналогичный формуле (3.57), а второй член — влияние неоднородности. Далее мы предполагаем в духе линейной термодинамики необратимых процессов Онсагера, что

$$\partial_t n(r, t) = -\frac{1}{\tau} \frac{\delta H[n]}{\delta n(r, t)}, \quad (4.15)$$

где τ — время релаксации, а δ означает вариационную производную. Из соотношения (4.15) следует уравнение типа реакции с диффузией

$$\begin{aligned} \partial_t n(r, t) &= W(n(r, t)) + D \Delta n(r, t), \\ W(n(r, t)) &= -\frac{1}{\tau} \frac{\partial V}{\partial n}, \quad D = \frac{D^*}{\tau}. \end{aligned} \quad (4.16)$$

Здесь $W(n)$ — как правило, нелинейная функция, нули которой

$$n_1, n_2, n_3, \dots$$

соответствуют стационарным и пространственно однородным решениям задачи. Иначе говоря,

$$n(r, t) = n_i \quad (4.17)$$

— решение уравнения (4.16), удовлетворяющее подходящим краевым условиям. Нелинейное дифференциальное уравнение в частных производных (4.16), выведенное для описания термодинамических структур, играет весьма важную роль в физике процессов эволюции. Первое фундаментальное исследование уравнения (4.16) было выполнено в основополагающей работе Колмогорова, Петровского и Пискунова «Изучение уравнения диффузии с источником вещества и его приложение к биологическим проблемам» (1937). Мы еще неоднократно вернемся к этому уравнению, а пока приведем лишь одно его частное решение, описывающее движение фронта между двумя фазами n_1 и n_3 . Предположим, что плотность свободной энергии имеет два минимума и что, соответственно, распределение источников имеет вид, показанный на рис. 4.3.

Решение, соответствующее движущемуся между двумя фазами в направлении положительной полуоси z плоскому фронту, имеет вид

$$n(z, t) = n_0(z - vt). \quad (4.18)$$

Здесь v — скорость фронта, $n_0(z)$ — переходная функция, вид которой схематически представлен на рис. 4.4. Приближенно выполняются следующие соотношения:

$$\begin{aligned} v &\approx v_0 = \frac{1}{2} \sqrt{2Dq(2n_2 - n_1 - n_3)}, \\ \Delta z &\approx l_0 = \frac{4}{q} \sqrt{2Dq(n_3 - n_1)^{-1}}, \\ q &= -\frac{1}{6} W'''(n_2). \end{aligned} \quad (4.19)$$

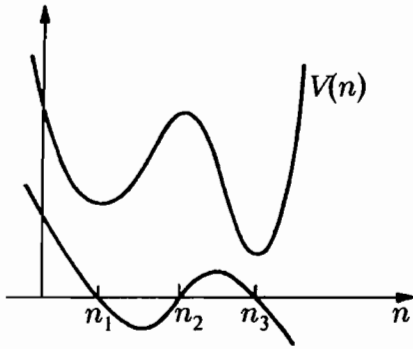


Рис. 4.3. Термодинамический потенциал и функция источника уравнения реакции с диффузией при изменении параметра порядка во времени и в пространстве

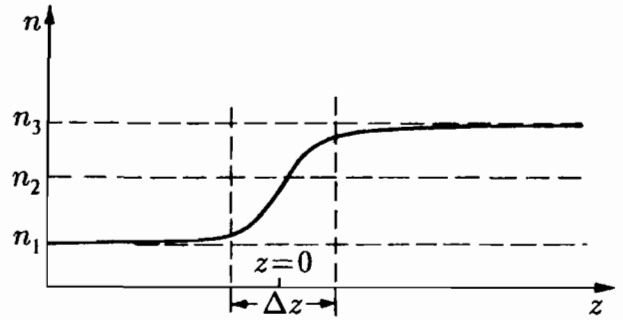


Рис. 4.4. Качественный ход расплывания плоского фронта, описывающего распространение границы фаз

Величины v_0 и l_0 соответствуют скорости фронта и ширине фронта для случая, когда W — полином 3-й степени. Приближения (4.19) могут, таким образом, быть применены в том случае, если W между нулями, по крайней мере приближенно, можно представить с помощью кубической зависимости. Для сферического ($d = 3$) или цилиндрического ($d = 2$) фронтов положение фронта описывается решением дифференциального уравнения

$$\dot{R} = v - (d-1)\frac{D}{R}. \quad (4.20)$$

Существует критический радиус, соответствующий остановке фронта

$$R_k = \frac{(d-1)D}{v}. \quad (4.21)$$

Локально распространение фронта в точке r можно описать с помощью локальной кривизны $K = R^{-1}$ и локального вектора нормали N , задав зависимость, соответствующую кубическому закону (Ebeling, 1986):

$$n(r, t) = n_1 + (n_3 - n_1) \left[1 + \exp \left\{ \frac{4}{l_0} [Nr - v_0 t + (d-1)DtK] \right\} \right]^{-1}. \quad (4.22)$$

Более точный анализ исходит из разложения в ряд по теории возмущений относительно предельного случая сильной нелинейности (т. е. больших W) (Diehl et al., 1980; Bausch et al., 1981; Kawasaki, Ohita, 1982; Engel et al., 1986). Введем переменную z — среднее направление распространения фронта и переменную x , отсчитываемую в перпендикулярном направлении (рис. 4.5). Тогда в главном порядке теории возмущений мы получаем для локального профиля фронта выражение

$$n(x, z, t) = n_0 \left[\frac{z - f(x, t)}{\sqrt{1 + (\nabla f)^2}} \right], \quad (4.23)$$

где

$$\frac{\partial_t f(x, t)}{\sqrt{1 + (\nabla f)^2}} = v + DK,$$

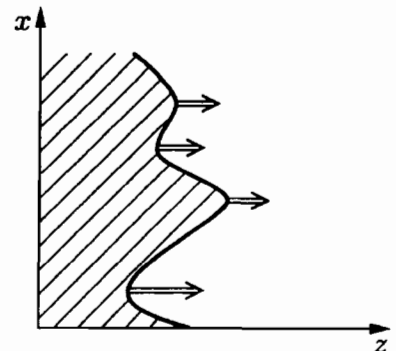


Рис. 4.5. Зависимость локальной скорости фронта от кривизны в случае искривленного фронта

причем

$$K = \nabla \frac{\nabla f}{\sqrt{1 + (\nabla f)^2}} \quad (4.24)$$

снова означает среднюю кривизну.

Как показывают приведенные выше уравнения, движение фронта между двумя фазами может иметь весьма сложный характер. В невозмущенном случае участок фронта, опережающий остальные, тормозится, а отстающий участок ускоряется, в результате чего в среднем наблюдается эффект сглаживания фронта (рис. 4.5). Влияние регулярных или стохастических, неупорядоченных возмущений можно учесть, включив в правую часть уравнения (4.16) дополнительные источники. Ряд авторов исследовал влияние таких возмущений (Ефетов, Ларкин, 1977; Nattermann, 1982, 1984; Mikhailov et al., 1983; Schimansky-Geier et al., 1983; Фейгельман, 1983; Гуревич, Минтц, 1984; Engel, Ebeling, 1987).

Мы изложим здесь кратко и в самых общих чертах основы теории движения фронтов фаз, поскольку такие движения фронтов представляют большой интерес для различных важных процессов эволюции. Возможные приложения простираются вплоть до космологических моделей (Coleman, 1977; Guth, 1983). Обобщения теории охватывают и гравитационные неустойчивости. Например, на этом пути удастся понять, как возникают паутинообразные структуры, которые были обнаружены в распределении галактик (Зельдович и Новиков, 1975; Новиков, 1979, 1983; Гуревич, Чернин, 1987).

4.2. Кинетические неустойчивости в жидких средах

В предыдущем разделе мы рассматривали только пассивные системы без внешней накачки. Структурообразование в таких системах основано на запасе высокоценной энергии, «хранилищем» которой служат начальные условия. Для процессов эволюции интерес представляют только такие начальные условия, которые весьма далеки от термодинамического равновесия.

Рассмотрим теперь жидкие системы, получающие высокоценную энергию за счет внешней накачки и находящиеся вдали от равновесия. Такие процессы сыграли и продолжают играть важную роль в эволюции галактик, звезд, планет, а также наружных слоев последних. Действующей силой в этом случае прежде всего выступают локальные градиенты давления, плотности, концентрации или температуры, а также локальные электромагнитные и гравитационные поля. В дальнейшем мы сначала ограничимся несжимаемыми жидкостями. На рис. 4.6 показаны три возможных варианта внешней накачки:

1) протягивание пластины по жидкой пленке индуцирует плоское течение Куэтта;

2) между вращающимся и неподвижным цилиндрами возникает течение Тейлора (цилиндрическое течение Куэтта);

3) разность давлений, приложенных к трубе, порождает течение Пуазейля.

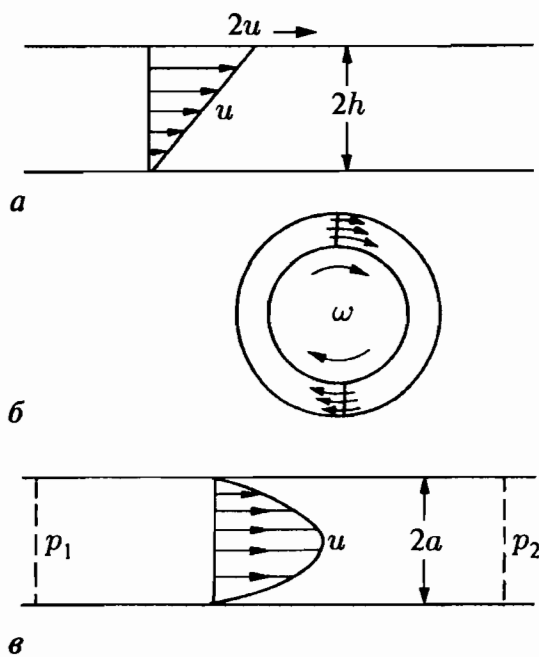


Рис. 4.6. Примеры конвективных систем, обнаруживающих неустойчивость при закритических вынуждающих силах: (а) плоское течение Куэтта; (б) течение Тейлора; (в) течение Пуазейля

Профиль течения во всех трех случаях получается при подходящих краевых условиях из уравнения Навье—Стокса

$$\partial_t u_i + u_k \partial_k u_i = -\rho^{-1} \partial_i p + \nu \partial_k (\partial_k u_i + \partial_i u_k), \quad (4.25)$$

следующего из уравнений (3.66).

Частные решения для течений Куэтта и Пуазейля имеют следующий вид:

$$u(z) = \frac{uz}{h}, \quad (4.26)$$

$$u(r) = u \left(1 - \frac{r^2}{a^2} \right). \quad (4.27)$$

Устойчивость течений Куэтта и Пуазейля определяется величиной числа Рейнольдса (Reynolds):

$$\text{Re} = \frac{ua}{\nu}, \quad \nu = \frac{\eta}{\rho}. \quad (4.28)$$

Плоское течение Куэтта, описываемое решением (4.26), устойчиво только при

$$\text{Re} \leq \text{Re}_{\text{крит}} \approx 1800. \quad (4.29)$$

Выше критического значения развивается турбулентное течение. Описываемое решением (4.27) течение Пуазейля устойчиво только при

$$\text{Re} \leq \text{Re}_{\text{крит}} \approx 2300 \quad (4.30)$$

и при больших числах Рейнольдса переходит в турбулентный режим. Турбулентный режим также допускает описание с помощью уравнения Навье—Стокса (4.25), но поле скоростей течения $u(r, t)$ должно быть случайным полем $\tilde{u}(r, t)$:

$$u(r, t) = [\tilde{u}(r, t)]. \quad (4.31)$$

Коэффициенты переноса, например, кинематическую вязкость, следует заменить их эффективными значениями:

$$\nu \rightarrow \nu_t = \nu + x, \quad (4.32)$$

причем дополнительные члены связаны с корреляторами случайных скоростей (Ландау, 1988; Ebeling, Klimontovich, 1984).

Рассмотрим теперь неустойчивость течения между вращающимися цилиндрами, открытую в 1923 г. одним из наиболее выдающихся гидродинамиков Тейлором (Chandrasekhar, 1961). Если пространство между двумя цилиндрами, из которых внутренний вращается с угловой скоростью ω , заполнить жидкостью, то при малых ω образуется ламинарное течение Куэтта. Выше некоторой критической угловой скорости ω_k наступает так называемая неустойчивость Тейлора, и вдоль продольной оси возникают периодически упорядоченные вихревые кольца (рис. 4.7), образующие правильный «узор». При больших значениях безразмерного числа Тейлора (Taylor)

$$\text{Ta} = \frac{(R_2 - R_1)R_1\omega}{\nu} \quad (4.33)$$

(R_1, R_2 — радиусы цилиндров) сначала наблюдаются волнообразные вихревые течения в азимутальном направлении, а затем, при превышении более высокого критического значения, — турбулентность (Yahata, 1978). Аналогичная картина наблюдается и в классической задаче об обтекании цилиндра (рис. 4.8). После превышения критического значения числа Рейнольдса $\text{Re} = dv/\nu$ (d — диаметр цилиндра, v — скорость) сначала образуется стоячая пара вихрей. Затем, при превышении более

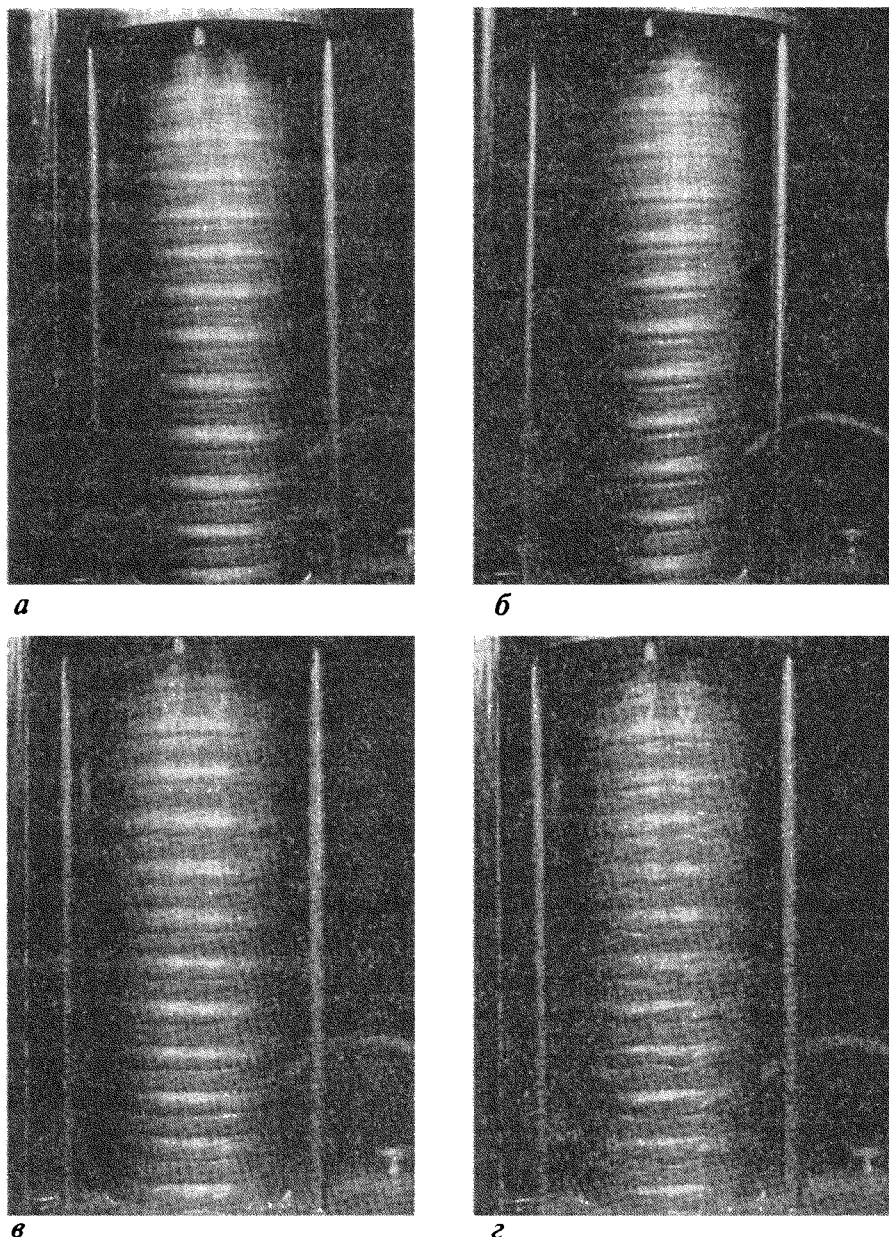


Рис. 4.7. Тейлоровская неустойчивость течения Куэтта для внутреннего цилиндра, вращающегося с угловой скоростью Ω внутри неподвижного цилиндра (а) $\Omega = 3,3 \text{ с}^{-1}$, (б) $\Omega = 8,6 \text{ с}^{-1}$, (в) $\Omega = 9,9 \text{ с}^{-1}$, (г) $\Omega = 19,2 \text{ с}^{-1}$

высокого критического значения вихри начинают попеременно осциллировать и, наконец, совершается переход к турбулентности. Спектр критических значений в гидродинамических проблемах всегда может быть найден как собственные значения корректно поставленных краевых задач, т. е. как решение уравнения Навье—Стокса при заданных краевых условиях. Характерная структура спектра критических значений представлена на рис. 4.9 (Ruelle, Takens, 1971; McLaughlin, Martin, 1975). На термодинамической ветви, как правило, появляется область высокоорганизованных ячеистых структур (течений жидкости), а при все возрастающем отклонении от равновесия — область колеблющихся структур и, наконец, при очень больших отклонениях от равновесия — область хаотических турбулентных движений. Простые математические модели с аналогичными спектрами были предложены Хакеном (Haken, 1980), Лоренцом (Lorenz, 1963) и Мэем (May, 1974). Из рассмотренных гидродинамических примеров следует, что кооперативное движение молекул жидкости может порождать высокоорганизованные структуры лишь в определенной области

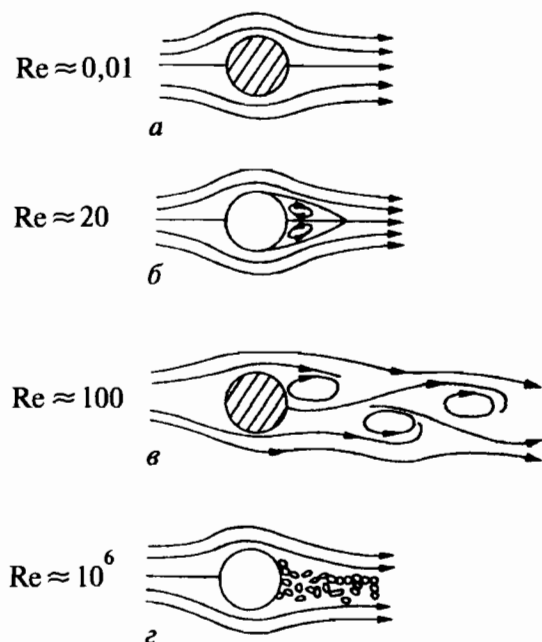


Рис. 4.8. Течение жидкости вокруг цилиндра: (а) ламинарное течение; (б) пара вихрей; (в) осциллирующий вихрь; (г) турбулентность

(рис. 4.12). При надкритических значениях разности температур режим теплопроводной покоящейся жидкости становится неустойчивым (штриховая линия на рис. 4.12) и сменяется устойчивым режимом конвективных ячеек. При больших разностях температур покоящаяся жидкость оказывается не в состоянии обеспечивать соответствующий интенсивный перенос тепла; поэтому теплопроводность уступает место более эффективному конвективному режиму. По сравнению с однородным распреде-

отклонений от равновесия. Типичным примером возникновения диссипативных структур вследствие тепловых неустойчивостей являются так называемые ячейки Бенара. Если слой жидкости (например, силиконового масла) сильно нагревать снизу (рис. 4.10), то между нижней поверхностью и верхней поверхностью возникает перепад температур ΔT , причем $T_1 > T_2$. При малой (подкритической) разности температур $(T_1 - T_2) < \Delta T_{\text{крит}}$ слой жидкости находится в покое, и перенос тепла осуществляется за счет механизма теплопроводности. Выше критической разности температур $(T_1 - T_2) > \Delta T_{\text{крит}}$, по Бенару, устанавливается конвекция, при этом образуется замкнутая система гексагональных ячеек (рис. 4.11).

Если в единицу времени с нижней поверхности к верхней поверхности должен быть перенесен суммарный поток тепла q , то зависимость его от разности температур имеет кусочно-линейный график с характерным изломом



Рис. 4.9. Спектр критических отклонений от равновесия для гидродинамических проблем

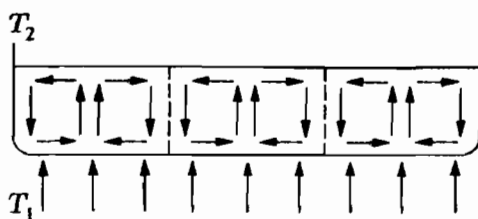


Рис. 4.10. Конвективные течения при подгреве слоя жидкости в форме ячеек-валов

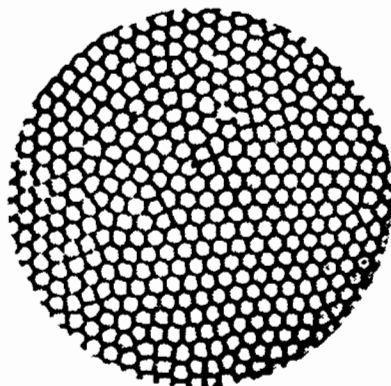


Рис. 4.11. Ячейки Бенара в спермацетовом масле (по оригинальной фотографии Бенара)



Рис. 4.12. Характерный излом в количестве переносимого тепла при переходе от теплопроводности к конвекции

лением скоростей конвективные ячейки представляют собой высокоорганизованную структуру, порожденную кооперативными движениями молекул жидкости.

Так как наша система обменивается с окружающей средой только теплом, полный поток энтропии через граничную поверхность системы определяется величиной

$$\frac{d_e S}{dt} = \frac{q}{T_1} - \frac{q}{T_2} = q \frac{T_2 - T_1}{T_2 T_1} < 0. \quad (4.34)$$

Следовательно, наша система действительно экспортирует энтропию, причем в стационарном случае ровно в таком объеме, в котором энтропия производится (за счет потерь на теплопроводность и трение) внутри системы.

При увеличении отклонения от равновесия подогреваемый снизу слой жидкости претерпевает ряд кинетических превращений (Chandrasekhar, 1961; Kahrig, Besserdich, 1977). В качестве меры отклонения от равновесия принято использовать безразмерное число Рэлея (Rayleigh)

$$Ra = \frac{g \alpha \beta h^4}{\kappa \nu}, \quad (4.35)$$

где g — ускорение силы тяжести, h — толщина слоя, β — градиент температуры, α — коэффициент объемного расширения, κ — коэффициент теплопроводности, ν — кинематическая вязкость. В качестве меры отклонения от линейного режима вводится число Нуссельта (Nusselt)

$$Nu = \frac{|q|}{|q_0|}. \quad (4.36)$$

Оно указывает отношение полного теплового потока к той части теплового потока, которая обусловлена чистой теплопроводностью. На рис. 4.13 показана зависимость числа Нуссельта от числа Рэлея в широком диапазоне по изменениям Силверстона. Нетрудно видеть, что первое критическое значение числа Рэлея лежит в окрестности $Ra \sim 1700$. Оно соответствует возникновению многоугольных ячеек, представленных на рис. 4.11. Следующее критическое значение числа Рэлея лежит вблизи $Ra \sim 3000$. При этом значении многоугольные ячейки переходят в цилиндрические валы (рис. 4.14). При следующем критическом значении, расположенном примерно при $Ra \sim 30000$, наблюдается появление дополнительной системы течений, направленных под прямым углом к конвективным валам (рис. 4.14). Зависимость производства энтропии от вызывающей его силы (числа Рэлея) представлена на рис. 4.15 (Linde, 1983). Мы видим, что все критические значения числа Рэлея соответствуют излому на графике производства энтропии. При этом после излома следующий участок отклоняется от предыдущего как вверх, так и вниз, т. е. производство энтропии может быть и больше, и меньше, чем аналитическое продолжение предыдущего звена (штриховой линией на рис. 4.15). Отчетливо видно, что в нелинейной области производство энтропии не следует никакому экспериментальному принципу.

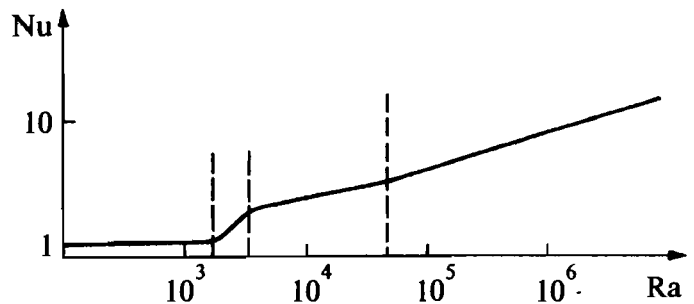


Рис. 4.13. Многократные изломы в зависимости числа Нуссельта от числа Рэлея при критических значениях последнего при качественных изменениях механизма переноса тепла

Более подробный анализ эффекта Бенара показал, что в нем большую роль играет также неустойчивость Марангони (Kahrig, Besserdich, 1977). Эта неустойчивость

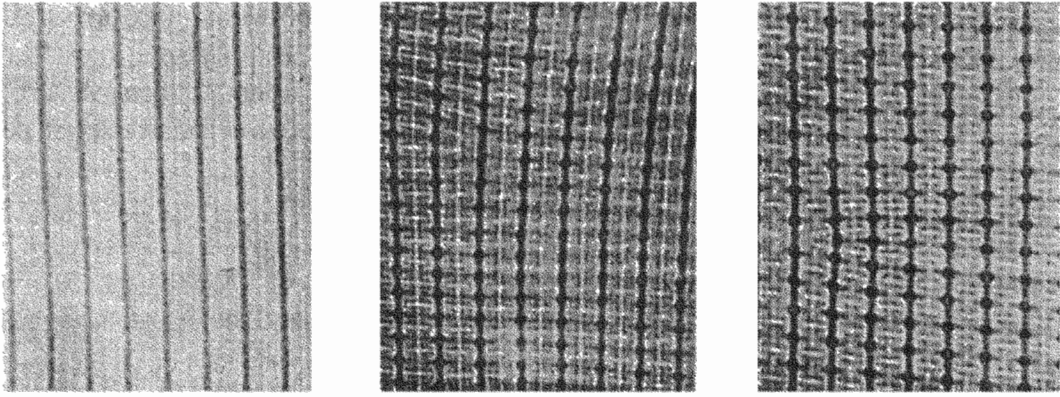


Рис. 4.14. Цилиндрические конвективные валы (слева — $Ra = 10\,000$) и наложение на них валов с перпендикулярными осями (в середине — $Ra = 25\,000$, справа — $Ra = 35\,000$) (Chen, Whitehead, 1968)

связана с силами поверхностного натяжения на границе раздела фаз жидкость—газ, которые вследствие зависимости поверхностного натяжения от концентраций и температуры могут приводить к образованию конвективных ячеек (Linde, Schwarz, 1963; Linde, Winkler, 1964).

Критический анализ неустойчивостей, возникающих при эффекте Бенара, был проведен Веларде (Velarde, 1976). Неустойчивость Рэлея проявляется в чистом виде только в том случае, если верхняя поверхность слоя жидкости ограничена твердой пластинкой. В этом случае, как показал Веларде, возникают только конвективные валы. Многоугольные ячейки образуются только тогда, когда неустойчивость отчасти обусловлена поверхностным натяжением (неустойчивостью Марантони). Такого рода поведение течения соответствует в обычных экспериментах (слой жидкости ограничен сверху воздухом) эффекту Бенара. Недавние экспериментальные исследования системы Бенара, заключенной между двумя горизонтальными твердыми пластинами, выявили ряд существенных обстоятельств, но ряд вопросов по-прежнему остаются открытыми (Ahlers, Behringer, 1978). Установлено, что конвекция всегда начинается при критическом значении Рэлея $Ra = 1708$. При больших числах Рэлея появляется еще один дополнительный фактор, влияющий на характер течения, — безразмерное число Прандтля $Pr = (\nu/\kappa)$, где ν — кинематическая вязкость, κ — коэффициент теплопроводности. Кроме того, вид течения определяется условиями на боковых торцах. В частности, условия на боковых торцах оказывают влияние на переход к непериодическим турбулентным течениям, возникающим при больших числах Рэлея. Алерс и Берингер (Ahlers, Behringer, 1978) провели измерения над нормально-текущим He^4 в цилиндрических сосудах при различных значениях так называемого аспектного отношения Γ (Γ — радиус системы / высота слоя).

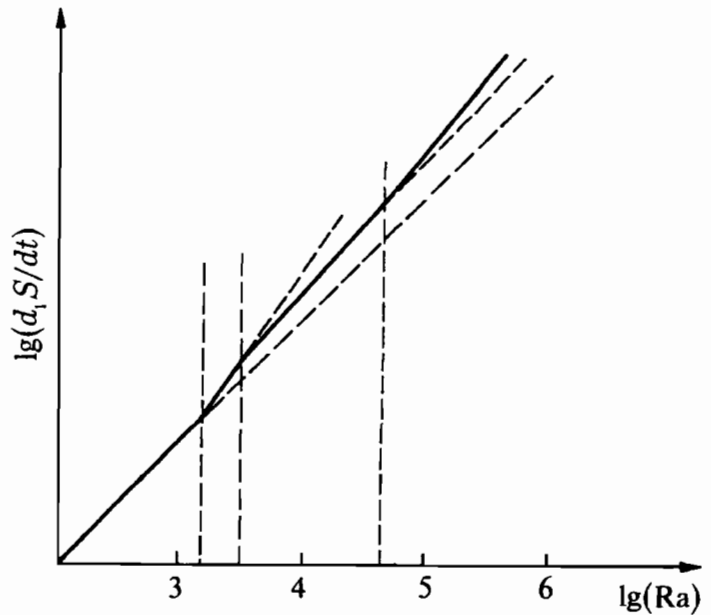


Рис. 4.15. Изломы в производстве энтропии при практических значениях числа Нуссельта (изломы происходят как вверх, так и вниз) Как видно, производство энтропии в этом случае не удовлетворяет экстремальному принципу

экспериментах (слой жидкости ограничен сверху воздухом) эффекту Бенара. Недавние экспериментальные исследования системы Бенара, заключенной между двумя горизонтальными твердыми пластинами, выявили ряд существенных обстоятельств, но ряд вопросов по-прежнему остаются открытыми (Ahlers, Behringer, 1978). Установлено, что конвекция всегда начинается при критическом значении Рэлея $Ra = 1708$. При больших числах Рэлея появляется еще один дополнительный фактор, влияющий на характер течения, — безразмерное число Прандтля $Pr = (\nu/\kappa)$, где ν — кинематическая вязкость, κ — коэффициент теплопроводности. Кроме того, вид течения определяется условиями на боковых торцах. В частности, условия на боковых торцах оказывают влияние на переход к непериодическим турбулентным течениям, возникающим при больших числах Рэлея. Алерс и Берингер (Ahlers, Behringer, 1978) провели измерения над нормально-текущим He^4 в цилиндрических сосудах при различных значениях так называемого аспектного отношения Γ (Γ — радиус системы / высота слоя).

Дно ячейки поддерживалось при постоянной температуре в несколько Кельвинов. Тепловой поток устанавливается постоянным, измеряются (с высоким разрешением — около 10^{-4}) возникающая флуктуирующая разность температур ΔT как функция времени. Наблюдаемый спектр ΔT позволяет сделать обратные заключения о спектре движения жидкости. Для ячейки с $\Gamma = 2,08$ в области $Ra_{\text{крит}} < Ra < 7,4 Ra_{\text{крит}}$ временные флуктуации не наблюдались. В области $7,4 Ra_{\text{крит}} < Ra < 9,9 Ra_{\text{крит}}$ возникает колебательный гармонический режим с одной основной частотой и с кратными частотами. В интервале между $Ra = 9,9 Ra_{\text{крит}}$ и $Ra = 10,97 Ra_{\text{крит}}$ существует колебательный режим с двумя основными частотами, комбинационной частотой и кратными частотами. При дальнейшем увеличении числа Рэлея Ra возникает третий колебательный режим, но спектр по-прежнему остается дискретным. Выше $Ra_i = 11,01 Ra_{\text{крит}}$ линии уширяются, и мы оказываемся в области широкополосного шума, который можно интерпретировать как наступление турбулентного движения. Для ячеек с большим аспектным отношением $\Gamma = 57$ Алерс и Берингер обнаружили широкополосный шум при всех числах Рэлея, лежащих выше критического значения $Ra_{\text{крит}}$. Таким образом, турбулентное число Ra_i при этих условиях совпадает с критическим числом $Ra_{\text{крит}}$. В случае конечного аспектного отношения на термодинамической ветви (чистая теплопроводность) сначала возникает режим стационарных конвективных ячеек, затем при переходе через следующее критическое значение — колебательный режим и, наконец, турбулентный режим.

Обратимся теперь к примерам из области физики плазмы. Из-за огромного числа разнообразных электромагнитных взаимодействий плазма является средой, особенно богатой нелинейными неравновесными структурами (Климонтович, 1975, 1995, 1999; Davidson, 1972; Кадомцев, 1976; Ebeling et al., 1987). Существует целый ряд волновых и шумовых процессов, которые могут возбуждаться в плазме. При этом возникающие процессы могут быть как регулярными, так и турбулентными — при возбуждении очень большого числа степеней свободы (Райзер, 1980; Велихов, Ковалёв, Рахимов, 1987). Особенно характерным примером является возбуждение плазменных волн электронными пучками (Leven, 1969, 1970; Karpman, 1977; (см. рис. 4.16)).

Рассмотрим плазму с ионной плотностью $n = \text{const}$ и электронной плотностью $n' = n'(x, t)$. В предположении малости возмущения электронной плотности мы можем записать уравнение движения, уравнения Пуассона и уравнение непрерывности в следующем виде:

$$\begin{aligned} m\partial_t v &= eE - \left(\frac{k_B T_e}{n} \right) \partial_x n' - m\zeta v, \\ \partial_x E &= 4\pi e n', \\ \partial_t n' + n\partial_x v &= 0. \end{aligned} \quad (4.37)$$

В уравнениях (4.37) T_e — электронная температура, ζ — частота столкновений нейтральных частиц. Исключая v и E из уравнений (4.37), получаем

$$\partial_t^2 n' + \zeta \partial_t n' + \omega_{\text{пл}}^2 n' = \frac{k_B T_e}{n} \partial_x^2 n'. \quad (4.38)$$

Здесь $\omega_{\text{пл}} = (4\pi n e^2 / m)^{1/2}$ — плазменная частота. Мы видим, что возмущение электронной плотности удовлетворяет телеграфному уравнению. Если пренебречь



Рис. 4.16. Схема генерации плазменных волн в электронном пучке, проникающем в плазму

столкновениями, то это уравнение допускает волновое решение, соответствующее продольной синусоиде в направлении оси x :

$$n'(x, t) = A \sin(\omega t - kx). \quad (4.39)$$

Дисперсионное уравнение таких, открытых Тонксом и Ленгмюром в 1929 г., плазменных волн, имеет вид:

$$\omega^2 = \omega_{\text{пл}}^2 + v_e^2 k^2 \quad (4.40)$$

(v_e — тепловая скорость электронов).

Как показали Ромпе и Штеенбек (*Rompe, Steenbeck, 1939*), наблюдаемые плазменные волны с длинами волн из сантиметрового диапазона соответствуют приведенным выше уравнениям. Учет взаимодействий при столкновениях приводит к затуханию волн. В том же направлении действует и так называемое затухание Ландау. Для генерации плазменных волн в системе необходима накачка энергии. Осуществить ее может электронный пучок, инжектируемый в плазму со скоростью v_1 . Распределение скоростей электронов в плазме становится при этом таким, как показано на рис. 4.17. Электроны пучка порождают дополнительный максимум распределения скоростей при $v = v_1$, который накладывается на распределение Максвелла в основной плазме. Рассмотрим теперь продольную плазменную волну в направлении оси x , распространяющуюся со скоростью $v_{\text{пл}}$. Электроны плазмы вступают во взаимодействие с электрическим полем волны. Электроны, движущиеся несколько быстрее волны, «подталкивают» волну и отдают ей энергию, в то время как более медленные электроны увлекаются волной и отбирают часть ее энергии. Обычно из-за убывающего характера распределения скоростей медленных электронов больше, чем быстрых, что и приводит к затуханию Ландау. Однако встречаются и такие плазменные волны, скорость которых чуть меньше v_1 (рис. 4.17). В этом случае вследствие возрастающего характера распределения быстрых электронов больше, чем медленных. Такой механизм приводит к генерации незатухающих плазменных волн, соответствующих структурообразованию за счет коллективного движения электронов.

Следующий пример, заимствованный из физики твердого тела, — эффект Ганна — также основан на плазменной неустойчивости. Рассмотрим специальный полупроводник с двумя зонами проводимости, минимумы энергии которых расположены в различных местах. Верхняя зона соответствует большей эффективной массе электронов и меньшей их подвижности (рис. 4.18). Обычно электроны, возникающие при возбуждении доноров (n_d — плотность доноров), сначала заполняют нижнюю зону проводимости. Если к нашему полупроводнику приложить электрическое поле, то скорость электронов возрастает с интенсивностью поля сначала линейно (рис. 4.19). При более высоких интенсивностях поля все большее число электронов туннелирует в верхнюю зону проводимости, где они в силу

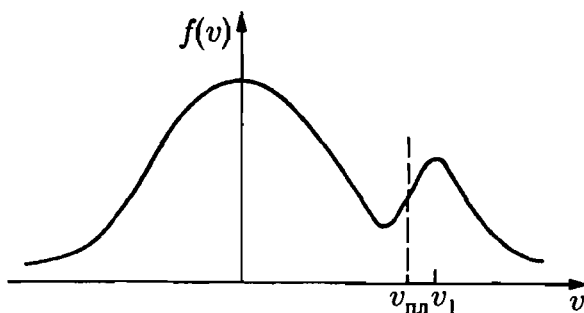


Рис. 4.17. Распределение скоростей в плазме, порождаемое плазменной волной, распространяющейся со скоростью $v_{\text{пл}}$

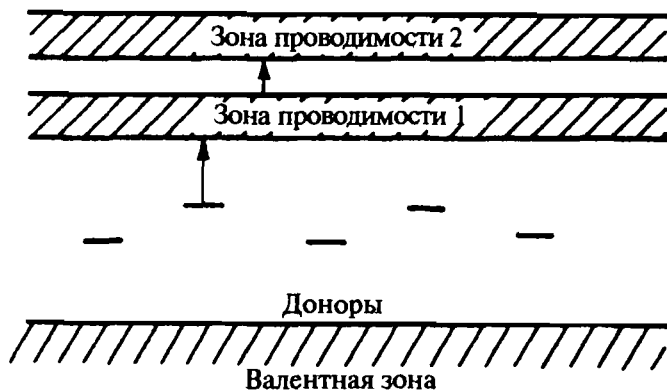


Рис. 4.18. Эффект Ганна: схема структуры полос в полупроводнике

электрическое поле, то скорость электронов возрастает с интенсивностью поля сначала линейно (рис. 4.19). При более высоких интенсивностях поля все большее число электронов туннелирует в верхнюю зону проводимости, где они в силу

меньшей подвижности могут развивать меньшую скорость. В результате средняя скорость электронов с увеличением интенсивности поля падает. Характерная зависимость для электронов $v_0(E)$ представлена на рис. 4.19. Если учесть еще диффузию электронов с коэффициентом диффузии D , то для тока справедливо выражение:

$$J = nv = nv_0(E) - D \partial_x n. \quad (4.41)$$

По аналогии с системой (4.37) запишем теперь уравнение Пуассона и уравнение непрерывности

$$\begin{aligned} \partial_x E &= \frac{4\pi e}{\varepsilon} (n - n_d), \\ \partial_t n + \partial_x (nv) &= 0. \end{aligned} \quad (4.42)$$

Дифференцируя уравнение Пуассона по времени и исключая плотность, получаем

$$\partial_x \left[\partial_t E + \frac{4\pi e}{\varepsilon} n_d v_0(E) + v_0(E) \partial_x E - D \partial_{xx} E \right] = 0. \quad (4.43)$$

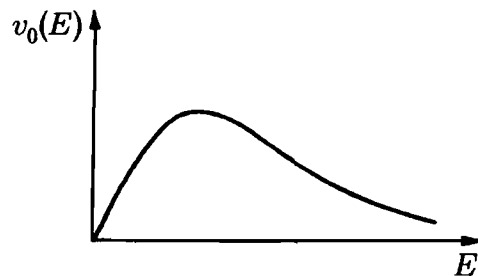


Рис. 4.19. Эффект Ганна: зависимость скорости от напряженности поля для полупроводника

Это уравнение обладает решениями, соответствующими, как показал Хакен (Хакен, 1980), незатухающим волнам. Генерация таких волн известна под названием эффекта Ганна. Экспериментально неустойчивость Ганна наблюдается, например, в монокристаллах Ge и Ga—As. Множество плазменных неустойчивостей столь разнообразно, что мы поневоле вынуждены ограничиться лишь малой их толикой. Плазменные неустойчивости играют существенную роль в процессах эволюции, но из-за необычайной сложности процессов исследованы лишь частично (Альфен, Аррениус, 1987).

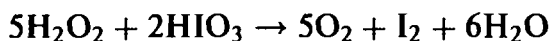
4.3. Кинетические неустойчивости в химических системах

В системах химических реакций особенно легко реализуются условия образования структур вдали от равновесия, так как эти системы, вообще говоря, характеризуются сильными нелинейностями внутренней динамики. Поэтому в тех особых внешних условиях, которые позволяют поддерживать реакцию вдали от равновесия, можно ожидать временной или пространственной самоорганизации. При этом принципиально существуют две возможности:

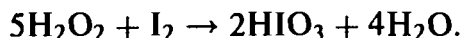
- 1) система в силу начальных условий находится вдали от равновесия и обладает большим запасом высококоценной химической энергии;
- 2) в систему (например, в проточном реакторе) непрерывно поступает приток высококоценной химической энергии.

Еще в XIX веке химикам были известны примеры колебательных реакций (Franck, 1978). В 1828 г. Фехнер сообщил о периодическом процессе на электродах, а в 1900 г. Оствальд опубликовал две знаменитые работы по колебаниям потенциала при коррозии хрома в соляной кислоте и железа в азотной кислоте. Оствальд указал на замечательную аналогию между электрохимическими и электрофизиологическими колебательными процессами и тем самым впервые обратил внимание на связь, вызывающую ныне большой интерес. Долгое время колебательные реакции, как правило, считались своего рода курьезами, кажущимися нарушениями второго начала термодинамики. В то время как в более старых примерах речь шла о гетерогенных реакциях с далеко не ясной кинетикой, Брею в 1921 г. удалось исследовать колебательную реакцию в гомогенном растворе H_2O_2 , HIO_3 и H_2SO_4 при температуре около $60^\circ C$ (рис. 4.20). Реакционный механизм основан на окислительном

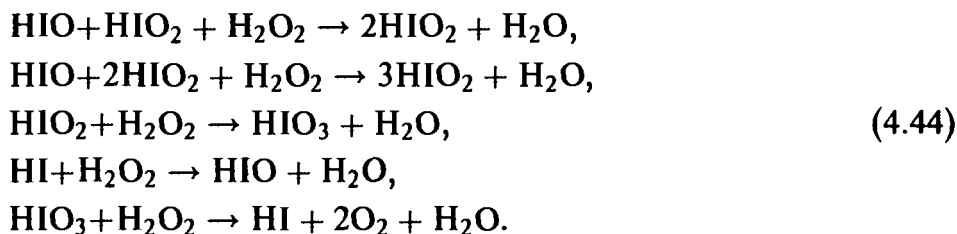
и в то же время восстановительном действии перекиси водорода и содержит стадию восстановления



и стадию окисления



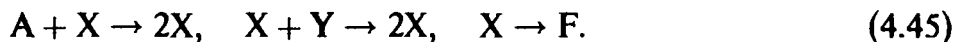
Возможный механизм второй стадии реакции Брея был предложен Мацузаки, Накаджимой и Либхафским (*Matsuzaki, Nakajima, Liebhaafsky*, 1974). Мы изменим механизм в некоторых деталях и воспользуемся моделью



Отличие от модели Мацузаки и др. состоит по существу в том, что мы ввели дополнительную автокаталитическую реакцию и обратные реакции.

Модификацией реакции Брея стали «иодные часы» Бриггса—Раушера. Реакция Бриггса—Раушера протекает в растворе перекиси водорода ($1,2 \text{ моль/дм}^3$), иодата калия ($0,05 \text{ моль/дм}^3$), малоновой кислоты ($0,04 \text{ моль/дм}^3$), сульфата марганца ($0,006 \text{ моль/дм}^3$), перхлорной или серной кислоты ($0,05 \text{ моль/дм}^3$) и крахмала ($0,01\%$). Создать иодные часы Бриггса—Раушера относительно просто, а периодические колебания окраски раствора от золотистой к синей и бесцветной производят сильное впечатление. Синий комплекс крахмала образуется, когда периодически колеблющаяся концентрация ионов йода приближается к максимальному значению около $10^{-4} \text{ моль/дм}^3$ (рис. 4.20). Детальная кинетика колебательной иодной реакции весьма сложна. Предложенный Деном (*Degn*, 1967) механизм реакции сводится к цепным реакциям радикалов. Подробные исследования реакции Бриггса—Раушера были проведены Де Кеппером, Пако и Росси (*De Kepper, Pacault, Rossi*, 1976). Наиболее важный из полученных ими экспериментальных результатов состоит в том, что в иодной реакции наряду с обычными периодическими колебаниями могут также наблюдаться режимы типа переключения, бистабильные состояния и гистерезисные явления.

Теоретические исследования реакций Брея и Бриггса—Раушера пока не позволили достичь полной ясности относительно их детальной кинетики. Однако можно сослаться на простые модели формальной кинетики, позволяющие качественно воспроизводить аналогичные режимы. Еще в 1921 г. Брей предложил воспользоваться для интерпретации открытой им реакции моделью, развитой за год до того Лоткой. Модель Лотки состояла из реакций



Для сравнения с реакцией Брея реагент А можно сопоставить перекиси водорода, а реагент F отождествить с иодом. Реагенты X и Y соответствуют гипотетическим промежуточным продуктам. Применяя закон Гульдберга—Вааге—Вант Гоффа к этой реакционной системе, Лотка получил следующие уравнения для концентраций X и Y:

$$\dot{X} = X(a - Y), \quad \dot{Y} = Y(X - b). \quad (4.46)$$

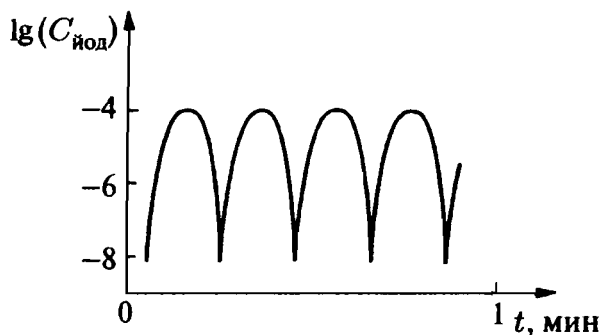


Рис. 4.20. Колебания концентрации йода в «иодных часах» Бриггса—Раушера

Вольтерра удалось получить решения этой системы в замкнутой форме и показать, что они соответствуют структурно неустойчивым колебаниям. Структурная неустойчивость означает в данном случае, что величина амплитуды неустойчива относительно внешних возмущений. Именно поэтому модель Лотки–Вальтерра не считается ныне реалистической моделью концентрационных колебаний.

Ниже мы исследуем другую модель, простую в математическом отношении и позволяющую воспроизводить наряду с устойчивыми колебаниями самовозбуждение, бистабильность и гистерезис. Рассмотрим введенную выше химическую модель реакции Брея (4.44). Полагая



запишем ее в виде системы



На первой стадии реакции из молекулы A автокаталитически образуется молекула X (вторая молекула X служит катализатором). На второй стадии A также превращается в X , но теперь уже в качестве катализатора выступают две молекулы X . (В реальных случаях такие тримолекулярные реакции возможны лишь как промежуточные стадии; поэтому уравнение второй реакции представляет собой некоторое упрощение). Третья реакция описывает переход промежуточного продукта X в продукт F . Уравнение четвертой «реакции» описывает превращение молекулы B в молекулу A . Аналогично, уравнение пятой реакции соответствует превращению F в B . Неопределенности в определении существенных стадий механизма реакции не позволяют нам считать уравнения (4.47) реалистической моделью реакции Брея. Однако мы можем показать, что модель реакции (4.47) качественно правильно передает ряд особенностей реальных реакций.

Чтобы записать дифференциальные уравнения для системы реакций (4.47), требуется знание лишь самых элементарных сведений по химической кинетике. Необходимо также знать, что скорость элементарной реакции пропорциональна произведению концентраций веществ, участвующих в реакции. Такое утверждение правдоподобно, поскольку произведение концентраций пропорционально вероятности столкновения соответствующих частиц. Если обозначить концентрации исходных и промежуточных веществ и продуктов реакции, соответственно, через A , X и F , то, например, скорости парциальных реакций 1 и 2 пропорциональны величинам

$$k_1AX \quad \text{и} \quad k_2AX^2,$$

а скорости обратных реакций пропорциональны величинам

$$k_{-1}X^2 \quad \text{и} \quad k_{-2}X^3.$$

Коэффициенты пропорциональности называются константами скоростей реакции. Если члены, описывающие увеличение концентраций, условиться брать со знаком плюс, а члены, описывающие убыль концентраций, — со знаком минус, то получатся следующие дифференциальные уравнения для изменений концентраций веществ, участвующих в реакции, в единицу времени:

$$\dot{A} = -k_1AX + k_{-1}X^2 - k_2AX^2 + k_{-2}X^3 + k_4B, \tag{4.48}$$

$$\dot{X} = k_1 AX - k_{-1} X^2 + k_2 AX^2 - k_{-2} X^3 - k_3 X + k_{-3} F, \quad (4.49)$$

$$\dot{F} = k_3 X - k_{-3} F - k_5 F, \quad (4.50)$$

$$\dot{B} = -k_4 B + k_5 F, \quad (4.51)$$

(k_4 — константа скорости образования вещества A ; k_5 — константа скорости распада вещества F ; обе эти константы мы считаем управляемыми). Дифференциальные уравнения описывают кинетику реакций системы на языке математики. Поскольку правые части уравнений (4.48)–(4.51) нелинейны, эти дифференциальные уравнения нелинейны. В отдельных частных случаях удастся найти их замкнутые аналитические решения. Простейший аналитически разрешимый случай соответствует выбору краевых условий $A = \text{const}$, $X = \text{const}$, $F = \text{const}$. Оно описывает простой релаксационный процесс

$$B(t) = B(0) \exp \{-k_4 t\} + \frac{k_5}{k_4} F [1 - \exp \{-k_4 t\}]. \quad (4.52)$$

Концентрация йода асимптотически стремится к стационарному значению

$$B(\infty) = \frac{k_5 F}{k_4}.$$

Как правило, для решения системы дифференциальных уравнений (4.48)–(4.51) приходится прибегать к компьютеру. Покажем теперь, что описанная выше система реакций позволяет воспроизвести ряд качественно различных режимов, носящих характер диссипативных структур. К этому следует добавить несколько частных случаев, возникающих при особом выборе констант скоростей реакций. Особый интерес для нас представляет временная зависимость X .

Самовозбуждение или кооперативное возбуждение реакций

Рассмотрим частный случай

$$A = \text{const}, \quad B = \text{const}, \quad k_2 = k_{-2} = k_{-3} = 0. \quad (4.53)$$

Такой выбор величин позволяет реализовать следующий режим: вторая реакция подавлена; концентрации A и B поддерживаются внешними параметрами управления на постоянном уровне; третья, обратная, реакция подавлена. Это соответствует дифференциальному уравнению

$$\dot{X} = (k_1 A - k_3) X - k_{-1} X^2. \quad (4.54)$$

Его решение, полученное с помощью разделения переменных, имеет вид

$$X(t) = \frac{(k_1 A - k_3) X(0)}{k_{-1} X(0) - [k_{-1} X(0) - (k_1 A - k_3)] \exp \{-k_1 A t + k_3 t\}}. \quad (4.55)$$

Характер зависимости решения (4.55) от времени носит совершенно различный характер при $A > A_{\text{крит}}$ и $A < A_{\text{крит}}$. Существует критическая концентрация исходного вещества $A_{\text{крит}} = k_3/k_1$. При $A < A_{\text{крит}}$ решение X всегда экспоненциально убывает; это означает, что со временем концентрация промежуточного продукта X убывает до нуля. Наоборот, при $A > A_{\text{крит}}$ концентрация $X(t)$ выходит на насыщение (рис. 4.21). Конечные состояния соответствуют стационарным (не зависящим от времени) решениям системы реакционно-кинетических уравнений. Если конечные состояния концентраций промежуточных продуктов превосходят заданную концентрацию A , то возникает зависимость, представленная на рис. 4.22. В под-

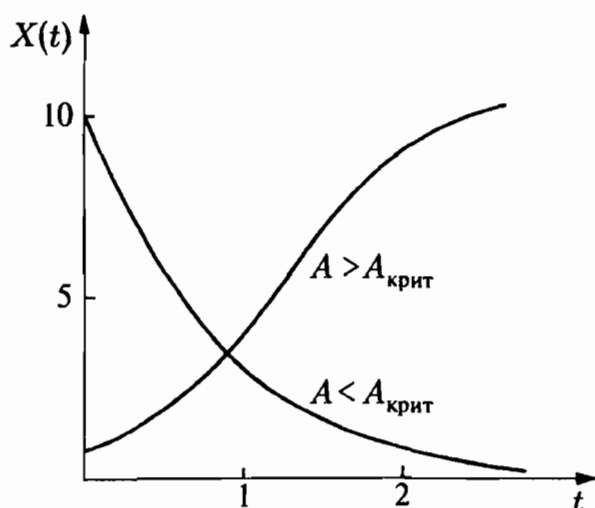


Рис. 4.21. Логистический рост в соответствии с формулой (4.55) в случае за- и подкритической концентрации исходного вещества (сырья) A

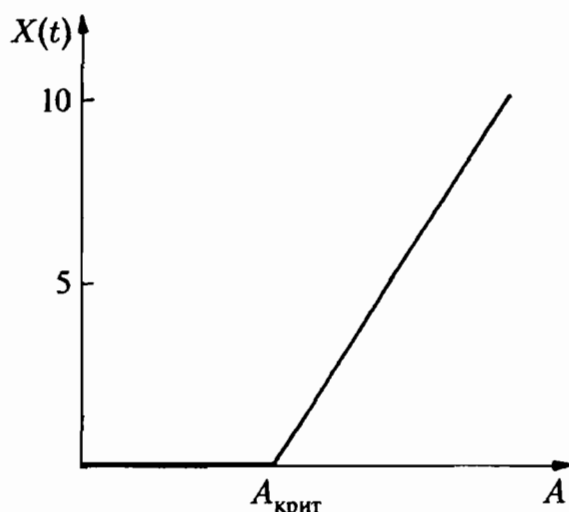


Рис. 4.22. Стационарная концентрация X в зависимости от концентрации сырья

критической области доминируют процессы распада; в надкритической области, наоборот, в результате кооперативного поведения молекул — процессы образования вещества X . Кооперативный характер обусловлен тем, что в каждом элементарном процессе образования вещества X участвуют по крайней мере две молекулы X , вследствие чего и возникает связь. Формальное сходство ситуации, представленной на рис. 4.22, с эффектом Бенара (рис. 4.12) и с лазером поистине поразительно. Переход от индивидуального к кооперативному поведению происходит во всех случаях скачкообразно при превышении некоторого критического значения параметра системы. Связан такой переход с (временным) нарушением симметрии и очень напоминает термодинамический фазовый переход второго рода. Поэтому мы говорим в подобных случаях о кинетическом фазовом переходе второго рода.

Бистабильный режим реакции

Рассмотрим теперь частный случай

$$A = \text{const}, \quad F = \text{const}, \quad k_1 = 0, \quad k_{-1} = 0, \quad B = \text{const}.$$

Он реализуется, если подавить первую реакцию и соответствующим образом управлять подводом и отводом веществ A и F . Подчеркнем, что и в этом случае существует критическое значение концентрации исходного вещества. При $A < A_{\text{крит}}$ решения $X(t)$ сходятся (при больших временах) к стационарной устойчивой конечной концентрации (рис. 4.23). С другой стороны, при $A > A_{\text{крит}}$ существуют два стационарных устойчивых конечных состояния, к которым система может стремиться в зависимости от заданной начальной концентрации $X(0)$ (рис. 4.24). Такая система бистабильна, как некоторые мультивибраторные схемы в электротехнике, и переход от одного стабильного состояния к другому требует значительного изменения управляющего параметра. На рис. 4.25 устойчивые стационарные состояния, а также лежащее между ними неустойчивое состояние при фиксированном значении A представлены в зависимости от F . Известно, как происходит

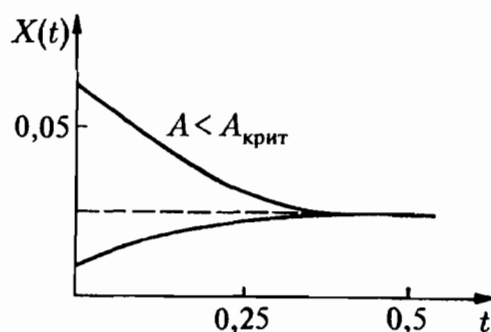


Рис. 4.23. Моностабильное поведение при подкритической концентрации сырья

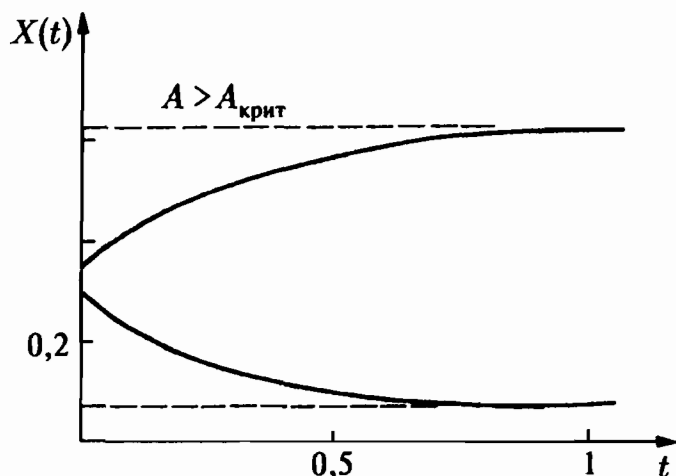


Рис. 4.24. Бистабильность при закритической концентрации сырья

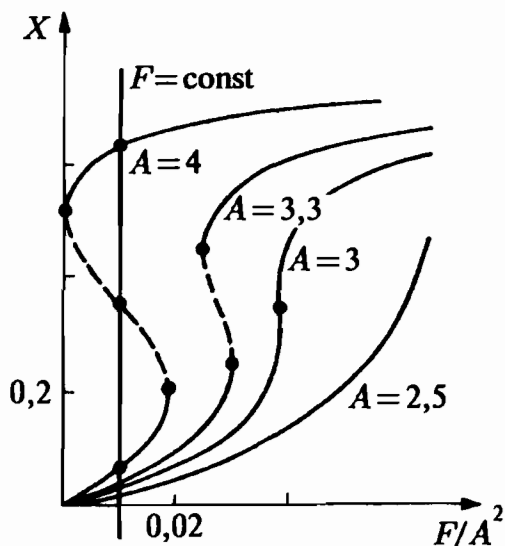


Рис. 4.25. «Провисание» графика зависимости стационарной концентрации X от F в бистабильной области

переход моностабильного режима в бистабильный с увеличением значений A . Химические системы с двумя стабильными состояниями представляют большой интерес для моделирования биологических механизмов, например, памяти, мембранных процессов и дифференциации при делении клеток.

Самовозбуждающиеся химические колебания

В качестве третьего частного случая рассмотрим

$$F = \text{const}, \quad B = \text{const}, \quad k_{-1} = k_{-2} = k_{-3} = 0.$$

Этот случай реализуется, когда все обратные реакции подавлены (в третьей элементарной реакции — путем быстрого удаления вещества F из реакционной смеси). В то время как скорость образования вещества A постоянна, концентрации A и X переменны. Возникает система двух связанных нелинейных дифференциальных уравнений, которые могут быть решены с помощью компьютера.

Для нас особый интерес представляют самовозбуждающиеся колебания, которые могут возникать при определенном критическом уровне концентрации исходных веществ (рис. 4.26): концентрации исходных веществ и продуктов реакции колеблются в противофазе между своими максимальными и минимальными значениями. Возможность самопроизвольного возникновения в химических системах незатухающих колебаний, естественно, представляет большой интерес для теоретической биологии и позволяет предложить одну из интерпретаций периодических биологических процессов.

Существование периодических колебаний в химических системах со сложной кинетикой может быть экспериментально доказано не только для упоминавшейся

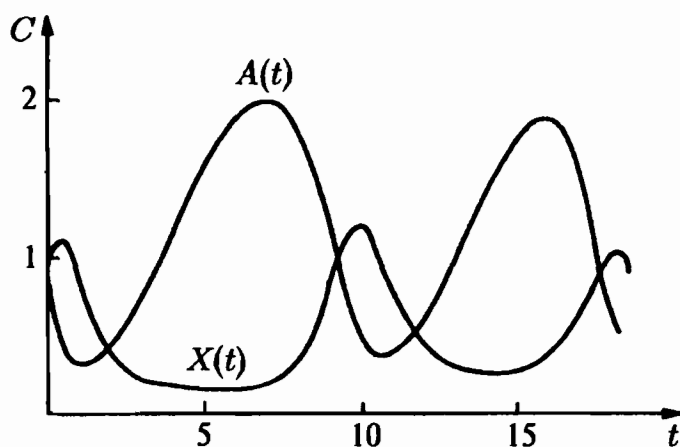
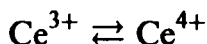


Рис. 4.26. Возникновение химических колебаний

выше реакции Брея, но и для ряда органических реакций. Наиболее известным примером может служить так называемая реакция Белоусова—Жаботинского (реакция БЖ) — сложная система элементарных реакций, включающая в себя восстановление ионов церия

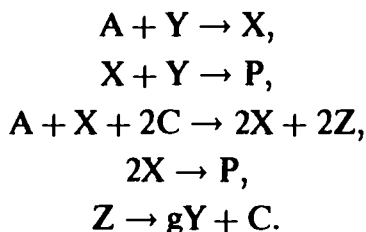


в присутствии малоновой кислоты, серной кислоты и бромата калия. Добавление окислительно-восстановительного индикатора (ферроина) позволяет следить за ходом реакции по изменению цвета реакционной смеси (красный цвет — избыток ионов Ce^{3+} , синий — избыток ионов Ce^{4+}). При определенных над-

критических концентрациях реагентов в растворе возникают временные или пространственные колебания (рис. 4.27). В реакционной системе Белоусова — Жаботинского, как видно из рис. 4.28, могут

возникать и существенно более динамические структуры, в которых можно различить кольцевые и спиральные волны. Спиральные волны, получившие название ревербераторов, образуются при искусственном разрыве плоских или кольцевых волн.

Так называемые ведущие центры являются спонтанно возникающими источниками кольцевых волн. Частота спонтанного возникновения ведущих центров на единицу площади поверхности зависит от концентраций реактивов, в то время как наличие неоднородностей или примесей сказывается на частоте лишь весьма слабо (Жаботинский, 1974). Возможно, что ведущие центры возникают в результате спонтанных флуктуаций концентраций. Вследствие большого числа элементарных реакций механизм реакции БЖ понят не до конца. Одна из наиболее известных моделей, первоначально предложенная Филдом и Нойесом (*Field, Noyes, 1974*), имеет следующий вид:



Здесь использованы следующие обозначения: $\text{A} = \text{BrO}_3^-$, $\text{X} = \text{HBrO}_2$, $\text{Y} = \text{Br}^-$, $\text{C} = \text{Ce}^{3+}$, $\text{Z} = \text{Ce}^{4+}$, $\text{P} = \text{HOBr}$ и бромированное органическое вещество. Эта модель получила название «Орегонатор».

Другие модели были предложены Жаботинским (1974) и Заикиным (*Zaikin et al., 1976*). Различными авторами были предложены также модели возникновения и распространения волн в реакции БЖ (*Zaikin, Kawczynski, 1976; Иваницкий, Кринский и Сельков, 1978*). Стали известны и различные модификации реакции БЖ (*Kuhnert, 1986; Pacault, Vidal, 1979*).

В исследованиях химических диссипативных структур реакция БЖ, несомненно, играет особую роль, так как она легко осуществима и позволяет наблюдать интересные экспериментаторов (и теоретиков) явления особенно. Ныне она служит

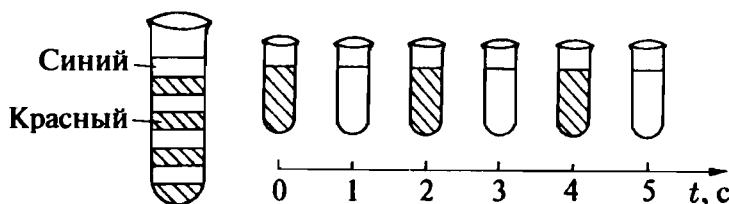


Рис. 4.27. Пространственная и временная периодичность в реакции Белоусова — Жаботинского



Рис. 4.28. Круговые и спиральные волны в реакции Белоусова—Жаботинского по Уинфри

излюбленной моделью при изучении спонтанных волновых процессов в активных средах. Понятие активной среды первоначально было связано с исследованием распространения сигнала по нервным волокнам, в головном мозгу, в сердце и в мышцах (Иваницкий, Кринский и Сельков, 1978). Ныне под активной средой принято понимать систему, обладающую следующими свойствами (Васильев и др., 1987).

Каждая пространственная точка системы

- 1) является источником свободной энергии;
- 2) является местом, где разыгрываются процессы, протекающие вдали от равновесия;
- 3) тесно связана с соседними точками с помощью процессов переноса.

Характерные процессы, наблюдаемые в активных средах, включают в себя следующие:

- 1) незатухающее распространение импульсов, фронтов и волн;
- 2) спонтанное возбуждение волн в определенных точках (ведущих центрах);
- 3) возникновение стохастических волн, приводящих к турбулентности;
- 4) возникновение пространственно синхронизированных временных колебаний;
- 5) возникновение стационарных неоднородных структур, например, стоячих фронтов и стоячих волн.

Теоретическая интерпретация этих явлений, как правило, опирается на уравнения реакции с диффузией вида

$$\partial_t X_i(r, t) = f_i(X_1, \dots, X_n) + \nabla \sum_{j=1}^n D_{ij} \nabla X_j(r, t).$$

Здесь $X_i(r, t)$ означает концентрацию i -го химического вещества в точке r в момент времени t , f_i — функция реакции и D_{ij} — коэффициенты диффузии, включающие в себя и перекрестные связи.

Глава 5

Самовоспроизведение

Жизнь есть то, что сохраняет форму в процессе обмена материей.

Томас Манн

5.1. Динамика простого самовоспроизведения

В естественных процессах самовоспроизведения и эволюции участвует множество различных элементарных процессов. К их числу относятся самопроизводство, бистабильность и мультистабильность, конкуренция, отбор, хранение и обработка информации, возникновение языка, распределение функций, дифференциация и т. д. При этом на первом месте по значению и простоте стоит самовоспроизводство, т. е. спонтанное «изготовление» копий рассматриваемого объекта, например, молекул.

Один из наиболее выдающихся математиков и физиков нашего столетия фон Нейман в своей фундаментальной работе 1951 г. показал, что система, наделенная способностью к самовоспроизведению, при определенных условиях может изготавливать копии, более сложные, чем исходная система (von Neumann, 1951, 1966).

К числу наиболее простых и наиболее известных процессов самовоспроизведения молекул относится химический автокатализ. В химии принято говорить об автокаталитических реакциях, если получение вещества X из исходного вещества A или из нескольких исходных веществ требует присутствия в реакционной системе вещества X или даже возможно только в присутствии вещества X (рис. 5.1). Схематически это выглядит так:

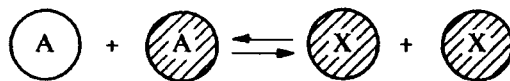
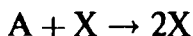
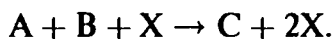


Рис. 5.1. Автокаталитическое репродуктивное



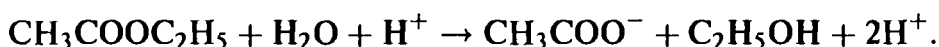
или



Как правило, в автокаталитических реакциях прежде всего образуется промежуточный комплекс, который обладает относительно малой энергией активации:



Примером автокаталитической реакции может служить реакция омыления эфира:



Наблюдается автокаталитическое размножение протонов.

Характерная особенность автокаталитических реакций состоит в том, что эффект усиления наблюдается макроскопически лишь выше определенной пороговой концентрации автокаталитических компонент. Ниже порогового значения доминируют реакции распада, в то время как выше порогового значения в результате

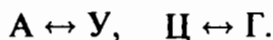
кооперативного действия автокаталитических компонент становится возможным макроскопический эффект усиления.

Еще одним примером процесса самовоспроизведения может служить испускание когерентных фотонов в лазере. Мы рассмотрим твердотельный лазер, состоящий из множества лазероактивных атомов, встроенных в кристаллическую решетку. На торцах лазера могут располагаться зеркала, которые отсортировывают фотоны, движущиеся в аксиальном направлении с дискретной характеристической частотой полости. В состоянии теплового равновесия энергетические уровни лазероактивных атомов E_n заполнены по закону Больцмана, т. е. числа заполнения пропорциональны величине

$$\exp \left\{ - \frac{E_n}{kT} \right\},$$

где T — температура среды. Если лазероактивные атомы возбуждать с помощью излучения накачки, то более высокие уровни заполняются сильнее. Такое «обратное» распределение можно охарактеризовать и с помощью отрицательной температуры. Если атомы, возбужденные излучением накачки, начинают испускать излучение, то при индуцировании новых актов испускания излучения воспроизводятся фотоны с характеристической частотой полости, движущиеся в аксиальном направлении. Максимальное усиление лазерной моды приводит к испусканию когерентного лазерного излучения. В лазерном режиме атомные излучающие диполи колеблются в фазе, т. е. речь идет о кооперативном поведении атомов. Лазерный эффект наблюдается лишь при условии, что интенсивность излучения накачки выше некоторого критического уровня, в то время как ниже порогового значения лазер функционирует как обыкновенная лампа, т. е. испускает некогерентное излучение.

Различные примеры самоорганизации известны из области макромолекулярных процессов. Молекулы рибонуклеиновой кислоты (сокращенно РНК) представляют собой гетерополимер, состоящий из огромного числа (от 10^3 до 10^4) звеньев. В состав РНК входят мономеры четырех сортов: А — аденин, Ц — цитозин, Г — гуанин, У — урацил. Молекулы РНК образуют длинные цепи, сложенные определенным образом (рис. 5.2). С помощью катализатора-фермента цепи РНК могут образовывать из элементарных звеньев-нуклеотидов новые цепи путем установления дополнительного соответствия — замены



В основе процесса лежат относительно прочные водородные связи, возникающие между А и У, а также между Ц и Г. Заданная цепь РНК склонна к тому, чтобы к ее звеньям пристраивались дополнительные звенья. Специальные ферменты (репликазы) позволяют звеньям, сначала лишь слабо связанным между собой, образовать с помощью прочных водородных связей новую дополнительную цепь. Так, в результате двукратной дополнительной репродукции самовоспроизводится исходная цепь РНК.

Аналогичным образом происходит и самовоспроизводство цепей дезоксирибонуклеиновой кислоты (ДНК). Эти цепи состоят из звеньев: А — аденина, Ц — цитозина, Г — гуанина, Т — тимина. На основе относительно прочных водородных

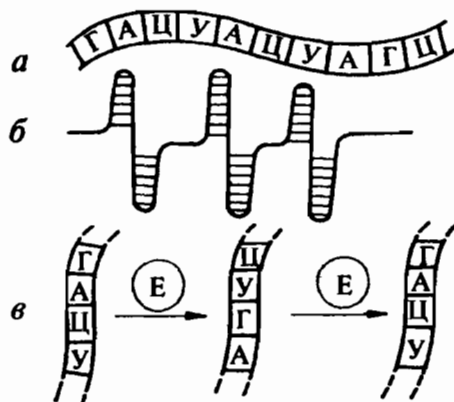


Рис. 5.2. Молекула РНК: (а) структура короткого участка, (б) пространственное складывание и фиксация молекулярной цепи силами связи, (в) репродукция под действием фермента

связей между А и Т, а также между Ц и Г образуется двойная цепь из взаимно дополнительных молекул (рис. 5.3). Под действием фермента эти цепи могут расцепиться, после чего к каждой из них пристраивается дополнительная цепь, поэтому в результате образуются два экземпляра исходного оригинала. Самовоспроизводство полинуклеотидов имеет фундаментальное значение для эволюции живых организмов.

Рассмотрим теперь систему, состоящую из N одинаковых физических, химических или биологических объектов (молекул, фотонов, живых организмов и т. д.), удваивающихся со скоростью B (скорость образования) и распадающихся со скоростью D (скорость распада). Плотность таких объектов в системе

$$X = \frac{N}{V} \quad (5.1)$$

(V — объем системы) удовлетворяет уравнению

$$\frac{d}{dt}X(t) = BX(t) - DX(t) = RX(t). \quad (5.2)$$

Величина $R = B - D$ характеризует результирующую скорость изменения. Решение уравнения (5.2) имеет вид

$$X(t) = X(0) \exp \{ (B - D)t \}. \quad (5.3)$$

Единственное стационарное состояние соответствует

$$X^{(1)} = 0. \quad (5.4)$$

Оно устойчиво при $B < D$ и неустойчиво при $B > D$.

Химическим примером системы такого рода может служить автокаталитическая реакция



где k_1 и k_2 — константы скоростей реакций. По законам формальной кинетики (Benson, 1960; Bittrich, Haberland, Just, 1973) для концентрации X выполняется уравнение

$$\frac{d}{dt}X(t) = k_1AX(t) - k_2X(t), \quad (5.6)$$

где A — концентрация исходного (сырьевого) вещества.

Рассмотрим теперь случай, когда при больших плотностях (концентрациях) в игру вступают сдерживающие факторы, полностью останавливающие дальнейший рост. Если X_0 — наибольшая достижимая плотность, то скорость изменения можно положить равной

$$R \left(1 - \frac{X}{X_0} \right).$$

Соответствующее дифференциальное уравнение имеет вид

$$\frac{dX}{dt} = RX \left(1 - \frac{X}{X_0} \right) = RX - SX^2, \quad (5.7)$$

где $S = R/X_0$.

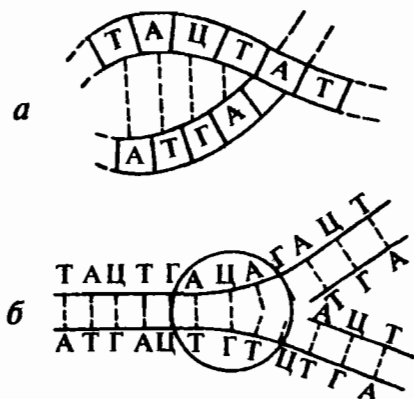
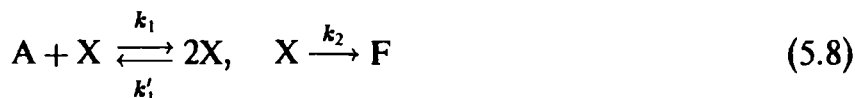


Рис. 5.3. Молекула ДНК: (а) структура, (б) репликация

В химическом примере



концентрация насыщения X_0 обусловлена возможностью обратной реакции. По правилам формальной кинетики реакция (5.8) приводит к дифференциальному уравнению

$$\frac{dX}{dt} = k_1 AX - k'_1 X^2 - k_2 X. \quad (5.9)$$

Сравнивая с уравнением (5.7), получаем

$$R = k_1 A - k_2, \quad X_0 = \frac{k_1 A - k_2}{k'_1}, \quad S = k'_1. \quad (5.10)$$

Уравнение (5.7) обладает двумя стационарными решениями

$$X^{(1)} = 0, \quad X^{(2)} = X_0 = \frac{R}{S}. \quad (5.11)$$

Как известно, второе стационарное решение имеет смысл только при $R > 0$, так как плотности и концентрации могут быть только положительными или равными нулю. Анализ устойчивости в случае уравнения (5.7) показывает, что

$R > 0$: $X^{(1)} = 0$ — неустойчиво; $X^{(2)} = R/S$ — устойчиво;

$R < 0$: $X^{(1)} = 0$ — устойчиво; $X^{(2)}$ — не существует.

Выводы относительно устойчивости непосредственно следуют из явного решения уравнения (5.7), которое имеет следующий вид:

$$\frac{X(t)}{X(0)} = \frac{X_0}{X(0) - [X(0) - X_0] \exp \{-Rt\}}. \quad (5.12)$$

Дифференциальное уравнение (5.7) и его решение (5.12) были получены еще в прошлом веке Ферхюльстом и Перлом, поэтому зависимость (5.12) принято называть законом роста Ферхюльста—Перла. На рис. 5.4 показано, как решение (5.12) зависит от времени. Мы видим, что при $t \rightarrow \infty$ плотность стремится к конечному предельному значению ($R > 0$) или к нулю ($R < 0$). Значение $R = 0$ соответствует бифуркации решения, так как при прохождении параметра через него решение претерпевает качественные изменения. На рис. 5.5 представлена зависимость стационарных решений от значения концентрации исходного вещества A :

$$X^{(1)} = 0, \quad X^{(2)} = \frac{k_1 A - k_2}{k'_1}. \quad (5.13)$$

Нетрудно видеть, что устойчивое стационарное значение всюду непрерывно, в том числе и в точке $A_{\text{крит}} = k_2/k_1$, где зависимость имеет характерный излом, который мы уже видели на рис. 4.12 в случае эффекта Бенара и лазерного эффекта. Из-за сходства с термодинамическими фазовыми переходами

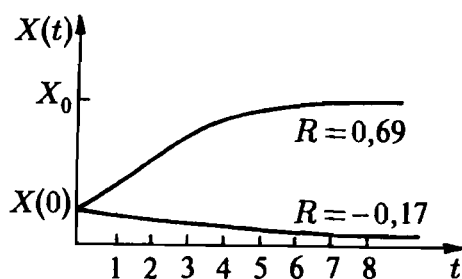


Рис. 5.4. Логистический рост (закон Ферхюльста—Перла)

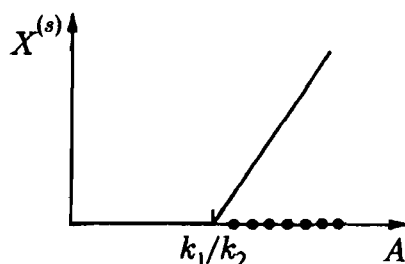


Рис. 5.5. Положение устойчивого стационарного решения в зависимости от значения управляющего параметра A

второго рода в этом случае также говорят о кинетических переходах второго рода. Типичным термодинамическим фазовым переходом второго рода является обращение в нуль намагничивания ферромагнетиков в точке Кюри $T_{\text{крит}}$. На рис. 5.6 представлен график зависимости намагничивания от обратной зависимости. Аналогия с кривой, представленной на рис. 5.5, очевидна. Из теории термодинамических фазовых переходов известно, что в окрестности критической точки флуктуации затухают лишь очень слабо (Ландау, Лифшиц, 1995).

Характерное явление слабого затухания флуктуаций наблюдается в окрестности критической точки и в случае химических реакций. При $R = 0$ получаем уравнение

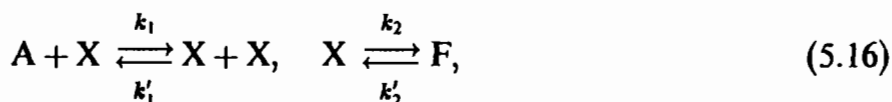
$$\dot{X} = -SX^2, \quad (5.14)$$

допускающее решение

$$X(t) = \frac{X(0)}{1 + X(0)St}. \quad (5.15)$$

В отличие от случая $R \neq 0$ флуктуации в окрестности устойчивого стационарного значения затухают не экспоненциально, а по закону (const/t) . В этой связи принято говорить о мягких модах.

Если в систему (5.8) ввести реакцию, обратную второй реакции, т. е.



то получится уравнение системы

$$\frac{dX}{dt} = Q + RX - SX^2, \quad (5.17)$$

где

$$R = k_1 A - k_2, \quad S = k'_1, \quad Q = k_2 F \geq 0.$$

При произвольном $Q > 0$ дискретный переход при $R = 0$ исчезает и заменяется непрерывным переходом. Всегда существует лишь стационарное решение, соответствующее положительному решению алгебраического уравнения

$$Q + RX - SX^2 = 0. \quad (5.18)$$

Находим:

$$X^{(1)} = \frac{R}{2S} + \left[\frac{R^2}{4S^2} + \frac{Q}{S} \right]^{1/2}. \quad (5.19)$$

Соответствующая кривая при $Q = 1$ представлена на рис. 5.7, на котором для сравнения показан предельный случай $Q = 0$. В окрестности точки излома при $R = 0$ происходит непрерывное изменение направления. Дискретный фазовый переход второго порядка при $Q = 0$ сменяется размытым фазовым переходом при $Q > 0$. Диффузные фазовые переходы играют большую роль и в термодинамике, они наблюдаются в твердых телах и в химическом равновесии, например, в ионизационном равновесии (Эбеллинг и др., 1979).

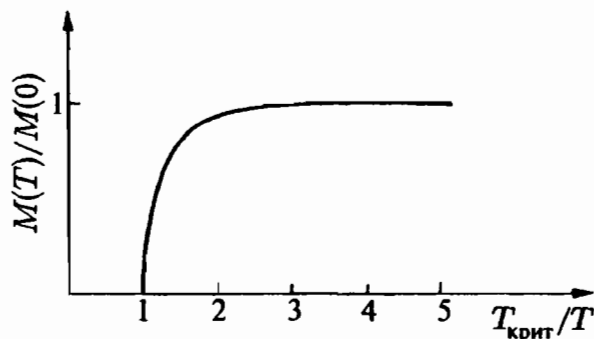


Рис. 5.6. Намагничивание ферромагнетика как функция температуры T ($T_{\text{крит}}$ — критическая температура)

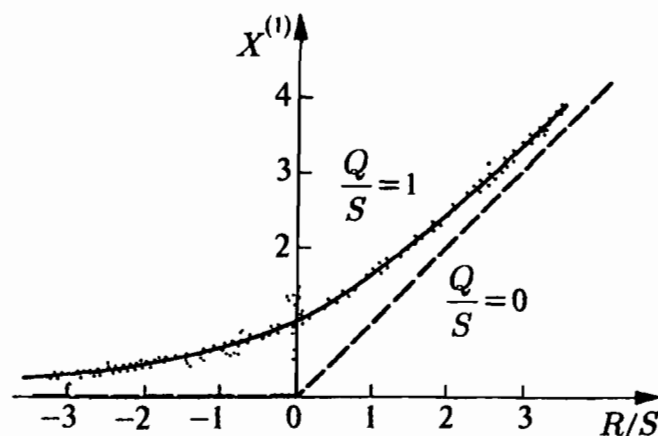


Рис. 5.7. Стационарное состояние автокаталитической реакционной системы. Точками показана плотность распределения вероятности в рамках стохастической теории

Прямую аналогию с уравнением (5.18) мы находим в случае ферромагнетика, находящегося в постоянном магнитном поле H . Для намагничивания M справедливо уравнение

$$H + \left[\frac{T}{T_{\text{крит}}} - 1 \right] M - M^2 = 0$$

Аналогия с уравнением (5.18) очевидна. При $H > 0$ дискретный термодинамический фазовый переход в точке Кюри «смягчается» и превращается в диффузный термодинамический фазовый переход.

Рассмотрим теперь временную зависимость решения уравнения (5.17). Интегрируя уравнение (5.17), получаем

$$X(t) = \frac{R}{2S} + \left[\frac{R^2}{4S^2} + \frac{Q}{S} \right]^{1/2} \cdot \frac{1 - C \exp \{-t/\tau\}}{1 + C \exp \{-t/\tau\}}. \quad (5.20)$$

Постоянную C находим из начального условия $X(0)$. Для времени релаксации τ справедливо соотношение

$$\tau = [R^2 + 4QS]^{-1/2},$$

где квадратный корень надлежит выбирать со знаком плюс. Из формулы (5.20) непосредственно видно, что при $Q > 0$ приближение к стационарному устойчивому состоянию неизменно происходит по экспоненциальному закону. Время релаксации τ конечно и достигает максимума при $R = 0$.

Тем самым мы отчетливо видим, что квадратичные члены в законе распада приводят к качественным изменениям в поведении решения, соответствующим кинетическим фазовым переходам второго рода или диффузным кинетическим фазовым переходам.

5.2. Стохастика простого самовоспроизведения

Попытаемся развить стохастическую теорию для реакции (5.16), которая была рассмотрена нами в разд. 5.1. В стохастической теории в качестве переменной вместо концентрации $X(t)$ выступает число частиц $N(t) = VX(t)$. По очевидным физическим причинам $N(t)$ может принимать только дискретные значения:

$$N(t) = \{0, 1, 2, \dots\} \quad (5.21)$$

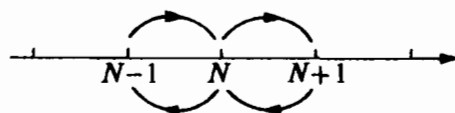


Рис. 5.8. Элементарные переходы ± 1 в одношаговых процессах

Определим вероятность $P(N, t)$ найти N частиц в момент времени t . Реакция (5.16) имеет четыре канала — в соответствии с двумя прямыми и двумя обратными реакциями. В каждом элементарном акте число частиц изменяется на ± 1 (рис. 5.8), т. е. речь идет о так называемых одношаговых процессах. Затем мы можем записать общее основное уравнение в форме баланса:

$$\frac{\partial}{\partial t} P(N, t) = W^+(N-1)P(N-1, t) + \\ + W^-(N+1)P(N+1, t) - [W^+(N) + W^-(N)]P(N, t). \quad (5.22)$$

Здесь $W^+(N)$ — вероятность перехода скачком из N в $N+1$, $W^-(N)$ — вероятность перехода скачком из N в $N-1$. Чтобы получить математическое выражение для вероятностей перехода, запишем детерминистическое уравнение (5.17) в форме, в которой оно вместо $X(t)$ содержит экстенсивную переменную — число частиц:

$$\langle N(t) \rangle = X(t)V. \quad (5.23)$$

Из уравнения (5.17) получаем

$$\frac{d}{dt} \langle N(t) \rangle = k'_2 FV + k_1 A \langle N(t) \rangle - k_2 \langle N \rangle - \frac{1}{V} k'_1 \langle N(t) \rangle^2. \quad (5.24)$$

В рамках стохастической теории мы должны отождествить величину $\langle N(t) \rangle$ со средним значением, т. е. положить

$$\langle N(t) \rangle = \sum_{N=0}^{\infty} NP(N, t). \quad (5.25)$$

Напрашивается мысль отождествить коэффициенты уравнения (5.24) с вероятностями перехода. Мы получаем:

$$W^+(N) = k_1 AN + k'_2 FV, \quad (5.26)$$

$$W^-(N) = k_2 N + \frac{1}{V} k'_1 N(N-1) \quad (5.27)$$

при $N \geq 0$ и $W^+(-1) = 0$. Член $N(N-1)$ в уравнении (5.27) вместо N^2 обусловлен тем, что после распада одной молекулы в системе остается только $(N-1)$ частиц. Проверим, действительно ли уравнение (5.22) вместе с уравнениями (5.26) и (5.27) согласуется с уравнением (5.24) для среднего значения (5.25). Для этого, умножая уравнение (5.22) на N и суммируя по N , получаем

$$\frac{\partial}{\partial t} \sum_{N=0}^{\infty} NP(N, t) = \sum_{N=0}^{\infty} N [k_1 A(N-1) + k'_2 FV] P(N-1, t) + \\ + \sum_{N=0}^{\infty} N \left[k_2 (N+1) + \frac{1}{V} k'_1 (N+1)N \right] P(N+1, t) - \\ - \sum_{N=0}^{\infty} N \left[k_1 AN + k'_2 FV + k_2 N + \frac{1}{V} k'_1 N(N-1) \right] P(N, t). \quad (5.28)$$

Заменим теперь в первой сумме, стоящей в правой части уравнения (5.28), переменную N на $N+1$, а во второй сумме — N на $N-1$. Некоторые члены взаимно

уничтожаются, и мы получаем уравнение

$$\frac{\partial}{\partial t} \sum N P(N, t) = k_1 A \sum N P(N, t) + k_2 F V \sum P(N, t) - k_2 \sum N P(N, t) - \frac{1}{V} k'_1 \sum N(N-1) P(N, t),$$

из которого следует, что

$$\frac{d}{dt} \langle N(t) \rangle = k'_2 F V + (k_1 A - k_2) \langle N(t) \rangle - \frac{1}{V} k'_1 \left\{ \langle [N(t)]^2 \rangle - \langle N(t) \rangle \right\}. \quad (5.29)$$

Если пренебречь дисперсией, т. е. положить

$$\langle [N(t)]^2 \rangle \approx [\langle N(t) \rangle]^2, \quad (5.30)$$

то уравнение (5.29) совпадает с уравнением (5.24) для всех систем с $\langle N(t) \rangle \gg 1$. Тем самым мы доказали, что стохастическая теория согласуется с детерминистической теорией.

Найдем теперь стационарное решение уравнения (5.22). Полагая $\frac{\partial P^0}{\partial t} = 0$, получаем

$$\begin{aligned} W^+(N-1)P^0(N-1) + W^-(N+1)P^0(N+1) = \\ = W^+(N)P^0(N) + W^-(N)P^0(N). \end{aligned} \quad (5.31)$$

Переставляя члены, приводим соотношение (5.31) к виду

$$\begin{aligned} W^-(N+1)P^0(N+1) - W^+(N)P^0(N) = \\ = W^-(N)P^0(N) - W^+(N-1)P^0(N-1). \end{aligned} \quad (5.32)$$

Так как в правой и левой частях равенства (5.32) стоит одна и та же функция с различными значениями аргумента, левая и правая части равенства не зависят от N . Так как $P^0(\infty) = 0$, значения левой и правой частей должны быть равны нулю. Таким образом, мы получаем

$$W^-(N+1)P^0(N+1) = W^+(N)P^0(N). \quad (5.33)$$

Равенство (5.33) известно под названием «условие детального равновесия» (Николис, Пригожин, 1979). Это условие утверждает, что переходы между N и $N+1$ и переходы между N и $N-1$ находятся в равновесии, т. е. соответствующие потоки вероятности в стационарном состоянии равны по абсолютной величине и противоположны по направлению. Из условия детального равновесия (5.33) с помощью простой итерации ($N \geq 1$) получаем стационарное решение:

$$P^0(N) = \frac{W^+(N-1)W^+(N-2) \dots W^+(0)}{W^-(N)W^-(N-1) \dots W^-(1)} P^0(0). \quad (5.34)$$

Подставляя соотношения (5.26) и (5.27), преобразуем решение (5.34) к виду

$$P^0(N) = \frac{[k_1 A(N-1) + k'_2 F V] \dots [k'_2 F V]}{[k_2 N + \frac{1}{V} k'_1 N(N-1)] \dots [k_2]} P^0(0). \quad (5.35)$$

Макнейл и Уоллс исследовали это решение более подробно для частного случая

$$k'_2 F = \frac{k_1 A}{V},$$

т. е. когда спонтанное и автокаталитическое образования вещества X происходят одинаково быстро, и обнаружили, что

$$P^0(N) = P^{(0)}(0) \left(\frac{k_1 A}{k_2} \right)^N \sum_{l=0}^{N-1} (1 + \mu l)^{-1},$$

$$P^0(0) = \left\{ {}_1F_1 \left(1; \frac{\mu + 1}{\mu}; \frac{k_1 A}{k_2 \mu} \right) \right\}^{-1}, \quad (5.36)$$

$$\mu = \frac{k'_1}{k_2 V},$$

где ${}_1F_1$ — гипергеометрическая функция. Ниже порогового значения, т. е. при

$$A < \frac{k_2}{k_1}, \quad (5.37)$$

наблюдается (приближенно) геометрическое распределение

$$P^0(N) \approx \left(1 - \frac{k_1 A}{k_2} \right) \left(\frac{k_1 A}{k_2} \right)^N. \quad (5.38)$$

Максимум этого распределения расположен при $N = 0$, а стандартное отклонение составляет величину

$$\sigma = \left(\frac{k_1 A}{k_2} \right)^{1/2} \left(1 - \frac{k_1 A}{k_2} \right)^{-1}. \quad (5.39)$$

Выше порогового значения, т. е. при

$$A > \frac{k_2}{k_1}, \quad (5.40)$$

устанавливается (приближенно) распределение Пуассона

$$P^0(N) \approx \frac{N_0^N}{N!} e^{-N_0}, \quad N_0 = \frac{k_1 A}{k_2}. \quad (5.41)$$

Максимум этого распределения лежит вблизи N_0 , а стандартное отклонение равно

$$\sigma = N_0^{-1/2} = \left(\frac{k_1 A}{k_2} \right)^{-1/2}. \quad (5.42)$$

Стандартное отклонение достигает своего наибольшего значения в критической точке $A = k_2/k_1$ и стремится к нулю при $A \rightarrow 0$ и $A \rightarrow \infty$. На рис. 5.7 пунктиром схематически показано стационарное распределение вероятности в окрестности детерминистической кривой.

Рассмотрим теперь другой предельный случай, в котором удастся найти даже зависящее от времени решение. Предположим, что при $t = 0$ в системе имеется небольшое число M молекул сорта X , т. е.

$$P(N, 0) = \begin{cases} 1 & \text{при } N = M, \\ 0 & \text{при } N \neq M. \end{cases} \quad (5.43)$$

Предположим далее, что обеими обратными реакциями можно пренебречь. Как следует из соотношений (5.26) и (5.27), для этого должны выполняться условия:

$$k'_2 FV \ll k_1 AM,$$

$$k'_1 \frac{M-1}{V} \ll k_2. \quad (5.44)$$

Основное уравнение (5.22) преобразуется при этом к виду

$$\begin{aligned} \frac{\partial}{\partial t} P(N, t) = & k_1 A(N-1)P(N-1, t) + \\ & + k_2(N+1)P(N+1, t) - (k_1 A + k_2)NP(N, t). \end{aligned} \quad (5.45)$$

Решение этого уравнения с начальным условием (5.42) имеет следующий вид (Bartholomay, 1958, 1959; Eigen, 1973):

$$\begin{aligned} P(N, t) = & \sum_{n=0}^{N \text{ или } M} (-1)^n \binom{M}{n} \binom{M+N-n-1}{N-n} (k_1 A)^{N-n} k_2^{M-n} \times \\ & \times [\exp \{(k_1 A - k_2)t\} - 1]^{M+N-2n} \times \\ & \times [k_1 A \exp \{(k_1 A - k_2)t\} - k_2]^{n-M-N} \times \\ & \times [k_2 \exp \{(k_1 A - k_2)t\} - k_1 A]^n. \end{aligned} \quad (5.46)$$

Верхний предел суммирования равен N при $0 \leq N < M$ и M при $M \leq N$. Как показывает распределение (5.46), начальное число частиц M сорта X со временем может как возрастать, так и убывать. Если в детерминистической картине «приговор» вполне однозначен: «Рост при $k_1 A > k_2$ и распад при $k_1 A < k_2$ », то в стохастической картине вывод носит гораздо более сложный вероятностный характер. В качестве особенно интересного случая рассмотрим вероятность полного исчезновения («вымирания») $N = 0$ сорта X по истечении времени t . Из формулы (5.46) следует, что

$$P(0, t) = \left\{ \frac{\exp \{(k_1 A - k_2)t\} - 1}{(k_1 A/k_2) \exp \{(k_1 A - k_2)t\} - 1} \right\}^M. \quad (5.47)$$

Для вероятности того, что исчезновение молекул сорта X вообще происходит ($t \rightarrow \infty$), мы получаем

$$\lim_{t \rightarrow \infty} P(0, t) = \begin{cases} \left(\frac{k_2}{k_1 A} \right)^M & \text{при } k_1 A > k_2, \\ 1 & \text{при } k_1 A < k_2. \end{cases} \quad (5.48)$$

В то время как при $k_1 A < k_2$ сорт X заведомо исчезает, при $k_1 A > k_2$ имеется некоторая вероятность того, что молекулы сорта X исчезнут. Соответственно, вероятность выживания этих молекул определяется величиной

$$\sigma = 1 - \lim_{t \rightarrow \infty} P(0, t) = \begin{cases} 1 - \left(\frac{k_2}{k_1 A} \right)^M & \text{при } k_1 A > k_2, \\ 0 & \text{при } k_1 A < k_2. \end{cases} \quad (5.49)$$

Следовательно, молекулы сорта X заведомо «вымирают» при $k_1 A < k_2$, но при $k_1 A > k_2$ «выживают» лишь с вероятностью, которая меньше единицы.

5.3. Самовоспроизводящиеся автоматы

Теория клеточных автоматов берет начало от лекции «Общая и логическая теория автоматов», прочитанной Джоном фон Нейманом в 1948 г. Ныне теория клеточных автоматов принадлежит к числу наиболее быстро развивающихся областей математики и естественных наук (Wolfram, 1984). Методы этой теории позволили

существенно продвинуться в анализе весьма сложных процессов и гидродинамических задач, казавшихся до того неразрешимыми. Особенно важное и интересное для нас направление теории автоматов было развито фон Нейманом (*von Neumann*, 1966). В своей монографии «Теория самовоспроизводящихся автоматов» фон Нейман предпринял попытку перенести на абстрактные автоматы некоторые фундаментальные свойства биологических систем. Придуманый фон Нейманом автомат весьма сложен. Поэтому мы начнем изложение теории клеточных автоматов с существенно более простого автомата, названного английским математиком Конуэем игрой «Жизнь». Игра Конуэя ныне входит в пакеты программ, придаваемых многим компьютерам, разворачивается на разделенной на клетки доске, напоминающей шахматную доску (рис. 5.9). Клетки нумеруются числами

$$\begin{aligned} i &= 1, 2, \dots, n, \\ j &= 1, 2, \dots, m. \end{aligned} \quad (5.50)$$

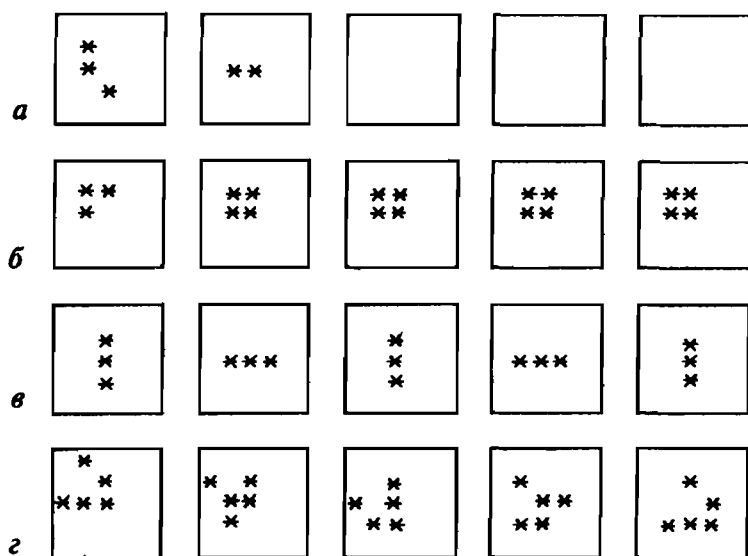


Рис. 5.9. Характерные конфигурации в игре «Жизнь» Конуэя: (а) вымиравшая популяция; (б) переход в стационарную популяцию; (в) осциллирующая популяция; (г) самовоспроизводящаяся дрейфующая популяция

Каждая клетка может находиться в состоянии 0 или 1. Развитие во времени начального (при $t = 0$) распределения занятых клеток происходит по следующим правилам.

1. Выживание. Занятая клетка выживает в следующем поколении, если заняты две или три соседние клетки.

2. Вымирание. Занятая клетка ($z = 1$) превращается в свободную ($z = 0$), если число занятых соседних клеток больше трех или меньше двух.

3. Рождение. Свободная (пустая) клетка превращается в занятую, если число соседних занятых клеток равно трем.

В ходе игры Конуэя на доске возникают весьма замысловатые и оригинальные конфигурации, которые легко моделируются на компьютере. Игра Конуэя позволяет моделировать значительную часть таких свойств биологических систем, как рост, гибель и взаимодействие (*Eigen, Winkler*, 1975, 1979).

Автомат фон Неймана, о котором пойдет речь, устроен гораздо сложнее. Он предназначен для того, чтобы сделать доступными математическому анализу процессы самовоспроизведения и эволюции. Речь идет о попытке найти ответы на следующие пять основных вопросов.

1. Логическая универсальность.

- При каких условиях определенный класс автоматов логически универсален?
- Существует ли логически универсальный автомат?

Конечный автомат с неограниченно продолжаемой лентой называется автоматом, или машиной, Тьюринга. Тьюринг показал, что предложенный им класс автоматов логически универсален, т. е. автоматы Тьюринга могут выполнить произвольный логический процесс (произвольное вычисление), если их снабдить конечным, но сколь угодно продолжаемым запоминающим механизмом (памятью). Тьюринг показал также, что существует универсальная машина Тьюринга, способная выполнять любые вычисления.

Тем самым Тьюринг дал положительный ответ на вопрос о логической универсальности, и фон Нейман поставил поэтому следующий вопрос.

2. Конструируемость.

- а) Может ли один автомат быть построен другим автоматом?
- б) Какой класс автоматов может быть построен каким-то автоматом?

3. Конструктивная универсальность.

- а) Существует ли конструктивно универсальный автомат (т. е. автомат, способный построить любой автомат)?

4. Самовоспроизведение.

- а) Существует ли самовоспроизводящийся автомат?
- б) Существует ли автомат, который помимо самовоспроизведения может решать и другие задачи?

5. Эволюция.

- а) Может ли при конструировании автомата автоматом происходить усложнение типа автомата?
- б) Может ли такая эволюция происходить в направлении от менее эффективного к более эффективному автомату (при надлежащем определении понятия эффективности)?

На все эти вопросы фон Нейман дал положительные ответы с помощью доказательства существования, построив специальный автомат, обладавший требуемыми свойствами. Фон Нейман воспользовался клеточным автоматом, у которого каждая клетка может находиться в 29 состояниях. Клеточный автомат состоит из многих однотипных автоматов, расположенных в узлах решетки; выход каждого автомата служит входом для соседних клеток. Нумерует клетки «радиус-вектор» (или вектор положения)

$$\vartheta = (i, j), \quad i, j = 0, \pm 1, \pm 2, \dots \quad (5.51)$$

(рис. 5.10).

Ближайшими соседями клетки (i, j) по определению считаются клетки $(i \pm 1, j)$ и $(i, j \pm 1)$, а ближайшими соседями — клетки $(i \pm 1, j \pm 1)$ (рис. 5.11).

В соответствии с этими определениями определим восемь различных векторов расстояния (рис. 5.12):

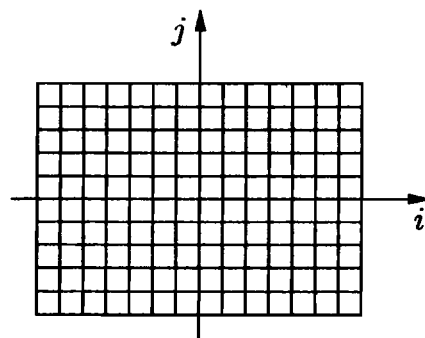
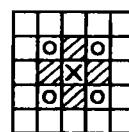


Рис. 5.10. Нумерация клеток клеточного автомата



○ близкие соседи
□ ближайшие соседи

Рис. 5.11. Соседние (по определению) клетки

$$\begin{aligned}
v^0 &= (1, 0), & v^1 &= (0, 1), \\
v^2 &= -v^0 = (-1, 0), & v^3 &= -v^1 = (0, -1), \\
v^4 &= (1, 1), & v^5 &= (-1, 1), \\
v^6 &= -v^4 = (-1, -1), & v^7 &= -v^5 = (1, -1).
\end{aligned} \tag{5.52}$$

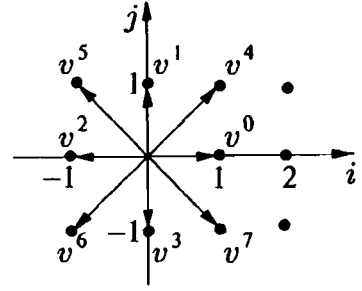


Рис. 5.12. Векторы разностей для обозначения соседей

Тогда $\vartheta + v^\alpha$ ($\alpha = 0, \dots, 3$) — ближайшие, а $\vartheta + v^\alpha$ ($\alpha = 4, \dots, 7$) — ближние соседи.

Время дискретно, т. е. изменяется по тактам:

$$t = 0, \pm 1, \pm 2, \tag{5.53}$$

на каждом такте каждая клетка ϑ находится в одном из n состояний $n = 0, 1, \dots, N - 1$, т. е. состояние на такте t есть

$$n_\vartheta^t. \tag{5.54}$$

Изменяются состояния по правилу перехода F , одинаковому для всех клеток (внутренняя однородность):

$$n_\vartheta^t = F(n_\vartheta^{t-1}; n_{\vartheta+v^\alpha}^{t-1} \mid \alpha = 0, \dots, 3). \tag{5.55}$$

В некоторых клеточных автоматах, например, в игре Конуэя «Жизнь» (*Eigen, Winkler, 1975*), F зависит еще и от ближних соседей, т. е. индекс α пробегает значения от 0 до 7.

В соотношении (5.55) F зависит от пяти переменных, которые могут принимать N значений, а поскольку F также принимает N различных значений при каждом значении аргумента, всего существует

$$m = N^{N^5} \tag{5.56}$$

различных передаточных функций F (различных моделей).

Автомат фон Неймана имеет $N = 29$ различных состояний (в игре Конуэя — только 2):

1. Транзитивные состояния («импульсы») $T_{u\alpha\epsilon}$, где $\tag{5.57}$

$$\begin{aligned}
u &= \begin{cases} 0 & \text{— обычное состояние,} \\ 1 & \text{— специальное состояние;} \end{cases} \\
\alpha &= \begin{cases} 0 & \text{— вправо,} \\ 1 & \text{— вверх,} \\ 2 & \text{— влево,} \\ 3 & \text{— вниз;} \end{cases} \\
\epsilon &= \begin{cases} 0 & \text{— состояние покоя,} \\ 1 & \text{— возбужденное состояние.} \end{cases}
\end{aligned}$$

2. Конфлюэнтные состояния $C_{\epsilon\epsilon'}$, где $\tag{5.58}$

$$\begin{aligned}
\epsilon &= \begin{cases} 0 & \text{— состояние покоя,} \\ 1 & \text{— возбужденное состояние;} \end{cases} \\
\epsilon' &= \begin{cases} 0 & \text{— состояние покоя на следующем такте,} \\ 1 & \text{— возбужденное состояние на следующем такте;} \end{cases}
\end{aligned}$$

3. Основное состояние U (невозбужденное). $\tag{5.59}$

4. Чувствительные (сенситивные) состояния S_Σ , где

(5.60)

$$\Sigma = \Theta, 0, 1, 00, 01, 10, 11, 000$$

и

$$S_{0000} = T_{000},$$

$$S_{0001} = T_{010},$$

$$S_{001} = T_{020},$$

$$S_{010} = T_{030},$$

$$S_{011} = T_{100},$$

$$S_{100} = T_{110},$$

$$S_{101} = T_{120},$$

$$S_{110} = T_{130},$$

$$S_{111} = C_{00}.$$

Из общего числа $m \approx 10^{30\,000\,000}$ различных передаточных функций F фон Нейман выбирает следующую.

Основное состояние U может оставаться неизменным или путем возбуждения переходить в чувствительное состояние S . В последнем случае последнее автоматически пробегает определенную последовательность чувствительных состояний, которая неизбежно заканчивается конфлюэнтным состоянием C или транзитивным состоянием T . Оба конечных состояния могут попеременно находиться в возбужденной и невозбужденной форме, оставаться неизменными или переходить снова в основное состояние (рис. 5.13).

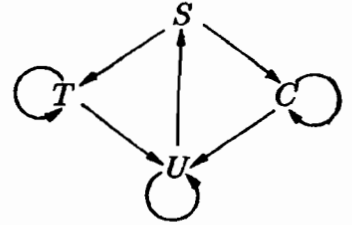


Рис. 5.13. Граф возможных переходов между четырьмя классами состояний

Более подробно передаточная функция F фон Неймана определяется следующими соотношениями.

1. Пусть $n_{\vartheta}^{t-1} = T_{u\alpha\epsilon}$.

а) $n_{\vartheta}^t = U$ в том и только в том случае, если $n_{\vartheta'}^{t-1} = T_{u'\alpha'1}$ для любой клетки ϑ' такой, что $\vartheta - \vartheta' = v^{\alpha'}$ и $u \neq u'$;

б) $n_{\vartheta}^t = T_{u\alpha 1}$ в том и только в том случае, если не выполняется условие (а) и выполняется условие (α) или (β):

α) $n_{\vartheta'}^{t-1} = T_{u\alpha'1}$ для любой клетки ϑ' , для которой $\vartheta - \vartheta' = v^{\alpha'} \neq -v^{\alpha}$,

β) $n_{\vartheta'}^{t-1} = C_1$ для любой клетки ϑ' , такой, что $\vartheta - \vartheta' = v^{\beta} \neq -v^{\alpha}$ ($\beta = 0, \dots, 3$);

в) $n_{\vartheta}^t = T_{u\alpha 0}$ в том и только в том случае, если не выполняется ни условие (а), ни условие (б).

2. Пусть $n_{\vartheta}^{t-1} = C_{\epsilon\epsilon'}$.

а) $n_{\vartheta}^t = U$ в том и только в том случае, если $n_{\vartheta'}^{t-1} = T_{1\alpha'1}$ при любом ϑ' , таком, что $\vartheta - \vartheta' = v^{\alpha'}$;

б) $n_{\vartheta}^t = C_{\epsilon'1}$ в том и только в том случае, если условие (а) не выполнено и выполняются условия (α) и (β):

$\alpha)$ $n_{\vartheta'}^{t-1} = T_{0\alpha'1}$ для любой клетки ϑ' , такой, что $\vartheta - \vartheta' = v^{\alpha'}$;

$\beta)$ $n_{\vartheta'}^{t-1} = T_{0\alpha'0}$ для всех клеток ϑ' , таких, что $\vartheta - \vartheta' = v^{\alpha'}$;

в) $n_{\vartheta}^t = C_{\epsilon'0}$ в том и только в том случае, если не выполняются условия (а) и (б).

3. Пусть $n_{\vartheta}^{t-1} = U$.

а) $n_{\vartheta}^t = S_0$ в том и только в том случае, если $n_{\vartheta'}^{t-1} = T_{u\alpha'1}$ для любой ячейки ϑ' , такой, что $\vartheta - \vartheta' = v^{\alpha'}$;

б) $n_{\vartheta}^t = U$ в том и только в том случае, когда условие (а) не выполняется.

4. Пусть $n_{\vartheta}^{t-1} = S_{\Sigma}$, $\Sigma = 0, \dots, 000$.

а) $n_{\vartheta}^t = S_{\Sigma 1}$ в том и только в том случае, если $n_{\vartheta'}^{t-1} = T_{u\alpha'1}$ для любой клетки ϑ' , такой, что $\vartheta - \vartheta' = v^{\alpha'}$;

б) $n_{\vartheta}^t = S_{\Sigma 0}$ в том и только в том случае, если условие (а) не выполняется.

Последнее правило можно было бы сформулировать следующим образом. Основное состояние U может быть переведено так называемым прямым процессом в чувствительное состояние и в конечном счете в $T_{u\alpha 0}$ или C_{00} . В обратном процессе состояния $T_{u\alpha 0}$ или $C_{\epsilon\epsilon'}$ разрушаются при переходе в основное состояние U . Конфлюэнтное состояние $C_{\epsilon\epsilon'}$, $\epsilon = 0, 1$, воспринимает сигналы, конъюнктивные состояниям $T_{0\alpha\epsilon}$, направленным на это состояние; с удвоенной задержкой оно возбуждает все состояния $T_{u\alpha\epsilon}$, не направленные к нему. При $\epsilon' = 0, 1$ конфлюэнтное состояние $C_{\epsilon\epsilon'}$ разрушается состоянием $T_{1\alpha 1}$, не направленным к нему, путем перехода в основное состояние U . Разрушение доминирует над восприятием импульсов. Транзитивное состояние $T_{0\alpha\epsilon}$ (рис. 5.14) воспринимает дизъюнктивные сигналы направленных к нему состояний $T_{0\alpha\epsilon}$ и состояний $C_{\epsilon\epsilon'}$. Оно передает импульс в исходном направлении с однократной задержкой:



Рис. 5.14. Транзитивные состояния первого рода

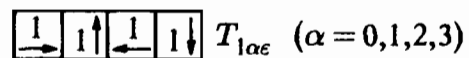


Рис. 5.15. Транзитивные состояния второго рода

а) состоянию $T_{0\alpha\epsilon}$, не направленному к нему, или состоянию $C_{\epsilon\epsilon'}$;

б) основному состоянию U или чувствительному состоянию в прямом процессе;

в) разрушает состояние $T_{1\alpha\epsilon}$ в обратном процессе.

Состояние разрушается состоянием $T_{1\alpha 1}$, которое не направлено на него, при переходе в основное состояние U . И в этом случае разрушение доминирует над восприятием импульса. Для транзитивного состояния $T_{1\alpha\epsilon}$ (рис. 5.15) выполняется такое же правило перехода, как для $T_{0\alpha\epsilon}$, только первый индекс 0 переставлен с индексом 1 и наоборот. Наконец, чувствительные состояния S_{Σ} — возникающие в прямом процессе промежуточные состояния $T_{u\alpha 1}$, направленные к основному состоянию U , — переводят U в S_{Σ} . Затем S_{Σ} переходит

а) в состояние $S_{\Sigma 1}$, если к последнему направлено любое состояние $T_{u\alpha 1}$;

б) в состояние $S_{\Sigma 0}$ в противном случае, причем так происходит до тех пор, пока прямой процесс не завершается в состоянии $T_{u\alpha 0}$ или C_{00} .

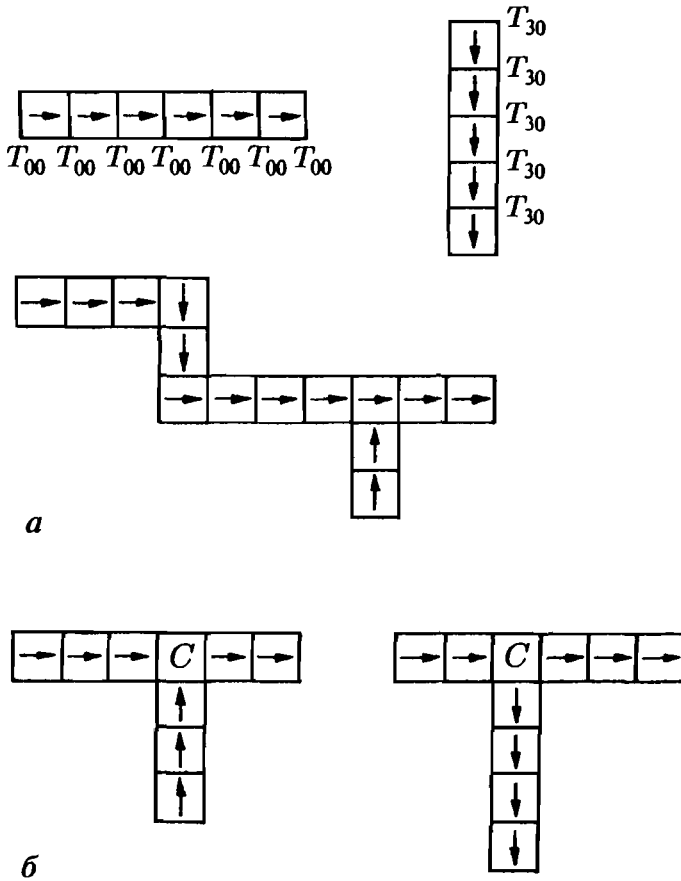


Рис. 5.16. Проведение и обработка импульса транзитивными и конфлюэнтными клетками

устройства, дискриминатора, конструирующего шупа и т. д., на структуре которых мы не будем останавливаться. Даже гораздо более простая игра Конуэя обладает интересными свойствами, позволяющими утверждать, что существуют автоматы, изготавливающие другие автоматы. Возможно ли на этом пути создание самовоспроизводящихся автоматов, окончательно еще не известно.

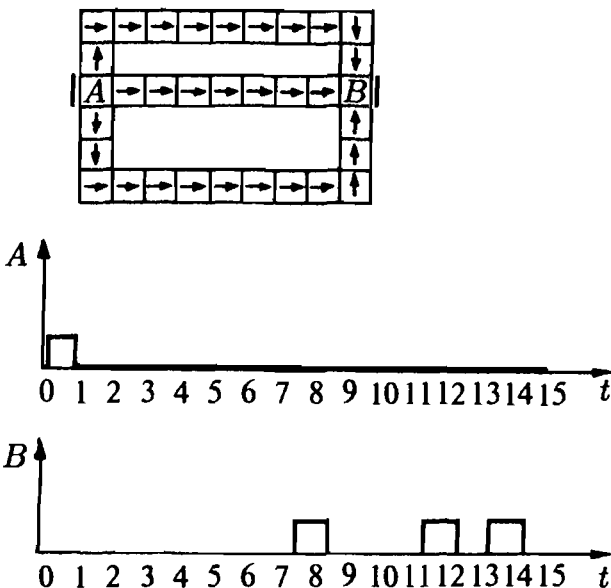


Рис. 5.17. Задержка импульса с помощью обходных путей

На рис. 5.16 а показаны простые транспортные состояния, т. е. цепочки клеток, способных проводить последовательность импульсов, а на рис. 5.16 в — такие цепочки с встроенными в них конфлюэнтными состояниями. С помощью проводящих путей различной длины из отдельных импульсов можно составлять определенные последовательности импульсов (рис. 5.17). Здесь импульс, проведенный при $t = 0$ клеткой А, вызывает возбуждение в клетке В при $t = 8, 12$ и 14 . Кроме того, такое «нервное волокно» может расти, как показывает прямой процесс, изображенный на рис. 5.18. Штриховкой показаны, какие клетки возбуждены.

Собственно самовоспроизводящийся автомат представляет собой весьма сложную конструкцию. Он состоит из множества узлов и деталей, таких, как лента памяти, считающего устройства, генераторов импульсов, сумматора, множительных устройств, декодера, распознающего

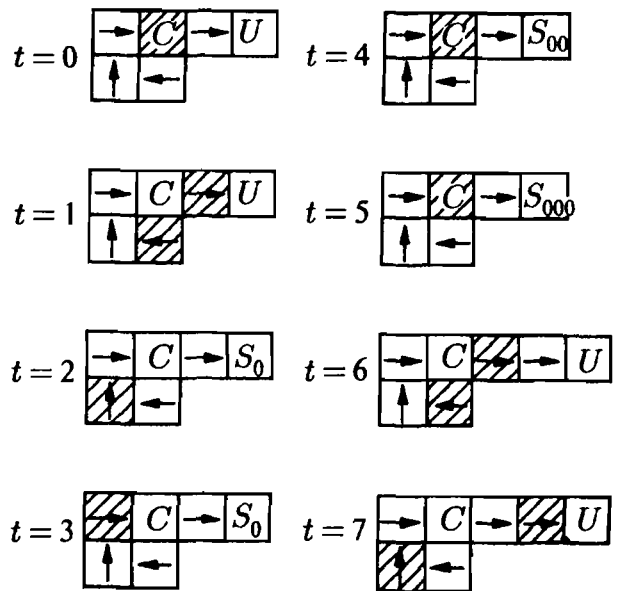


Рис. 5.18. «Прямой процесс», моделирующий растущее нервное волокно

5.4. Динамика нелинейного самовоспроизведения

Рассмотренные выше законы роста самовоспроизведения имеют линейный характер, т. е. скорость роста прямо пропорциональна числу уже имеющихся элементов (молекул, живых организмов и т. д.). Существует ряд важных процессов, в которых рост происходит быстрее или медленнее. В первом случае говорят о самоускорении, во втором — о самоторможении. И в том, и в другом случае, если самовоспроизводство оказывает ускоряющее или тормозящее обратное действие на процесс (а такое бывает очень часто), законы роста должны быть нелинейными. В качестве примера можно было бы сослаться на рост новой моды: сначала мы наблюдаем надлинейный (т. е. более быстрый, чем линейный) рост, а затем в области насыщения — слабый подлинейный (т. е. более медленный, чем линейный) рост до предельного значения. Детерминистическая теория процессов нелинейного роста была всесторонне исследована Пешелем и Менде (*Peschel, Mende, 1983, 1986*). Мы ограничимся поэтому рассмотрением нескольких специальных случаев, в частности, исследуем системы, обладающие несколькими стационарными состояниями. Наличие мультистабильности придает динамической системе большую гибкость, так как в зависимости от своей предыстории она может переходить в различные состояния. Это свойство памяти может послужить основой для хранения информации. Такого рода хранение информации, по-видимому, играло определенную роль на самых ранних стадиях биологической эволюции (*Decker, 1975, Heidmann, Decker, Pohlmann, 1977*). Деккер предложил называть открытые химические системы, обладающие несколькими стационарными состояниями, биоидами. Простейшей моделью биоида может служить так называемая реакция Шлёгля



обладающая двумя устойчивыми стационарными состояниями для концентрации катализатора X . Хайдман, Деккер и Польман (*Heidmann, Decker, Pohlmann, 1977*) показали, что реакции типа (5.61) встречаются при образовании сахара из формальдегида в присутствии ионов кальция и ионов гидроксила. Экспериментально установлено, что эта реакция (так называемая формоль-реакция) представляет собой автокаталитическую реакцию второго порядка относительно продукта (сахара):



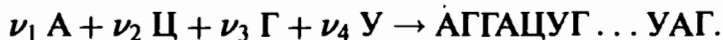
Здесь A — формальдегид, X — сахар. При интерпретации этой реакции исходят из предположения о комплексе из иона Ca^{++} и двух лигандов сахара, реагирующем с формальдегидом. Хайдеман, Деккер и Польман предполагают, что формоль-реакция могла играть в ходе биологической эволюции роль первого биоида. Свою гипотезу они обосновывают среди прочего тем, что нуклеонпротеиновые реакции (репликация полинуклеотидов в присутствии белков) в силу их сложности имеют гораздо меньшую вероятность на спонтанную реализацию.

Как бы ни происходило хранение информации на самых ранних стадиях эволюции, ныне в основном используется принцип хранения последовательностей нуклеотидов в дезоксирибонуклеиновой кислоте (ДНК) или в некоторых случаях в рибонуклеиновой кислоте (РНК). Различные возможные первичные структуры полинуклеотидов с одной и той же брутто-формулой (изомеры) мы можем рассматривать как мультистабильные консервативные структуры полинуклеотидов, а реакцию образования тех или иных первичных структур — как мультистабильную реакцию.

Рассмотрим в качестве схематического примера реакцию, приводящую к синтезу молекулы полинуклеотида (РНК) из мономеров аденина (A), цитозина (C), гуанина (G) и урацила (U):



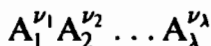
Этот синтез можно рассматривать как мультистабильную реакцию, так как при другом ходе реакции могла бы возникнуть другая последовательность мономеров, например,



Различные первичные структуры (т. е. структуры с различной последовательностью мономеров) практически не могут спонтанно превращаться друг в друга, поскольку они разделены высокими потенциальными барьерами. Их можно рассматривать как различные уровни возбуждения одной и той же молекулы и по аналогии с ядерной физикой называть изомерами (*Schrödinger*, 1944, 1951). Линейная полимерная молекула, состоящая из ν_1 мономеров сорта A_1 , ν_2 мономеров сорта A_2 и ν_λ мономеров сорта A_λ и имеющая длину

$$\nu = \nu_1 + \nu_2 + \dots + \nu_\lambda,$$

обладает $W = \nu! / (\nu_1! \nu_2! \dots \nu_\lambda!)$ различными первичными структурами (изомерами). Так как длина биополимеров может быть очень большой (ν от 10^2 до 10^8), число изомеров обычно астрономически велико. Соответственно, синтез биополимера с брутто-формулой



представляет мультистабильную реакцию с чудовищно большим числом каналов (рис. 5.19):

$$\nu_1 A_1 + \nu_2 A_2 + \dots + \nu_\lambda A_\lambda \rightarrow \begin{cases} W \text{ каналов,} \\ W = \frac{(\nu_1 + \nu_2 + \dots + \nu_\lambda)!}{\nu_1! \nu_2! \dots \nu_\lambda!} \end{cases}$$

Термодинамическое равновесие между различными изомерами устанавливаться не может, так как энергия активации вследствие сильных ковалентных связей между мономерами очень велика. С точки зрения термодинамики макроскопическая система, в которой имеются первичные структуры, определяемые биополимерами, находится в метастабильном состоянии с чрезвычайно большим временем жизни. Однако неравновесный характер присущ не только первичным структурам биополимеров, но довольно часто вторичным и третичным структурам (*Волькенштейн*, 1978, 1981; *Blumenfeld*, 1977).

Неравновесные структуры биополимеров, как правило, являются не диссипативными структурами, так как для их поддержания не требуется экспорт энтропии (или импорт свободной энергии), а консервативными структурами.

В последующих разделах мы анализируем ряд метастабильных систем, характер которых обусловлен экспортом энтропии, в силу чего их надлежит считать диссипативными структурами.

Начнем с исследования автокаталитических реакций высокого порядка. Мы говорим об автокатализе высокого порядка, если скорость образования вещества X возрастает с увеличением концентрации этого вещества быстрее, чем линейная

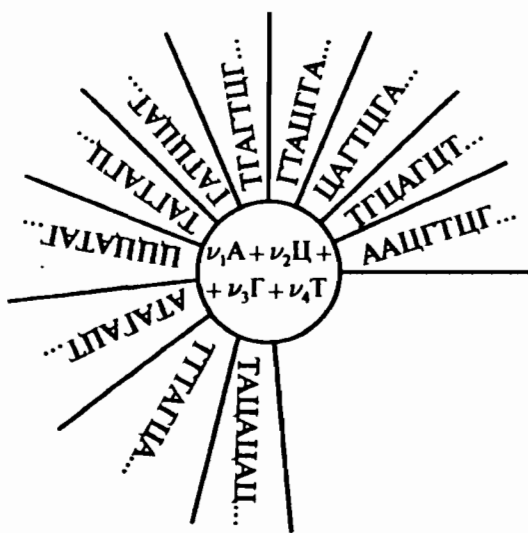
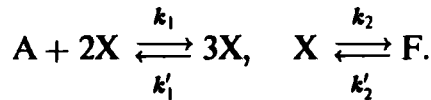


Рис. 5.19. Мультистабильность при синтезе биополимеров из четырех оснований: аденина, тимина, гуанина и цитозина

функция. Шлёгль впервые показал, что такие системы могут вести себя, как бистабильные. Реакция Шлёгля состоит из четырех элементарных реакций:



Эту систему реакций можно рассматривать как модель формоль-реакции (Heidmann, Decker, Pohlmann, 1977). В предположении, что элементарные реакции протекают по законам формальной кинетики, мы получаем следующее реакционно-кинетическое уравнение

$$\frac{d}{dt}X(t) = k_1AX^2 - k'_1X^3 - K_2X + K'_2F. \quad (5.63)$$

Напомним, что аналогичное уравнение было рассмотрено нами в разд. 4.3. Вводя единицы

$$X_0 = \frac{k_1A}{k'_1}, \quad t_0 = \frac{k'_1}{k_1^2A^2},$$

коэффициенты при двух старших членах в правой части уравнения (5.63) можно сделать равными 1. Уравнение (5.63) переходит при этом в следующее:

$$\begin{aligned} \dot{X} &= -X^3 + X^2 - \beta X + \gamma, \\ \beta &= k_2t_0, \quad \gamma = k'_2t_0 \frac{F}{X_0}. \end{aligned} \quad (5.64)$$

В стационарном состоянии правая часть уравнения (5.64) должна обращаться в нуль, т. е.

$$\gamma = [X^{(1)}]^3 - [X^{(1)}]^2 + \beta X^{(1)}. \quad (5.65)$$

На рис. 5.20 стационарные решения представлены как функции параметров β и γ . С первого взгляда бросается в глаза сходство с кривыми Ван-дер-Ваальса из термодинамики реальных газов.

Нетрудно видеть, что при $\beta > 1/3$ система всегда моностабильна, так как при каждом значении γ существует только одно устойчивое стационарное решение. При $\beta < 1/3$ в интервале $0 < \gamma < 1/27$ всегда найдется такая область значений γ , для которой существуют три стационарных состояния

$$X^{(1)} < X^{(2)} < X^{(3)},$$

состояния $X^{(1)}$ и $X^{(3)}$ устойчивы, а состояние $X^{(2)}$ неустойчиво. Зависящее от времени решение уравнения (5.64) можно найти с помощью трех корней. Интегрируя уравнение (5.64), получаем

$$t = - \int_{X(0)}^{X(t)} \frac{dx}{(X - X^{(1)})(X - X^{(2)})(X - X^{(3)})}. \quad (5.66)$$

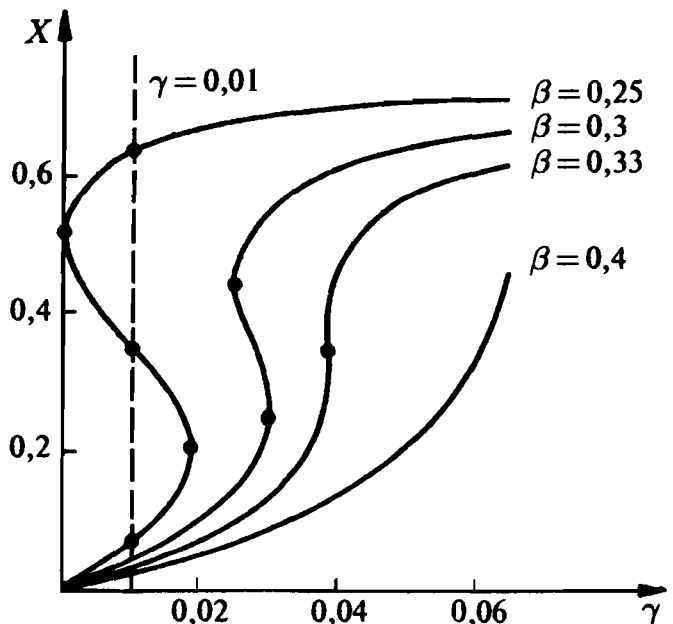


Рис. 5.20. Стационарные состояния X в зависимости от значений управляющих параметров β и γ модели Шлёгля

Как показывает анализ выражения (5.66), неустойчивое стационарное решение образует сепаратрису в фазовом пространстве (рис. 5.21), в то время как устойчивые стационарные решения $X^{(1)}$ и $X^{(3)}$ являются аттракторами для траекторий. Чтобы перевести систему из одного устойчивого состояния в другое, необходимо отклониться от сепаратрисы $X^{(2)}$. Спонтанные переходы в рамках детерминистической теории исключаются.

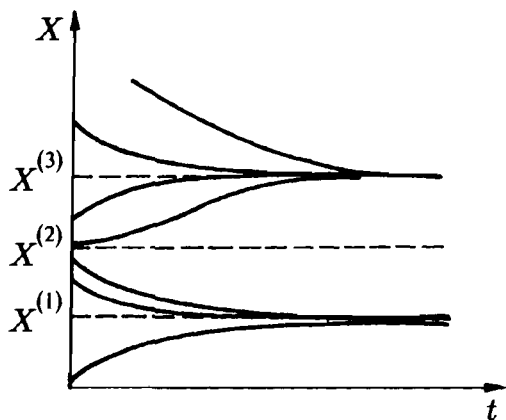


Рис. 5.21. Временная зависимость решений в бистабильной модели Шлёгеля

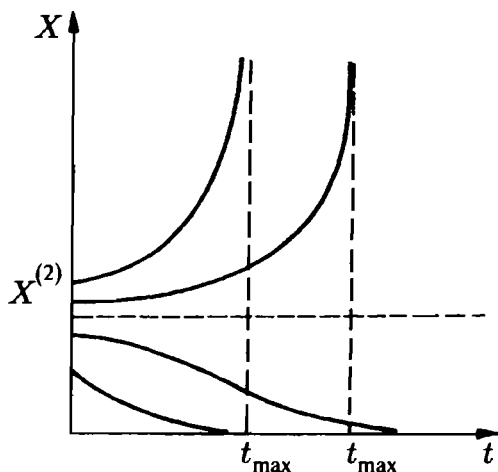
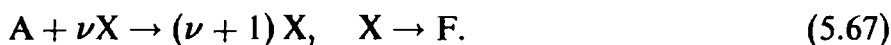


Рис. 5.22. Гиперболический рост приводит за конечные времена к вымиранию или к расходимости (поведение типа «всё или ничего»)

Рассмотрим теперь общий случай: для того, чтобы из молекулы исходного вещества A образовалась одна молекула X , должны сойтись вместе ν молекул вещества X ($\nu > 1$ — произвольное целое число):



Для простоты мы пренебрегаем обратными реакциями. В рамках формальной кинетики получается кинетическое уравнение

$$\dot{X} = kAX^\nu - k'X. \quad (5.68)$$

При постоянном запасе исходного вещества ($A = \text{const}$) и $\nu > 1$ (случай $\nu = 1$ мы рассмотрели в разд. 5.1) существуют два стационарных состояния

$$X^{(1)} = 0, \quad X^{(2)} = \left[A \frac{k}{k'} \right]^{1/(1-\nu)}. \quad (5.69)$$

Как видно из рис. 5.22, $X^{(2)}$ представляет собой критическую концентрацию, ниже которой преобладает скорость распада, а выше — скорость синтеза. То, что истекает из этого для временной зависимости $X(t)$, мы увидим, проинтегрировав уравнение (5.68). Так как эта зависимость бернуллиевского типа, мы, производя замену переменной

$$z(t) = [X(t)]^{1-\nu}, \quad (5.70)$$

легко приходим к цели:

$$X(t) = X^{(2)} \left[1 + \left[\left(\frac{X(0)}{X^{(2)}} \right)^{\nu-1} - 1 \right] \times \exp \{ (\nu - 1) k' t \} \right]^{(1-\nu)/2}. \quad (5.71)$$

Поведение некоторых из решений (5.71) представлено на рис. 5.22; $X(t) = X^{(2)}$ оказывается неустойчивым стационарным состоянием. Оно соответствует состоянию $X^{(2)}$ в реакции Шлёгеля. Если начальная концентрация лежит ниже $X^{(2)}$, то

в конце концов происходит распад всех молекул. Если же начальная концентрация превышает $X^{(2)}$, то за время

$$t_{\max} = -\frac{\nu - 1}{2k'} \ln \left[1 - \left(\frac{X^{(2)}}{X(0)} \right)^{\nu-1} \right], \quad (5.72)$$

зависящее от начальной концентрации $X(0)$, концентрация становится бесконечно большой. Такого рода режимы принято называть режимами типа «все или ничего» (Schuster, 1972; Эйген, Шустер, 1982); они соответствуют «бесповоротному» выбору одной из двух возможностей и представляют особый интерес при анализе процессов конкуренции и отбора. Расходимость формально является следствием пренебрежения обратными реакциями, т. е. $X^{(3)}$ вырождается в бесконечность. Расходимость концентраций, разумеется, является нереалистической чертой нашей модели, в природе ничего подобного не наблюдается. Но эта трудность легко исключается, если мы учтем в теории обратные реакции.

Еще одна возможность состоит в том, чтобы связать концентрации двух различных веществ при условии постоянства их суммарной концентрации $X_1 + X_2$. Для этого вместо постоянной скорости распада следует ввести управляемую скорость разрежения:

$$\dot{X}_i = k_i A X_i^\nu - \Phi(t) X_i, \quad i = 1, 2. \quad (5.73)$$

Из условия $X_1 + X_2 = C$ получаем:

$$\Phi(t) = \frac{A}{C} [k_1 X_1^\nu + k_2 X_2^\nu]. \quad (5.74)$$

Рассмотрим теперь решения уравнений (5.73). После некоторых вычислений находим три стационарных состояния:

$$X_1^{(1)} = 0, \quad X_2^{(1)} = C, \quad (5.75)$$

$$X_1^{(2)} = C \left[1 + \left(\frac{k_1}{k_2} \right)^{1/(\nu-1)} \right]^{-1}, \quad (5.76)$$

$$X_2^{(2)} = C \left[1 + \left(\frac{k_2}{k_1} \right)^{1/(\nu-1)} \right]^{-1},$$

$$X_1^{(3)} = C, \quad X_2^{(3)} = 0. \quad (5.77)$$

Здесь $C = X_1 + X_2$. При анализе устойчивости могут представиться три качественно различные ситуации (мы предполагаем, что $k_1 > k_2$).

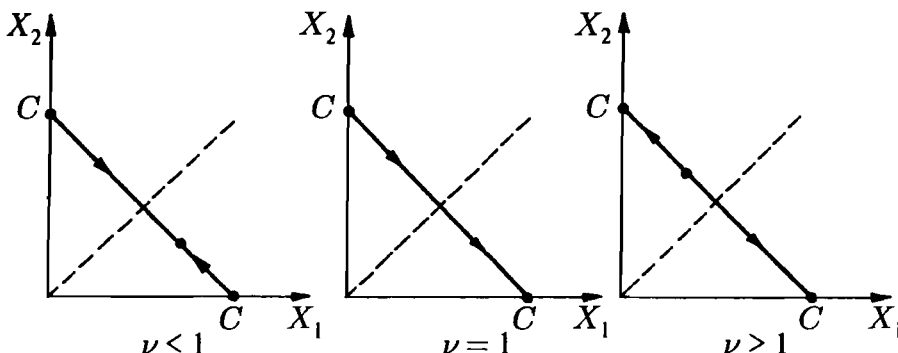


Рис. 5.23. Стабилизация при гиперболическом росте концентраций двух компонент и постоянной суммарной концентрации

При $\nu < 1$ решения (5.75) и (5.77) соответствуют неустойчивым состояниям, а решение (5.76) — устойчивому состоянию. При $\nu > 1$ ситуация прямо противоположна. В промежуточном случае $\nu = 1$ решения (5.76) и (5.77) совпадают, и, как уже известно, решение (5.75) неустойчиво, а решение (5.77) устойчиво. На рис. 5.24 показано, как зависят положение и устойчивость стационарных состояний от ν .

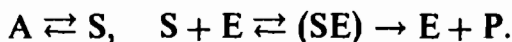
В области автокатализа высшего порядка, т. е. при $\nu > 1$, мы обнаруживаем тем самым бистабильное поведение. Состояние сосуществования (5.76) неустойчиво, в зависимости от начальных условий выживает то один, то другой сорт молекул.

Бифуркационная диаграмма (рис. 5.24) имеет в этом случае совершенно другой вид, чем в предыдущем разделе (рис. 5.5). Однако тесная взаимосвязь между той и другой бифуркационной диаграммой становится очевидной, если мы рассмотрим кинетическое уравнение (5.68) для X_1 и $\nu = 2$. Заменяя X_2 на $C - X_1$, получаем

$$\dot{X}_1 = \frac{A}{C} X_1 \{ -k_2 C^2 + (k_1 + 2k_2) C X_1 - (k_2 + k_1) X_1^2 \}. \quad (5.78)$$

График зависимости скорости реакции качественно не отличается от кривой для реакции Шлёгля (рис. 5.20).

С явлением бистабильности мы встречаемся не только в случае автокаталитических реакций, но и при изучении определенного класса ферментативных реакций (Жаботинский, 1974; Ebeling, Czajkowski, 1976; Czajkowski, Ebeling, 1977). Рассмотрим реакцию между ферментом E и субстратом S, протекающую по следующей схеме:



Здесь A — исходное вещество, или резервуар для субстрата S, P — продукт, отводимый из системы. На промежуточной ступени образуется комплекс фермент-субстрат, который распадается на свободный фермент и на продукт (рис. 5.25). Действие фермента можно представить себе, как работу фабрики, на которую поступает сырье A и с которой вывозятся продукты. За подробностями ферментативной кинетики мы отсылаем к обширной специальной литературе (Hofmann, 1979, 1980).

По поводу моделирования сложных ферментативных цепей мы также вынуждены ограничиться отсылкой к литературе (Reich, Selkov, 1981; Романовский, Степанова, Чернавский, 1984; Иваницкий, Кринский, Сельков, 1978).

Мы проанализируем только простейший случай ферментативной реакции с одной-единственной степенью свободы:

$$\frac{dS}{dt} = A - BS - v(S). \quad (5.79)$$

Здесь S — концентрация субстрата, рассматриваемая в качестве переменной, A — приток субстрата, B — скорость распада субстрата, $v(S)$ — скорость реакции фермент-субстрат. Наблюдаемую эмпирически зависимость скорости реакции

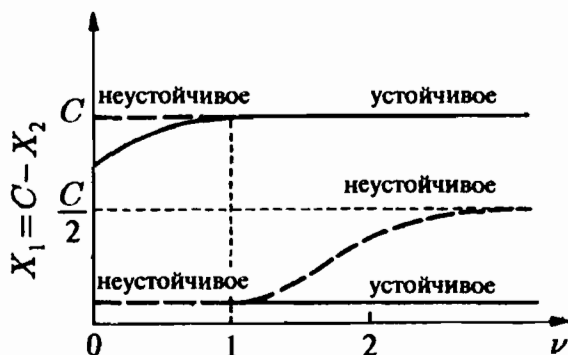


Рис. 5.24. Бифуркационная диаграмма для двух конкурирующих компонент в зависимости от порядка ν закона роста

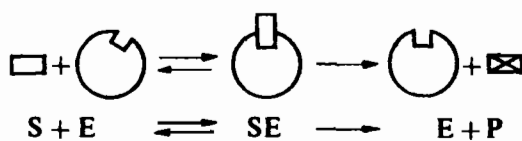


Рис. 5.25. Механизм Михаэлиса—Ментен в ферментативной кинетике

от концентрации субстрата в хорошем приближении можно представить в виде отношения двух многочленов (Clealand, 1963):

$$v(S) = \frac{\sum_{n=1}^p e_n S^n}{1 + \sum_{m=1}^q g_m S^m}. \quad (5.80)$$

Здесь всегда выполняется неравенство $p \leq q$, причем более частый случай $p = q$ соответствует насыщению скорости реакции с увеличением концентрации субстрата (кинетика Михаэлиса—Ментен), а $p < q$ — торможению реакции при возрастании концентрации субстрата. Различные возможные зависимости скорости реакции представлены на рис. 5.26. Заданные условием стационарные концентрации на рис. 5.26 соответствуют точкам пересечения

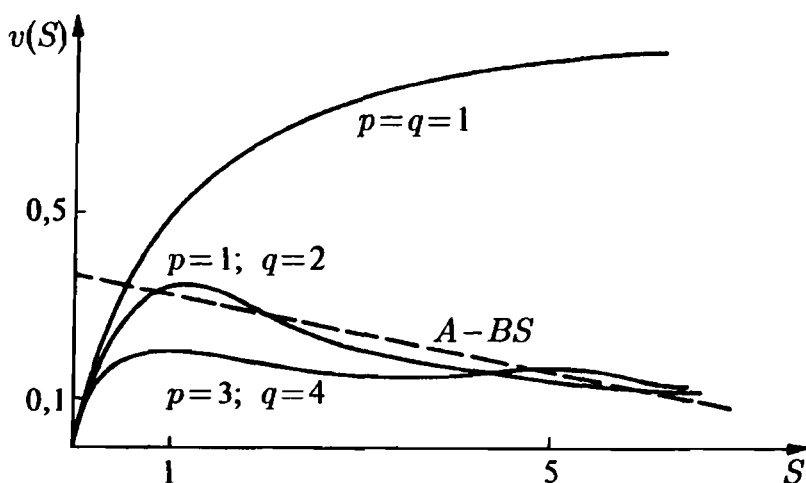


Рис. 5.26. Различные зависимости скорости ферментативной реакции v от концентрации субстрата S

$v(S)$ с прямой $A - BS$. Нетрудно видеть, что график $v(S)$ в случае кинетики насыщения имеет только одну точку пересечения с прямой $A - BS$, но в случае кинетики торможения ($p < q$) число точек пересечения может быть равно трем или большему нечетному числу. Это означает, что можно найти определенную область значений параметров реакции, которые приводят к бистабильности или даже к мультистабильности (Czajkowski, Ebeling, 1977). В простейшем случае $p = 1$, $q = 2$ поверхность стационарных состояний

$$A - BS = \frac{e_1 S}{1 + g_1 S + g_2 S^2} \quad (5.81)$$

имеет форму сборки, или «нависающего утеса» (рис. 5.27). Соответствующая бифуркационная диаграмма (рис. 5.28) имеет форму острия — на ней существует точка

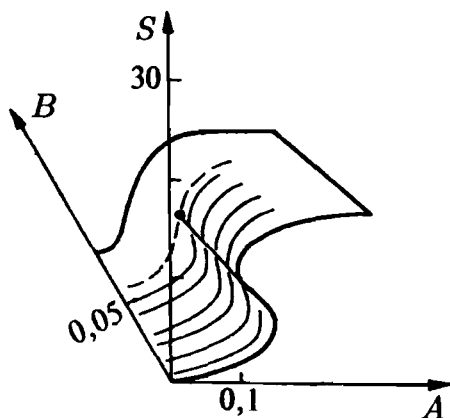


Рис. 5.27. Сборка («нависающий утес») стационарных состояний заторможенной ферментативной реакции

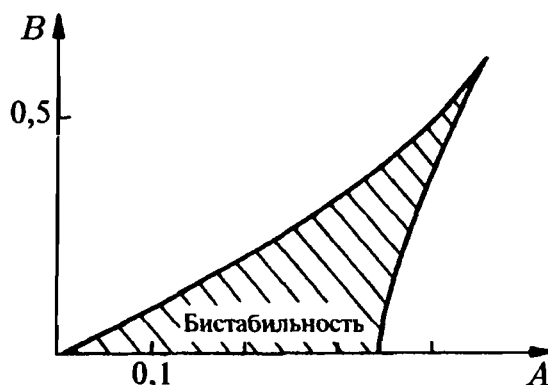


Рис. 5.28. Бифуркационная картина бистабильной ферментативной реакции

возврата. В смысле Тома (*Thom*, 1972) речь идет о типичной катастрофе типа острия (*Арнольд и др.*, 1984; *Арнольд*, 1990). При больших p и q наблюдаются катастрофы более высокого порядка (*Ebeling*, 1978). Например, при $p = 3$, $q = 4$ наблюдается катастрофа бабочки.

Рассмотрим теперь симметричную бистабильную систему, выбранную Моно и Жакобом в качестве модели управляемого синтеза белка в клетке (*Романовский, Степанова, Чернавский*, 1984). Почти все клетки обладают полным набором хромосом. Так как генетический аппарат и тем самым генетическая информация клеток одинаковы, наблюдаемое различие клеток и их распределение по классам связаны с блокировкой одной части информации и активацией другой части информации. Синтез белков, соответствующих блокированной части ДНК, не происходит, в то время как синтез белков, соответствующих активированной части ДНК, происходит запрограммированным образом. В качестве переключательного механизма между активацией и блокировкой Жакоб и Моно (*Jacob, Monod*, 1961) предложили модель, которая в упрощенном виде представлена на рис. 5.29. Синтез белка E_1 происходит с помощью транспортной РНК, образующейся на гене G_1 . Продукт P_1 этого синтеза действует как репрессор на второй ген G_2 , причем (косвенно) действует на оперонную часть этого гена O_2 . Продукт P_2 синтеза, управляемого геном G_2 , воздействует на оперон O_1 гена G_1 . По Чернавскому и др., временное развитие концентраций продуктов P_1 и P_2 описывается следующей системой уравнений:

$$\frac{d}{dt}P_2 = \frac{A_2}{B_2 + P_1^m} - q_2P_2. \quad (5.82)$$

Для показателя m справедливо условие взаимного торможения $m \geq 2$. Как видно из фазового портрета, изображенного на рис. 5.30, система (5.82) обладает двумя устойчивыми стационарными состояниями

- 1) $P_1 > P_2$,
- 2) $P_2 > P_1$.

В первом из этих двух состояний синтезируется только белок E_1 , а синтез белка E_2 блокирован, в то время как во втором состоянии синтезируется белок E_2 , а синтез белка E_1 блокирован. По современным представлениям, регулировочный механизм по Жакобу и Моно играет большую роль при дифференциации клеток (*Николис, Пригожин*, 1979; *Babloyantz, Nicolis*, 1972).

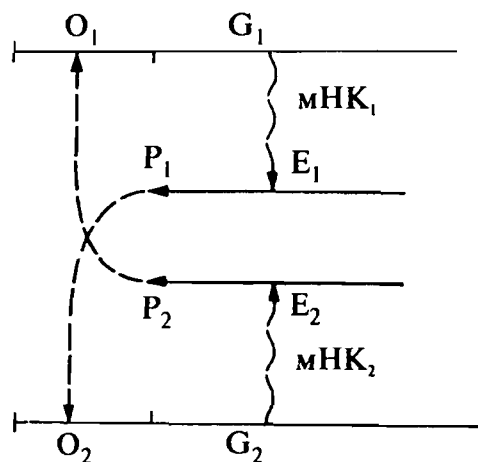


Рис. 5.29. Упрощенная модель механизма индуктор-репрессор по Жакобу—Моно

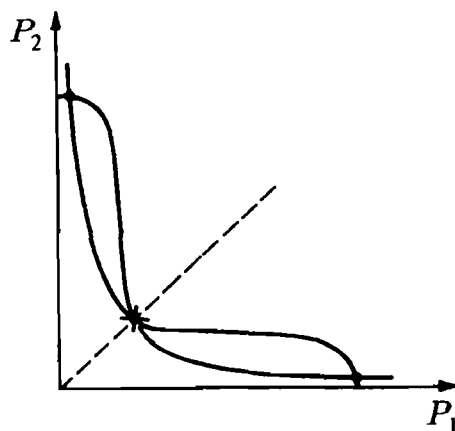
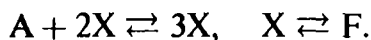


Рис. 5.30. Фазовый портрет динамики концентраций продуктов в модели взаимного торможения (5.82). Сплошные линии соответствуют главным изоклинам $P_1 = 0$ и $P_2 = 0$, штриховая линия — сепаратрисе, две жирные точки — устойчивым стационарным состояниям, звездочка — неустойчивому стационарному состоянию

5.5. Стохастика нелинейного самовоспроизводства

Разовьем теперь стохастическую теорию для реакции Шлёгеля



Для анализа химических реакций мы будем следовать методу, изложенному в разд. 5.2. В качестве стохастической переменной мы выбираем, как в разд. 5.2, число частиц сорта X:

$$N(t) = \{0, 1, 2, \dots\}.$$

Речь идет только об одношаговых процессах $N \rightarrow N \pm 1$.

Основное кинетическое уравнение имеет следующий вид:

$$\begin{aligned} \frac{\partial}{\partial t} P(N, t) = & W^+(N-1)P(N-1, t) + \\ & + W^-(N+1)P(N+1, t) - [W^+(N) + W^-(N)]P(N, t). \end{aligned} \quad (5.83)$$

Вероятности перехода для парциальных процессов в четырех каналах реакции мы снова будем считать пропорциональными феноменологическим коэффициентам:

$$\begin{aligned} N \rightarrow N+1: \quad W^+(N) &= V \left[\frac{k_1 A N(N-1)}{V^2} + k'_2 F \right], \\ N \rightarrow N-1: \quad W^-(N) &= V \left[\frac{k'_1 N(N-1)(N-2)}{V^3} + \frac{k_2 N}{V} \right]. \end{aligned} \quad (5.84)$$

Переходя к средним значениям, нетрудно показать, как в разд. 5.2, что стохастические уравнения (5.83) и (5.84) согласуются с уравнением (5.63), если положить $X(t) = \langle N(t) \rangle / V$. Стационарное решение уравнения (5.83) имеет следующий вид:

$$P^0(N) = P^0(0) \prod_{n=1}^N \frac{W^+(n-1)}{W^-(n)}. \quad (5.85)$$

Часто решение (5.85) записывают в экспоненциальной форме:

$$P^0(N) = P^0(0) \exp \left\{ - \sum_{n=1}^N \ln \frac{W^-(n)}{W^+(n-1)} \right\},$$

что позволяет с меньшими трудностями вычислять значения, принимаемые $P^0(N)$ при заданном N , и использовать формулу суммирования Эйлера—Маклорена

$$\begin{aligned} \sum_{n=1}^N \ln \frac{W^-(n)}{W^+(n-1)} &= \int_1^N dn \ln \frac{W^-(n)}{W^+(n-1)} + \\ &+ \frac{1}{2} \left[\ln \frac{W^-(N)}{W^+(N-1)} + \ln \frac{W^-(1)}{W^+(0)} \right] + \dots \end{aligned}$$

Вводя сокращенные обозначения

$$a = \frac{k_1 A}{k'_1}, \quad b = \frac{k_2}{k'_2}, \quad c = \frac{k'_2 F}{k_1 A V},$$

мы получаем приближение Эйлера—Маклорена к стационарному распределению (Ebeling, Schimansky-Geier, 1979):

$$\begin{aligned} P_{\text{ЭМ}}^0 &= P^0(0) \frac{3Vab}{2cN} \left\{ \frac{Va[(N-1)(N-2) + cV^2]}{N[(N-1)(N-2) + bV^2]} \right\}^{N-1} \times \\ &\times \exp \left\{ N - 2V \left[\sqrt{b} \operatorname{arctg} \left(\frac{N-1}{\sqrt{b}V} \right) - \sqrt{c} \operatorname{arctg} \left(\frac{N-1}{\sqrt{c}V} \right) \right] \right\} \end{aligned} \quad (5.86)$$

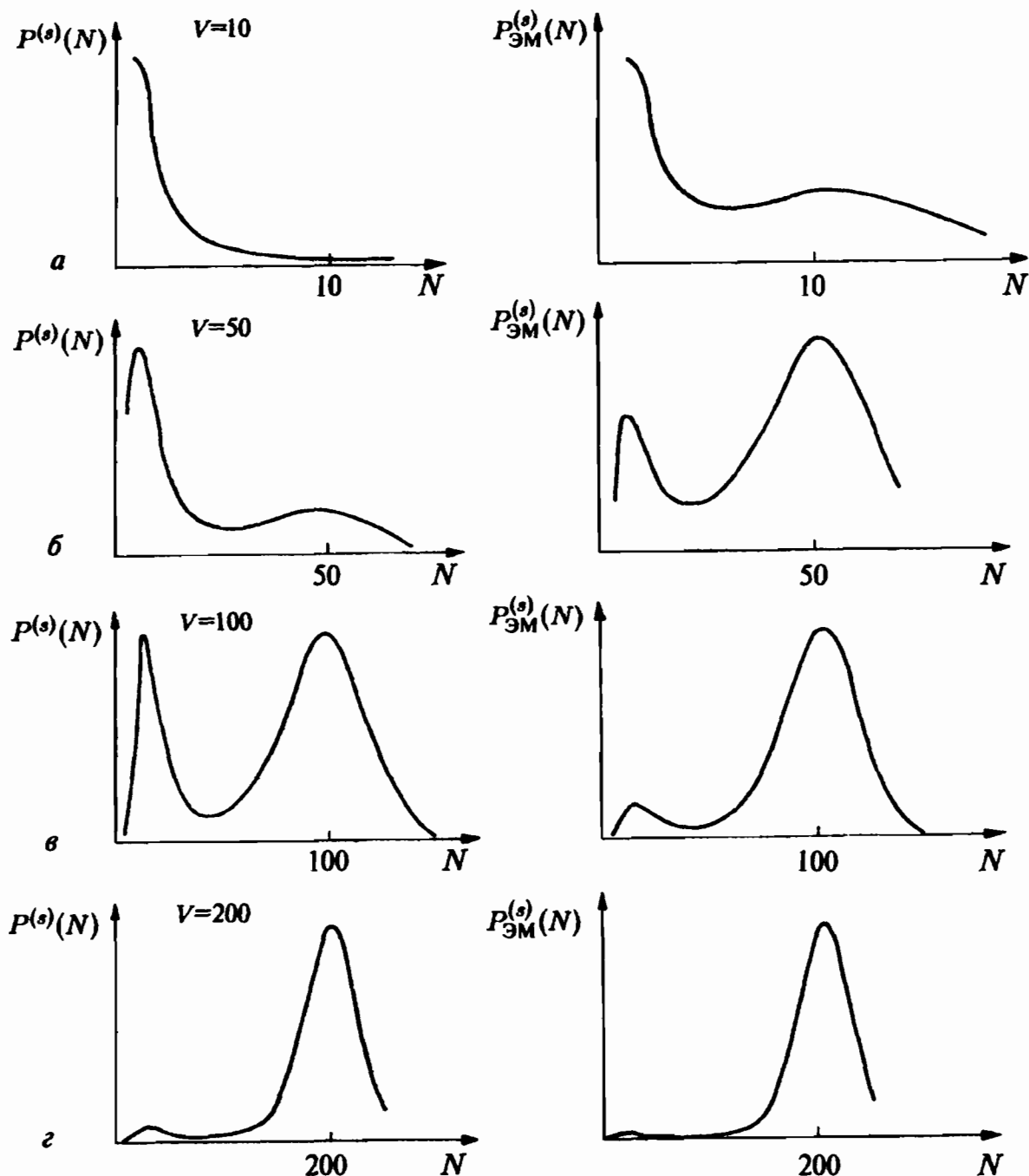


Рис. 5.31. Стационарные распределения вероятностей в стохастической модели Шлёгеля для различных объемов ($V = 10; 50; 100; 200$) и сравнение точного решения с приближением Эйлера–Маклорена

В термодинамическом пределе $V \rightarrow \infty$ приближенное решение (5.86) переходит в точное асимптотическое решение Николиса и Тернера. Формула (5.85) позволяет вычислять значение $P^0(N)$ при произвольном V . Вычисления легко производятся с помощью программируемого микрокалькулятора. На рис. 5.31 представлены различные стационарные распределения при значениях параметров $a = 1,55$; $b = 0,595$; $c = 0,03903$ и объемах $V = 10; 50; 100; 500$, полученные численными методами Эбелингом и Шиманским-Гайером (Ebeling, Schimansky-Geier, 1979). Детерминистическая теория предсказывает при этих значениях параметров бистабильный режим с устойчивыми состояниями $X_1 = 0,17$; $X_3 = 1,03$ и неустойчивым состоянием $X_2 = 0,36$. Стохастическая теория предсказывает при $V \geq 50$ существование бимо-

дальнего распределения с двумя максимумами и одним минимумом примерно там, где находятся детерминистически предсказанные стационарные состояния. Устойчивым состояниям соответствуют максимумы, неустойчивому состоянию — минимум. Форма и ширина максимумов весьма чувствительно зависят от объема; при $V > 100$ доминирует правый максимум, при $V < 100$ — левый, при $V = 100$ оба максимума имеют примерно равную высоту. Различие в высоте и ширине отражает различную устойчивость двух состояний. При $V \lesssim 25$ свойство бимодальности исчезает, так как правый максимум полностью утрачивается. Для сравнения на рис. 5.31 приведены решения в приближении Эйлера—Маклорена. Как видно из сравнения, приближение дает удовлетворительные результаты только при $V \geq 200$, а при $V \leq 100$ искажает даже качественную картину.

Рассмотрим теперь бифуркационное поведение решения. Назовем множеством катастроф K' те точки пространства параметров C , в которых претерпевают изменение качественные свойства распределения. В нашем случае речь идет о переходе от унимодального распределения (с 1 максимумом) к бимодальному распределению (с 2 максимумами). Для непрерывных распределений множество катастроф определяется одновременным обращением в нуль первой и второй производных:

$$\frac{dP^0(x, u)}{dx} = 0, \quad \frac{d^2P^0(x, u)}{dx^2} = 0. \quad (5.87)$$

В случае дискретных распределений того типа, который мы здесь рассматриваем, условие (5.87) заменяется условием

$$P^0(N, u) = P^0(N - 1, u). \quad (5.88)$$

Введем безразмерные переменные

$$v = \frac{k_1 A}{k'_1} V, \quad \beta = \frac{k_2 k'_1}{k_1^2 A^2}, \quad \gamma = \frac{k'_2 (k'_1)^2 F}{k_1^3 A^3}$$

Детерминистическое множество бифуркаций имеет вид острия (точки возврата) с двумя ветвями:

$$\gamma = \frac{\beta}{3} - \frac{2}{27} \pm \sqrt{\left(\frac{1}{9} - \frac{\beta}{3}\right)^2} \quad (5.89)$$

и вершиной в точке $\beta = 1/3$, $\gamma = 1/27$.

Множество катастроф стохастического распределения переходит при $v \rightarrow \infty$ в детерминистическое острие.

На рис. 5.32 помимо этой кривой показаны также множества катастроф при $v = 100$ и $v = 10$. Бимодальные распределения соответствуют значениям параметров внутри острия, а мономодальные распределения — значениям параметров, лежащих вне острия. С уменьшением объема V влияние стохастических эффектов возрастает. Поразительно, что флуктуации оказывают весьма сильное влияние на область бистабильности.

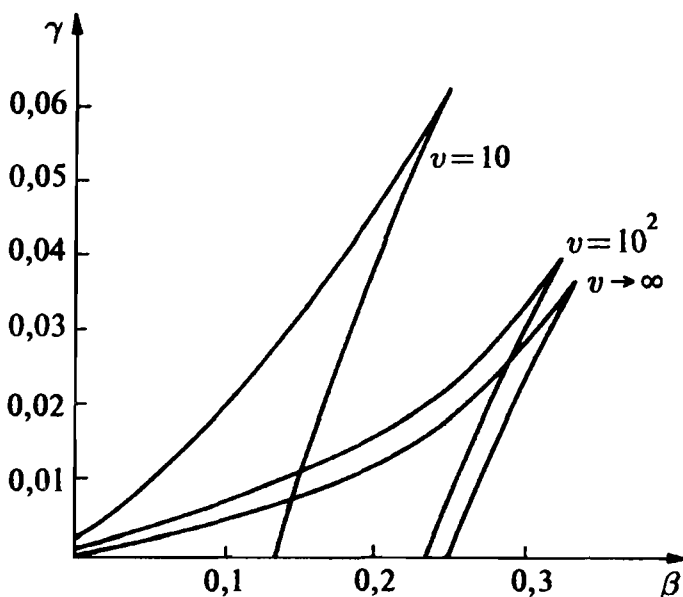


Рис. 5.32. Изменение бифуркационной картины в зависимости от величины системы

Примером могут служить хотя бы кривые, представленные на рис. 5.20: набор параметров $\beta = 0,2477$ и $\gamma = 0,01624$, соответствующих заданным значениям a , b и c , определяет точку, лежащую внутри детерминистического острия. При уменьшении V острие смещается влево, и наступает момент, когда представляющая точка оказывается снаружи острия; это соответствует исчезновению правого максимума на рис. 5.31 при $V \lesssim 25$. Другим примером может служить система с параметрами $\beta = 0,13$ и $\gamma = 0,02$, которая в рамках детерминистической теории моностабильна. При уменьшении объема до $V = 10$ под действием начинающихся сильных флуктуаций возникает бистабильность. Таким образом, основной результат этих исследований состоит в том, что флуктуации числа частиц могут приводить даже к качественным изменениям макроскопического поведения. Бистабильность может быть и порождена, и уничтожена флуктуациями.

Рассмотрим теперь релаксацию распределения вероятности к стационарному равновесию и воспользуемся для этого методами стохастической динамики и эргодической теоремой (Ebeling, Schimansky-Geier, 1979). Мы предполагаем, что эргодическая теорема выполняется в данном случае, поскольку все состояния $N(t) = \{0, 1, 2, \dots\}$ связаны между собой возможными переходами в обоих направлениях. В случае эргодической системы исследование достаточно длинных траекторий $N(t)$ с произвольным начальным состоянием $N(0)$ дает полную информацию относительно вероятности $P^0(N)$, возникающей как предельное значение относительной частоты попадания в точку N . Стохастическое моделирование траекторий помимо распределения вероятностей позволяет получить и другую важную информацию относительно стохастического процесса, например, корреляционную функцию. Методы стохастической динамики были применены Файстелем (Feistel, 1977, 1979),

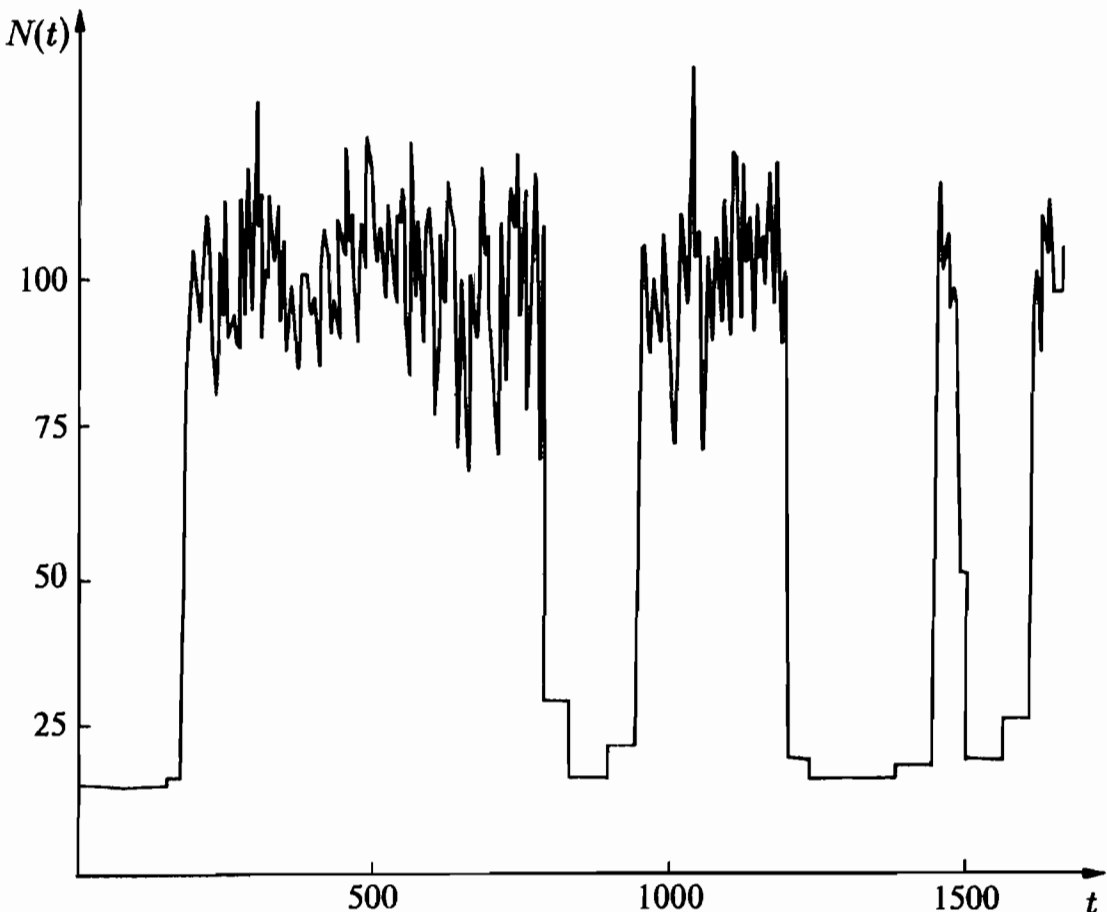


Рис. 5.33. Стохастическая траектория в модели Шлёлля

Таблица 5.1. Параметры стохастического моделирования реакций Шлёгля

	Система I	Система II
V	100	100
a	1,550	0,34114
b	0,595	0,02197
c	0,03903	0,001
Устойчивое состояние	$N_1 = 13$; $N_3 = 99$	$N_1 = 1$; $N_3 = 24$
Неустойчивое состояние	$N_2 = 44$	$N_2 = 10$
Начальное состояние	а) $N(0) = 5$ б) $N(0) = 110$	а) $N(0) = 0$ б) $N(0) = 30$
Среднее время первой релаксации	а) 3,21 б) 2,52	а) 36,2 б) 13,3
Среднее время перехода в неустойчивое состояние	а) 83,1 б) 166	а) 1 350 б) 600
Среднее время туннелирования в другое устойчивое состояние	а) 166 б) 433	а) 2 270 б) 2 030

Файстелем и Эбелингом (*Feistel, Ebeling, 1978*), а также Николисом и Малек-Мансуром (*Nicolis, Malek-Mansour, 1978*) к колебательным химическим реакциям, а Эбелингом и Шиманским-Гайером (*Ebeling, Schimansky-Geier, 1979*) — к бистабильным химическим реакциям. Мы следуем здесь последней из упомянутых работ. На рис. 5.33, 5.34 и 5.35 показаны три примера индивидуальных траекторий системы Шлёгля. Чтобы не перегружать графики, мы наносили на них не результаты расчетов для отдельной реализации, а средние по нескольким сотням реализаций реакции. Траектории флуктуируют относительно одного устойчивого состояния и за большие временные интервалы выходят на другое устойчивое состояние. В табл. 5.1 приведены некоторые характеристики двух рассмотренных систем, а именно параметры V , a , b , c устойчивого и неустойчивого состояний, начальное состояние и среднее время релаксации до достижения соседнего устойчивого состояния, среднее время до первого выхода на неустойчивое состояние и среднее время туннелирования в другое устойчивое состояние. Система I выбрана с таким расчетом, что оба максимума вероятности $P^0(N)$ имеют одинаковую высоту, а в системе II оба максимума имеют одинаковую площадь. Следует отметить, что средние времена туннелирования зависят от направления и совпадают только для максимумов с одинаковой площадью.

Мы установили также, что среднее время туннелирования из одного максимума в другой существенно (примерно в 2 раза) больше среднего времени достижения минимума. Это связано с тем, что большое число траекторий по достижении минимума возвращается.

Если известна средняя частота, с которой система оказывается в отдельных состояниях $N = \{0, 1, 2, \dots\}$, то можно вычислить распределение вероятностей. Выберем параметры системы I из табл. 5.1 и будем исходить из начальных условий

$$N(0) = 97 \quad \text{и} \quad N(0) = 15.$$

При $t = 0$ распределение получается δ -образным:

$$P(N, 0) = \delta_{N,97} \quad \text{и} \quad P(N, 0) = \delta_{N,15}.$$

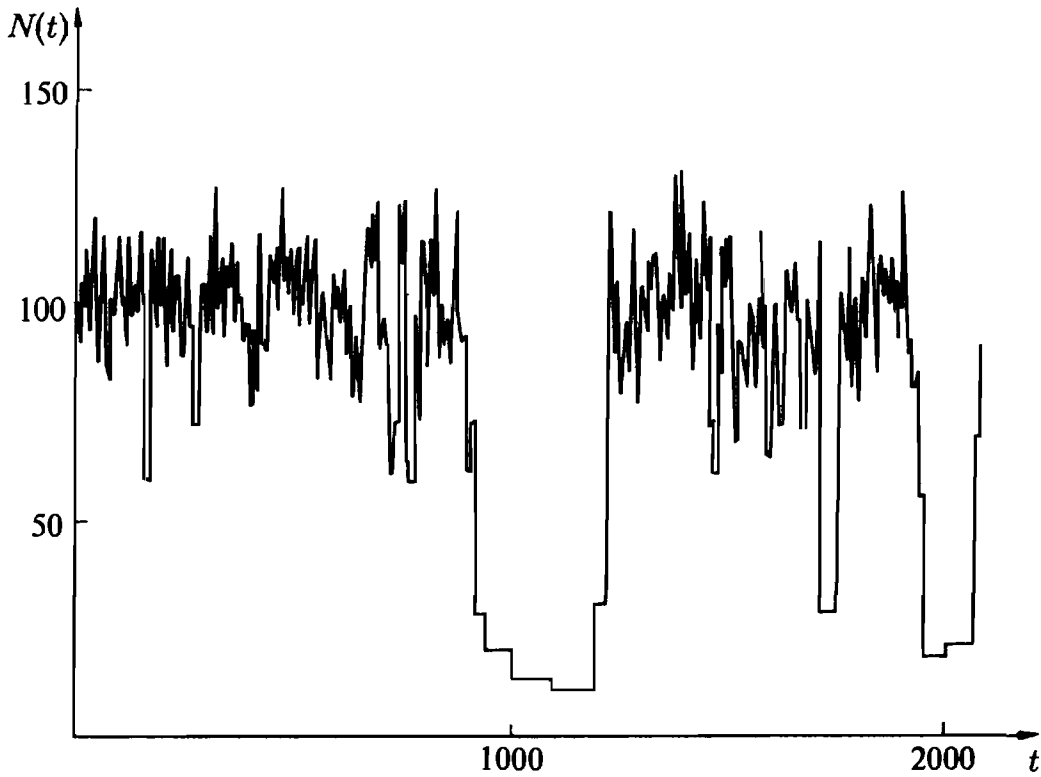


Рис. 5.34. Стохастическая траектория в модели Шлёгля

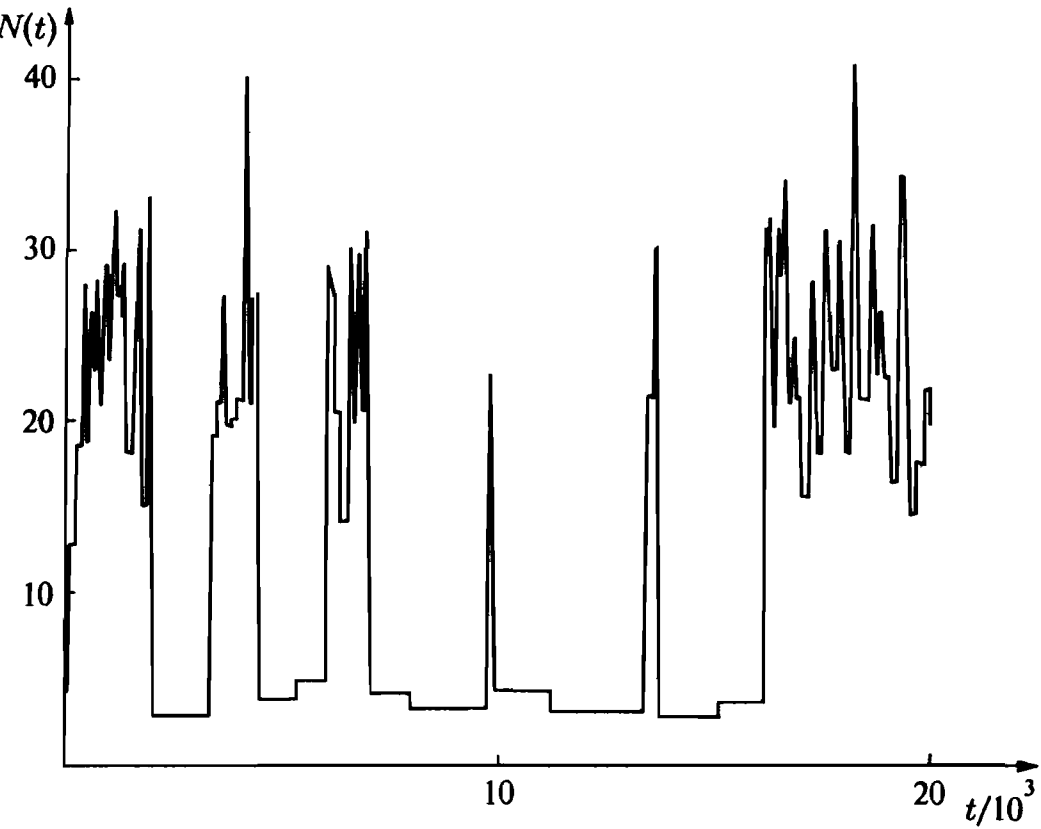


Рис. 5.35. Стохастическая траектория в модели Шлёгля

Результаты численных расчетов распределений вероятности при трех конечных временах $t > 0$ представлены на рис. 5.36. Нетрудно видеть, что временная эволюция распределения складывается из двух периодов. На первом относительно коротком периоде вокруг ближайшего устойчивого состояния формируется распределение

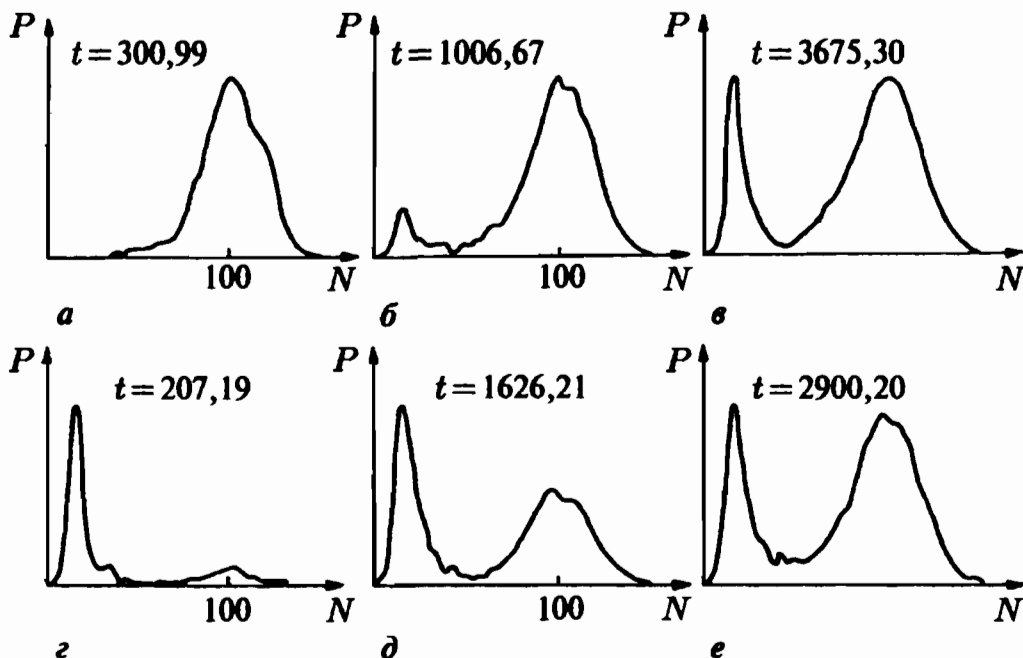


Рис. 5.36. Временная эволюция распределения вероятности, вычисленная с помощью стохастической динамики

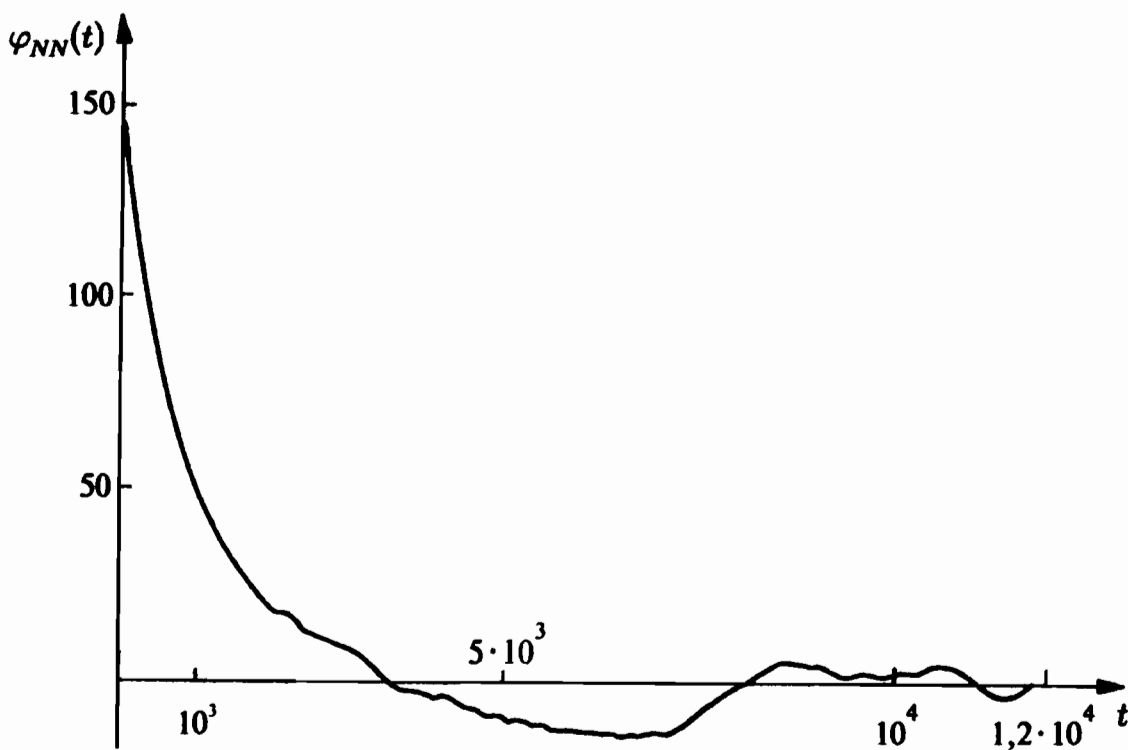


Рис. 5.37. Автокорреляция в стохастической модели Шлёгеля

Гаусса. На втором более продолжительном периоде образуется второй максимум. На больших временах распределение сходится к бимодальной кривой, которая находится в хорошем согласии с аналитически вычисленными распределениями (рис. 5.31).

В качестве примера разносторонности методов стохастической динамики приведем вычисления автокорреляционных функций (Ebeling, Schimansky-Geier, 1979):

$$\varphi_{NN}(t) = \langle N(t)N(0) \rangle - \langle N(t) \rangle^2. \quad (5.90)$$

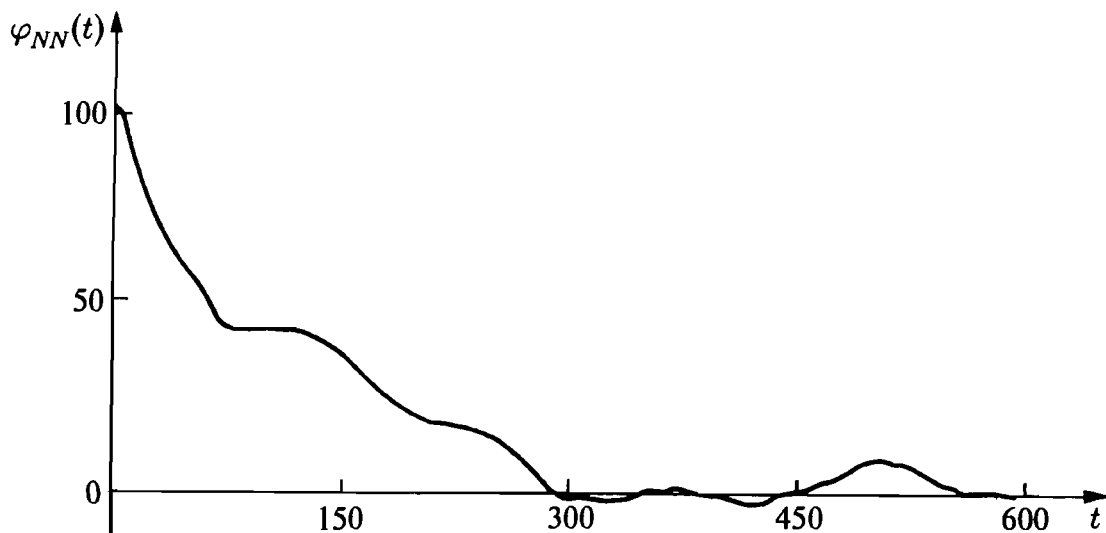


Рис. 5.38. Автокорреляция при малых временах корреляции

Прежде всего разрежем на части длинную траекторию и построим ансамбль траекторий с различными начальными условиями $N(0)$. Вычислим для каждого члена этого ансамбля величину $N(t)N(0)$ и усредним полученные результаты по ансамблю. Мы вычислили автокорреляционную функцию для системы II в соответствии с табл. 5.1 для двух различных масштабов времени. На рис. 5.37 показана автокорреляционная функция φ_{NN} для временной шкалы $t \sim 10^4$. Характерными флуктуациями являются переходы между двумя устойчивыми состояниями, а характерное время соответствует среднему времени туннелирования. На рис. 5.38 представлена автокорреляционная функция для временной шкалы $t \approx 500$. Эта корреляционная функция описывает флуктуации относительно устойчивого состояния и, следовательно, характеризует время первой релаксации.

5.6. Самовоспроизводство в сложных системах

При рассмотрении 12 эпох эволюции, которые мы бегло перечислили в гл. 2, становится отчетливо ощутимым особый характер фактора самовоспроизводства. Самовоспроизводство вступает в игру лишь в последнюю, 12-ю, эпоху эволюции и приводит к возникновению жизни и всех связанных с нею структур. Нет никаких сомнений в том, что эволюция всех высших ступеней связана с процессом самовоспроизводства (*Эбелинг и др.*, 1986). Под самовоспроизводством мы понимаем здесь в весьма общем смысле способность системы при определенных краевых условиях (например, притоке свободной энергии и необходимых исходных веществ, наличии соответствующего «аппарата») репродуцировать себя как органическое целое. Копии исходной системы могут быть как тождественными ей, так и модифицированными. Самовоспроизведение при таком подходе выступает как специальная, несколько изощренная форма самоорганизации. Формы самовоспроизведения весьма разнообразны, они охватывают и уже обсуждавшиеся нами автокаталитические химические процессы, и умножение числа фотонов в лазере, и размножение живых организмов, и репродукцию научных результатов. Использование термина «самовоспроизводство» применительно к научным процессам необычно, вместо этого, следуя Марксу, принято говорить о процессах воспроизводства. Мы не считаем разумным проведение резкой границы между «самовоспроизводством в узком смысле», без помощи, и «воспроизводством» с вспомоществованием. «Чистое самовоспроизведение» существует только в специальных физических и химических процессах

(например, в автокатализе и в лазерном эффекте). Даже «самовоспроизведение» полинуклеотидов в клетке требует чрезвычайно сложного аппарата, мало чем уступающего заводу по «воспроизводству» автомобилей.

Научные основы исследования такого рода процессов в мире живого и в человеческом обществе заложили в прошлом веке почти одновременно Дарвин и Маркс. Правда, многие детали удалось уточнить и выяснить лишь в нашем столетии. Показательна история возникновения понятия «ген» (*Sander, 1988*). Генами мы называем репродуктивные единицы биологической эволюции. Еще в 1909 г. датский исследователь Иоганнсен ввел понятие «ген» для обозначения «счетного устройства», принимающего участие в процессах наследственности. В 1922 г. Мёллер назвал в качестве важнейшей характеристики гена его способность к саморепликации. Ближайшее открытие, имевшее решающее значение, было совершено в 1944 г. Эвери и его сотрудниками, обнаружившими, что носителем наследственного материала является дезоксирибонуклеиновая кислота (ДНК) (*Judson, 1979*). Следующим решающим шагом стало объяснение структуры и механизма самовоспроизведения ДНК Уотсоном и Криком. Эти исследователи охарактеризовали молекулы ДНК в клетке как «молекулярную основу матриц, необходимых для генетической репликации». Не имея возможности вдаваться в дальнейшие детали роли самовоспроизводства в биологической эволюции, мы вынуждены отослать интересующихся читателей к обширной литературе (*Тимофеев-Рессовский и др., 1969; Корнберг, 1977; Hofmann, 1979, 1980; Mayr et al., 1978; Хессин, 1985; Greenwood et al., 1985*).

Существенно более сложными по сравнению с процессами самовоспроизводства в биологических системах являются процессы самовоспроизводства в технических и экономических системах. В некоторых случаях заключения о поведении таких систем могут быть получены по аналогии на основании исследования моделей более простых систем. Такие заключения по аналогии надлежит считать вполне законным методологическим вспомогательным средством исследования (*Hörz, 1974*). Естественно, при этом не следует забывать о том, что простая модель сложной системы всегда намного беднее, чем реальность, и в лучшем случае правильно отражает лишь некоторые ее аспекты.

Проиллюстрируем эти общие соображения на конкретном примере самовоспроизведения технологий. Мы будем следовать работам Фишера и Прая (*Fisher, Pry, 1970, 1971*), Мэнсфилда (*Mansfield, 1971*), а также Хименеса-Монтаньо и Эбелинга (*Jiménez-Montaño, Ebeling, 1980*). Слово «технология» употребляется здесь в весьма широком смысле. Процесс технологической эволюции протекает так, что устаревшие технологии заменяются новыми, лучшими технологиями. Опираясь на экспериментальные исследования внедрения новых технологий, Фишер и Прай разработали сильно упрощенную модель, дающую необычайно хорошее описание имеющихся данных. Они исходили из упрощающего предположения о том, что человеческое общество обладает рядом элементарных потребностей (в питании, одежде, крове, транспорте, коммуникации и образовании), которые требуется удовлетворить. Технологическая эволюция (если рассматривать ее в сильно упрощенном виде) состоит в создании новых форм удовлетворения этих элементарных потребностей. Решающее значение имеет (по Фишеру и Праю) процесс замещения, или вытеснения, старой технологии новой технологией. Обозначим через f долю новой технологии. Таким образом, $f = 1$ означает полное замещение, а $f = 0,5$ — пятидесятипроцентное замещение. Основное допущение, обоснованное эмпирическими данными (*Fisher, Pry, 1970*), состоит в том, что скорость изменения доли f пропорциональна как доле нового f , так и доле старого $(1 - f)$, т. е.

$$\frac{df}{dt} = \alpha f(1 - f) + \beta. \quad (5.91)$$

Здесь β — относительно малая частота спонтанного возникновения новых технологий в результате научных исследований и конструкторских разработок, α — константа скорости. Входящая в правую часть уравнения (5.91) пропорциональность доле старой технологии $(1 - f)$ означает, что обновление технологии предполагает наличие определенной доли старой, способной к обновлению технологии. Пропорциональность доле f мы интерпретируем как эффект имитации: чем больше доля уже имеющейся новой технологии, тем быстрее происходит замещение старой технологии. Имитация — важная элементарная стадия в общественных процессах: воспроизводятся и повторяются успешные или широко распространенные действия. Подходящим примером имитационного поведения может служить распространение моды на ту или иную одежду. При замещении технологии простое подражание уже введенной и успешной технологии также играет важную роль. Ряд других социологических и экологических примеров процесса имитации исследован в работе (Callen, Shapero, 1974).

Но вернемся к уравнению (5.91). Сравнение его с уравнением (5.17) обнаруживает полную аналогию с дифференциальным уравнением автокаталитического процесса. Поэтому мы можем воспользоваться решением (5.20) и перенести его на интересующий нас сейчас случай:

$$f(t) = \frac{1}{2} + \left[\frac{\beta}{\alpha} + \frac{1}{4} \right]^{1/2} \cdot \frac{1 - C \exp \{-t/\tau\}}{1 + C \exp \{-t/\tau\}},$$

$$\tau = \frac{1}{\alpha} \left[1 + 4 \frac{\beta}{\alpha} \right]^{-1/2}.$$
(5.92)

Выясним теперь, что дает стохастический подход к проблеме (Jiménez-Montaño, Ebeling, 1980). Рассмотрим N фабрик, удовлетворяющих потребность в определенных товарах. Понятие «фабрика» здесь трактуется в смысле наименьшей производственной единицы, способной изготавливать эти товары. Предположим, что N_1 фабрик используют старую технологию, а N_2 фабрик — новую. Общее число фабрик $N = N_1 + N_2$ в рассматриваемый период времени можно считать постоянным. Пусть E_1 — скорость роста числа «старых» фабрик, а E_2 — скорость роста числа «новых» фабрик. Если новые технологии обеспечивают более высокую производительность, то $E_2 > E_1$. Элементарным стохастическим процессом является замещение старой технологии новой технологией на одной отдельной фабрике. С математической точки зрения речь идет о переходе

$$N_1 \rightarrow N_1 - 1, \quad N_2 \rightarrow N_2 + 1.$$

Для соответствующей вероятности перехода мы полагаем

$$W^+(N_2) = E_2 N_1 \frac{N_2}{N} + E_{21}.$$
(5.93)

Иначе говоря, мы предполагаем, что вероятность перехода пропорциональна числу фабрик, использующих старую технологию, и доле фабрик, использующих новую технологию (тенденция к имитации). Пусть далее E_{21} — скорость спонтанного перехода из 1 в 2 (от старой технологии к новой). Для противоположного процесса, также возможного, но, разумеется, менее вероятного, мы полагаем

$$W^-(N_2) = E_1 N_2 \frac{N_1}{N} + E_{12}.$$
(5.94)

Определим теперь $P(N_2, t)$ как вероятность того, что N_2 фабрик используют новую технологию, а N_1 фабрик — старую. Следуя стандартным методам теории дискретных марковских процессов, которые были изложены в разд. 5.2, мы приходим

к основному кинетическому уравнению:

$$\begin{aligned} \frac{\partial}{\partial t} P(N_2, t) = & \left[E_{21} + \frac{E_2}{N} (N_2 - 1)(N - N_2 + 1) \right] P(N_2 - 1, t) + \\ & + \left[E_{12} + \frac{E_1}{N} (N_2 + 1)(N - N_2 - 1) \right] P(N_2 + 1, t) - \\ & - \left[E_{12} + E_{21} + \frac{1}{N} (E_1 + E_2) N_2 (N - N_2) \right] P(N_2, t). \end{aligned} \quad (5.95)$$

Четыре члена в правой части уравнения (5.95) соответствуют двум «прибыльным процессам» $(N_2 - 1) \rightarrow N_2$ и $(N_2 + 1) \rightarrow N_2$ и двум «убыточным процессам» $N_2 \rightarrow (N_2 - 1)$ и $N_2 \rightarrow (N_2 + 1)$. Определим среднее значение

$$\langle N_2(t) \rangle = \sum_{N_2=0}^{\infty} N_2 P(N_2, t). \quad (5.96)$$

Умножая уравнение (5.95) на N_2 и суммируя, получаем

$$\frac{d}{dt} \langle N_2(t) \rangle = \frac{1}{N} (E_2 - E_1) \langle N_2(N - N_2) \rangle + (E_{21} - E_{12}). \quad (5.97)$$

Если воспользоваться приближением $\langle N_2^2 \rangle \approx \langle N_2 \rangle^2$ и ввести обозначения

$$f = \frac{\langle N_2 \rangle}{N}, \quad \alpha = (E_2 - E_1), \quad \beta = \frac{E_{21} - E_{12}}{N}, \quad (5.98)$$

то мы снова получим исходное детерминистическое уравнение

$$\dot{f} = \alpha f(1 - f) + \beta. \quad (5.99)$$

Стационарное решение основного кинетического уравнения (5.95) можно привести в явном виде. Запишем сначала

$$W^-(N_2) P^0(N_2) = W^+(N_2 - 1) P^0(N_2 - 1). \quad (5.100)$$

Интегрируя, получаем:

$$P^0(N_2) = \frac{W^+(N_2) W^+(N_2 + 1) \dots W^+(N - 1)}{W^-(N_2 + 1) W^-(N_2 + 2) \dots W^-(N)} P^0(N). \quad (5.101)$$

Если в детерминистической картине через определенное время всегда устанавливается состояние $f = 1$, т. е. $N_2 = N$, то в стохастической картине мы приходим к распределению вероятности с $P^0(N_2) \neq 0$ при $N_2 \neq N$, т. е. между старой и новой технологией (взятых в определенном соотношении) устанавливается сосуществование при условии, что величина $(E_2 - E_1)$ принимает не слишком большое значение.

Интересный предельный случай возникает в том случае, если вероятности спонтанных переходов от старой технологии к новой и от новой технологии к старой равны нулю, т. е.

$$E_{21} = E_{12} = 0. \quad (5.102)$$

В этом случае состояния $N_2 = 0$ и $N_2 = N$ являются поглощающими состояниями, которые система не может миновать, и все распределения

$$P^0(N_2) = \sigma_1 \delta_{0N_2} + \sigma_2 \delta_{NN_2}, \quad \sigma_1 + \sigma_2 = 1, \quad (5.103)$$

— стационарные решения уравнения (5.95), где σ_i — произвольное действительное число, заключенное в интервале от нуля до единицы. Таким образом, существует

не однозначное стационарное распределение, а множество стационарных распределений. Утрата однозначности связана с тем, что поглощательные состояния $N_2 = 0$ и $N_2 = N$ становятся недостижимыми из остальных состояний. Какие стационарные режимы системы реализуются, зависит от начальных условий. Предположим, что начальное распределение имеет вид:

$$P(N_2, 0) = \delta_{N_2 N_2^0}, \quad (5.104)$$

и пусть

$$P(N_2, t \rightarrow \infty) = \sigma_1 \delta_{0N_2} + \sigma_2 \delta_{NN_2}, \quad \sigma_1 + \sigma_2 = 1, \quad (5.105)$$

— искомое распределение. Найдем пока не известное действительное число σ_2 . Для этого воспользуемся тем, что величина

$$I(t) = \left\langle \left(\frac{E_1}{E_2} \right)^{N_2} \right\rangle = \sum_{N_2=0}^{\infty} \left(\frac{E_1}{E_2} \right)^{N_2} P(N_2, t) \quad (5.106)$$

при $E_{12} = E_{21} = 0$ является инвариантом уравнения (5.95), т. е.

$$I(t) = \text{const}, \quad (5.107)$$

$$I(0) = I(\infty). \quad (5.108)$$

Доказать соотношение (5.107) совсем нетрудно: для этого нужно лишь умножить уравнение (5.95) на $(E_1/E_2)^{N_2}$ и просуммировать по всем N_2 . Из равенства (5.108) с учетом соотношений (5.104)–(5.106) получаем

$$\begin{aligned} \left(\frac{E_1}{E_2} \right)^{N_2^0} &= \sigma_1 + \sigma_2 \left(\frac{E_1}{E_2} \right)^N = 1 - \sigma_2 \left[1 - \left(\frac{E_1}{E_2} \right)^N \right], \\ \sigma_2 &= \frac{1 - (E_1/E_2)^{N_2^0}}{1 - (E_1/E_2)^N}. \end{aligned} \quad (5.109)$$

В предельном случае при очень большом числе N ($N \gg 1$) и $N_2^0 \ll N$ находим:

$$\sigma_2 = \begin{cases} 0 & \text{при } E_2 < E_1, \\ 1 - \left(\frac{E_1}{E_2} \right) & \text{при } E_2 > E_1. \end{cases} \quad (5.110)$$

Эта формула соответствует соотношению (5.49), выведенному другим способом. Величина σ_2 есть вероятность того, что система с начальным состоянием $N_2 = N_2^0$ переходит при $t \rightarrow \infty$ в состояние $N_2 = N$, а $\sigma_1 = 1 - \sigma_2$ — вероятность того, что система переходит в состояние $N_2 = 0$. В рамках проблемы введения новой технологии формулы (5.109) и (5.110) допускают следующую интерпретацию.

Если новая технология первоначально вводится на очень небольшом числе фабрик N_2^0 , то, как следует из формулы (5.110), она заведомо обречена на замещение старой технологией при условии, что скорость роста E_2 новой технологии меньше скорости роста E_1 старой технологии. Если же скорость роста E_2 новой технологии существенно превосходит скорость роста E_1 старой технологии, то, как следует из той же формулы (5.110), новая технология с большей вероятностью

$$1 - \left(\frac{E_1}{E_2} \right)^{N_2^0} \quad (5.111)$$

замещает старую технологию и с меньшей вероятностью

$$\left(\frac{E_1}{E_2}\right)^{N_2^0} \quad (5.112)$$

уступает место старой технологии. Четкий «выбор» в пользу новой технологии возникает только при условии, если

$$\begin{aligned} \left(\frac{E_1}{E_2}\right)^{N_2^0} &< \frac{1}{2}, \\ N_2^0 \log \left(\frac{E_2}{E_1}\right) &> \log 2. \end{aligned} \quad (5.113)$$

Новая, более прогрессивная технология реально замещает старую технологию, только если новая технология существенно превосходит старую, например, при $N_2^0 = 1$ примерно вдвое, или если число N_2^0 достаточно велико. Если это условие не выполнено, то случайные флуктуации мешают замещению старой технологии новой, лучшей, технологией. Таким образом, стохастическая теория дает существенно более тонкую картину, чем детерминистическая теория. В то время как в детерминистической картине при $E_2 > E_1$ новая технология безусловно замещает старую, то в стохастической картине замещение старой технологии новой технологией происходит лишь с определенной вероятностью; при выполнении условия (5.113) эта вероятность доходит почти до уровня достоверности.

Мы изложили модель воспроизведения технологии несколько подробнее потому, что квадратичные (логистические) законы воспроизводства и роста весьма часто встречаются в общественных науках. Приведенная выше схема позволяет без особого труда построить для таких процессов стохастическую теорию, учитывающую и влияние флуктуаций.

Интересным и спорным остается вопрос о том, может ли нематериальный объект служить примером самовоспроизведения. Доукинс (*Dawkins, 1978*) высказал гипотезу о том, что и в культурной эволюции существуют элементы самовоспроизведения. Доукинс называет эти элементы мемами. В качестве примеров мемов он приводит мелодии, мысли, лозунги, модные силуэты и навыки. Подобно тому, как гены воспроизводятся в генофонде и через спермии или яйца переносятся от тела к телу, мемы, по Доукинсу, воспроизводятся в своем фонде и перескакивают из головы в голову. При распространении мемов большую роль играет имитация и взаимное обучение. По мнению авторов этой книги, гипотезы Доукинса содержат рациональное зерно, поскольку они опираются на наблюдения. Каждый знает из повседневного опыта, что слухи, анекдоты, страхи и т. д. с формальной точки зрения множатся и распространяются так же, как вирусы и бактерии во время эпидемии. В рамках такого сравнения аналогия вполне обоснована. Впрочем, отнюдь не обязательно заходить так далеко, как Доукинс, и даже живые существа рассматривать как машины для воспроизводства генов или мемов. Правда, здесь возникает диалектическое взаимодействие между двумя противоположностями — между самовоспроизводящейся информацией (хранящейся в гене или в меме) и системой, к которой эта информация относится.

В монографии Ламсдена и Уилсона (*Lumsden, Wilson, 1981*) так называемый культур-ген вводится как самовоспроизводящаяся единица культурных процессов. Ратнер и Шанин (*Ratner, Shanin, 1983*) предложили называть единицы самовоспроизведения сисерами (системами самовоспроизведения) (*Ратнер, 1986*).

Независимо от того, привыкнут ли эти термины в научной литературе, ныне твердо установлено, что процессы самовоспроизведения играют центральную роль

во всех высших эволюционных процессах. Нет никаких сомнений в том, что фундаментальные исследования этих процессов на основе простых базовых моделей, введенных фон Нейманом и усовершенствованных Эйгеном, Шустер и Ратнером, а также другими авторами, составляют ядро теории эволюции.

Продолжительность процессов самовоспроизведения до сих пор никак не учитывалась. Совершенно ясно, что изготовление копии заданного оригинала (матрицы, архетипа, эталона, конструктивного образца и т. д.) требует времени, причем тем большего, чем сложнее оригинал (*Эбелинг и др.*, 1986). Простая молекула может быть скопирована с помощью автокатализа за миллисекунды, отливка литеры с помощью формы требует секунд, а изготовление копии картины старого мастера — по крайней мере одной недели. Бабочки-однодневки завершают свой жизненный цикл за один день, человеку для той же цели требуется более полувека. У всех социальных существ «самовоспроизводство» включает в себя не только биологический процесс продолжения рода, но и более продолжительный процесс роста и созревания. Чем сложнее система, тем в меньшей степени она может быть воспроизведена с помощью однократного акта. Этой проблеме мы посвящаем особую главу (гл. 7), так как анализ ее требует довольно сложного математического аппарата.

Глава 6

Процессы конкуренции и отбора

Бежать быстрее всех
Еще не значит всех опередить.

Луи Арагон

6.1. Эволюция как проблема оптимизации

Конкуренция и отбор — основные понятия самоорганизации и эволюции в сложных системах, состоящих из многих подсистем. Конкуренция в таких системах означает состязание подсистем, качественно одинаковых, но обладающих различными количественными свойствами и пытающихся достичь общей цели, достижимой по правилам игры всей системы лишь для одной или нескольких подсистем. С математической точки зрения такая целеустановка, как правило, означает оптимизацию параметров, функций или функционалов. Понятие конкуренции используется во многих научных дисциплинах, в частности, оно принадлежит к числу основных понятий политической экономии. В экологии под конкуренцией принято понимать совместное использование жизненно необходимого фактора двумя или более организмами.

Конкуренция, как правило, связана с селекцией, т. е. отбором «лучшей» из конкурирующих подсистем в смысле заданного критерия цели. Понятие естественного отбора было впервые сформулировано в 1859 г. Чарлзом Дарвином, который понимал естественный отбор как выживание наиболее приспособленного — «*survival of the fittest*». Однако понятие «наиболее приспособленного», его математическое определение и возможность количественного измерения на практике и поныне до конца не выяснены и остаются предметом споров и обсуждений, в особенности среди биологов, популяционных и молекулярных генетиков. Упомянем лишь родившийся в этих спорах термин «недарвиновская эволюция» (*Kimura, 1968; King, Jukes, 1969*).

В литературе встречаются различные интерпретации понятия отбора. Специалисты по популяционной биологии понимают под отбором «различные изменения относительных частот генотипов на основе различных способностей их фенотипов к передаче в ближайшем поколении» или «процесс, обуславливающий вероятность того, что определенные индивиды достигают возраста, в котором они обретают способность к размножению» (*Тимофеев-Рессовский, Воронцов, Яблоков, 1969*). Под генотипом живого организма мы понимаем совокупность наследственных признаков, хранящихся в ядре клетки, а под фенотипом — совокупность внешних признаков живого организма. Различные авторы (например, Холдейн, Райти, Шмальгаузен, Тимофеев-Рессовский и др.) по-разному классифицируют процессы отбора, выделяя, например, отбор по выживанию или по размножению, отбор динамический, подвижный, направленный, стабилизирующий, отбор внутривидовой и межвидовой, отбор между территориями и т. д. Все эти определения сводят процесс отбора к объектам, обладающим генотипами, фенотипами или способностью к размножению, и исключают отбор в чисто физических процессах. Исходя из этого, многие

исследователи (Тимофеев-Рессовский, Воронцов, Яблоков, 1969) оценивают процесс отбора только по его результатам; иначе говоря, с точки зрения этих авторов, отбор есть выживание того, что выживает. Но тогда принцип Дарвина сводился бы к тавтологии. По Эйгену (*Eigen*, 1973), существует, по крайней мере, при специально выбранных условиях, принципиально вычислимая «наиболее приспособленная», оптимальная селекционная ценность, позволяющая предсказывать результат отбора. Соответственно, все системы подразделяются на системы «с дарвиновским поведением» (выживание с преимуществом при отборе) и «с недарвиновским поведением» (выживание без преимущества при отборе) (*Эйген, Шустер*, 1982).

Распространение «эволюционных стратегий» на решение практических проблем или на компьютерные эксперименты (*Bremermann*, 1970; *Rechenberg*, 1973; *Schweifel*, 1977) требует математического определения отбора (*Papentin*, 1973). По этому определению, отбор состоит в том, что после вычисления вполне определенного «функционала приспособленности Φ » для каждого из сортов популяции к репликации допускается только подмножество привилегированных сортов с максимальным Φ . Иначе говоря, отбор состоит в последовательном замещении сортов, образующих множество с фиксированным числом элементов, сортами, обладающими большей приспособленностью в конкуренции за выживание. Именно в конкуренции между генетически различными членами популяции, взаимодействующими с окружающей средой, главным образом, с организмами другого вида, Камшилов (*Камшилов*, 1974) усматривает самую основу отбора. Таким образом, при поверхностном подходе отбор сводится к выживанию одних и вымиранию других.

Если мы захотим ответить на вопрос о возможности отбора в общих системах, например, описываемых системой дифференциальных уравнений

$$\dot{x}_i = f_i(x_1, \dots, x_n),$$

то прежде всего напрашивается определение, которое требует, чтобы при $t \rightarrow \infty$ некоторые x_i обращались в нуль. Однако такая точка зрения чрезмерно широка, поскольку позволяет интерпретировать распадающиеся радиоактивные изотопы как пример отбора. Более многообещающим представляется включение конкурентного аспекта как существенного элемента отбора (*Эбелинг*, 1979). С точки зрения авторов этой книги, отбор является частной формой кооперативного поведения подсистем (*Feistel, Ebeling*, 1976). Ограничение отбора процессами с конкурентными свойствами кажется неприемлемым лишь на первый взгляд. Не следует ли рассматривать как отбор уничтожение способных к размножению особей или вымирание части популяции в условиях изменяющейся окружающей среды (температура, влажность, свет и т. д.)? С точки зрения популяционной генетики, речь в этом случае идет об отборе, хотя и без конкуренции. Однако с точки зрения теории систем, исчезновение сортов соответствует лишь упомянутому выше радиоактивному распаду: в обоих случаях объекты исчезают независимо от того, существуют ли одновременно с ними другие объекты или не существуют. Поэтому мы считаем предметом отбора только такие объекты, которые в принципе могут существовать при заданных краевых условиях (в заданной окружающей среде) и существованию или выживанию которых может угрожать только конкуренция с одновременно существующими объектами (того же качества). Простой выбор между годными и негодными кандидатами мы не считаем отбором в собственном смысле.

Тем самым мы приходим к следующему определению: отбором называется особая форма когерентного поведения подсистем (сорт) сложной системы. Она встречается в том случае, когда в результате конкурентной ситуации между способными в принципе к существованию подсистемами (сортами) возникает процесс, ведущий к исчезновению по крайней мере одной из подсистем (сорт) или класса подсистем.

Критерием для отбора служит поведение системы при добавлении новой подсистемы (сорта). Мы говорим об отборе в том случае, когда добавление новых подсистем (сортов) к уже имеющимся подсистемам (чистым культурам) приводит при $t \rightarrow \infty$ к исчезновению по крайней мере одного класса имеющихся систем (чистых культур).

На основе рассмотрения процессов пребиологической эволюции Романовский, Степанова и Чернавский (Романовский, Степанова, Чернавский, 1984) обсуждали наряду с простым отбором (отбором благоприятных вариантов) еще два отбора:

1) один объект возникает раньше другого, что не позволяет последнему принять участие в происходящих событиях (по Кастлеру);

2) выбор из эквивалентных вариантов (в смысле мультистабильности).

Мы рассматриваем такие формы отбора с той точки зрения, что результат их действия зависит от начальных условий (начальных концентраций) и что существенную роль играет нелинейный закон роста (гиперболический рост). Следуя Деккеру (Decker, 1975), мы называем такие формы отбора «гиперотбором». Наоборот, результат рассмотренных выше процессов отбора зависит не от количества начальных условий, а лишь от некоторых совокупностей параметров, характеризующих сорта (от абстрактного фенотипа). Такой отбор называется «простым отбором» и в дальнейшем играет важную роль, поскольку допускает неограниченную последовательность сортов «все более высокого качества». Как мы увидим в дальнейшем, «простота» такой системы отнюдь не гарантирует существования селекционной ценности, стремления к оптимуму, действия обобщенного экстремального принципа.

Простой отбор и гиперотбор являются основными принципами самоорганизации. Они сыграли важную роль в процессе эволюции жизни на Земле.

Экстремальные принципы имеют особое значение в теории процессов отбора и эволюции. В математике понятие экстремального принципа трактуется весьма широко. Рассматриваемые экстремальные принципы носят специфический характер. В последующих разделах мы обсудим различные виды процессов конкуренции и отбора. Для этого полезно классифицировать важнейшие величины, экстремизируемые при отборе. Для наглядности мы ограничимся системами, содержащими только изолированные стационарные состояния; в принципе наш подход допускает обобщение на квазистационарные состояния и предельные циклы. Особенность рассматриваемых нами процессов состоит в том, что вектор состояния не может иметь отрицательных компонент: $\mathbf{x}(t) = \{x_i(t)\}$, $x_i(t) \geq 0$, так как концентрации и плотности не могут быть отрицательными. Вследствие этого динамика системы должна развиваться так, чтобы вектор состояния не выходил из так называемого положительного конуса.

Пусть R — односвязная область в положительном конусе n -мерного векторного пространства E^n . Пусть на R определена векторная функция $\mathbf{f}$, описывающая динамику вектора состояния и обладающая следующим свойством, о котором говорилось выше:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}), \quad (6.1)$$

где $\mathbf{u}$ — множество параметров. В R могут оказаться особые точки $\mathbf{x}^{(i)}$; $i = 1, 2, \dots, m$; пусть ровно одна из них ($\mathbf{x}^{(s)}$) асимптотически устойчива в смысле Ляпунова:

$$\mathbf{f}(\mathbf{x}^{(i)}, \mathbf{u}) = 0 \quad \text{при} \quad i = 1, 2, \dots, m. \quad (6.2)$$

Для классификации экстремальных принципов процессов отбора мы предлагаем следующую терминологию: если на R определена непрерывно дифференцируемая по $\mathbf{x}$ функция $G(\mathbf{x}, \mathbf{x}^0, \mathbf{u})$, определенная так, что

$$а) \quad L(\mathbf{x}, \mathbf{x}^{(s)}, \mathbf{u}) = [G(\mathbf{x}, \mathbf{x}^{(s)}, \mathbf{u}) - G(\mathbf{x}^{(s)}, \mathbf{x}^{(s)}, \mathbf{u})] < 0 \quad \text{при всех} \quad \mathbf{x} \neq \mathbf{x}^{(s)};$$

$$б) \quad \frac{d}{dt} L(\mathbf{x}, \mathbf{x}^{(s)}, \mathbf{u}) > 0 \quad \text{при всех } \mathbf{x} \neq \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(m)}, \quad (6.3)$$

то G можно назвать *динамическим экстремальным принципом на R* . Функция L здесь по существу есть не что иное, как функция Ляпунова. Если для особых точек $\mathbf{x}^{(i)}$ из R функция $F(\mathbf{x}^{(i)}, \mathbf{u})$, $i = 1, 2, \dots, m$, определена так, что неравенство

$$F(\mathbf{x}^{(s)}, \mathbf{u}) > F(\mathbf{x}^{(i)}, \mathbf{u}) \quad (6.4)$$

выполняется при всех $i \neq s$, то F можно назвать *статическим экстремальным принципом на R* . Если на R существует динамический экстремальный принцип G , такой, что

$$F(\mathbf{x}^{(i)}, \mathbf{u}) = G(\mathbf{x}^{(i)}, \mathbf{x}^{(i)}, \mathbf{u}) \quad (6.5)$$

— статический экстремальный принцип на R , то G можно назвать *полным экстремальным принципом на R* .

Для исследования процессов отбора наибольшее значение имеет вопрос о том, равна ли концентрация особей некоторого сорта нулю или она больше нуля, то есть представлен ли рассматриваемый сорт в системе или отсутствует. С математической точки зрения этот вопрос сводится к выяснению того, лежит ли конец вектора состояния на границе или внутри положительного конуса (рис. 6.1).

Важным подпространством положительного конуса является p -мерное *положительное подпространство* $R_p^>$, $\dim(R_p^>) = p$, в котором одни компоненты $x_1, x_2, \dots, x_p > 0$ всегда положительны, а все остальные считаются несуществующими. Рассмотрение такого подпространства представляет интерес и имеет смысл только тогда, когда

$$f_i(\mathbf{x}, \mathbf{u})_{x_{p+1}=\dots=x_n=0} = 0 \quad \text{при } i = p+1, \dots, n. \quad (6.6)$$

В этом случае

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u})$$

— замкнутая система уравнений в подпространстве $R_p^>$, где $\mathbf{u}$ — набор параметров, осуществляющих взаимосвязь компонент $x_1, \dots, x_p$ в $\mathbf{f}$.

Сказанное относится и к неотрицательному подпространству $R_p^{\geq}$, $\dim(R_p^{\geq}) = p$, в котором одни компоненты $x_1, x_2, \dots, x_p \geq 0$ неотрицательны, а все остальные считаются несуществующими.

Условимся называть экстремальный принцип

а) *принципом сегрегации*, если он распространяется на все имеющие смысл положительные подпространства $R_p^>$; (6.7)

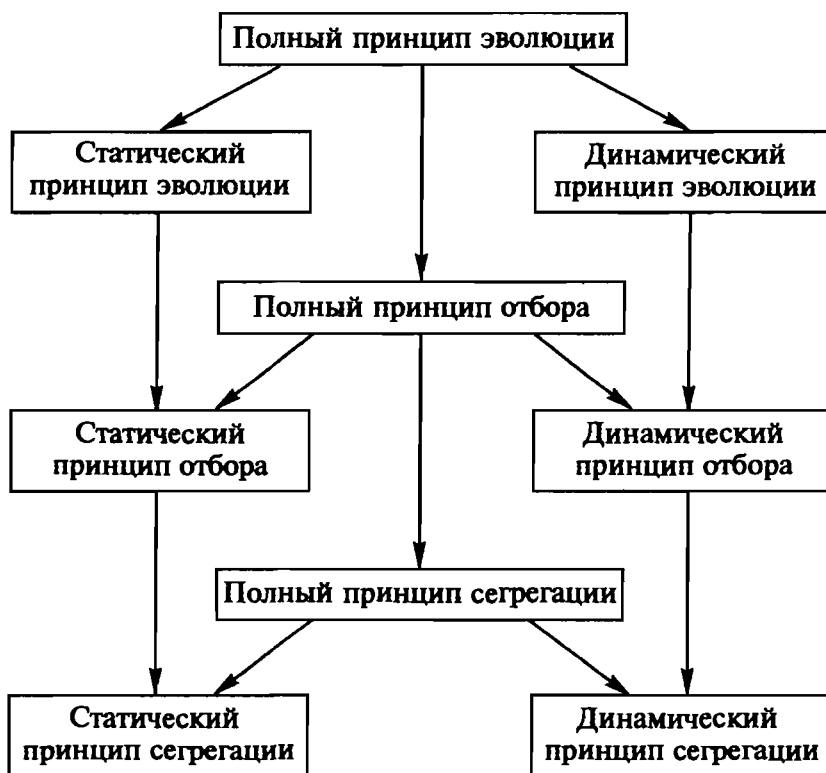
б) *принципом отбора*, если он распространяется на все имеющие смысл положительные подпространства $R_p^>$ и их объединения с неотрицательными пространствами $R_p^{\geq}$, соответствующими сортам, прошедшим отбор; (6.8)

в) *принципом эволюции*, если он распространяется на все имеющие смысл неотрицательные подпространства $R_p^{\geq}$. (6.9)



Рис. 6.1. Пространство концентрации химического вещества соответствует положительному конусу в евклидовом векторном пространстве

Определения (6.3)–(6.6) и (6.7)–(6.9) можно произвольным образом комбинировать друг с другом, в результате чего мы различаем девять различных принципов — от динамического принципа сегрегации до полного принципа эволюции. Некоторые из этих принципов включают в себя другие (см. схему).



Таким образом, существование любого из этих принципов гарантирует существование всех подчиненных ему принципов. Например, доказательство несуществования одного из принципов влечет за собой вывод о несуществовании всех предшествующих ему общих принципов. Если установлен полный принцип эволюции, то тем самым известны качественные свойства всех возможных процессов отбора и эволюции рассматриваемой модели.

Такой полный принцип эволюции в некоторых случаях приводит к термодинамике. Обусловленное силами изменение $d_x P$ производства энтропии для нашей однородной системы является просто пфаффовой формой относительно любых переменных:

$$d_x P = \sum_i A_i dx_i. \quad (6.10)$$

Если эта пфаффова форма — полный дифференциал:

$$A_i = -\frac{\partial G}{\partial x_i},$$

то по теореме Гленсдорфа—Пригожина (см. гл. 3) $G(x)$ представляет собой по крайней мере динамический, а быть может даже полный принцип эволюции.

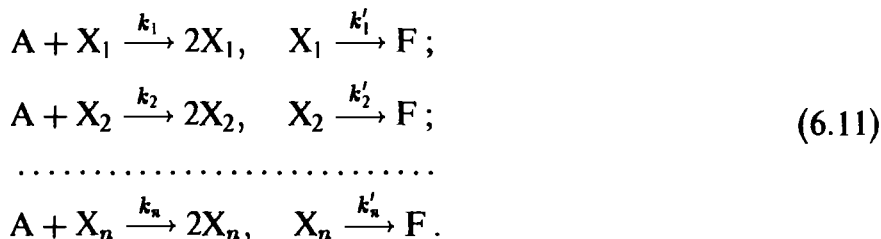
6.2. Отбор в случае простой конкуренции

В качестве простейшей стандартной модели конкуренции и отбора мы рассмотрим в этом разделе реакции, протекающие параллельно при условиях конкуренции. Простейшие условия конкуренции задаются для химических реакций

следующими краевыми условиями (Eigen, 1973; Schuster, 1972; Eigen, Winkler, 1975; Küppers, 1975, 1983; Ebeling, Feistel, 1974, 1976; Эбелинг, 1979):

- а) ограниченный подвод исходных веществ, расходуемых в ходе реакции;
- б) ограниченное общее число частиц конкурирующих сортов;
- в) ограниченное общее число особей конкурирующих видов.

В качестве модели в случае (а) мы рассмотрим параллельное автокаталитическое образование веществ X_i из исходного вещества A и распад веществ X_i в конечный продукт F :



В эквивалентной экологической модели происходит размножение (репродукция) видов X_i (с временем жизни $\frac{1}{k'_i}$) на основе пищевого ресурса A и их естественная смерть. Формальные кинетические уравнения для описания этих процессов имеют следующий вид:

$$\begin{aligned} \dot{A} &= \Phi - \sum_i k_i A X_i, \\ \dot{X}_i &= k_i A X_i - k'_i X_i. \end{aligned} \quad (6.12)$$

Здесь $\Phi > 0$ — постоянный подвод питательных веществ извне. Точные аналитические решения уравнений (6.12) неизвестны, но временная зависимость решений может быть исследована компьютерными методами (рис. 6.2).

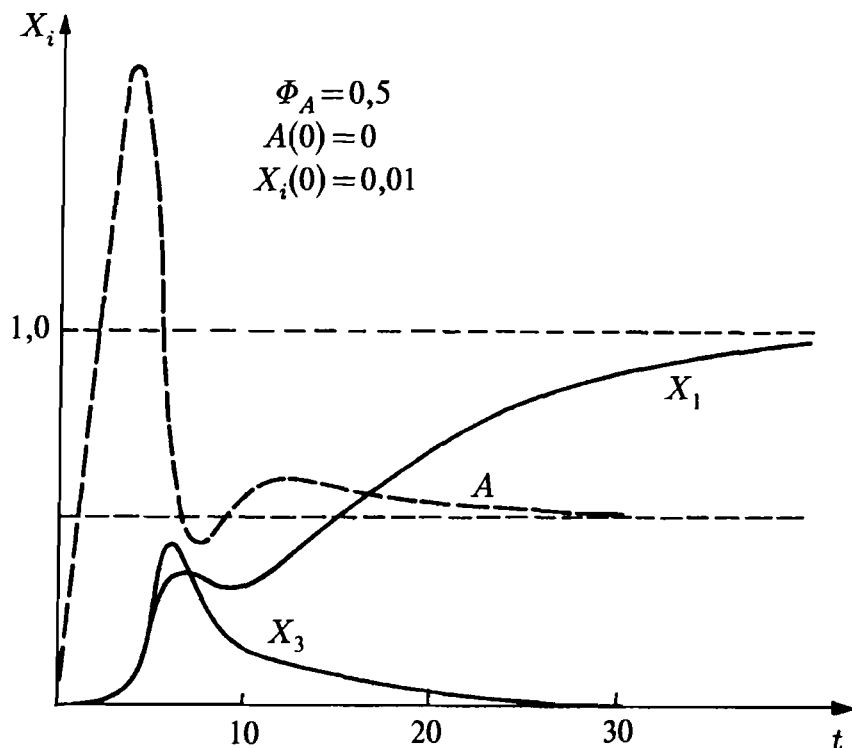


Рис. 6.2. Численное решение системы уравнений (6.12)

В пределе при больших временах решение асимптотически сходится к устойчивому стационарному решению. Свойства этого стационарного решения могут

быть исследованы аналитически. Для краткости введем для скорости репликации обозначение

$$E_i = k_i A - k'_i. \quad (6.13)$$

Эта величина удовлетворяет уравнениям (6.12):

$$\dot{A} = \Phi - A \sum_i k_i X_i, \quad \dot{X}_i = E_i X_i. \quad (6.14)$$

Для стационарных решений, по определению, выполняются равенства $\dot{X}_1 = \dot{X}_2 = \dots = \dot{X}_n = \dot{A} = 0$. Равенство $\dot{X}_i = 0$ выполняется только при $X_i = 0$ или $E_i = 0$. Из условия $\dot{A} = 0$ следует, что по крайней мере одна величина X_i должна быть положительной: $X_i > 0$. Так как E_i зависит только от A , в общем случае в нуль может обратиться самое большее одна величина E_i . Предположим, что нулю равна величина, соответствующая виду с номером s , т. е.

$$\begin{aligned} E_s &= 0, \quad X_s \neq 0; \\ E_i &\neq 0, \quad X_i = 0 \quad \text{при всех } i \neq s. \end{aligned} \quad (6.15)$$

Из $\dot{A} = 0$ следует, что в стационарном состоянии $A^{(s)} = k'_s/k_s$ и $X_s^{(s)} = \Phi/k'_s$. Для анализа устойчивости состояния, определяемого соотношениями (6.15), вычислим собственные значения матрицы Якоби системы (6.12):

$$\begin{aligned} p_i &= k_i \left(\frac{k'_s}{k_s} - \frac{k'_i}{k_i} \right) \quad \text{при всех } i \neq s, \quad i \leq n, \\ p_s &= p_{n+1} = -\frac{\Phi k_s}{2 k'_s} \left[1 \pm \sqrt{1 - \frac{4 k'_s k'_s}{\Phi k_s}} \right]. \end{aligned}$$

Условие устойчивости гласит:

$$p_i = k_i \left(\frac{k'_s}{k_s} - \frac{k'_i}{k_i} \right) < 0 \quad \text{при всех } i \neq s, \quad i \leq n. \quad (6.16)$$

Таким образом, выживают те сорта частиц, образование которых может начинаться с минимальных стационарных концентраций исходных веществ $A^{(s)} = k'_s/k_s$. Иначе говоря, выживают те сорта, для которых максимально произведение времени жизни $1/k'_s$ и числа «потомков» k_s на единицу количества исходного вещества и единицу времени, т. е. сорта, каждая «молекула» которых в течение своей «жизни» порождает максимальное число «потомков» (Эбелинг, 1979).

Число, позволяющее вычислить свойства сорта и предопределяющее исход отбора, мы, следуя Эйгену, называем «селекционной ценностью». Как мы убедимся, существование такой величины обеспечено отнюдь не во всех случаях.

В смысле нашей классификации экстремальных принципов, приведенной в разд. 6.1, речь идет о статическом принципе эволюции для функции

$$F(\mathbf{x}^{(i)}, \mathbf{u}) = \frac{1}{A^{(i)}} = \sum_j \frac{k_j}{k'_j} \delta_{ij}, \quad (6.17)$$

поскольку она определена для всех стационарных состояний (в том числе и для стационарных состояний на границе положительного конуса) и в асимптотически устойчивом состоянии принимает наибольшее значение. Так как концентрация исходного вещества со временем изменяется не монотонно (рис. 6.2), функция (6.17)

не порождает динамический экстремальный принцип. Тем не менее на определенных гиперповерхностях в фазовом пространстве функция $G = 1/A$ может выступать в качестве динамического и тем самым полного принципа эволюции (Feistel, 1983, 1986).

Рассмотрим теперь, как происходит отбор в системе n сортов, находящейся в асимптотически устойчивом состоянии

$$A^{(s)} = \frac{k'_s}{k_s}, \quad X_i^{(s)} = 0 \quad \text{при} \quad i \neq s, \quad X_s^{(s)} = \frac{\Phi}{k'_s} \quad (6.18)$$

до прививки некоторого нового сорта μ или мутанта старого вида. Пока концентрация прививаемого сорта мала, рост происходит по простому закону

$$\begin{aligned} \dot{X}_\mu &= k_\mu \left(\frac{k'_s}{k_s} - \frac{k'_\mu}{k_\mu} \right) X_\mu, \\ X_\mu(t) &= X_\mu(0) \exp \left\{ \left(\frac{k'_s}{k_s} - \frac{k'_\mu}{k_\mu} \right) k_\mu t \right\}. \end{aligned} \quad (6.19)$$

В соответствии со сформулированным выше экстремальным принципом и предположением о том, что $A^{(\mu)} < A^{(s)}$, протекающий в системе процесс отбора в конце концов сходится к новому асимптотически устойчивому состоянию

$$A^{(\mu)} = \frac{k'_\mu}{k_\mu}, \quad X_i^{(\mu)} = 0 \quad \text{при} \quad i \neq \mu, \quad X_\mu^{(\mu)} = \frac{\Phi}{k'_\mu}. \quad (6.20)$$

Таким образом, в предположении $(k'_\mu/k_\mu) < (k'_s/k_s)$ прививаемый сорт μ вытесняет доминировавший прежде сорт s . Если добавить к системе такие мутанты или новые сорта μ , то при условии

$$E_\mu = k_\mu A^{(s)} - k'_\mu > 0,$$

т. е.

$$\frac{k'_\mu}{k_\mu} = A^{(\mu)} < A^{(s)} = \frac{k'_s}{k_s}, \quad (6.21)$$

они могут расти. Собственные значения p_s, p_{n+1} (6.16) позволяют заключить, что при малых значениях подвода исходного вещества ($\Phi < 4k_\mu'^2/k_\mu$) режим выходит на стационарное состояние, а не на затухающие колебания. Наша модель обладает еще одной интересной особенностью, а именно: результат конкуренции не зависит от краевых условий (величины Φ), то есть независимо от того, поступает ли исходное вещество в избытке или подвод его стремится к нулю, оптимален один и тот же сорт s и устанавливается одна и та же концентрация исходного вещества.

Рассмотрим теперь второе из упоминавшихся в самом начале условий конкуренции — постоянство общего числа частиц. Концентрацию исходного числа мы поддерживаем постоянной, управляя надлежащим образом величиной Φ и реализуя с помощью управляемого разрежения системы постоянство суммарного состава (Eigen, 1973):

$$\sum_{i=1}^n X_i = C = \text{const}. \quad (6.22)$$

Геометрически это условие означает, что все траектории лежат на поверхности, которая пересекает положительный конус при $X_i = C$. Эйген и Шустер (Эйген, Шустер, 1982) называют многогранную симметричную поверхность, образующую

в данном случае пространство состояний, симплексом. Кинетические уравнения для движения системы по симплексу имеют следующий вид:

$$\dot{X}_i = E_i X_i - k' X_i. \quad (6.23)$$

По предположению, для всех сортов частиц одна и та же скорость разбавления k' должна вводиться с таким расчетом, чтобы выполнялось условие (6.22). Суммируя по всем сортам частиц, получаем

$$\sum \dot{X}_i = \frac{d}{dt} \sum X_i = \sum E_i X_i - k' \sum X_i.$$

Так как левая часть обращается в нуль в силу условия (6.22), находим

$$k' = \frac{\sum E_i X_i}{\sum X_i} = \langle E \rangle, \quad (6.24)$$

откуда следует уравнение Фишера—Эйгена

$$\dot{X}_i = X_i (E_i - \langle E \rangle), \quad (6.25)$$

где $\langle E \rangle = \sum E_i X_i / C$ — среднее избыточное производство.

Соотношение (6.22) можно интерпретировать и как возможность протекания в замкнутой системе только процессов превращения

$$X_i \leftrightarrow X_j. \quad (6.26)$$

При такой интерпретации можно попытаться вывести принцип эволюции на термодинамическом пути. Нас интересует только число частиц в системе, в соответствии с чем мы используем в качестве исходного пункта энтропию смешения идеальных частиц (Ландау, Лифшиц, 1995):

$$S = -kV \sum_i X_i \left(\ln \frac{X_i}{X_i^0} - 1 \right). \quad (6.27)$$

Поскольку мы не допускаем обмен частицами, производство энтропии составляет величину

$$-k \sum_i X_i (E_i - \langle E \rangle) \ln \frac{X_i}{X_i^0}. \quad (6.28)$$

Здесь $\dot{X}_i$ — термодинамический поток, $-k \ln(X_i/X_i^0)$ — термодинамическая сила (сродство); для вызванного изменением силы изменения производства энтропии σ справедливо соотношение

$$d_x \sigma = -k \sum_{i=1}^n (E_i - \langle E \rangle) dX_i. \quad (6.29)$$

Это пфафдова форма есть полный дифференциал:

$$G = kC \langle E \rangle + \text{const}, \quad d_x \sigma = -d\sigma. \quad (6.30)$$

В силу принципа Гленсдорфа—Пригожина $d_x \sigma \leq 0$, для этой величины выполняется динамический экстремальный принцип. Чтобы убедиться в сказанном, рассмотрим выражение

$$X_i(t) = z_i(t) \exp \left\{ - \int_0^t \langle E \rangle dt \right\}. \quad (6.31)$$

Нетрудно видеть, что

$$X_i(t) = C \frac{X_i(0) \exp \{E_i t\}}{\sum_j X_j(0) \exp \{E_j t\}} \quad (6.32)$$

— аналитическое решение. График его представлен на рис. 6.3.

Совокупность n различных стационарных состояний $s = 1, 2, \dots, n$ характеризуется тем, что в системе выживает только сорт X_s :

$$X_1^{(s)} = \dots = X_{s-1}^{(s)} = X_{s+1}^{(s)} = \dots = X_n^{(s)} = 0; \quad (6.33)$$

$$X_s^{(s)} = C.$$

Из всех этих состояний асимптотически устойчиво именно то, которое соответствует наибольшему значению E_i :

$$E_s > E_i \quad \text{при всех } i \neq s. \quad (6.34)$$

Следовательно, рассматриваемая система обладает только одним-единственным устойчивым стационарным состоянием $X_s = C$; $X_1 = \dots = X_n = 0$, расположенным в одной из вершин симплекса. Траектория системы при $t \rightarrow \infty$ асимптотически стремится к этому устойчивому состоянию. Проверим теперь, полон ли наш экстремальный принцип. Согласно соотношению (6.3), неравенство

$$L(\mathbf{x}, \mathbf{u}) = \langle E \rangle - E_s < 0 \quad (6.35)$$

должно выполняться при всех $\mathbf{x} \neq \mathbf{x}^{(s)}$. Оно выполняется в силу неравенства (6.34), так как среднее значение всегда меньше усредняемой величины. Проверим теперь условие (6.36):

$$\begin{aligned} \frac{d}{dt} L &= \frac{d}{dt} \langle E \rangle = \frac{1}{C} \sum_i E_i \dot{X}_i = \\ &= \frac{1}{C} \sum_i (E_i^2 - E_i \langle E \rangle) X_i = \langle E^2 \rangle - \langle E \rangle^2 \geq 0. \end{aligned} \quad (6.36)$$

Наконец, для

$$F(\mathbf{x}^{(i)}, \mathbf{u}) = E_i, \quad (6.37)$$

как следует из неравенства (6.34), выполняется статический экстремальный принцип. Так как F и G определены и на всех границах положительного конуса, для величины

$$G(\mathbf{x}, \mathbf{u}) = \langle E \rangle \quad (6.38)$$

действительно выполняется полный экстремальный принцип.

Рассмотрим в заключение еще одно важное обобщение конкурентной ситуации, о которой шла речь выше. Предположим, что воспроизведение молекул сорта i осуществляется в результате каталитического действия имеющихся молекул сорта j . Уравнения реакций имеют вид:

$$\dot{X}_i = E_i X_i + X_i \sum_j b_{ij} X_j - k' X_i. \quad (6.39)$$

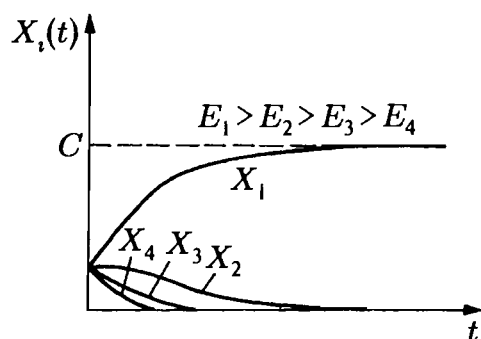


Рис. 6.3. Конкурентный процесс при постоянной суммарной концентрации

Предполагается, что $b_{ii} = 0$ при всех i и $b_{ij} > 0$ по крайней мере при некоторых i и j . Из условия постоянства полного числа частиц $\sum X_i = C$ следует постоянство величины k , и мы получаем

$$k' = \sum_i \left\{ E_i X_i + X_i \sum_j b_{ij} X_j \right\} \frac{1}{C}. \quad (6.40)$$

В зависимости от матрицы связи $\{b_{jk}\}$ система уравнений (6.39) может обнаруживать чрезвычайно сложное поведение отбора (см. также гл. 8). В то время как простая система (6.21), т. е. $b_{jk} = 0$, обладает только единственным устойчивым стационарным состоянием, лежащим в одной из вершин симплекса, уравнение (6.39), как правило, обладает несколькими устойчивыми стационарными состояниями. Симплекс разделен сепаратрисами на несколько «бассейнов», каждый из которых представляет собой область притяжения какого-то аттрактора. Теория системы отбора (6.39) была разработана в связи с теорией гиперциклов Эйгена—Шустера (Эйген, Шустер, 1982; Küppers, 1983, 1987; Hofbauer, Sigmund, 1984).

В реальных процессах эволюции отбор, как правило, происходит между установившимся чистым сортом и вновь возникающим, более приспособленным в конечном счете мутантом. Последний впервые появляется в виде всего лишь нескольких экземпляров (частиц), вследствие чего решение о выживаемости мутанта «принимается» в условиях крайне малого числа частиц. Поэтому поведение системы носит существенно стохастический характер.

Для анализа этих эффектов сформулируем в рамках рассмотренной выше модели отбора основные уравнения (Ebeling, Feistel, 1977; Ebeling, Feistel, Jiménez-Montaño, 1977; McCaskill, 1984). Начнем, как и прежде, с краевого условия — ограниченного подвода исходного вещества, и обозначим число частиц исходного вещества А через N_0 , число частиц сорта X_1 — через N_1 и т. д., и через V — объем рассматриваемой системы. Для трех различных каналов реакции (рис. 6.4.) находим следующие вероятности перехода W (см. гл. 5):

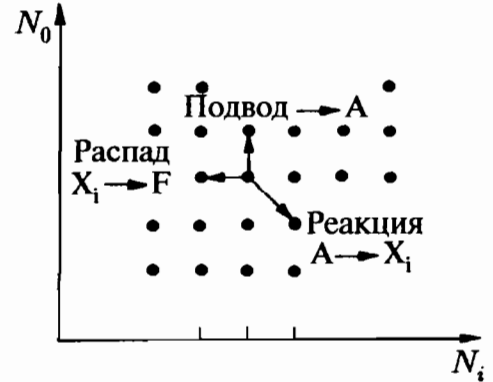


Рис. 6.4. Состояния и переходы в модели отбора при постоянном подводе сырья

$$\begin{aligned} R \rightarrow A: \quad N_0 &\rightarrow N_0 + 1, & W_1(N_0, \dots, N_n) &= \Phi V, \\ A \rightarrow X: \quad \begin{pmatrix} N_0 \\ N_i \end{pmatrix} &\rightarrow \begin{pmatrix} N_0 - 1 \\ N_i + 1 \end{pmatrix}, & W_2(N_0, \dots, N_n) &= \frac{k}{V} N_0 N_i, \\ X_i \rightarrow F: \quad N_i &\rightarrow N_i - 1, & W_3(N_0, \dots, N_n) &= k'_i N_i \end{aligned} \quad (6.41)$$

(R — резервуар, F — конечный продукт). Тем самым основное уравнение для вероятности $P(N_0, N_1, \dots, N_n, t)$ того, что в момент времени t в системе имеется ровно N_0 частиц исходного вещества, N_1 частиц сорта X_1 и т. д., имеет следующий вид:

$$\begin{aligned} \frac{\partial}{\partial t} P(N_0, \dots, N_n, t) &= \\ &= \Phi V [P(N_0 - 1, N_1, \dots, N_n, t) - P(N_0, N_1, \dots, N_n, t)] + \\ &+ \sum_i \frac{k_i}{V} [(N_0 + 1)(N_i - 1)P(N_0 + 1, N_1, \dots, N_i - 1, \dots, N_n, t) - \end{aligned}$$

$$\begin{aligned}
& - N_0 N_i P(N_0, N_1, \dots, N_i, \dots, N_n, t) \Big] + \\
& + \sum_i k'_i [(N_i + 1) P(N_0, N_1, \dots, N_i + 1, \dots, N_n, t) - \\
& - N_i P(N_0, N_1, \dots, N_i, \dots, N_n, t)].
\end{aligned} \tag{6.42}$$

К сожалению, эти уравнения слишком сложны для того, чтобы аналитическими методами можно было найти решение $P(N_0, N_1, \dots, N_n, t)$. Рассмотрим прежде всего, как ведет себя система при больших временах. В пределе при $t \rightarrow \infty$ только стоки (поглощающие состояния) графика, описывающие переходы между состояниями, имеют вероятность, отличную от нуля. Поэтому мы рассмотрим график в пространстве $N_i - N_j$ (рис. 6.5).

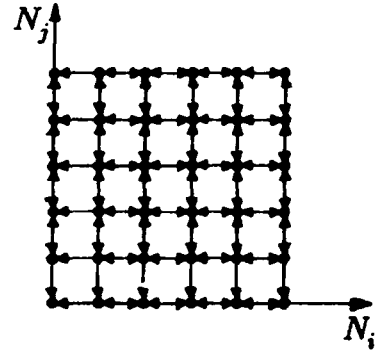


Рис. 6.5. Граф процессов перехода при условиях отбора обладает стоками на границе (поглощающими состояниями)

Стоками для каждого сорта N_i ($i > 0$) являются состояния $N_i = 0$. В частности, при $t \rightarrow \infty$ система стремится также к состоянию $N_1 = N_2 = \dots = N_n = 0$, а количество частиц исходного вещества N_0 неограниченно возрастает.

Это означает вымирание всех интересных сортов — так называемую «флуктуационную катастрофу». Характерный для этой катастрофы промежуток времени при не слишком малом среднем числе частиц все же очень велик, в частности в термодинамическом пределе (при $V \rightarrow \infty$) стремится к бесконечности. Наоборот, характерное время протекания процессов отбора всегда конечно и через константы реакций определяет участвующие в реакции сорта частиц. Если объем системы также выбран достаточно большим, то мы имеем два процесса с весьма различными временными масштабами:

- 1) стремление отбора к равновесию ($N_s > 0$, $N_i = 0$ при $i \neq s$, $i > 0$);
- 2) вымирание последних сортов вследствие флуктуационной катастрофы.

Поскольку последний процесс протекает очень медленно, условимся рассматривать распределение вероятности сразу же после наступления равновесия в отборе как стационарное.

Если основное уравнение (6.42) умножить по очереди на N_0 , N_i и гармоническую сумму

$$H(N_0) = \sum_{k=1}^{N_0} \frac{1}{k} \tag{6.43}$$

и каждый раз суммировать по числу частиц каждого сорта, то получим соотношения

$$\frac{\partial}{\partial t} \langle N_0 \rangle = \Phi V - \frac{1}{V} \sum_i k_i \langle N_0 N_i \rangle, \tag{6.44}$$

$$\frac{\partial}{\partial t} \langle N_i \rangle = \frac{1}{V} k_i \langle N_0 N_i \rangle - k'_i \langle N_i \rangle, \tag{6.45}$$

$$\frac{\partial}{\partial t} \langle H(N_0) \rangle = \Phi V \left\langle \frac{1}{N_0 + 1} \right\rangle - \frac{k_i}{V} \langle N_i \rangle. \tag{6.46}$$

Гармоническая сумма играет в стохастической картине примерно такую же роль, какую логарифмическая функция играет в детерминистической картине, и в термодинамическом пределе переходит в последнюю. Мы рассмотрим случай, когда имеется один сорт s с положительным числом частиц N_s . Из соотношений (6.44)–(6.46),

если их левые части положить равными нулю, получаем:

$$\left\langle \frac{V}{N_0 + 1} \right\rangle = \frac{k_s}{k'_s} \quad \text{и} \quad \left\langle \frac{N_s}{V} \right\rangle = \frac{\Phi}{k'_s}, \quad (6.47)$$

что согласуется с детерминистическим результатом (6.18). В стохастической картине однозначно не определено, какой сорт s побеждает в конкуренции, поэтому каждый сорт i имеет определенную вероятность σ_i на выживание. Эти вероятности выживания представляют особый интерес в том случае, когда к имеющемуся чистому сорту s примешивается мутант i . В предположении, что чистый сорт находится в стационарном состоянии

$$N_0 \approx V \cdot \frac{k'_s}{k_s}, \quad N_s \approx V \cdot \frac{\Phi}{k'_s} \quad (6.48)$$

и что число частиц мутанта мало

$$N_i \ll N_s, \quad (6.49)$$

вероятность выживания σ_i может быть вычислена приближенно. Для этого упростим основное уравнение (6.42): воспользуемся соотношениями (6.48) и выберем N_i в качестве стохастической переменной:

$$\begin{aligned} \frac{\partial}{\partial t} P(N_i; t) = & k_i \frac{k'_s}{k_s} (N_i - 1) P(N_i - 1; t) + \\ & + k'_i (N_i + 1) P(N_i + 1; t) - \left(k_i \frac{k'_s}{k_s} + k'_i \right) N_i P(N_i; t). \end{aligned} \quad (6.50)$$

Решение этого уравнения может быть найдено методом производящей функции (см. разд. 5.2). Для вероятности того, что в момент времени t сорт N_i вымирает, получаем:

$$P(0; t) = \left[\frac{\exp \{ (k'_s/k_s - k'_i/k_i) k_i t \} - 1}{(k_i/k'_i)(k_s/k'_s) \exp \{ (k'_s/k_s - k'_i/k_i) k_i t \} - 1} \right]^{N_i^0}. \quad (6.51)$$

Зная ее, мы вычисляем вероятность перехода:

$$\sigma_i = 1 - \lim_{t \rightarrow \infty} P(0; t) = \begin{cases} 0 & \text{при } \frac{k_i}{k'_i} < \frac{k_s}{k'_s}; \\ 1 - \left(\frac{k_s k'_i}{k'_s k_i} \right)^{N_i^0} & \text{при } \frac{k_i}{k'_i} > \frac{k_s}{k'_s}. \end{cases} \quad (6.52)$$

В частности, мы получаем для стандартной ситуации, когда мутант встречается лишь как одна частица ($N_i^0 = 1$),

$$\sigma_i = \frac{k_i/k'_i - k_s/k'_s}{k_i/k'_i} \quad \text{при} \quad \frac{k_i}{k'_i} > \frac{k_s}{k'_s}, \quad (6.53)$$

т. е. вероятность выживания нового сорта равна его относительному преимуществу при отборе по сравнению с чистым сортом. (Отношения k_i/k'_i встречались нам в детерминистической теории под названием селекционной ценности.) Отсюда мы можем сделать вывод относительно общего случая: как правило, сорт s с наивысшей селекционной ценностью

$$\frac{k_s}{k'_s} > \frac{k_i}{k'_i} \quad \text{при всех} \quad i \neq s \quad (6.54)$$

обладает наибольшей вероятностью выживания и поэтому чаще всего оказывается победителем в конкурентной борьбе. Это означает также, что средняя обратная концентрация исходного вещества (6.47), как правило, стремится к максимальному значению:

$$\left\langle \frac{V}{N_0 + 1} \right\rangle = \frac{k_s}{k'_s} \rightarrow \max. \quad (6.55)$$

Этот результат известен нам из детерминистической теории простых конкурирующих реакций.

Анализ вероятностей выживания выступает как стохастическая противоположность детерминистическому исследованию устойчивости стационарных состояний на границе положительного конуса. Оба подхода приводят к согласующимся результатам. Следует подчеркнуть, что в детерминистической картине более приспособленный мутант всегда вытесняет предшествующий сорт, в то время как в стохастической картине всегда существует отличная от нуля вероятность того, что чистый сорт сохранит свое доминирующее положение.

Для вероятности перехода

$$\begin{pmatrix} N_i \\ N_j \end{pmatrix} \rightarrow \begin{pmatrix} N_i - 1 \\ N_j + 1 \end{pmatrix}$$

положим

$$\begin{aligned} W(N_1, \dots, N_i - 1, \dots, N_j + 1, \dots, N_n \mid N_1, \dots, N_i, \dots, N_j, \dots, N_n) = \\ = A_{ji} N_i + B_{ji} N_j N_i \end{aligned} \quad (6.56)$$

и определим коэффициенты B_{ji} , A_{ji} из условия, что в случае исчезающих флуктуаций для средних значений выполняются уравнения (6.23). Тогда

$$A_{ji} = 0, \quad B_{ji} = \frac{1}{N} E_j \quad (E_j > 0). \quad (6.57)$$

Заметим, что противоположный выбор $B_{ji} = E_i/N$ представляет интерес для биологических и экономических проблем (Ebeling, Feistel, 1976; Feistel, Ebeling, 1976; Ebeling, 1979). Из соотношений (6.56) и (6.57) следует основное кинетическое уравнение:

$$\frac{\partial}{\partial t} P(N_1, \dots, N_n; t) = \frac{1}{N} \sum_i \sum_{j \neq i} E_i \{ (N_i - 1)(N_j + 1) P(N_i - 1, N_j + 1) - N_i N_j P \} \quad (6.58)$$

(не указанные явно аргументы вероятности P в правой части уравнения (6.58) совпадают с аргументами, указанными в левой части уравнения). Для средних значений

$$\langle N_i \rangle = \sum_{N_1} \sum_{N_2} \dots \sum_{N_n} N_i P(N_1, \dots, N_i, \dots, N_n; t) \quad (6.59)$$

получаем из уравнений движения

$$\frac{d}{dt} \langle N_i \rangle = E_i \langle N_i \rangle - \frac{1}{N} \sum_j E_j \langle N_i N_j \rangle. \quad (6.60)$$

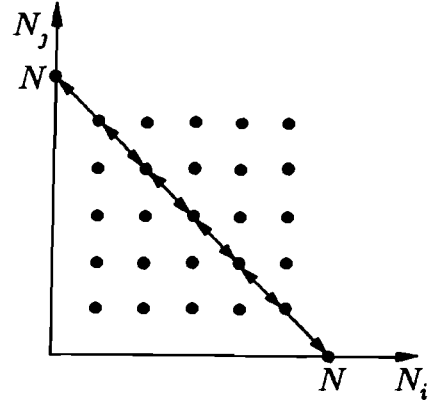


Рис. 6.6. Стохастическая реализация условия постоянства суммарного числа частиц

Дифференцируя $\langle E \rangle$ по времени и используя уравнения (6.60), получаем:

$$\frac{d}{dt} \langle E \rangle = \frac{1}{N} \left\langle \sum_i N_i \left(E_i - \frac{1}{N} \sum_j E_j N_j \right)^2 \right\rangle \geq 0, \quad (6.61)$$

т. е. $\langle E \rangle$ также максимизируется. Нетрудно видеть, что равенство в соотношении (6.61) может выполняться только для стационарного распределения вида

$$P^0(N_1, \dots, N_n) = \sum_{s=1}^n \sigma_s \cdot \delta_{N_1 0} \cdot \dots \cdot \delta_{N_s N} \cdot \dots \cdot \delta_{N_n 0}, \quad (6.62)$$

где

$$\sigma_s \geq 0, \quad \sum_s \sigma_s = 1$$

— вероятности выживания сортов s . И снова стационарное распределение характеризуется поглощающими состояниями (стоками) при $N_i = 0$, типичными для процессов отбора. Они разрешают полное исчезновение уже отобранных сортов и допускают будущие «лучшие» сорта, которые образуются в ходе дальнейшей эволюции, без влияния на ныне существующую систему.

Величина $\langle E \rangle$ не достигает своего абсолютного максимума, как в детерминистическом случае, а принимает лишь значение

$$\lim_{t \rightarrow \infty} \langle E \rangle = \sum_i E_i \sigma_i, \quad (6.63)$$

где величины σ_i зависят от начальных условий.

В случае двух сортов вероятности выживания легко определимы. Рассмотрим снова чистый сорт s и мутант i . Так как вследствие нормировки выполняется равенство $\sigma_s + \sigma_i = 1$, нам для вычисления σ_i вместо полного решения необходимо знать только интеграл движения, т. е. сохраняющее среднее значение функции $f(N_i)$:

$$\frac{d}{dt} \langle f(N_i) \rangle = 0. \quad (6.64)$$

Искомый интеграл имеет следующий вид:

$$\langle f(N_i) \rangle = \sum_{N_i} f(N_i) P(N_i; t) = \sum_{N_i} \left(\frac{E_s}{E_i} \right)^{N_i} P(N_i; t) = G \left(\frac{E_s}{E_i}; t \right). \quad (6.65)$$

В случае распределений

$$\begin{aligned} P(N_i; t=0) &= \delta_{N_i N_i^0}, \\ P(N_i; t=\infty) &= \sigma_i \delta_{N_i N} + (1 - \sigma_i) \delta_{N_i 0} \end{aligned} \quad (6.66)$$

из равенства

$$\langle f(N_i) \rangle_{t=0} = \langle f(N_i) \rangle_{t=\infty}$$

следует уравнение для σ_i , допускающее решение

$$\sigma_i = \frac{[1 - (E_s/E_i)^{N_i^0}]}{[1 - (E_s/E_i)^N]}. \quad (6.67)$$

Вероятности выживания зависят от селекционных ценностей и одновременно от начальных условий. Для случая мутантов ($N_i^0 = 1$) и большого общего числа частиц $N \rightarrow \infty$ получаем:

$$\sigma_i = \begin{cases} 0 & \text{при } E_i < E_s; \\ \frac{(E_i - E_s)}{E_i} & \text{при } E_i > E_s, \end{cases} \quad (6.68)$$

т. е. и в этом случае вероятность выживания равна относительному преимуществу при отборе.

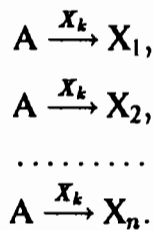
В заключение мы рассмотрим стохастическую теорию для случая, когда между двумя конкурирующими сортами имеются каталитические связи, соответствующие соотношениям (6.33) в детерминистической картине. Вместо вероятностей (6.56) и (6.57) мы получаем вероятности переходов (Ebeling, Jiménez-Montaño, 1980):

$$\begin{aligned} W(N_1, \dots, N_i - 1, \dots, N_j + 1, \dots, \\ N_k, \dots, N_n \mid N_1, \dots, N_i, \dots, N_j, \dots, N_k, \dots, N_n) = \\ = \frac{1}{N} E_j N_i N_j + \frac{1}{NV} b_{jk} N_i N_j N_k. \end{aligned} \quad (6.69)$$

Соответствующие стохастические процессы обсуждаются в других работах (Эбелинг и др., 1986; Bruckner et al., 1989).

6.3. Конкуренция и отбор в сетях

При рассмотрении простой конкуренции мы ограничились параллельными и независимо протекающими автокаталитическими реакциями. Связи между этими реакциями, ответственные за конкурентное поведение, были заменены условием постоянного притока исходных веществ и условием постоянства суммарной концентрации. Введем теперь перекрестные каталитические связи, состоящие в том, что на производство вещества одного сорта оказывает каталитическое влияние существование вещества другого сорта. В общем случае концентрация каждого из имеющихся сортов X_k может оказывать влияние на производство сортов $X_1, X_2, \dots, X_n$:



Автокатализ можно рассматривать как частный случай перекрестной каталитической связи. В дальнейшем мы ограничимся исключительно линейной каталитической связью. Начнем с рассмотрения систем с постоянным притоком исходных веществ.

В рамках формальной кинетики нашу модель можно сформулировать следующим образом: наряду с реакциями



мы допускаем реакцию типа



Запишем соответствующие дифференциальные уравнения

$$\begin{aligned}\dot{A} &= \Phi - A \sum_{i,j} k_{ij} X_j, \\ \dot{X}_i &= A \sum_j k_{ij} X_j - k'_i X_i\end{aligned}\tag{6.72}$$

в матричном виде

$$\begin{aligned}\dot{A} &= \Phi - A \cdot I^T k X, \\ \dot{X} &= A k X - k' X = E X,\end{aligned}\tag{6.73}$$

где (индекс Т означает транспонирование)

$$\begin{aligned}X &= (X_1, X_2, \dots, X_n)^T, \quad I = (1, 1, \dots, 1)^T, \\ k &= \{k_{ij}\}, \quad k' = \{k'_i \delta_{ij}\}, \quad E = A k - k' .\end{aligned}$$

Описание автокаталитических процессов репликации приводит к уравнениям вида (6.73) и в случае учета мутаций (Eigen, 1971; Jones et al., 1976; Ebeling, Feistel, 1975, 1977; Эйген, Шустер, 1982). Действительно, если реакции (6.71) протекают достаточно медленно по сравнению с реакциями (6.70), т. е.

$$\sum_{j \neq i} k_{ij} < k_{ii}, \quad \sum_{j \neq i} k_{ji} < k_{ii} \quad \text{при всех } i,\tag{6.74}$$

то реакции (6.71) следует рассматривать не как катализ, а как производство ошибочных копий. Различие между двумя интерпретациями уравнений (6.73) математически выражается в том, что матрица k в последнем случае вследствие неравенств (6.74) обладает свойством диагональной доминантности и поэтому принадлежит к классу монотонных матриц.

Попытаемся теперь найти стационарные решения системы (6.73). Из $\dot{A} = 0$, $\dot{X} = 0$ мы заключаем (Ebeling, Feistel, 1976), что

$$\Phi = I^T k X\tag{6.75}$$

и

$$k'^{-1} k X = \frac{1}{A} X,\tag{6.76}$$

так как матрица k' — невырожденная диагональная матрица и поэтому обратима. В соответствии со сказанным стационарные значения величин, обратных концентрациям исходных веществ, являются собственными значениями матрицы $k'^{-1} k$, а соответствующие концентрации X_i образуют собственный вектор, норма которого определяется соотношением (6.76). Тем самым исследование стационарных состояний системы реакций сводится к решению задачи на собственные значения.

Огромным преимуществом для анализа этой задачи на собственные значения является то обстоятельство, что матрица $k'^{-1} k$ не содержит отрицательных элементов. Это позволяет использовать общую теорию свойств разложимости и спектральных свойств неотрицательных матриц. Кроме того, анализ структуры матрицы $k'^{-1} k$ может быть произведен весьма наглядно с помощью соответствующих ориентированных графов. Так как k' — диагональная матрица, свойства разложимости матриц $k'^{-1} k$ и k полностью совпадают, что позволяет нам ограничиться рассмотрением матрицы k .

Назовем граф, соответствующий матрице $\mathbf{k}$, производящим графом системы (6.73). Ориентированные ребра производящего графа соответствуют матричным элементам $k_{ij} > 0$. Ребро, ориентированное от j к i , означает, что сорт j производит сорт i . Структура производящего графа тем самым отражает сеть реакционных потоков в системе.

Предположим сначала, что матрица $\mathbf{k}$ неразложима. В качестве примера выберем матрицу

$$\mathbf{k} = \begin{bmatrix} k_{11} & k_{12} & k_{13} & 0 & 0 & 0 & 0 & k_{18} \\ k_{21} & k_{22} & 0 & k_{24} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & k_{52} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & k_{65} & 0 & k_{67} & 0 \\ 0 & 0 & 0 & 0 & 0 & k_{76} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & k_{87} & 0 \end{bmatrix}$$

(см. рис. 6.7).

В этом случае существует ровно одно стационарное состояние $\mathbf{X}^0$, причем ни один сорт не исчезает, так как неотрицательная неразложимая матрица всегда обладает действительным положительным невырожденным собственным значением, которое остальные собственные значения не превосходят по абсолютной величине, и соответствующим собственным вектором с положительными компонентами (теорема Фробениуса). Поскольку неразложимая неотрицательная матрица не имеет двух линейно независимых неотрицательных собственных векторов, стационарное состояние однозначно определено.

Частным случаем является диагонально доминантная неразложимая (в большинстве случаев даже положительная) матрица $\mathbf{k}$ (6.14). Таким образом, соответствующий положительный собственный вектор обладает весьма большими компонентами по сравнению со всеми остальными собственными векторами; это соответствует селективному замещению чистого сорта при одновременном присутствии мутантов в малой концентрации (*Jones et al.*, 1976).

Другим частным случаем неразложимой матрицы служит простое кольцо (цикл). Рассмотрим в качестве примера матрицу

$$\mathbf{k} = \begin{bmatrix} 0 & 0 & 0 & 0 & k_{15} \\ k_{21} & 0 & 0 & 0 & 0 \\ 0 & k_{32} & 0 & 0 & 0 \\ 0 & 0 & k_{43} & 0 & 0 \\ 0 & 0 & 0 & k_{54} & 0 \end{bmatrix}$$

(см. рис. 6.8).

Для таких матриц $\{k_{ij}\}$ характеристический многочлен имеет вид

$$p^n = k_{n,n-1} k_{n-1,n-2} \cdots k_{32} k_{21} k_{1n}, \quad (6.77)$$

где n — длина цикла, поэтому максимальное собственное значение равно среднему геометрическому из перекрестно-каталитических коэффициентов (*Eigen*, 1973; *Feis-*

Производящий граф для $\mathbf{k}$:

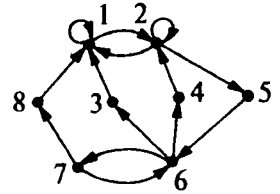


Рис. 6.7. Неразложимость каталитической сети

Производящий граф для $\mathbf{k}$:

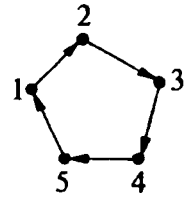


Рис. 6.8. Простой каталитический цикл

tel, 1976). Матрица k в математике называется импримитивной, или циклической, индекса n .

Компоненты соответствующих собственных векторов матрицы k удовлетворяют уравнениям

$$pX_i = k_{ij}X_j, \quad \frac{X_i}{X_j} = \frac{1}{p}k_{ij}, \quad (6.78)$$

где i и j — соседние индексы в цикле.

Таким образом, система с неразложимой матрицей k не обнаруживает селективного поведения в том смысле, что отдельные сорта не исчезают из системы.

Перейдем теперь к рассмотрению полностью разложимой матрицы k . Примером ее может служить следующая матрица:

$$k = \begin{bmatrix} 0 & 0 & 0 & k_{14} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ k_{21} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & k_{32} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & k_{43} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & k_{55} & k_{56} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & k_{65} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_{78} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_{8,12} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_{98} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_{10,9} & 0 & k_{10,11} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_{11,12} \\ 0 & 0 & 0 & 0 & 0 & 0 & k_{12,7} & 0 & 0 & k_{12,10} & 0 & 0 \end{bmatrix}$$

(см. рис. 6.9).

Производящий граф состоит из нескольких неразложимых компонент. Сорта, принадлежащие одной такой компоненте, образуют один кластер; анализируя неразложимые матрицы, мы заключаем, что сорта, входящие в один кластер, могут существовать или исчезать только все вместе.

Максимальному собственному значению полностью разложимой матрицы соответствует положительный собственный вектор, если максимальные собственные значения кластеров равны; в противном случае для каждого кластера существует собственный вектор, компоненты которого для индексов кластера положительны, а для всех остальных индексов равны нулю. Это означает, что кластеры сортов с различными максимальными собственными значениями не могут сосуществовать стационарно.

Если существует n кластеров, то из всех n возможных стационарных состояний устойчивы те, которые соответствуют наибольшему собственному значению $1/A$ по формуле (6.76). Ситуация полностью аналогична той, которая возникает в случае конкуренции отдельных автокаталитических сортов; единственное отличие состоит в том, что теперь в качестве наименьших единиц выступают кластеры. И снова стационарная обратная концентрация исходного вещества оказывается величиной, для которой выполняется статический принцип эволюции, а ее зависимость от параметров для различных стационарных состояний неявно задается положительными

Производящий граф для k :

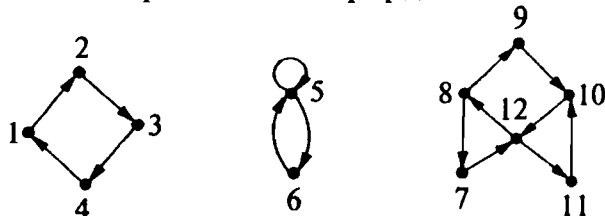


Рис. 6.9. Полностью разложимая каталитическая система

решениями характеристического уравнения

$$\det \left(\mathbf{k}'^{-1} \mathbf{k} - \frac{1}{A} \mathbf{I} \right) = 0, \quad (6.79)$$

где $\mathbf{I}$ — единичная матрица.

Соответствие между отдельными автокаталитическими сортами и кластерами не совсем случайно, оно выражает возможность образования иерархических структур. В статистической физике предметом изучения являются элементарные частицы, но соответствующие уравнения показывают, что так называемые связанные состояния частиц можно рассматривать как элементарные частицы (*Ebeling, 1974; Эбелинг и др., 1979*). Рассмотрим пример: водородная плазма состоит только из электронов и протонов, но возникающие при более низких температурах связанные состояния H и H_2 также можно рассматривать как самостоятельные частицы с внутренними степенями свободы. В статистической физике в этом случае говорят о переходе от «физической картины» к «химической картине».

Аналогичным образом обстоит дело и в случае конкуренции кластеров сортов. Если считать, что сорта X_i соответствуют молекулам, то кластеры можно рассматривать как комплексы функций молекул, т. е. как простейшие «организмы». Мы попытаемся поэтому совершить переход к «экологической картине», в которой между собой взаимодействуют уже не комплексы молекул, а новые виды с внутренними степенями свободы (*Feistel, Ebeling, 1978, 1989*).

Запишем в качестве «вида» l

$$S_l = \sum_{\text{кластер } l} X_i, \quad (6.80)$$

а в качестве его «внутренней структуры» введем величину

$$\rho_{il} = \begin{cases} \frac{X_i}{S_l}, & \text{если } i \text{ принадлежит кластеру } l, \\ 0 & \text{в противном случае,} \end{cases} \quad (6.81)$$

так что $\sum_i \rho_{il} = 1$ для всех видов l . Определим теперь среднее по кластеру:

$$\langle k_i \rangle_l = \sum_j k_{ij} \rho_{jl} \quad \text{и} \quad \langle k'_i \rangle_l = \sum_i k'_i \rho_{il}. \quad (6.82)$$

Уравнения (6.72) переходят в иерархическую систему уравнений

$$\begin{aligned} \dot{\rho}_{il} &= A \left[\langle k_i \rangle_l - \sum_j \langle k_j \rangle_l \rho_{il} \right] - (k'_i - \langle k'_i \rangle_l) \rho_{il}, \\ \dot{S}_l &= \left[\sum_i \langle k_i \rangle_l \cdot A - \langle k'_i \rangle_l \right] S_l, \\ \dot{A} &= \Phi - A \sum_l \sum_i \langle k_i \rangle_l S_l, \end{aligned} \quad (6.83)$$

если матрица $\mathbf{k}$ вполне разложима. Для упрощения предположим, что скорости распада k'_i сортов, принадлежащих одному кластеру, различаются несущественно, то есть

$$k'_i \approx \langle k'_i \rangle_l$$

для всех i из кластера l .

Таким образом, «уравнение внутренней структуры» (6.83) сводится к уравнению

$$\dot{\rho}_{il} = A \left[\langle k_i \rangle_l - \sum_j \langle k_j \rangle_l \rho_{il} \right]. \quad (6.84)$$

Важно, что уравнение (6.84) не зависит от l и только через A зависит от отбора вида S_l . В частности, A в уравнении (6.84) влияет не на положение, а лишь на скорость установления внутреннего равновесия.

Теорема Тихонова утверждает, что решения дифференциального уравнения сходятся друг к другу независимо от того, принята или отвергнута гипотеза о псевдостационарном состоянии, если постоянные времени уравнений, рассматриваемых как стационарные, достаточно велики по сравнению с постоянными времени остальных уравнений. Предположим, например, что A и S_l очень велики, но одного порядка. Тогда ρ_{il} стремится к стационарному состоянию гораздо быстрее, чем S_l . Это означает, что мы можем изучать отбор кластеров, оставаясь в рамках кластерного приближения и рассматривая структуру кластеров как равновесную.

Из уравнения (6.84) с учетом равенства $\dot{\rho}_{il} = 0$ следует, что распределение сортов в кластере l удовлетворяет уравнению на собственные значения

$$\sum_j k_{ij} \rho_{jl} = k_i^* \rho_{il}, \quad k_i^* = \sum_j \langle k_j \rangle_l. \quad (6.85)$$

Следовательно, ρ_{il} есть собственный вектор подматрицы матрицы $\mathbf{k}$, соответствующей кластеру l . Разумеется, все компоненты ρ в кластере положительны, так как, по определению, соответствующая подматрица матрицы $\mathbf{k}$ неразложима. Собственный вектор нормирован на 1. Собственное значение k_i^* положительно и однозначно определено по теореме Фробениуса.

Уравнения принимают теперь следующий вид:

$$\begin{aligned} \dot{S}_l &= (k_i^* A - k'_l) S_l, \\ \dot{A} &= \Phi - A \sum_l k_i^* S_l, \end{aligned} \quad (6.86)$$

где мы положили $k'_l = \langle k' \rangle_l$. Эти уравнения в точности совпадают с уравнениями (6.12), которые были рассмотрены в разделе о простой конкуренции. Из этого мы можем заключить, что «экологическая картина» допускает аналогичное рассмотрение кластеров сортов (видов) и автокаталитических веществ.

Проиллюстрируем изложенный выше метод на численном решении системы дифференциальных уравнений (6.72). Выберем $\Phi = 2$, в качестве $\mathbf{k}'$ — единичную матрицу и в качестве $\mathbf{k}$ — полностью разложимую матрицу

$$\mathbf{k} = \begin{bmatrix} 0 & 3/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 0 \\ 4/3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/8 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

(см. рис. 6.10).

Имеются два вида: $S_1 = X_1 + X_2 + X_3$ и $S_2 = X_4 + X_5 + X_6$. Задача об определении внутренней структуры (6.84) приводит к следующему результату:

$$X_1 : X_2 : X_3 = 3 : 2 : 4, \quad X_4 : X_5 : X_6 = 2 : 1 : 4;$$

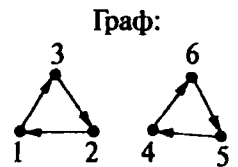


Рис. 6.10. Два несвязанных каталитических кольца

$$k_1^* = \sqrt[3]{\frac{4}{3} \frac{3}{2} \frac{1}{2}} = 1, \quad k_2^* = \sqrt[3]{1 \frac{1}{8} 1} = \frac{1}{2}.$$

В ходе отбора цикл S_2 с селекционной ценностью $1/2$ оказывается побежденным циклом S_1 с селекционной ценностью 1 ; при $t \rightarrow \infty$ получаем:

$$S_1 = \frac{\Phi}{k_1'} = 2, \quad S_2 = 0, \quad A = \frac{1}{k_1^*} = 1.$$

Теоретические результаты подтверждаются результатами компьютерных вычислений (рис. 6.11). Следует иметь в виду, что уже при $t = 4$ внутренняя структура становится стационарной, в то время как отбор длится до $t \approx 10$.

Обратимся теперь к следующему частному случаю разложимой матрицы — элементарной разложимой матрице. В качестве примера выберем матрицу

$$k = \begin{pmatrix} k_{11} & 0 & 0 & 0 \\ k_{21} & 0 & 0 & 0 \\ 0 & k_{32} & k_{33} & 0 \\ k_{41} & 0 & 0 & 0 \end{pmatrix}$$

(см. рис. 6.12).

В этом случае граф не обладает другими циклами, кроме петель, и сорта можно всегда перенумеровать так, что возникает нижняя треугольная матрица k :

$$k = \begin{bmatrix} k_{11} & 0 & 0 & \dots & 0 \\ k_{21} & k_{22} & 0 & \dots & 0 \\ k_{31} & k_{32} & k_{33} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ k_{n1} & k_{n2} & k_{n3} & \dots & k_{nn} \end{bmatrix}, \quad k_{ij} \geq 0.$$

Сорта с $k_{ii} > 0$ автокаталитические. Из уравнения (6.76) для концентрации исходного вещества следует возможное стационарное значение

$$\frac{1}{A} = \frac{k_{ii}}{k_i'}, \quad (6.87)$$

так как собственные значения треугольной матрицы совпадают с элементами, стоящими на главной диагонали. Анализ соответствующих собственных векторов несложен, но громоздок. Поэтому мы ограничимся лишь тем, что приведем результаты: в конкурентной борьбе выживает автокаталитический сорт s с наибольшим значением k_{ss}/k_s' . Вместе с ним выживают все сорта, следующие за сортом s в производящих графах, причем независимо от того, автокаталитические сорта или нет. Все другие сорта вымирают.

Теперь мы можем перейти к изучению наиболее общего случая разложимой матрицы k . В качестве примера выберем матрицу

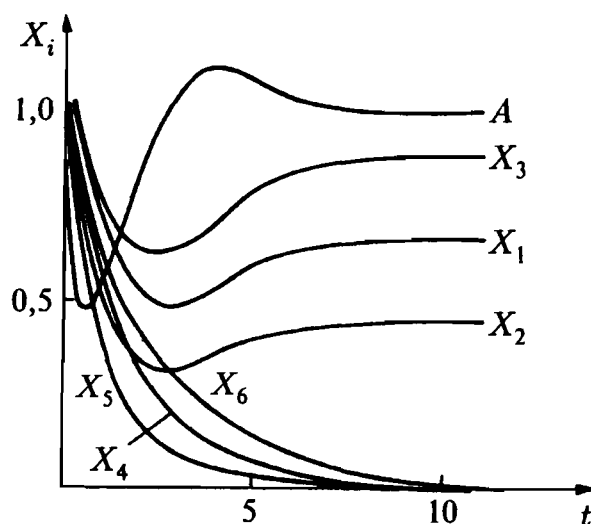


Рис. 6.11. Отбор между двумя каталитическими соединениями

Производящий граф для k :

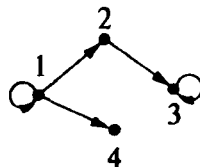
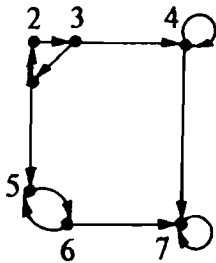


Рис. 6.12. Каталитический каскад (элементарный разложимый)

$$\mathbf{k} = \begin{bmatrix} 0 & 0 & k_{13} & 0 & 0 & 0 & 0 \\ k_{21} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & k_{32} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & k_{43} & k_{44} & 0 & 0 & 0 \\ k_{51} & 0 & 0 & 0 & 0 & k_{56} & 0 \\ 0 & 0 & 0 & 0 & k_{65} & 0 & 0 \\ 0 & 0 & 0 & k_{73} & 0 & k_{76} & k_{77} \end{bmatrix}$$

(см. рис. 6.13).

Производящий граф:



Граф разложения:

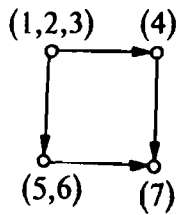


Рис. 6.13. Граф и уплотненный граф (разложения) разложимой системы

Преобразуем уравнения (6.83) в соответствии с экологической картиной к виду

$$\begin{aligned} \dot{\rho}_{il} &= \frac{A}{S_l} \sum_m S_m \left[\langle k_i \rangle_m - \sum_{j \in S_l} \langle k_j \rangle_m \rho_{il} \right] - \rho_{il} (k'_i - \langle k' \rangle_l), \\ \dot{S}_l &= A \sum_m \sum_{j \in S_l} \langle k_j \rangle_m S_m - \langle k' \rangle_l S_l, \\ \dot{A} &= \Phi - A \sum_l \sum_m \sum_{j \in S_l} \langle k_j \rangle_m S_m. \end{aligned} \quad (6.88)$$

Это представление обладает преимуществом только тогда, когда взаимодействие между кластерами не сводится к разнообразным взаимодействиям между отдельными сортами, образующими кластеры, а обусловлено действием одного кластера как целого на другой. В этом случае внутренняя структура каждого кластера не зависит от других кластеров, поэтому в уравнениях (6.88) все слагаемые с $m \neq l$ обращаются в нуль, если внутренние структуры кластеров m и l стационарны, то есть являются собственными векторами соответствующих подматриц (мы по-прежнему предполагаем, что $k'_i \approx \langle k' \rangle_l$ при $i \in S_l$). Это позволяет нам проделывать с уравнениями (6.88) любые операции, которые мы проделали с уравнениями (6.84), и подставить вместо величин ρ_{il} в первое из уравнений (6.84) их стационарные значения. В результате мы получаем систему

$$\begin{aligned} \dot{S}_l &= A \sum_m k_{lm}^* S_m - k'_l S_l, \\ \dot{A} &= \Phi - A \sum_l \sum_m k_{lm}^* S_m, \end{aligned} \quad (6.89)$$

обладающую такой же структурой, как исходная система (6.72), только $\{k_{lm}^*\}$ — теперь всегда элементарная разложимая матрица, поскольку соответствует разложимому графу. Случай элементарной разложимой матрицы уже рассмотрен нами выше. Мы знаем, что при $t \rightarrow \infty$ в системе остается кластер сортов s , обладающий наибольшим собственным значением $1/A = k_{ss}^*/k'_s$, и все кластеры, следующие за ним в графах, возникающих при разложении исходного графа. Этот результат остается

в силе и без приведенных выше ограничений на взаимодействие между кластерами. К доказательству утверждения в общем случае лучше подходить, отправляясь от задачи на собственные значения (6.75) и (6.76).

В качестве примера мы рассмотрим систему реакций с производящим графом, изображенным на рис. 6.14. Численные значения коэффициентов матрицы k выбраны такими же, как в примере, соответствующем рис. 6.10. Соответственно, мы заменяем $k_{63} = 0$ на $k_{63} = 1/2$. Значения Φ и k' мы получаем из того же примера. Воздержимся от подробного обсуждения результатов, поскольку теоретическое объяснение следует из приведенных выше соображений (рис. 6.15).

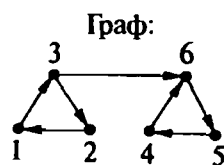


Рис. 6.14. Два связанных каталитических кольца

Рассмотрим теперь модель с неизменной организацией в целом. Для этого обобщим уравнения (6.21) на случай перекрестно каталитических процессов:

$$\dot{X}_i = \sum_j E_{ij} X_j - \langle E \rangle X_i, \quad (6.90)$$

где

$$\langle E \rangle = \frac{1}{C} \sum_{i,j} E_{ij} X_j.$$

В матричных обозначениях уравнения (6.90) записываются так:

$$\begin{aligned} \dot{X} &= \langle E - \langle E \rangle I \rangle X, \\ \langle E \rangle &= \frac{1}{C} I^T E X. \end{aligned} \quad (6.91)$$

Нетрудно видеть, что решение уравнения $\dot{X} = 0$ приводит к задаче на собственные значения для матрицы E с собственным вектором X и собственным значением $\langle E \rangle$. Однако поведение системы может быть исследовано и в явном виде с помощью аналитического решения, которое мы находим, полагая

$$X = Z \exp \left\{ - \int_0^t \langle E \rangle dt \right\} \quad (6.92)$$

(Thompson, McBride, 1974; Jones et al., 1976; Jones, 1977; Ebeling, Feistel, 1976, 1977; Feistel, Ebeling, 1976, 1978).

Подставляя выражение (6.92) в уравнение (6.91), получаем уравнение

$$\dot{Z} = EZ,$$

допускающее решение

$$Z(t) = T \exp \{Pt\} T^{-1} Z(0). \quad (6.93)$$

Здесь T — модальная матрица (матрица собственных векторов) матрицы E , а P — диагональная матрица собственных значений матрицы E :

$$ET = TP, \quad P = \{p_i \delta_{ij}\}. \quad (6.94)$$

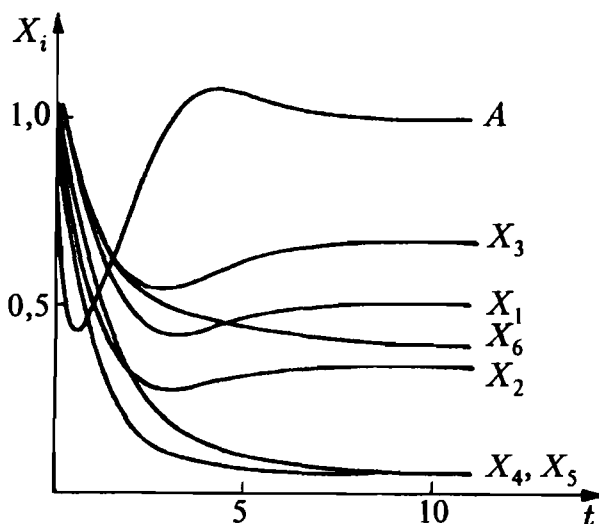


Рис. 6.15. Численное исследование реакционной системы с двумя связанными каталитическими соединениями

Используя соотношение

$$\begin{aligned} \int_0^t \langle E \rangle dt &= \int_0^t \frac{\mathbf{I}^T \mathbf{E} \mathbf{Z}}{\mathbf{I}^T \mathbf{X}} dt = \int_0^t \frac{\mathbf{I}^T \mathbf{E} \mathbf{Z}}{\mathbf{I}^T \mathbf{Z}} dt = \\ &= \int_0^t \frac{\mathbf{I}^T \dot{\mathbf{Z}}}{\mathbf{I}^T \mathbf{Z}} dt = \ln(\mathbf{I}^T \mathbf{Z}) - \ln C, \end{aligned} \quad (6.95)$$

получаем из формулы (6.92) решение для $\mathbf{X}(t)$:

$$\mathbf{X}(t) = C \frac{\mathbf{T} \exp \{ \mathbf{P} t \} \mathbf{T}^{-1} \mathbf{X}(0)}{\mathbf{I}^T \mathbf{T} \exp \{ \mathbf{P} t \} \mathbf{T}^{-1} \mathbf{X}(0)}. \quad (6.96)$$

При $t \rightarrow \infty$ выживают все сорта, у которых собственный вектор матрицы $\mathbf{E}$, соответствующий собственному вектору $\lambda_{\max}$ с наибольшей действительной частью, содержит ненулевые компоненты.

По определению, матрица $\mathbf{E}$ может содержать отрицательные элементы самое большее на главной диагонали. Пусть ε — наименьший элемент на главной диагонали. Тогда

$$\mathbf{E}' = \mathbf{E} - \varepsilon \mathbf{I}$$

— неотрицательная матрица. Если подставить решение (6.96) в уравнение (6.91), то это уравнение преобразуется в себя, т. е. матрицу $\mathbf{E}$ можно считать неотрицательной.

В полной аналогии с рассмотренной выше матрицей $\mathbf{k}'^{-1} \mathbf{k}$ или $\mathbf{k}$ мы можем теперь проанализировать свойства разложимости и спектральные свойства матрицы $\mathbf{E}$. Введение экологической картины для описания конкуренции между кластерами сортов в этой модели даже проще, чем при условии постоянного притока исходного вещества, поскольку отпадает связь через исходное вещество (*Feistel, Ebeling, 1978*).

Так как для стационарных состояний $\langle E \rangle$ всегда принимает значение, равное собственному значению матрицы $\mathbf{E}$, а для устойчивых состояний — максимальному собственному значению, $\langle E \rangle$ и в рассматриваемом случае порождает статический принцип эволюции:

$$\lim_{t \rightarrow \infty} \langle E \rangle = \lambda_{\max}. \quad (6.97)$$

Однако для перекрестно каталитической системы не выполняется динамический и тем самым полный экстремальный принцип, в чем нетрудно убедиться на примере:

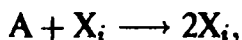
$$\mathbf{E} = \begin{pmatrix} 0 & \frac{1}{8} \\ 2 & 0 \end{pmatrix}, \quad X_1(0) = C, \quad X_2(0) = 0;$$

$$\langle E \rangle_{t=0} = 2, \quad \lim_{t \rightarrow \infty} \langle E \rangle = \frac{1}{2}.$$

Убывание $\langle E \rangle$ является следствием установления равновесия внутри кластеров, а отбор между кластерами, наоборот, всегда приводит к возрастанию $\langle E \rangle$.

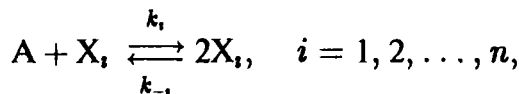
6.4. Сосуществование и гиперотбор

Уравнения реакций



рассмотренные в разделе о процессах простой конкуренции, с химической точки зрения представляют собой идеализацию, поскольку они не учитывают обратные

реакции. Поэтому сейчас мы переходим к вопросу о том, приводит ли учет обратной реакции



к существенным последствиям для поведения системы в ходе отбора. Для этого мы рассмотрим формально кинетические уравнения

$$\begin{aligned} \dot{A} &= \Phi - \sum_i k_i A X_i + \sum_i k_{-i} X_i^2, \\ \dot{X}_i &= k_i A X_i - k_{-i} X_i^2 - k'_i X_i \end{aligned} \quad (6.98)$$

при условии постоянного притока исходного вещества. Вследствие сильной нелинейности от модели (6.98) следует ожидать более сложного поведения, чем от рассмотренных ранее моделей. Прежде чем входить в математические детали, проанализируем несколько более подробно несколько частных случаев.

В качестве первого рассмотрим случай, когда концентрация исходного вещества поддерживается постоянной с помощью подходящего управления параметром Φ . В этих условиях уравнения для X_i при $k_{-i} > 0$ имеют вид

$$\dot{X}_i = k_{-i}(C_i - X_i)X_i, \quad (6.99)$$

где

$$C_i = \frac{k_i A - k'_i}{k_{-i}}. \quad (6.100)$$

Уравнение (6.99) есть не что иное, как логистический закон роста, который при малых X_i происходит экспоненциально, а при больших X_i выходит на насыщение (X_i асимптотически стремится к C_i — «емкости» сорта X_i).

Сказанное предполагает, что емкость C_i положительна — в противном случае X_i вымирает. Из соотношения (6.100) следует:

$$C_i > 0, \quad A > \frac{k'_i}{k_i} = A_i^{\text{крит}},$$

т. е. для каждого сорта существует определенная критическая концентрация исходного вещества $A_i^{\text{крит}}$, своего рода «минимум существования», величина которого по сравнению с истинным притоком исходного вещества A определяет существование или вымирание сорта i . Представляется разумным упорядочить сорта в порядке возрастания их минимумов существования:

$$\begin{aligned} A_1^{\text{крит}} &\leq A_2^{\text{крит}} \leq A_3^{\text{крит}} \leq \dots \leq A_n^{\text{крит}}, \\ \frac{k'_1}{k_1} &\leq \frac{k'_2}{k_2} \leq \frac{k'_3}{k_3} \leq \dots \leq \frac{k'_n}{k_n}. \end{aligned}$$

Соответственно, мы можем расположить упорядоченные сорта на числовой оси (рис. 6.16).

Если текущее значение подвода исходного вещества обозначить через A (см. рис. 6.16), то ясно, что все сорта слева от A выживают, а все сорта справа от A вымирают. Таким же остается поведение и в том случае, если мы не будем задавать значение A , а оно будет определяться самими сортами X_i ; это означает, что если некоторый сорт X_l выживает, то выживают также все сорта X_i с $i < l$.

Выясним теперь, где именно стабилизируется концентрация исходного вещества, сначала — при наличии одного сорта X_i . Нетрудно определить стационарное значение

$$\begin{aligned} A_i^{(1)} &= \frac{\Phi k_{-i} + k_i'^2}{k_i k_i'}, \\ X_i^{(1)} &= C_i^{(1)} = \frac{\Phi}{k_i'} \end{aligned} \quad (6.101)$$

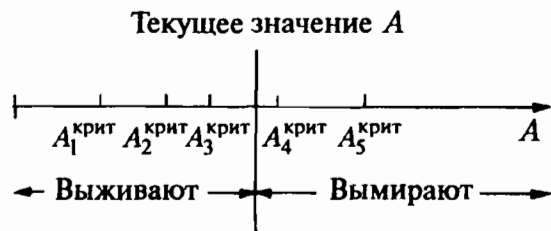


Рис. 6.16. Упорядочение сортов по их минимуму существования на числовой оси

и доказать, что оно устойчиво. Где расположено стационарное значение $A_i^{(1)}$ относительно $A_i^{\text{крит}}$? Разность этих значений составляет величину

$$\Delta A_i^{(1)} = A_i^{(1)} - A_i^{\text{крит}} = \frac{\Phi k_{-i}}{k_i k_i'} = \frac{k_{-i}}{k_i} C_i^{(1)} = \frac{k_{-i}}{k_i} X_i^{(1)},$$

т. е. следствие логистического роста (т. е. из-за обратной реакции) сорта X_i образуется резерв исходного вещества ΔA_i , пропорциональный стационарной концентрации $X_i^{(1)}$ (рис. 6.17). В случае простой конкуренции без учета обратной реакции ($k_{-i} = 0$) этот резерв всегда равен нулю, т. е. $A_i^{\text{крит}}$ и $A_i^{(1)}$ совпадают на числовой оси.

«Зазор» ΔA_i образует экологическую нишу для «худших» сортов A_j ($A_j^{\text{крит}} > A_i^{\text{крит}}$), если минимум существования лежит в следующем интервале (Ebeling, Feistel, 1976):

$$A_i^{\text{крит}} < A_j^{\text{крит}} < A_i^{(1)}.$$

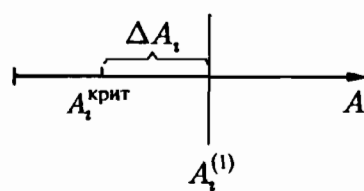


Рис. 6.17. Положение стационарной концентрации сырья $A^{(1)}$ относительно минимума существования $A_i^{\text{крит}}$

Переходим теперь к общему случаю n сортов. Ясно, что сосуществование, если оно вообще наступает, всегда затрагивает первые l сортов с самыми низкими минимумами существования. Число l неявно задано устанавливающейся стационарной концентрацией исходного вещества $A^{(1)}$ (Ebeling, Schelzer, 1980):

$$A_i^{\text{крит}} < A^{(1)} < A_{l+1}^{\text{крит}}. \quad (6.102)$$

Эта стационарная концентрация следует из уравнений (6.98), оказывается равной

$$A^{(1)} = \frac{\Phi + \sum_{i=1}^l \frac{k_i'^2}{k_{-i}}}{\sum_{i=1}^l \frac{k_i'}{k_i k_{-i}}}$$

и определяет стационарные значения отдельных сортов:

$$X_i^{(1)} = C_i^{(1)} = \frac{k_i A^{(1)} - k_i'}{k_{-i}} \quad \text{при } i = 1, \dots, l \quad (6.103)$$

и

$$X_i^{(1)} = 0 \quad \text{при } i = l + 1, \dots, n.$$

Для анализа устойчивости мы вычисляем матрицу Якоби

$$J = \left(\begin{array}{c|c} J_{ll} & J_{ln} \\ \hline 0 & J_{nn} \end{array} \right), \quad (6.104)$$

где

$$J_{ll} = \begin{bmatrix} -\sum_{i=1}^l k_i X_i^{(1)}; & k_{-1} X_1^{(1)} - k_1 A^{(1)}; & 2k_{-2} X_2^{(1)} - k_2 A^{(1)}; & \dots; & 2k_{-l} X_l^{(1)} - k_l A^{(1)} \\ k_1 X_1^{(1)}; & k_1 A^{(1)} - 2k_{-1} X_1^{(1)} - k'_1; & 0; & \dots; & 0 \\ k_2 X_2^{(1)}; & 0; & k_2 A^{(1)} - 2k_{-2} X_2^{(1)} - k'_2; & \dots; & 0 \\ \dots & \dots & \dots & \dots & \dots \\ k_l X_l^{(1)}; & 0; & 0; & \dots; & k_l A^{(1)} - 2k_{-l} X_l^{(1)} - k'_l \end{bmatrix}$$

и

$$J_{nn} = \{ (k_i A^{(1)} - k'_i) \delta_{ij}, \quad i, j = l+1, \dots, n \} \quad (6.105)$$

— квадратные блок-матрицы; блок-матрица J_{ln} для анализа устойчивости несущественна. Для устойчивости должны выполняться условия

$$\operatorname{Re}(p) < 0 \quad \text{для всех решений характеристического уравнения} \quad \det(J_{ll} - pI) = 0 \quad (6.106)$$

и

$$k_i A^{(1)} - k'_i < 0 \quad \text{при} \quad i = l+1, \dots, n. \quad (6.107)$$

В то время как условия (6.107) вследствие неравенств (6.102) выполняются всегда, прямая проверка условий (6.106) довольно затруднительна. Однако характеристический многочлен (6.106) может быть приведен к виду

$$\sum_{i=1}^l \left\{ \frac{k'_i + p}{k_{-i} X_i^{(1)} + p} k_i X_i^{(1)} \right\} + p = 0 \quad (6.108)$$

(Ebeling, Schmelzer, 1980).

Если записать $p = \lambda + i\omega$ и разложить уравнение (6.108) на уравнения для действительной и мнимой части, то непосредственно ясно, что в качестве решений могут получаться только отрицательные λ , поскольку в силу неравенств $X_i^{(1)''} > 0$ все коэффициенты многочлена положительны.

Резюмируя, мы можем сформулировать полученные результаты следующим образом. В качестве селекционной ценности i -го сорта выступают минимумы существования

$$A_i^{\text{крит}} = \frac{k'_i}{k_i},$$

задающие концентрацию исходного вещества, которая требуется для выживания. Устанавливаемая концентрация исходного вещества допускает одновременное существование l сортов с наименьшими значениями $A_i^{\text{крит}}$, где l определяется значениями Φ и всех параметров k_i, k'_i и k_{-i} l сортов. В каждом положительном подпространстве пространства концентраций $X_1, \dots, X_n$ находится самое большое одно стационарное состояние, которое всегда асимптотически устойчиво. В каждом m -мерном неотрицательном подпространстве находятся самое большое $2^m - 1$ стационарных состояний, из них ровно одно асимптотически устойчиво — то, которое соответствует наименьшей стационарной концентрации исходного вещества $A^{(1)}$.

Следовательно, величина $1/A^{(1)}$ порождает статический принцип эволюции в смысле определения, приведенного в разд. 6.1. С помощью этого принципа решение проблемы сосуществования может быть описано следующим образом.

Вычислим для всех 2^n возможных подмножеств сортов суммы

$$S_1 = \sum_i \frac{k'_i}{k_i k_{-i}}, \quad S_2 = \sum_i \frac{k_i'^2}{k_{-i}}. \quad (6.109)$$

Эти подмножества сортов образуют устойчивое состояние сосуществования. Величина $1/A^{(1)}$ принимает на соответствующих суммах свое наибольшее значение:

$$\frac{1}{A^{(1)}} = \frac{S_1}{\Phi + S_2}. \quad (6.110)$$

Разумеется, все эти утверждения легко могут быть модифицированы, если допустить к рассмотрению различные сорта с определенными равными значениями k_i или $A_i^{\text{крит}}$; однако имея в виду требование структурной устойчивости, можно предположить, что эти параметры для различных сортов никогда не бывают строго равными.

В заключение продемонстрируем возможность сосуществования на одном компьютерном примере (рис. 6.18). Как видно из графиков на рис. 6.18, с увеличением притока исходного вещества возрастает число сосуществующих сортов.

В другой работе (Ebeling, Schmelzer, 1980) проблема сосуществования была исследована и для следующей кинетики:

$$\begin{aligned} \dot{X}_i &= \frac{\mu_i A X_i}{k_i + A} - k'_i X_i - k_{-i} X_i^2, \quad i=1,2,\dots,n, \\ \dot{A} &= \Phi - \sum_{i=1}^n \frac{1}{Y_i} \frac{\mu_i A X_i}{k_i + A}. \end{aligned} \quad (6.111)$$

В этой экологически ориентированной модели Y_i^{-1} означает коэффициент прироста, служащий мерой того, какое количество исходного вещества требуется для увеличения концентрации X_i . Кинетика роста в данном случае — типа Михаэлиса—Ментен, т. е. при высокой плотности исходного вещества наступает насыщение. Кинетическое уравнение содержит член $k_{-i} X_i^2$, моделирующий уменьшение роста при высоких плотностях сорта i в случае перенаселенности, болезни и т. д. Необходимое и достаточное условие сосуществования l сортов в рассматриваемом случае имеет вид

$$\frac{k_l k'_l}{\mu_l - k'_l} < A^{(1)} < \frac{k_{l+1} k'_{l+1}}{\mu_{l+1} - k'_{l+1}}. \quad (6.112)$$

Отсюда ясно, что полученные в этом разделе результаты относятся не только к специальной модели, но и могут быть обобщены.

Еще одна возможность обобщения состоит в рассмотрении системы каталитических сортов с несколькими различными исходными веществами:

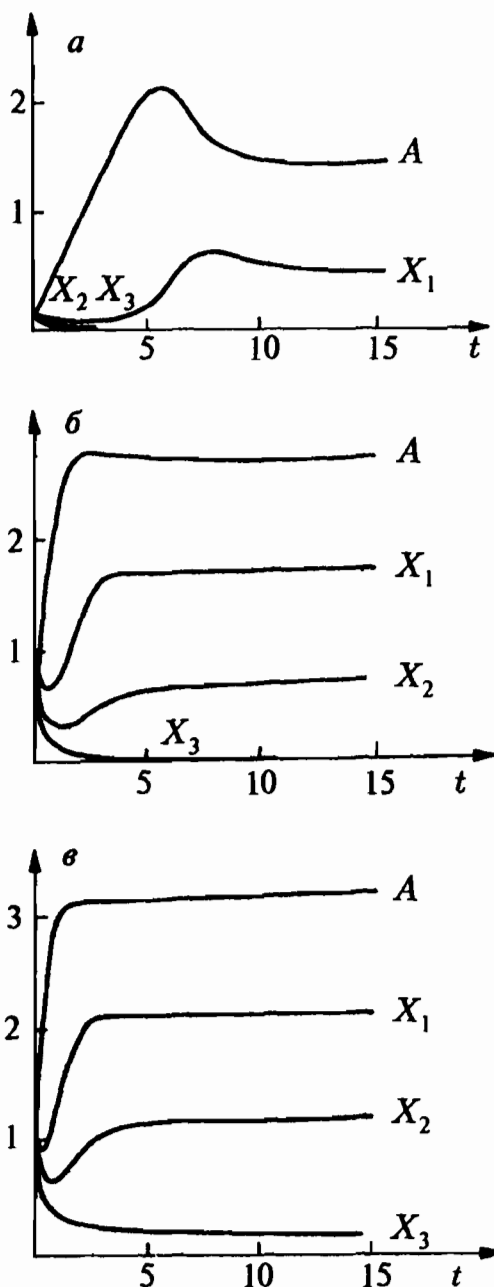
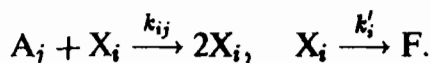


Рис. 6.18. Рост числа сосуществующих сортов в зависимости от подвода сырья: $k_i = k_{-i} = 1$; $k'_1 = 1$; $k'_2 = 2$; $k'_3 = 3$; (а) $\Phi = 0,5$; (б) $\Phi = 3,0$; (в) $\Phi = 5,0$

В модели с постоянным подводом исходных веществ соответствующие реакционно-кинетические уравнения имеют следующий вид:

$$\begin{aligned} \dot{A}_i &= \Phi_i - \sum_{j=1}^n X_j k_{ji} A_i \quad \text{при } i = 1, \dots, m, \\ \dot{X}_i &= \sum_{j=1}^m X_i k_{ij} A_j - k'_i X_i \quad \text{при } i = 1, \dots, n. \end{aligned} \quad (6.113)$$

Здесь $k_{ij} \geq 0$ — скорости, с которыми X_i воспроизводятся на основе исходных веществ A_j . Если требуется найти стационарные состояния системы, то должны выполняться равенства

$$X_i = 0 \quad \text{или} \quad \sum_{j=1}^m k_{ij} A_j = k'_i \quad \text{при } i = 1, \dots, n. \quad (6.114)$$

Если мы рассмотрим состояние, в котором концентрации s сортов X_i отличны от нуля и перенумеруем эти сорта последовательными числами от 1 до s , то все стационарные концентрации могут быть определены из уравнений

$$\begin{aligned} \sum_{j=1}^m k_{ij} A_j &= k'_i \quad \text{при } i = 1, \dots, s, \\ X_i &= 0 \quad \text{при } i = s + 1, \dots, n, \\ \sum_{j=1}^s X_j k_{ji} &= \frac{\Phi_i}{A_i} \quad \text{при } i = 1, \dots, m. \end{aligned} \quad (6.115)$$

Рассмотрим теперь, как расположено решение в пространстве концентраций исходных веществ A_i . Уравнения (6.115) определяют в этом пространстве $(m - s)$ -мерную гиперплоскость (рис. 6.19). Если исключить s величин X_j , то останется $(m - s)$ линейных однородных уравнений относительно $1/A_i$, определяющих s -мерный гиперboloид в пространстве концентраций A_i . Стационарному состоянию соответствует «точка пересечения» гиперплоскости и гиперboloида. Разумеется, это состояние имеет смысл только в том случае, если оно соответствует конечным положительным A_i и неотрицательным X_i .

Кроме того, ясно, что уравнения (6.115), вообще говоря, выполняются лишь при $s \leq m$. Это означает, что в системе, содержащей m различных исходных веществ, могут сосуществовать самое большее m различных автокаталитических сортов; существует самое большее m «экологических ниш».

Исследуем теперь устойчивость. Матрица Якоби, соответствующая уравнениям (6.115), является сверхматрицей, распадающейся на три блок-строки:

$$J = \begin{pmatrix} J_{ss} & J_{sn} & J_{sm} \\ J_{ns} & J_{nn} & J_{nm} \\ J_{ms} & J_{mn} & J_{mm} \end{pmatrix}. \quad (6.116)$$

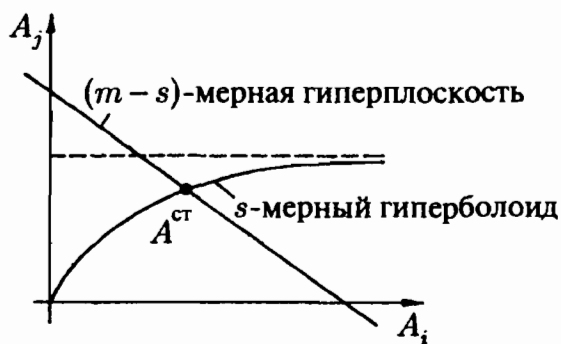


Рис. 6.19. Стационарное состояние существования s сортов в пространстве сырья

Здесь блок-матрицы J_{mm} и J_{nn} диагональные, а J_{nm} , J_{ns} и J_{ss} — нулевые матрицы. Если из характеристического уравнения для матрицы J выделить блоки с индексами s и m , то для первых $m + s$ собственных значений p характеристический многочлен имеет вид

$$\det(M) = 0,$$

где

$$M_{ij} = \frac{p^{s/m}}{A_i} \left(p + \frac{\Phi_i}{A_i} \right) \delta_{ij} + S_{ij} p^{s/m-1}. \quad (6.117)$$

Здесь

$$S_{ij} = \sum_{l=1}^s X_l k_{li} k_{lj}, \quad i, j = 1, \dots, m, \quad (6.118)$$

— неотрицательная симметричная матрица. Ее элементы S_{ij} служат мерой того, насколько связан расход исходных веществ A_i и A_j совместными пользователями X_l , $l = 1, \dots, n$. Для системы специализированных X_l матрица S диагональна. Нетрудно показать, что при $s < m$ матрица S вырождена и положительно полуопределена, а при $s = m$ линейно независимые векторы-строки матрицы k положительно определены. Зная это, можно спросить, может ли симметрическая, но не обязательно эрмитова матрица M для неотрицательных действительных частей собственных значений p иметь нулевое собственное значение. Применение к собственным значениям некоторых теорем о двусторонних оценках показывает, что уравнение (6.117) выполняется только для собственных значений p с отрицательными действительными частями.

Из блок-матрицы J_{nn} следуют последние $(n - s)$ условий устойчивости:

$$\sum_{j=1}^m k_{ij} A_j < k'_i \quad \text{при} \quad i = s + 1, \dots, n. \quad (6.119)$$

Смысл этих неравенств станет ясен, если рассмотреть их в пространстве сортов исходных веществ A_i , в котором мы проведем все n гиперплоскостей размерности $(m - 1)$

$$\sum_{j=1}^m k_{ij} A_j = k'_i, \quad i = 1, \dots, n, \quad (6.120)$$

(рис. 6.20).

Вокруг начала координат существует выпуклый полиэдр, который не пересекается ни с одной гиперплоскостью. На всех гиперплоскостях, в их точках пересечения, на линиях пересечения и т. д. могут лежать стационарные состояния, но устойчивые стационарные состояния располагаются только на поверхности выпуклого полиэдра. Это m -мерное обобщение требования минимизации концентрации исходного вещества, с которым мы неоднократно встречались в качестве условия устойчивости.

В частности, все вершины полиэдра с положительными координатами соответствуют состояниям сосуществования m сортов X_i , точки на $(m - s)$ -мерных ребрах

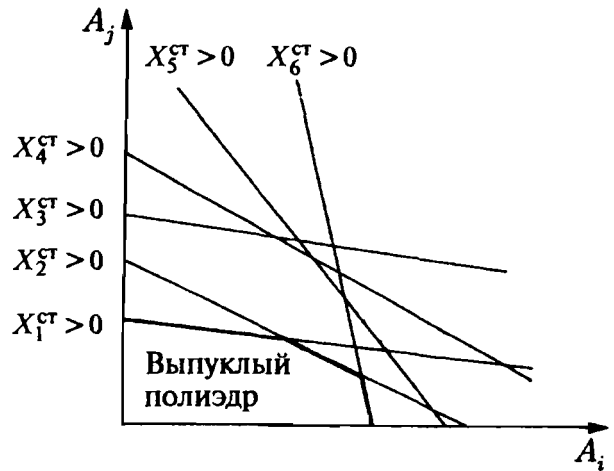


Рис. 6.20. Устойчивые состояния могут располагаться на поверхности выпуклого полиэдра

и на границах соответствуют дополнительным условиям сосуществования s сортов, которые все устойчивы. Существует

$$1 \leq N \leq \sum_{s=1}^m \binom{n}{s}$$

устойчивых стационарных состояний. В какие из них перейдет система, зависит от начальных условий. В этом и заключается упоминавшееся еще во введении свойство гиперотбора: результат процесса отбора зависит не только от того, какие сорта имеются в системе при $t = 0$, но и от концентрации сортов. Понять это можно, приняв во внимание следующие соображения. В простейшей модели отбора каждый сорт X_i имеет селекционную ценность k_i/k'_i ; побеждает сорт, имеющий наибольшую ценность. В рассматриваемой модели каждому сорту соответствует «вектор отбора»

$$\left(\frac{k_{i1}}{k'_i}, \frac{k_{i2}}{k'_i}, \dots, \frac{k_{im}}{k'_i} \right),$$

и отношение «больше» между такими векторами, вообще говоря, не определено. Если один вектор отбора «больше» другого, т. е. если все компоненты первого вектора больше соответствующих компонент второго вектора, то побеждает лучший сорт. Если же векторы не сравнимы, то в принципе оба сорта сосуществуют.

В реальных системах часто бывает так, что большие коэффициенты k_{ij} достижимы лишь в результате специализации. Что же касается неспециализированных сортов, то их коэффициенты невелики, хотя некоторые из коэффициентов пренебрежимо малы, но по величине уступают коэффициентам специализированных сортов (рис. 6.21). Как видно из схемы на рис. 6.22, при этих условиях специализированные сорта одерживают верх над неспециализированными, т. е. специализированные сорта, как правило, занимают все m ниш. Что же касается конкуренции специализированных видов, то она протекает в условиях простого отбора.

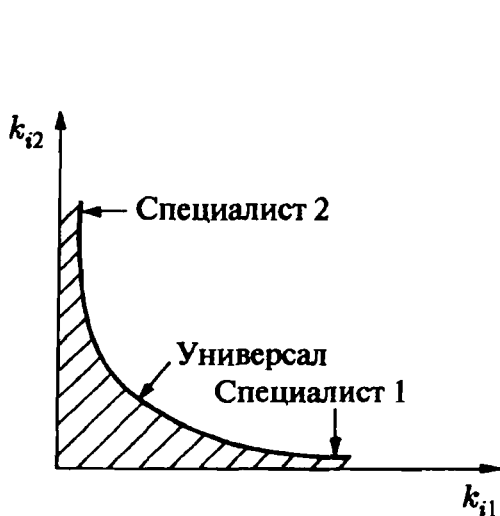


Рис. 6.21. «Множество приспособленности к внешним условиям» (множество реализуемых фенотипических свойств)

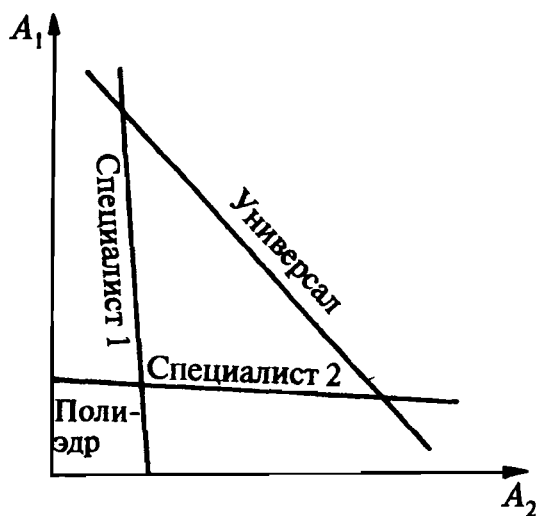


Рис. 6.22. В ходе отбора специалисты, как правило, превосходят универсалов

Несмотря на возникающий гиперотбор, изложенная выше модель способна эволюционировать, т. е. в ней принципиально существует возможность появления новых сортов, способных вырастать при минимальнейших концентрациях начальных сортов и вытеснять чистые сорта (см. рис. 6.23).

С принципиально иной ситуацией мы сталкиваемся в модели со свойствами гиперотбора:

$$A + \nu_i X_i \xrightarrow{k_i} (\nu_i + 1) X_i, \quad X_i \xrightarrow{k'_i} F.$$

При $\nu_i > 1$ возникает концентрационная волна для X_i , ниже порога волнообразования X_i всегда отбираются. Однажды появившийся сорт не может быть вытеснен новыми сортами с любыми коэффициентами k_i и меньшими начальными концентрациями. Мы имеем здесь дело с отбором типа «все или ничего», или «раз и навсегда». Так происходит и в случае уравнений конкуренции гиперциклов (Eigen, 1973; Эйген, Шустер, 1982), поэтому гиперциклы не допускают последовательного улучшения своих свойств с помощью механизма отбора мутаций (см. гл. 8).

Гиперотбор может возникать и в системе при обобщении процессов простой конкуренции на нелинейный перекрестный катализ (например, при производстве копий некоего эталона, или архетипа), как будет кратко показано в разд. 7.3 (см. также разд. 12.3). С другой стороны, гиперотбор в такой модели может реализоваться только в зависимости от численных значений кинетических коэффициентов. В частности, при особом выборе этих коэффициентов, как в схеме Лотки—Вольтерры, может реализоваться и простой отбор.

6.5. Отбор в сложных системах

В этом разделе мы совсем кратко остановимся на роли отбора в экологических и экономических системах. Отбор является фундаментальным механизмом изменения экологических систем; естественно, что в этой области теория процессов отбора развита весьма широко (см. Goel et al., 1971; May, 1973; Saunders, 1976; Свирижев, Логофет, 1978; Sonntag et al., 1981). Мы ограничимся рассмотрением так называемой схемы Лотки—Вольтерры (Peschel, Mende, 1983, 1986)

$$\frac{dN_i}{dt} = N_i \left(\varepsilon_i - \sum_{j=1}^n \gamma_{ij} N_j \right), \quad i = 1, 2, \dots, n. \quad (6.121)$$

Это — стандартная модель теоретической экологии. Здесь N_i — число особей i -го вида; коэффициенты ε_i описывают скорости размножения на автотрофной основе, а также естественную смертность:

$$\begin{array}{lll} \text{автотрофные виды:} & \varepsilon_i > 0, \quad \gamma_{ij} \geq 0 & \text{при } j = 1, \dots, n; \\ \text{гетеротрофные виды:} & \varepsilon_i < 0, \quad \gamma_{ij} < 0 & \text{при наименьшем } j. \end{array}$$

Величины γ_{ij} образуют так называемую матрицу популяции, или сообщества. Они описывают изменение популяции в результате парного взаимодействия особей двух видов (отношение «хищник—жертва», самопомехи и взаимные помехи, конкуренция).

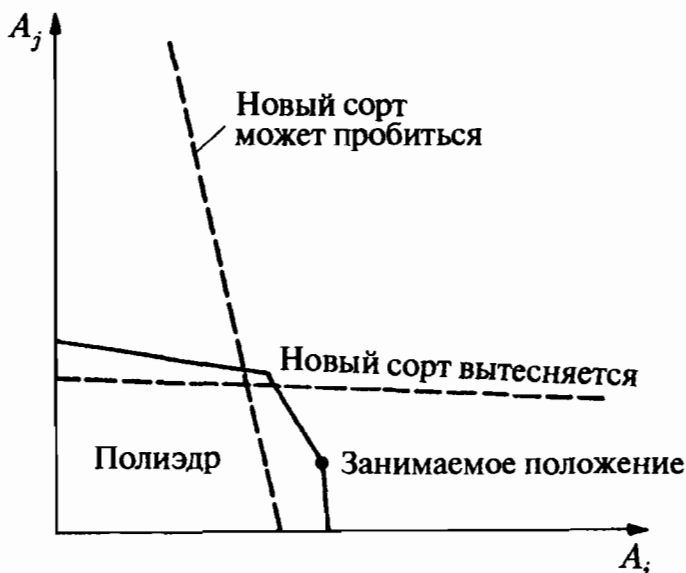


Рис. 6.23. Отбор происходит, когда соответствующая мутантам гиперплоскость отсекает предыдущее состояние полиэдра

Несмотря на простой вид уравнений (6.121), эта схема содержит множество более специальных моделей, в частности, модель простой конкуренции при постоянной организации в целом (разд. 6.2). Вместо модели (6.121) охотно используют также описание с помощью парциальных свободных энтальпий

$$g_i = \mu_i N_i,$$

где μ_i — химический потенциал сорта i (см. разд. 3.7). Для них справедливы уравнения

$$\frac{dg_i}{dt} = N_i \frac{d\mu_i}{dt} + \mu_i \frac{dN_i}{dt} = g_i \left(a_i - \sum_j b_{ij} g_j \right), \quad (6.122)$$

где

$$a_i = \left(1 + \frac{kT}{\mu_i} \right) \epsilon_i, \quad b_{ij} = \left(1 + \frac{kT}{\mu_i} \right) \frac{1}{\mu_j} \gamma_{ij}. \quad (6.123)$$

Поскольку, как правило, выполняются неравенства $kT \ll \mu_i$ (см. разд. 3.7), выражения в скобках можно считать равными 1. Экологи используют также вместо g_i «биомассы» видов, однако это понятие физически определено недостаточно строго и поэтому представляется несколько проблематичным. В дальнейшем мы используем также химические потенциалы и считаем (приближенно), что они не зависят от N_i . Так как парные взаимодействия не приводят к выигрышу свободной энергии, должны выполняться неравенства

$$g_i g_j (b_{ij} + b_{ji}) \geq 0,$$

т. е. симметрическая часть матрицы $\mathbf{b}$ является неотрицательной матрицей

$$\mathbf{b}^{\text{сим}} = \frac{1}{2}(\mathbf{b} + \mathbf{b}^T) \geq 0,$$

описывающей коллективную диссипацию популяции. Аналогично, антисимметрическая часть

$$\mathbf{b}^{\text{асим}} = \frac{1}{2}(\mathbf{b} - \mathbf{b}^T)$$

описывает обмен свободной энергией между видами и тем самым трофическую структуру экосистемы. Более точное утверждение сводится к следующему: если в матрице $(-\mathbf{b}^{\text{асим}})$ вычеркнуть все отрицательные элементы, то получающаяся в результате остаточная матрица сопряжена с ориентированным графом, характеризующим поток свободной энергии через экосистему и называемым «трофической структурой системы», или «структурой питания» (рис. 6.24).

Этот граф обычно не содержит циклов и задает полуупорядочение, допускающее классификацию по трофическим уровням. Самый нижний уровень (источники) образуют автотрофные растения (травы), следующий уровень занимают травоядные животные, а более высокие уровни — хищники (плотоядные животные) (Stugren, 1978). Стоки графа соответствуют так называемым высшим хищникам. Если трофическая структура не содержит циклов, то сорта всегда можно перенумеровать так, что в матрице $\mathbf{b}^{\text{асим}}$ под главной диагональю будут стоять неотрицательные,

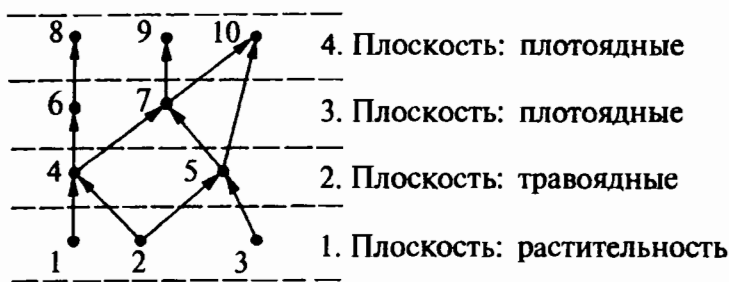


Рис. 6.24. Трофическая структура экосистемы

а выше главной диагонали — неположительные элементы. Таким образом, травы, травоядные и плотоядные могут быть сгруппированы в уровни с соседними индексами, и принадлежность к более высокому трофическому уровню соответствует большому значению индекса (рис. 6.24).

Относительно нашей системы мы предполагаем, что в ходе ее временной эволюции ни одна из энергий g_i не может обращаться в бесконечность. Например, в случае логистического роста это достигается за счет положительной определенности матрицы $\mathbf{b}$.

Если a_i и b_{ij} описывают схему Лотки—Вольтерры с конечным числом частиц каждого из n сортов, то всегда существует набор положительных чисел $x_1, \dots, x_s$, $s \leq n$ (при надлежащей нумерации n сортов), обладающих следующими свойствами:

$$\begin{aligned} \sum_{j=1}^s b_{ij} x_j &= a_i \quad \text{при } i = 1, \dots, s, \\ \sum_{j=1}^s b_{ij} x_j &> a_i \quad \text{при } i = s+1, \dots, n. \end{aligned} \quad (6.124)$$

Если $s = n$, т. е. если уравнения (6.122) допускают положительное стационарное состояние $g_i^0 > 0$, $i = 1, \dots, n$, и матрица $\mathbf{b}^{\text{сим}}$ не только не отрицательная, но и положительно определенная (что означает диагональную доминантность), то

$$L(g_i, g_i^0) = \sum_i \left\{ g_i - g_i^0 \left(\ln \frac{g_i}{g_i^0} + 1 \right) \right\} \quad (6.125)$$

— функция Ляпунова.

Если воспользоваться нашей терминологией (разд. 6.1), то речь идет о том, что функция

$$G(g_i, g_i^0) = - \sum_i \{ g_i - g_i^0 \ln g_i \} \quad (6.126)$$

порождает неполный динамический принцип сегрегации, так как эта функция определена только в положительных подпространствах ($g_i > 0$), причем только в таких подпространствах, которые содержат положительное стационарное состояние g_i^0 .

Можно показать, что функция (6.125) порождает не только принцип сегрегации, но и динамический принцип отбора. Для неотрицательного стационарного состояния с $g_i^0 = x_i$ при $i = 1, \dots, s$ и $g_i^0 = 0$ при $i = s+1, \dots, n$ функция G имеет вид

$$G(g_i, g_i^0) = - \sum_{i=1}^s \{ g_i - g_i^0 \ln g_i \} - \sum_{i=s+1}^n g_i.$$

Так как $g_i^0 = 0$ при $i = s+1, \dots, n$ и $\ln x \leq (x-1)$, всегда выполняется неравенство

$$G(g_i, g_i^0) - G(g_i^0, g_i^0) < - \sum_{j=s+1}^n g_j < 0 \quad \text{при } g_i \neq g_i^0.$$

Для производной по времени, используя положительную определенность матрицы $\mathbf{b}^{\text{сим}}$ и неравенства (6.124), получаем

$$\frac{dG}{dt} = \sum_{i,j=1}^n (g_i - g_i^0) b_{ij} (g_j - g_j^0) - \sum_{i=s+1}^n g_i \left(a_i - \sum_{j=1}^n b_{ij} g_j^0 \right) \geq 0. \quad (6.127)$$

Равенство в (6.127) выполняется только при $g_i = g_i^0$, $i = 1, \dots, n$. Из доказательства существования (6.124) следует, что функция G задает динамический экстремальный принцип, определенный во всех положительных подпространствах, а также на тех частях границы положительного конуса, которые соответствуют сортам, прошедшим отбор. Следовательно, мы имеем дело с динамическим принципом отбора, но не с принципом эволюции, так как на границах $g_i = 0$ с $g_i^0 > 0$ функция G не существует. Установленный принцип показывает, что не только существует по крайней мере решение типа (6.124), а при положительно определенной матрице $\mathbf{b}^{\text{сим}}$ существует, причем только одно, такое решение, соответствующее асимптотически устойчивому состоянию.

Чтобы найти такое состояние в заданной системе, необходимо решить систему уравнений

$$\sum_j b_{ij} x_j = a_i \quad (i, j \text{ принадлежат полупространству}) \quad (6.128)$$

и проверить свойства (6.124) в каждом из 2^n возможных подпространств. Если решение системы в любом подпространстве имеет вид вектора $\mathbf{x}$ только с отрицательными компонентами, то по теореме Штробека (*Strobeck*, 1973) соответствующие сорта всегда претерпевают отбор.

Подчеркнем следующее важное обстоятельство: доказательство существования решения (6.124) и единственности и устойчивости соответствующего состояния с G -функцией при положительно определенной матрице $\mathbf{b}^{\text{сим}}$ показывает, что такие экологические системы удовлетворяют простому отбору, т. е. ответ на вопрос о том, какие виды выживают при $t \rightarrow \infty$, зависит от того, какие виды были представлены при $t = 0$, но не зависит от чисел $N_i(0) > 0$.

Необходимые и достаточные условия для существования принципа эволюции в смысле дарвиновского «выживания наиболее приспособленного вида» в схеме Лотки—Вольтерры до сих пор неизвестны. Не подлежит сомнению, однако, что решающее значение имеет структура трофической системы.

Еще один важный вопрос — физический или экологический смысл функции L или G . Некоторые авторы обсуждают сходство функции Ляпунова (6.125) с энтропией смешения в идеальной системе из-за входящего в (6.125) логарифмического члена, однако такого рода интерпретация, по-видимому, носит чисто формальный характер. Представляет интерес предпринятая Кернером попытка вероятностной интерпретации функции L (*Kerner*, 1957). Рассматривались только так называемые консервативные системы, для которых матрица $\mathbf{b}^{\text{сим}}$ тождественно равна нулю, т. е. взаимодействие не сопровождается диссипацией. Если никакого процесса отбора не происходит, то L — сохраняющаяся величина, интеграл движения. Зная такие аддитивные сохраняющиеся величины, можно с помощью формализма Хинчина (*Khinchin*, 1943) построить (формальную) статистическую механику так, что, например,

$$P \sim e^{-\beta L} \quad (6.129)$$

— вероятность замещения состояния с «экологической» энергией L и «экологической температурой» $T \sim 1/\beta$. Однако интерпретация возникающих «термодинамических функций» оказалась столь трудной оттого, что условие $\mathbf{b}^{\text{сим}} = 0$ представляет собой частный случай структурной неустойчивости, обладающей рядом нереалистических особенностей (*May*, 1973). В общем случае L изменяется со временем, и вероятности замещения отнюдь не находятся в каком-либо равновесии, а приближенно определяются основными кинетическими уравнениями.

Мы принимаем во внимание следующие изменения чисел заполнения:

Автотрофное увеличение	$N_i \rightarrow N_i + 1$	Скорость $W = A_i N_i$,	
Гетеротрофное увеличение	$N_i \rightarrow N_i + 1$	Скорость $W = \sum_{j \neq i} B_{ij} N_i N_j$,	
Самоторможение	$N_i \rightarrow N_i - 1$	Скорость $W = C_{ii} N_i (N_i - 1)$,	(6.130)
Смерть добычи	$N_i \rightarrow N_i - 1$	Скорость $W = \sum_{j \neq i} C_{ij} N_i N_j$,	
Естественная смерть	$N_i \rightarrow N_i - 1$	Скорость $W = k_i N_i$.	

Основное кинетическое уравнение получается следующим:

$$\begin{aligned}
 \frac{\partial P}{\partial t} = & \sum_i A_i [(N_i - 1)P(N_i - 1) - N_i P(N_i)] + \\
 & + \sum_i k'_i [(N_i + 1)P(N_i + 1) - N_i P(N_i)] + \\
 & + \sum_i C_{ii} [(N_i + 1)N_i P(N_i + 1) - N_i(N_i - 1)P(N_i)] + \\
 & + \sum_i \sum_{j \neq i} C_{ij} N_j [(N_i + 1)P(N_i + 1, N_j) - N_i P(N_i, N_j)] + \\
 & + \sum_i \sum_{j \neq i} B_{ij} N_j [(N_i - 1)P(N_i - 1, N_j) + N_i P(N_i, N_j)].
 \end{aligned} \quad (6.131)$$

Тем самым мы получаем для производящей функции

$$\begin{aligned}
 G(s_1, \dots, s_n, t) &= \langle s_1^{N_1} \cdot s_2^{N_2} \cdot \dots \cdot s_n^{N_n} \rangle = \\
 &= \sum_{N_1} \dots \sum_{N_n} s_1^{N_1} \dots s_n^{N_n} P(N_1, \dots, N_n, t).
 \end{aligned} \quad (6.132)$$

уравнение

$$\begin{aligned}
 \frac{\partial G}{\partial t} = & \sum_i \left\{ (A_i s_i - k'_i)(s_i - 1) \frac{\partial G}{\partial s_i} + \right. \\
 & \left. + \sum_j s_j (B_{ij} s_i - C_{ij})(s_i - 1) \frac{\partial^2 G}{\partial s_i \partial s_j} \right\}.
 \end{aligned} \quad (6.133)$$

Как видно из представления функции G , функция G при малых $s \gtrsim 0$, $s \ll 1$ определяется исключительно вероятностями малых чисел частиц; вероятности больших чисел частиц начинают влиять на G лишь при $s = 1 - \epsilon$, $\epsilon \ll 1$. Соответственно, приближенные решения в этих двух интересных предельных случаях могут быть получены с помощью разложения по s (при малых s) и $(1 - s)$ (при больших s). Нас интересует функция G при больших числах частиц, поэтому мы линеаризуем G по $(1 - s)$:

$$\frac{\partial G}{\partial t} \approx \sum_i (s_i - 1) \left\{ \epsilon_i \frac{\partial G}{\partial s_i} - \sum_j \gamma_{ij} \frac{\partial^2 G}{\partial s_i \partial s_j} \right\}, \quad (6.134)$$

где

$$\epsilon_i = A_i - k'_i, \quad \gamma_{ij} = C_{ij} - B_{ij}.$$

Полагая

$$G(s_1, \dots, s_n; t) = \exp \left\{ \sum_i (s_i - 1) f_i(t) \right\}, \quad (6.135)$$

получаем

$$\frac{df_i}{dt} = f_i \left(\epsilon_i - \sum_j \gamma_{ij} f_j \right), \quad (6.136)$$

т. е. $f_i(t) = N_i(t)$ — аналитическое решение уравнений Лотки—Вольтерры (6.121). Выражение (6.135) дает достаточно общее решение уравнений (6.134), так как суперпозиция решений (6.135) при различных начальных значениях $f_i(0)$ позволяет удовлетворить всем разумным начальным условиям (Gardiner, Chaturvedi, 1977). Выражению (6.135) соответствует мультипуассоновское распределение

$$P(N_1^0, \dots, N_n^0; t) = \prod_{i=1}^n \frac{[f_i(t)]^{N_i^0} e^{-f_i(t)}}{N_i^0!}. \quad (6.138)$$

Последнее выражение получается при использовании формулы Стирлинга

$$P(N_1^0, \dots, N_n^0; t) = \exp \{ -L(f_i(t), N_i^0) \}. \quad (6.139)$$

Таким образом, уменьшение L вплоть до минимума означает возрастание вероятности найти систему в стационарном состоянии. Со стохастической точки зрения также ясно, почему функция L расходится на определенных частях границы положительного конуса: вероятность достичь состояние, лежащее внутри положительного конуса, из поглощающего состояния на границе равна нулю.

Помимо экологических и химических систем, при моделировании которых весьма часто используются близкие математические выражения, конкурентное поведение встречается и в системах совсем другой природы, например, в лазере, где наблюдается конкуренция мод (Хакен, 1980), или в экономике, где конкуренция между производителями при капитализме играет главную роль — двигателя развития новых технологий и регулятора пропорциональности между различными областями производства. Маркс охарактеризовал эти действия как «стимулирующую и регулирующую функцию закона стоимости». Марксова теория стоимости позволяет установить аналогию между экономическими и химическими процессами отбора (Feistel, Ebeling, 1976; Ebeling, 1977; Feistel, 1977; Ebeling, 1978). Аналогичные математические выражения возникают при описании процессов общественного воспроизводства (Bruckner, 1978). Мы изложили в общих чертах основные идеи лишь первых из названных выше работ.

Множество производителей в условиях капитализма свободной конкуренции перенумеровано натуральными числами $i = 1, 2, 3, \dots$. Производитель i за время dt производит количество товара dx_i индивидуальной стоимостью $dW_i^* = \omega_i dx_i$. Индивидуальная стоимость на количество товара ω_i есть мера индивидуально необходимого рабочего времени, то есть $1/\omega_i$ есть мера производительности труда производителя i . В процессе общественного обмена один и тот же товар обладает стоимостью $dW_i = p dx_i$, где p — единая рыночная цена, выражающая общественно признанное необходимое рабочее время. Так как стоимость возникает в процессе труда, для процессов обмена выполняется закон сохранения

$$\sum_i dW_i = \sum_i dW_i^*. \quad (6.140)$$

Отсюда при скорости производства $y_i = \frac{dx_i}{dt}$ товара мы получаем уравнение для цены:

$$\sum_i \omega_i y_i = p \sum_i y_i, \quad p = \langle \omega \rangle = \frac{\sum_i \omega_i y_i}{\sum_i y_i}. \quad (6.141)$$

При продаже товара dx_i производитель получает в форме денег стоимость $\langle \omega \rangle dx_i = dW_i$; на производство такого же количества товара он ассигнует переменный капитал dv_i (заработную плату) и постоянный капитал dc_i (машины), в сумме дающие затраты κ_i :

$$dv_i + dc_i = \kappa_i dx_i. \quad (6.142)$$

Разность между вложенной и реализуемой стоимостью дает прибыль:

$$dm_i = (\langle \omega \rangle - \kappa_i) dx_i. \quad (6.143)$$

Эта прибыль может быть израсходована на расширение производства:

$$\frac{dy_i}{dt} \sim \frac{dm_i}{dt} = (\langle \omega \rangle - \kappa_i) y_i. \quad (6.144)$$

Если принять в качестве простейшей модели линейную зависимость

$$\frac{dy_i}{dt} = \alpha y_i (\langle \omega \rangle - \kappa_i), \quad (6.145)$$

то, как нетрудно видеть, выжить могут только производители с $\kappa_i < \langle \omega \rangle$. Но и те производители, которые оказываются в более благоприятных условиях по своим затратам, испытывают конкурентное давление, так как рыночная цена при положительной корреляции затраты—стоимость

$$\frac{d}{dt} \langle \omega \rangle = \alpha \langle \omega \rangle \langle \kappa \rangle - \alpha \langle \omega \kappa \rangle \leq 0 \quad (6.146)$$

постоянно падает. То, что процесс носит характер отбора, становится ясно, если исследовать долевое участие i -го производителя в рынке $\rho_i = y_i / \sum y_i$ (величины κ_i и ω_i считаются фиксированными):

$$\frac{d\rho_i}{dt} = \alpha \rho_i (\langle \kappa \rangle - \kappa_i). \quad (6.147)$$

Такое увеличение было подробно исследовано в разд. 6.2. Как было показано, в этой модели должен наблюдаться относительный успех производителей. Мы можем придать модели реалистичность, если учесть, что более высокая скорость производства y_i , как правило, приводит к падению удельных затрат

$$\frac{dx_i}{dy_i} < 0. \quad (6.148)$$

При таком предположении становится заметной новая качественная особенность: появляются нелинейные функции роста, приводящие к суперотбору и абсолютному успеху по истечению конечного периода и отбору типа «все или ничего».

Уравнение (6.145) может быть выведено из стохастической теории, исходящей из анализа поведения покупателя на рынке (Ebeling, 1978, 1979). Основная идея состоит в том, что вероятность перехода покупателя от производителя l к производителю k определяется следующим выражением:

$$W(N_k + 1, N_l - 1 | N_k, N_l) = \alpha d(\omega_k, \omega_l) N_l \left(\frac{N_k}{N} \right). \quad (6.149)$$

Коэффициент скорости d зависит от двух значений ω_k и ω_l и может быть выбран следующим образом:

$$\begin{aligned} d &= \gamma\omega_l + (1 - \gamma)(\omega_l - \omega_k) \quad \text{при} \quad \omega_l > \omega_k, \\ d &= \gamma\omega_l \quad \text{при} \quad \omega_l \leq \omega_k. \end{aligned} \tag{6.150}$$

Здесь γ — доля покупателей, не предпринимающих осознанного сравнения стоимости. Значение $\alpha(\omega_k - \kappa)$ этой доли может быть подставлено вместо истинного роста производства. Для соответствующего процесса завоевания новых покупателей мы предполагаем, что

$$W(N_k + 1 | N_k) = \alpha(\omega_k - \kappa)N_k.$$

Тем самым стохастический процесс полностью определен. Можно убедиться в том, что с такими вероятностями перехода в среднем мы приходим к детерминистическому поведению (6.145). Соответствующие модели с успехом применялись и в науковедении для описания подвижности ученых (*Bruckner et al.*, 1989).

Нельзя не отметить тот факт, что обе схематические модели, рассмотренные нами выше, являются сильным упрощением реальной ситуации, и поэтому их можно уподобить лишь «моделям идеального газа». Анализ технического и экономического развития в реальных условиях, например, в условиях социального рыночного хозяйства, требует существенно более сложной модели. Рассмотрение эволюционной теории подобных процессов можно найти, например, в работах Дози, Фримана, Нельсона, Силверберга и Соете (*Dosi et al.*, 1988).

Глава 7

Индивидуальное развитие как новая стратегия эволюции

Все, что рождается, умирает.

Квинтилиан

7.1. Роль онтогенеза в эволюции

Живые организмы, появившиеся 3–4 миллиарда лет назад в условиях Праземли, были еще весьма примитивны. Как было показано во второй главе, «протоклетка» была своего рода микрореактором, содержащим в растворе полинуклеотиды и полипептиды, а также другие вещества. Этот реактор был отделен от окружающей среды мембраной и одновременно связан с окружающей средой путем обмена водой и растворенными в ней веществами.

Протоклетки обладали способностью к самовоспроизводству с небольшими ошибками, конкурировали друг с другом и были подвержены давлению отбора. В силу уникальных физико-химических свойств полинуклеотидов протоклетки были наделены способностью к хранению и передаче информации, закодированной в последовательности оснований. Новые протоклетки возникали в результате деления клеток. Дочерние клетки после непродолжительной физико-химической фазы роста становились неотличимыми от материнских клеток по всем свойствам, дочерние клетки не претерпевали индивидуального развития и по крайней мере в принципе были бессмертны. Многие из ныне живущих простейших микроорганизмов, например, вирусы и бактерии, лишены индивидуального развития, они «рождаются в готовом виде» и живут столько, сколько позволяют условия окружающей среды. У более сложных одноклеточных организмов, например, у дрожжевых клеток, наблюдаются отчетливо выраженные жизненные циклы с фазами юности, зрелости, размножения и старости. С возникновением многоклеточных организмов индивидуальное развитие занимает все более важное место.

Наиболее ярким примером может служить человек. Между грудным ребенком и зрелым взрослым человеком имеются существенные качественные различия не только с точки зрения морфологии. Младенец и взрослый отличаются и по своей работоспособности, и по своим интеллектуальным возможностям.

На ранней фазе эволюции живого первые многоклеточные организмы появились лишь как результат случайного образования кластеров. По своим свойствам они еще несущественно отличались от одноклеточных организмов. Индивидуальное развитие как новая стратегия складывалось лишь постепенно, и последующий успех этой стратегии опирался на ее значительные эволюционные преимущества. Онтогенез, характеризуемый как процесс индивидуального развития, является большим достижением эволюции, типичным лишь для более поздних стадий филогенеза — исторического развития таксономических групп. Онтогенез — неперемнная предпосылка не только образования более сложных морфологических структур, но и процессов

обучения, на которых основывается работоспособность всех высших организмов. Поскольку ген вследствие ограниченной емкости его «памяти» может передавать только информацию, важную для выживания, основной поток информации осуществляется через процессы обучения. Возможность обучения является важным преимуществом при отборе (Klix, 1985). Процессы обучения открывают совершенно новые возможности обработки информации (Völz, 1983). Поскольку онтогенез занимает в эволюции центральное положение, его эволюционное моделирование имеет большое значение.

7.2. Дискретные модели циклов жизни

Важные свойства систем, элементы которых претерпевают индивидуальное развитие, могут быть изучены даже на простых дискретных моделях. В простейшем случае жизненный цикл состоит только из двух стадий — молодость и зрелость, которые мы обозначим номерами 1 и 2.

Рассмотрим в качестве примера популяцию клеток прокариотов, например, бактерий. Индивидуальное развитие клеток прокариотов можно разделить на две стадии.

1) Молодость клетки (*C*-период); на этой стадии, продолжающейся около 40 минут, делится ДНК клетки.

2) Зрелость клетки (*D*-период).

Развитие популяций эукариотов, например, дрожжевых клеток, может быть разделено на следующие фазы.

1) Молодость клетки (*G*-фаза): клетка сосредоточивается на синтезе белка.

2) Зрелость клетки: синтез ДНК (*S*-фаза); после различных дальнейших этапов, наконец, наступает

3) Митоз (*M*-фаза).

Если ограничиться двумя стадиями, то процесс моделируется следующими уравнениями (Романовский и др., 1984);

$$\begin{aligned} \frac{d}{dt} x(1, t) &= b(2)x(2, t) - d(1)x(1, t) - D(1)x(1, t) - \Phi x(1, t), \\ \frac{d}{dt} x(2, t) &= D(1)x(1, t) - d(2)x(2, t) - \Phi x(2, t). \end{aligned} \quad (7.1)$$

Здесь $b(2)$ характеризует размножение, а также порождение молодых клеток зрелыми клетками, $d(i)$ — смертность и D — переход молодость—зрелость.

Величина Φ означает разбавление потока для создания давления отбора. В простейшем случае неумирающих клеток имеем:

$$\begin{aligned} d(1) &= 0, \quad d(2) = \frac{1}{T_2}, \quad D(1) = \frac{1}{T_1}, \\ b(2) &= \frac{2}{T_2}, \end{aligned}$$

где T_i — продолжительность i -й фазы.

Скорости размножения могут в определенных пределах замедляться. Такое торможение регулируется производством чалонов — ингибиторов клеточного деления. Это приводит к зависимости скорости роста от концентрации (Романовский и др., 1984). Если ингибиторы выпускаются только на стадии зрелости, то скорость роста можно описать следующим выражением:

$$\frac{1}{T_2} = w_0 [1 + c_2 x(2)^n]^{-1}. \quad (7.2)$$

В этом случае в системе существует и колебательный режим. Если же ингибиторы производятся главным образом на стадии молодости клетки, то скорость роста описывается следующим выражением:

$$\frac{1}{T_2} = w_0 [1 + c_1 x(1)^n]^{-1}, \quad (7.2a)$$

где n — надлежащим образом выбранный показатель.

Конкуренция и отбор могут быть введены следующим образом. Рассмотрим систему, состоящую из различных сортов, в две стадии:

$$x_i(k, t), \quad i = 1, \dots, s, \quad k = 1, 2,$$

и введем конкуренцию, как в гл. 5 и 6, с помощью зависящего от времени разбавления потока:

$$\begin{aligned} \frac{d}{dt} x_i(1, t) &= b_i(2)x_i(2, t) - [d_i(1) + D(1) + \Phi(t)]x_i(1, t), \\ \frac{d}{dt} x_i(2, t) &= D(1)x_i(1, t) - [d_i(2) + \Phi(t)]x_i(2, t). \end{aligned} \quad (7.3)$$

При условии постоянного общего числа частиц получаем:

$$\Phi(t) = \frac{\sum_i [b_i(2) - d_i(2)]x_i(2, t) - \sum_i d_i(1)x_i(1, t)}{\sum_i x_i(1) + \sum_i x_i(2)}. \quad (7.4)$$

Вторая возможность состоит в том, чтобы создать давление конкуренции посредством связи через субстрат, т. е.

$$\begin{aligned} b_i(2) &= 2\mu_i(S), \\ d_i(2) &= d_i^0 + \mu_i(S). \end{aligned}$$

Здесь d_i^0 — естественная смертность зрелых клеток, $\mu_i(S)$ — скорость деления в зависимости от концентрации субстрата. Для описания расхода субстрата используется дополнительное уравнение:

$$\frac{d}{dt} S = D(S_0 - S) = \sum_i x_i(t)\mu_i(S)Y_i^{-1},$$

где Y_i — мера эффективности использования субстрата.

Более подробное обсуждение моделей с двумя и более стадиями индивидуального развития, а также их приложений можно найти в литературе (Полужков и др., 1980; Романовский и др., 1984; Feistel, Ebeling, 1989).

7.3. Модель старения Маккендрика—фон Фёрстера

Для математического описания индивидуального развития Маккендрик и фон Фёрстер разработали рекуррентную модель, в которой непрерывный процесс старения рассматривается как последовательность отдельных, изолированных, актов. С математической точки зрения речь идет о том, чтобы вывести систему дифференциальных уравнений для функции плотности $x(t, \tau)$, описывающей, сколько особей возраста τ существуют в системе в момент времени t . Рассмотрим такую систему

в момент времени $t + \Delta t$. Число особей, которые к этому времени достигнут возраста τ , равно числу особей, которые к моменту времени t достигли возраста $\tau - \Delta t$, за вычетом особей, умерших за интервал времени Δt . В результате мы получаем:

$$x(t + \Delta t, \tau) = x(t, \tau - \Delta t) - D(x, t, \tau)\Delta t, \quad (7.5)$$

где $D(x, t, \tau)$ — смертность. Совершая предельный переход при $\Delta t \rightarrow 0$, приходим к дифференциальному уравнению

$$\partial_t x(t, \tau) + \partial_\tau x(t, \tau) = -D(x, t, \tau). \quad (7.6)$$

Важную роль в динамике системы играют также процессы воспроизводства, характеризующиеся рождаемостью $B(x, t, \tau)$. Эта величина показывает, сколько потомков производят в момент времени t особи, достигшие возраста τ . В качестве начального условия по τ для уравнения (7.6) мы получаем величину

$$x(t, 0) = \int_0^\infty d\tau B(x, t, \tau). \quad (7.7)$$

Оно дополняется заданием второго начального условия

$$x(0, \tau) = \varphi(\tau). \quad (7.8)$$

Уравнения (7.5), (7.6) с начальными условиями (7.7), (7.8) задают модель Маккендрика—фон Фёрстера для популяций с возрастной структурой. Эта модель позволяет определить временную эволюцию функции плотности $x(t, \tau)$ по заданному начальному распределению $\varphi(\tau)$, если известны скорость воспроизведения и смертность. Решения, получаемые в рамках этой модели при различных конкретных функциях $D(x, t, \tau)$ и $B(x, t, \tau)$, подробно изучены (McKendrick, 1926; von Foerster, 1959; Полуэктов, 1974; Полуэктов и др., 1980; Романовский и др., 1984). Смертность D и рождаемость B , вообще говоря, сложным образом зависят от плотности $x(t, \tau)$ (см. соотношения (7.2) и (7.3)), что приводит к существованию разнообразных структур решений. Однако уравнения в этом случае также становятся весьма сложными, и их аналитическое решение становится невозможным. Поэтому мы сначала рассмотрим простейший случай, когда рождаемость и смертность линейны по плотности $x(t, \tau)$ и не зависят явно от времени:

$$\begin{aligned} D(x, t, \tau) &= d(\tau)x(t, \tau), \\ B(x, t, \tau) &= b(\tau)x(t, \tau). \end{aligned} \quad (7.9)$$

Реалистические примеры функций $d(\tau)$ и $b(\tau)$ представлены на рис. 7.1. Смертность по прошествии первого времени после рождения стабилизируется на относительно низком уровне и начинает снова возрастать лишь при больших τ ; в дальнейшем мы, кроме того, принимаем предположение

$$\int_0^\infty d\tau d(\tau) = \infty. \quad (7.10)$$

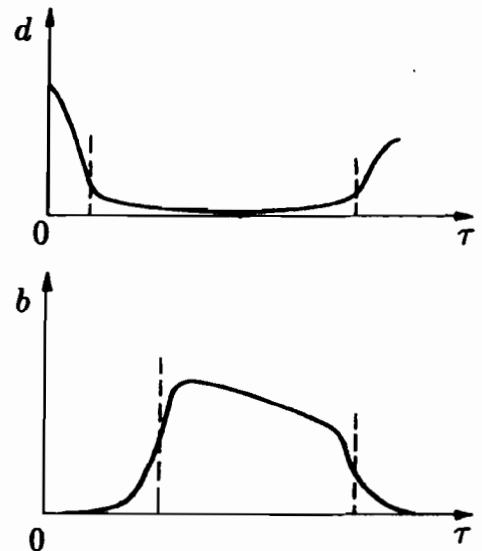


Рис. 7.1. Типичный ход рождаемости и смертности в линейной модели Маккендрика—фон Фёрстера

Рождаемость, как правило, достигает своего максимального значения лишь по истечении определенного периода после начала фазы зрелости и в старости снова понижается.

Учитывая соотношения (7.9), мы получаем из уравнений (7.6)–(7.8) систему уравнений

$$\begin{aligned}\partial_t x(t, \tau) + \partial_\tau x(t, \tau) &= -d(\tau)x(t, \tau), \\ x(t, 0) &= \int_0^\infty d\tau b(\tau)x(t, \tau), \\ x(0, \tau) &= \varphi(\tau).\end{aligned}\tag{7.11}$$

Поведение решений системы уравнений (7.11) вполне очевидно. Из первого уравнения (7.11) мы получаем, прежде всего, вероятность $W(\tau)$ того, что особь достигает возраста τ :

$$W(\tau) \sim \exp \left\{ - \int_0^\tau d\xi d(\xi) \right\}.$$

Следовательно,

$$K = \int_0^\infty d\tau b(\tau) \exp \left\{ - \int_0^\tau d\xi d(\xi) \right\} \tag{7.12}$$

(K — среднее число потомков, порожденных особью за всю ее жизнь). В зависимости от величины K возможны три качественно различных случая. При $K > 1$ число новорожденных в единицу времени больше числа умерших, доминируют процессы воспроизводства, и при $t \rightarrow \infty$ мы получаем при всех τ расходящуюся плотность $x(t, \tau) \rightarrow \infty$. Наоборот, при $K < 1$ воспроизводство слишком слабо, и при $t \rightarrow \infty$ мы получаем $x(t, \tau) \rightarrow 0$ при всех τ , т. е. вид вымирает. Наконец, при $K = 1$ оба процесса находятся в равновесии, соответственно, существует бесконечно много стационарных состояний, и только от начального условия $\varphi(\tau)$ зависит, какое из них реализуется. Разумеется, в случаях $K > 1$ и $K < 1$ результат не зависит от начального условия $\varphi(\tau)$.

Тем самым мы получаем качественную характеристику динамики индивидуального развития внутри отдельного вида при упрощающем предположении (7.9). Исследуем теперь, какие модификации возникают в том случае, когда n видов развиваются в соответствии с уравнениями, аналогичными уравнению (7.11), и, кроме того, взаимодействуют между собой посредством процесса отбора. Затронутая проблема связана с вопросом оптимальной стратегии старения, сложившимся в ходе эволюции.

7.4. Процессы отбора в моделях с непрерывным старением

Прежде всего систему n не взаимодействующих между собой видов можно описать уравнениями, обобщающими уравнение (7.11):

$$\begin{aligned}\partial_t x_i(t, \tau) + \partial_\tau x_i(t, \tau) &= -d_i(\tau)x_i(t, \tau), \\ x_i(t, 0) &= \int_0^\infty d\tau b_i(\tau)x_i(t, \tau), \\ x_i(0, \tau) &= \varphi_i(\tau), \quad i = 1, \dots, n.\end{aligned}\tag{7.13}$$

В качестве простого метода создания давления отбора мы по аналогии с моделью Эйгена (см. разд. 6.2) потребуем постоянства общего числа особей в системе:

$$\sum_i \int_0^{\infty} d\tau x_i(t, \tau) = N = \text{const}. \quad (7.14)$$

Чтобы условие (7.14) выполнялось, необходимо модифицировать систему уравнений 7.13, что можно осуществить различными способами. Особый интерес представляют две возможности.

1. *Введение потоковых членов в модель Эйгена.*

Такая операция соответствует подстановкам

$$\begin{aligned} d_i(\tau) &\rightarrow d_i(t) + \Phi(t), \\ b_i(\tau) &\rightarrow b_i(\tau) + \Psi(t) \end{aligned} \quad (7.15)$$

в уравнение (7.13), причем во избежание патологии, например, отрицательных концентраций, должны выполняться неравенства $\Phi(t) \geq 0$ и $\Psi(t) \geq 0$.

2. *Регуляция скорости воспроизведения.*

Регуляция достигается с помощью подстановки

$$b_i(\tau) \rightarrow A(t)b_i(\tau). \quad (7.16)$$

И в том, и в другом случае существенно, что модификации либо видо-, либо возрастоспецифичны.

Если равенство (7.14) продифференцировать по времени и воспользоваться уравнением (7.13) с подстановками (7.15), то получится следующее:

$$\Phi(t) - \Psi(t) = \langle b \rangle - \langle d \rangle, \quad (7.17)$$

где по определению

$$\langle z \rangle = \frac{1}{N} \sum_i \int_0^{\infty} z_i(\tau) x_i(t, \tau) d\tau. \quad (7.18)$$

Учитывая положительность $\Phi(t)$ и $\Psi(t)$, мы получаем

$$\begin{aligned} \Phi(t) &= \langle b \rangle, \\ \Psi(t) &= \langle d \rangle \end{aligned} \quad (7.19)$$

и, наконец, приходим к системе уравнений

$$\begin{aligned} \partial_t x_i(t, \tau) + \partial_\tau x_i(t, \tau) &= -(d_i(\tau) + \langle b \rangle) x_i(t, \tau), \\ x_i(t, 0) &= \int_0^{\infty} d\tau (b_i(\tau) + \langle d \rangle) x_i(t, \tau), \\ x_i(0, \tau) &= \varphi_i(\tau). \end{aligned} \quad (7.20)$$

Проводя аналогичные вычисления с использованием подстановок, получаем, полагая

$$A(t) = \frac{\langle d \rangle}{\langle b \rangle}, \quad (7.21)$$

систему уравнений

$$\partial_t x_i(t, \tau) + \partial_\tau x_i(t, \tau) = -d_i(\tau)x_i(t, \tau),$$

$$x_i(t, 0) = \frac{\langle d \rangle}{\langle b \rangle} \int_0^\infty d\tau b_i(\tau)x_i(t, \tau), \quad (7.22)$$

$$x_i(0, \tau) = \varphi_i(\tau).$$

Уравнения (7.20) и (7.22) описывают временную эволюцию систем стареющих конкурирующих между собой видов и тем самым удобны для математического анализа индивидуального развития и отбора.

В отличие от системы уравнений (7.13) для независимых видов дифференциальные уравнения (7.20) и (7.22) связаны между собой через определенное соотношение (7.18) среднее значение. С одной стороны, эта связь выступает как математическое выражение взаимодействия между видами, а с другой — исключает возможность получения аналитических решений и обуславливает тем самым весьма широкое применение численных методов.

Ряд интересных утверждений может быть высказан и без явного решения системы уравнений. В частности, необходимо выяснить, каким образом, зная функции $d_i(\tau)$ и $b_i(\tau)$, можно определить те виды, которые замещают другие и поэтому доминируют при больших временах.

Необходимый для этого качественный анализ динамики удастся осуществить с помощью подстановок

$$x_i(t, \tau) = n_i(t)\rho_i(t, \tau), \quad (7.23)$$

где

$$n_i(t) = \int_0^\infty x_i(t, \tau) d\tau \quad (7.24)$$

— общее число частиц,

$$\rho_i(t, \tau) = \frac{x_i(t, \tau)}{n_i(t)}$$

— нормированная возрастная структура i -го рода. Рассмотрим сначала ситуацию, описываемую уравнением (7.20). Пользуясь подстановкой (7.23), получаем следующие уравнения для $n_i(t)$ и $\rho_i(t, \tau)$:

$$\partial_t n_i(t) = \left[\int_0^\infty [b_i(\tau) - d_i(\tau)] \rho_i(t, \tau) d\tau - \langle b - d \rangle \right] n_i(t), \quad (7.25)$$

$$n_i(0) = \int_0^\infty \varphi_i(\tau) d\tau$$

и

$$\partial_t \rho_i(t, \tau) + \partial_\tau \rho_i(t, \tau) = - \left[d_i(\tau) + \int_0^\infty (b_i(\tau) - d_i(\tau)) \rho_i(t, \tau) d\tau + \langle d \rangle \right] \rho_i(t, \tau),$$

$$\rho_i(t, 0) = \int_0^\infty (b_i(\tau) + \langle d \rangle) \rho_i(t, \tau) d\tau, \quad (7.26)$$

$$\rho_i(0, \tau) = \frac{\varphi_i(\tau)}{\int_0^{\infty} \varphi_i(\tau) d\tau}.$$

Уравнение (7.25) имеет в точности такую же структуру, как уравнение Эйгена (6.21), с тем лишь различием, что теперь приспособленность

$$E_i = \int_0^{\infty} (b_i(\tau) - d_i(\tau)) \rho_i(t, \tau) d\tau \quad (7.27)$$

— функционал нормированной возрастной структуры $\rho_i(t, \tau)$ и поэтому может изменяться во времени.

Его временная эволюция определяется изменением во времени возрастной структуры $\rho_i(t, \tau)$, которая в свою очередь зависит от динамики чисел n_i через уравнение (7.26). Зависимость приспособленности нормированной возрастной структуры приводит к тому, что виды могут повысить свои шансы на успех в ходе отбора за счет подходящего распределения особей по возрастным группам; иначе говоря, в ходе эволюции происходит замещение одних видов другими с оптимальной возрастной структурой. Каким образом по заданным функциям $b_i(\tau)$ и $d_i(\tau)$ можно определить, какие виды выживут в конце концов? Чтобы ответить на этот вопрос, рассмотрим стационарные решения n_i^0 и $\rho_i^0(\tau)$ уравнений (7.25) и (7.26) при больших временах. Мы получаем

$$0 = [E_i^0 - \langle E^0 \rangle] n_i^0, \quad E_i^0 = E_i[\rho_i^0] \quad (7.28)$$

и

$$\begin{aligned} \partial_{\tau} \rho_i^0(\tau) &= -[d_i(\tau) + E_i^0 + \langle d \rangle^0] \rho_i^0(\tau), \\ \rho_i^0(0) &= \int_0^{\infty} [b_i(\tau) + \langle d \rangle^0] \rho_i^0(\tau) d\tau. \end{aligned} \quad (7.29)$$

При рассмотрении уравнения Эйгена мы обнаружили, что уравнение (7.28) допускает n различных стационарных решений вида

$$n_i = N \delta_{ij}, \quad j = 1, \dots, n, \quad (7.30)$$

т. е. стационарны только такие ситуации, в которых все N особей представлены одним видом. С учетом соотношения (7.30) из формулы (7.18) следует, что

$$\langle d \rangle^0 = \int_0^{\infty} d_j(\tau) \rho_j^0(\tau) d\tau,$$

поэтому уравнение для $\rho_j^0(\tau)$ замыкается (уравнения для $\rho_j^0(\tau)$ при $i \neq j$ не представляют интереса, так как $n_i = 0$, $i \neq j$). Таким образом, мы получаем:

$$\partial_{\tau} \rho_j^0(\tau) = - \left[d_j(\tau) + \int_0^{\infty} b_i(\tau) \rho_j^0(\tau) d\tau \right] \rho_j^0(\tau), \quad (7.31)$$

$$\rho_j^0(0) = \int_0^{\infty} [b_j(\tau) + d_j(\tau)] \rho_j^0(\tau) d\tau. \quad (7.32)$$

Вводя сокращенные обозначения

$$c_j = \int_0^{\infty} b_i(\tau) \rho_j^0(\tau) d\tau, \quad (7.33)$$

запишем уравнение (7.31) в виде

$$\rho_j^0(\tau) = \rho_j^0(0) \exp \left\{ - \int_0^{\tau} d\xi d_j(\xi) - c_j \tau \right\} \quad (7.34)$$

и далее, с помощью соотношения (7.32),

$$1 = \int_0^{\infty} [b_j(\tau) + d_j(\tau)] \exp \left\{ - \int_0^{\tau} d\xi d_j(\xi) - c_j \tau \right\} d\tau, \quad (7.35)$$

а также

$$\rho_j^0(0) = \left[\int_0^{\infty} \exp \left\{ - \int_0^{\tau} d\xi d_j(\xi) - c_j \tau \right\} d\tau \right]^{-1} \quad (7.36)$$

вследствие того, что по определению ρ_i

$$\int_0^{\infty} \rho_i^0(\tau) d\tau = 1.$$

Тем самым мы полностью охарактеризовали стационарные решения систем уравнений (7.25), (7.26). Можно показать (Ebeling, Engel, Mazenko, 1988), что при заданных $b_j(\tau)$ и $d_j(\tau)$ уравнение (7.35) всегда допускает ровно одно решение c_j . Величины c_j в силу соотношения (7.36) определяют $\rho_i^0(0)$, поэтому $\rho_i^0(\tau)$ однозначно определяются соотношением (7.34).

Из стационарных решений (7.30) устойчиво только одно, и при $t \rightarrow \infty$ именно оно описывает поведение системы. Для этого решения справедливы неравенства

$$E_j^0 > E_i^0 \quad \forall i \neq j,$$

т. е. выживает вид, обладающий наибольшей приспособленностью. Соотношения (7.34)–(7.36) позволяют определить этот вид посредством формулы

$$E_j^0 = \max_i E_i^0 = \max_i \int_0^{\infty} d\tau [b_i(\tau) - d_i(\tau)] \rho_i^0(\tau) \quad (7.37)$$

по известным функциям $d_i(\tau)$, $b_i(\tau)$, $i = 1, \dots, n$. Тем самым для системы конкурирующих видов с возрастной структурой, описываемой уравнением (7.20), становится возможным определять на основе заданных зависящих от возраста скорости воспроизведения и смертности тот из видов, который побеждает в ходе отбора. Используя соотношения (7.33)–(7.35), получаем

$$E_i^0 = \int_0^{\infty} d\tau [b_i(\tau) - d_i(\tau)] \rho_i^0(\tau) = 2c_i - \rho_i^0(0). \quad (7.38)$$

Можно показать (Ebeling, Engel, Mazenko, 1986), что

$$\frac{\partial E_i^0}{\partial \tau_i} > 0 \quad \text{при} \quad E_i^0 > 0.$$

Таким образом, если мы ограничимся видами с $E_i^0 > 0$, т.е. такими, которые в отсутствие отбора сами не вымирают (см. соотношения (7.27) и (7.20)), то максимум величины E_i^0 достигается для того вида j , для которого величина c_j^0 также максимальна. Следовательно, доминирующий вид может определяться вместо соотношения (7.36) соотношением

$$c_j = \max_i c_i, \quad (7.39)$$

где c_i определяется формулой (7.35). Аналогичные соображения применимы и к модели (7.22) (Эбелинг и др., 1986). Следует отметить, что в рассматриваемом случае соотношение (7.39) получается и как определяющее уравнение для выживающего вида. Если воспользоваться определением (7.33) параметров c_i , то соотношение (7.39) можно рассматривать как разумное обобщение соотношения (7.12).

В заключение продемонстрируем важность внутривидовой возрастной структуры для процесса отбора на простом примере динамики, описываемой уравнением (7.20). Рассмотрим два вида с одинаковой и постоянной смертностью

$$d_1(\tau) = d_2(\tau) = d \quad (7.40)$$

и рождаемостью

$$b_i(\tau) = \begin{cases} b_0 \pm \Delta, & \text{если } 0 \leq \tau < \frac{\tau_0}{2}, \\ b_0 \mp \Delta, & \text{если } \frac{\tau_0}{2} \leq \tau < \tau_0, \\ 0 & \text{в противном случае,} \end{cases} \quad (7.41)$$

где в (7.41) и в последующих соотношениях верхние знаки относятся к первому виду ($i = 1$), а нижние — ко второму виду ($i = 2$). Выбранные рождаемости представлены на рис. 7.2. Чтобы выяснить, какой вид побеждает в процессе отбора, необходимо исследовать, какая из определяемых соотношением (7.35) величин c_i больше. В рассматриваемом случае равенство (7.35) упрощается до

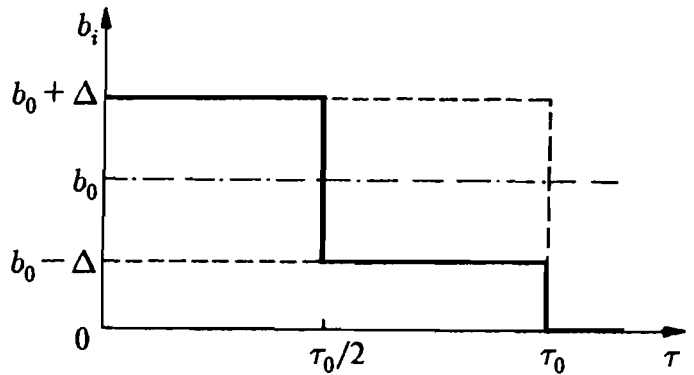


Рис. 7.2. Темп размножения в простом примере

$$1 = \int_0^{\infty} (b_i(\tau) + d) \exp \{-(d + c_i)\tau\} d\tau,$$

или

$$c_i = (c_i + d) \int_0^{\infty} b_i(\tau) \exp \{-(d + c_i)\tau\} d\tau.$$

Отсюда с учетом формулы (7.41) получаем

$$c_i = b_0 \left[1 - \exp \{-(c_i + d)\tau_0\} \right] \pm$$

$$\pm \Delta \left[1 + \exp \left\{ -(c_i + d)\tau_0 \right\} - 2 \exp \left\{ -(c_i + d) \frac{\tau_0}{2} \right\} \right]. \quad (7.42)$$

Подставляя $x_i = \exp \left\{ -(c_i + d)\tau_0/2 \right\}$, находим

$$-\frac{2}{\tau_0} \ln x_i - d + b_0(x_i^2 - 1) = \pm \Delta(1 - x_i^2). \quad (7.43)$$

Графическое решение этого уравнения представлено на рис. 7.3. Из хода кривой мы заключаем, что $x_1 < x_2$ вследствие монотонности подстановки (7.42) приводит к неравенству $c_1 > c_2$. Таким образом, первый вид выживает, а второй вымирает; грубо говоря, те преимущества, которыми второй вид обладает в старости, не перевешивают его недостатков в юности. Этот пример наглядно демонстрирует упрощенное описание процесса отбора с учетом возрастной структуры. В случае динамики Эйгена оба вида характеризовались бы усредненной приспособленностью

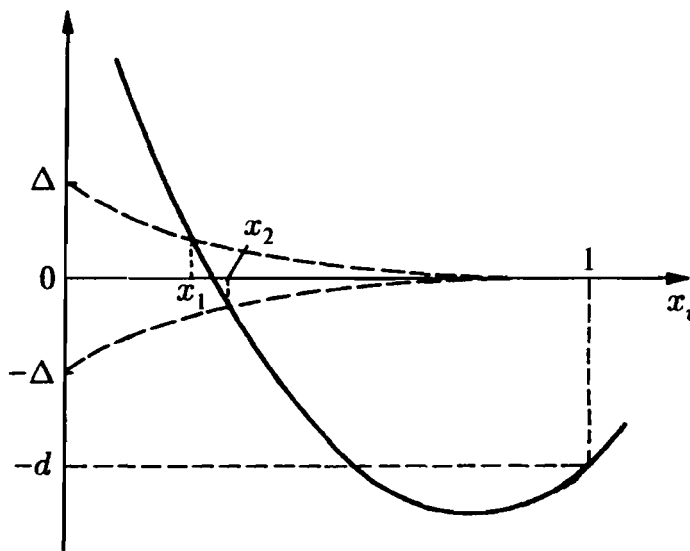


Рис. 7.3. Графическое решение уравнения (7.43)

$$E_1 = E_2 = (b_0 - d),$$

в результате чего никакого отбора не происходило бы, и оба вида могли бы сосуществовать.

Действительно, при одинаковых начальных условиях

$$x_1(0, \tau) = x_2(0, \tau) = \begin{cases} \frac{N}{2} & \text{при } \tau \leq \tau_0, \\ 0 & \text{в противном случае} \end{cases}$$

мы получаем из соотношения (7.27)

$$E_1(t = 0) = E_2(t = 0).$$

Однако динамика внутривидовой возрастной структуры $\rho_i(t, \tau)$ приводит к временному изменению приспособленности:

$$E_1(t \rightarrow \infty) > E_2(t \rightarrow \infty).$$

Тем самым даже в простейших моделях индивидуальное развитие внутри видов имеет решающее значение для исхода протекающих процессов отбора. Другие простые примеры для функций $b_i(\tau)$ и $d_i(\tau)$ (Эбелинг и др., 1986) показывают, что обычно высокая скорость воспроизведения и низкая смертность в сравнительно молодом возрасте является преимуществом в борьбе за отбор. Это обстоятельство тесно связано с тем, что стационарные возрастные структуры $\rho_i^0(\tau)$ монотонно затухают по τ .

7.5. Сложные возрастные структуры

В двух предыдущих разделах мы с помощью простых математических моделей, решения которых по крайней мере качественно могут быть охарактеризованы достаточно наглядно, показали важность индивидуального развития для процесса отбора.

Однако можно указать несколько случаев, когда более сложные и, следовательно, более реалистические возрастные структуры удастся описать с помощью модели Маккендрика—фон Фёрстера и ее обобщения. Мы не будем предпринимать попыток решить соответствующие уравнения, а ограничимся изложением возможностей, присущих формализму.

Прежде всего напомним наиболее общую форму (7.5)–(7.8) модели в случае одного отдельного вида:

$$\begin{aligned}\partial_t x(t, \tau) + \partial_\tau x(t, \tau) &= -D(x, t, \tau), \\ x(t, 0) &= \int_0^\infty B(x, t, \tau) d\tau, \\ x(0, \tau) &= \varphi(\tau).\end{aligned}\tag{7.44}$$

Множество самых различных решений определяется выбором функций $D(x, t, \tau)$ и $B(x, t, \tau)$. В дальнейшем мы обсудим лишь несколько принципиальных вариантов. Соображения, развиваемые ниже относительно функции $D(x, t, \tau)$, могут быть по аналогии перенесены на рождаемость $B(x, t, \tau)$.

Прежде всего мы предположим существование явной зависимости от времени. Это позволит учитывать изменения внешних условий. Колебания могли бы моделировать годовые или более длительные изменения, чередования теплых и холодных периодов, а скачкообразные изменения — влияние природных катаклизмов на экосистемы. Существуют и другие разнообразные условия, приводящие как к положительным, так и к отрицательным последствиям, но их вряд ли уместно классифицировать более подробно.

В дальнейшем мы всегда будем предполагать, что система всегда находится в стационарном окружающем поле. Явная зависимость от времени в этом случае не возникает, но изменение величины $D(x, \tau)$ в зависимости от плотности x может быть весьма разнообразным. Рассмотрим сначала простейший случай — зависимость от x , локальную по τ . Простейший вариант

$$D(x, \tau) = d(\tau) x(t, \tau)$$

рассмотрен в разд. 7.2. Несколько более реалистичным является учет ограничивающего члена в виде

$$D(x, \tau) = d_1(\tau) x(t, \tau) + d_2(\tau) x^2(t, \tau),\tag{7.45}$$

препятствующем для $K > 1$ (см. формулу ((7.12)) расходимости $x(t, \tau) \rightarrow \infty$ при $t \rightarrow \infty$. Если число особей очень велико, то смертность сильно возрастает и система стабилизируется на уровне, зависящем от функций $d_1(\tau)$ и $d_2(\tau)$. В $D(x, \tau)$ могут быть включены члены более высокого (по сравнению с формулой (7.45)) порядка. Включение кубической нелинейности может привести к возникновению бистабильных ситуаций, в которых число особей $x(t, \tau)$ может устанавливаться на двух, вообще говоря, не зависящих от τ устойчивых значениях (см. гл. 5).

Однако возможные режимы этим отнюдь не исчерпываются. Особенно большой интерес наряду с перечисленными выше вариантами нелокальных зависимостей представляют такие, которые в простейшей форме могут быть записаны следующим образом:

$$D(x, \tau) = d_1(\tau) x(t, \tau) + \int_0^\infty d\tau' d_2(\tau, \tau') x(t, \tau) x(t, \tau').\tag{7.46}$$

Ядро интеграла $d_2(\tau, \tau')$ описывает взаимодействие между особями самых различных возрастов внутри одного вида. При $d_2(\tau, \tau') < 0$ эффективная смертность особей, достигших возраста τ , понижается за счет наличия особей возраста τ' . Примером может служить, например, особенно тщательная забота о потомстве на протяжении первых отрезков жизни или поддержка старых, весьма опытных особей. Соответственно, $d_2(\tau, \tau') > 0$ приводит к внутривидовой конкуренции, проявляющейся в борьбе за верховенство или в «конфликтах поколений». Как и в соотношении (7.45), в формулу (7.46) могут входить члены высшего порядка, описывающие взаимодействие между тремя и более различными возрастными группами.

Введение давления отбора с помощью условия постоянства организации в целом (7.14) позволяет моделировать только простейшую форму межвидовой конкуренции. Взаимодействие хищник—жертва можно рассматривать как обобщение взаимодействия (7.46) вида

$$D_i(x, \tau) = d_i(\tau) x_i(t, \tau) + \sum_j \int_0^\infty d\tau' d_{ij}(\tau, \tau') x_i(t, \tau) x_j(t, \tau'), \quad (7.47)$$

учитывающее то обстоятельство, что довольно часто особи определенного возраста (например, молодняк животных) оказываются в роли жертв. Если считать, что все перечисленные выше возможности выбора функций $B(x, t, \tau)$ и $D(x, t, \tau)$ могут быть объединены в произвольные комбинации, то становится ясно, сколь разнообразные процессы эволюции могут быть описаны в рамках рассмотренной нами модели.

Глава 8

Мутации и спирали эволюции

Развитие и характер мира определяют не скалы, а мелкие камешки и песчинки.

Б. Травен

8.1. Частота ошибок при последовательной репликации

Процессы конкуренции и отбора, исследованию которых мы посвятили две предыдущие главы, имеют фундаментальное значение для процессов эволюции, но сами по себе они еще не приводят к эволюции. Естественно возникает вопрос: какие механизмы необходимо привести в действие для того, чтобы система, обладающая свойствами самовоспроизведения, мультистабильности, конкуренции и отбора, обрела способность к эволюции? Таким механизмом служат мутации, т. е. случайное производство альтернативных возможностей. Одна отдельная мутация, как правило, означает ухудшение скорее, чем улучшение системы в смысле эволюции. Но если случайно возникает новая благоприятная возможность и она усиливается, то эволюция делает шаг вперед. Такие благоприятные возможности называются инновациями. Механизм мутаций присущ самовоспроизводящимся системам. Причина заключается в физических законах квантовой механики и статистической термодинамики, лежащих в основе каждого процесса воспроизведения. Эти законы гарантируют возможность появления ошибки, первопричиной которой могут быть как квантово-механические, так и термодинамические флуктуации. Потенциальные возможности возникновения новых первичных структур при репродукции биополимеров практически неисчерпаемы. Число возможностей составить биополимер длины ν из λ различных сортов мономеров составляет

$$W = \lambda^\nu. \quad (8.1)$$

Число возможных первичных структур для полинуклеотида длиной 1000 составляет величину

$$W = 4^{1000} \approx 10^{600}. \quad (8.2)$$

Это — астрономически большое число. Некоторое представление о его величине дает следующее замечание: все водоемы Земли не могут вместить более 10^{40} молекул биополимера. Тем самым резервуар возможностей практически неисчерпаем.

Как показал Эйген (*Eigen*, 1973, 1976), мутации являются единственным источником новых структур и новой информации (*Eigen*, *Winkler*, 1975; *Эйген*, *Шустер*, 1982). Тейяр де Шарден был неправ, когда он писал об эволюции: «... Ничто в мире не может возникнуть внезапно после ряда эволюционных процессов, если его не было ранее в зародыше». Такая точка зрения приводит только к процессам конкуренции и отбора, представляющим собой чистое «развитие» имеющихся возможностей, но не к мутационным процессам. С помощью случайных процессов эволюция может порождать совершенно новые структуры: благодаря им эволюция обретает творческий и скачкообразный характер. Возникновение новой сущности всегда происходит скачком,

в некий дискретный момент времени (Эбелинг, 1979). Например, возникновение нового сорта действительно является «внезапным» скачком. Мутация, как правило, является результатом одного-единственного элементарного молекулярного процесса.

Различают микромутации (точечные мутации), затрагивающие лишь один или несколько нуклеотидов, и макромутации (макроповреждения), охватывающие значительные участки генома (Günther, 1978; Хессин, 1985). Примером точковых мутаций может служить замещение оснований (одного или нескольких оснований какими-то другими основаниями) и сдвиг раstra (вставка или выпадение нескольких нуклеотидов). Примером макроповреждений может служить дупликация (удвоение фрагмента последовательности нуклеотидов), делеция (утрата фрагмента последовательности), инверсия (обращение последовательности нуклеотидов в фрагменте) и инсерция (вставка чуждого гену фрагмента). Различают хромосомные и пloidные мутации (Günther, 1978). Типичным и в том и в другом случае является элементарный молекулярный характер мутации. Преобладающим действием мутаций являются отрицательные изменения в фенотипе, ведущие к вымиранию варианта. В относительно редких случаях, когда мутации ведут к улучшениям, этот микроскопический эффект претерпевает макроскопическое усиление в результате отбора новых вариантов. В этом заключается содержание фундаментального принципа усиления (Möglich et al., 1944; Патнер, 1983).



Рис. 8.1. Схема действия мутаций в процессе эволюции: принцип усиления

Взаимодействие мутаций и отбора в процессах эволюции разыгрывается по следующей общей схеме. Каждая мутация означает возмущение в установившемся на некоторое время равновесии (рис. 8.1). Мутант приводит к проверке системы на устойчивость относительно появления нового сорта. Если мутация не дает никаких преимуществ по сравнению с существовавшими ранее сортами, то новый сорт исчезает в результате процесса отбора. Система оказывается устойчивой относительно возмущения. Отбор ведет к уничтожению нового сорта, и система возвращается в исходное состояние. Но если оказывается, что мутант обладает определенным преимуществом по сравнению с первоначально имевшимися штаммовыми сортами, то отбор ведет к росту нового сорта. Система оказывается неустойчивой относительно возмущения. Усиливая эти возмущения (мутанты), система постепенно переходит в новое селекционное равновесие, соответствующее переходу на более высокий уровень эволюции (рис. 8.1). Следовательно, процесс отбора описывает замкнутые петли и открытые спирали, причем последние выводят систему на более высокий уровень эволюции (рис. 8.2). Эта картина напоминает нарисованную Гегелем принципиальную схему диалектического подхода к эволюции. Согласно гегелевской схеме, эволюцию надлежит воспринимать как процесс, развивающийся по восходящей линии, переход от простого к сложному, от низких свойств к высшим,

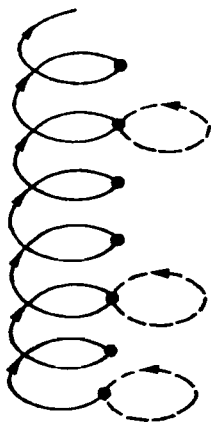


Рис. 8.2. Схема эволюционной спирали чередование замкнутых петель и незамкнутых дуг спиралей

как единство непрерывных и скачкообразных изменений, и путем повторения стадий процесса на более высоком уровне обнаруживается его спиралеобразная структура. Согласно гегелевской точке зрения, эволюция не означает автоматически развития по возрастающей, она содержит также элементы стагнации и развития вспять, но при этом отличительная особенность процессов эволюции заключается в возникновении новых фундаментальных свойств (инноваций) (Horz, 1974). Тем самым эволюция означает возникновение новых качеств по крайней мере как тенденции. Развита выше схема игры мутаций и отбора в ходе эволюционных процессов весьма удачно вписывается в общую концепцию гегелевской теории эволюции. Следует заметить, однако, что рассматриваемые в этой книге процессы протекают в основном на самом низком — молекулярном — уровне эволюции. Основным элементом служат биологические макромолекулы, которые, как пока-

зывают соотношения (8.1) и (8.2), могут образовывать невообразимо большое число изомерных форм молекул с различными первичными структурами.

Ныне мы знаем (и это можно доказать экспериментально), что 3–4 миллиарда лет назад в условиях Праземли спонтанно могло возникать огромное число биомолекул — полипептидов, полинуклеотидов, сахаров и жирных кислот. Однако вследствие невообразимо большого числа различных первичных структур практически исключено случайное образование биологически значимого множества первичных структур. Даже если бы вся масса Вселенной, составляющая около 10^{79} масс протона, превратилась бы в биомолекулы с молекулярным весом около 10^6 , то и тогда было бы испытано около 10^{73} вариантов. Это число до смешного мало в сравнении с числом возможных первичных структур, которое, как было показано, для одних лишь полинуклеотидов составляет величину около 10^{600} . В то время как остальные этапы биогенеза и в особенности период химической эволюции ныне известны достаточно досконально (см гл. 2), самоорганизацию биополимеров, т. е. возникновение биологически значимых структур, до сих пор удалось понять лишь отчасти. Именно этой проблеме посвящены фундаментальные работы Эйгена (Eigen, 1973). По Эйгену, самоорганизация биополимеров является результатом процессов эволюции в условиях Праземли. Гипотеза Эйгена служит основой физической интерпретации процессов, которые происходили в глубокой древности на Земле и привели к возникновению жизни. В этом гипотеза Эйгена намного превосходит различные другие гипотезы. Например, Вигнер, Йордан и Эльзассер, исходя из оценки чрезвычайно малой вероятности спонтанной самоорганизации биополимеров, пришли к заключению о том, что биогенез должен был содержать нефизические акты или, иначе говоря, что в рамках биогенеза должны были протекать процессы, прямо противоречащие законам физики. Разумеется, такое предположение противоречило бы развиваемому нами основному тезису, согласно которому законы физики никогда и нигде не нарушаются. Не может удовлетворить физика и выдвинутая Моно (Monod, 1971) гипотеза, согласно которой возникновение жизни есть однократное, совершенно невероятное случайное событие. Поэтому мы придерживаемся гипотезы Эйгена, которая, несмотря на все неясности, содержит «прозрачную» физическую концепцию возникновения жизни.

Гипотеза Эйгена исходит из предпосылки, согласно которой в «протосупе», образовавшемся 3–4 миллиарда лет назад в водах, омывавших Праземлю, образовались

условия для молекулярной эволюции. Как следует из соображений, изложенных нами в предыдущих главах, процессы эволюции могут протекать только в том случае, если выполняются следующие условия:

- способность к экспорту энтропии, например, с помощью постоянного притока свободной энергии;
- способность к самовоспроизведению структур;
- мультистабильность как предпосылка хранения информации и введения историчности;
- конкуренция вариантов и отбор благоприятных структур;
- мутации, приводящие к возникновению все новых и новых структур.

Условия на Праземле были проанализированы в гл. 2 (*Körner*, 1974, 1979; *Kaplan*, 1978; *Kämpfe*, 1980). Показано, что на Праземле действительно установились перечисленные выше физические условия (см. также разд. 2.4). Под воздействием ультрафиолетового излучения, высоких температур и электрических разрядов, выступавших в качестве источников свободной энергии, в восстановительной протопланетной атмосфере, содержащей водород, азот, водяной пар, метан, аммиак, двуокись углерода и т. д., могли образовываться аминокислоты, пиримидиновые и пуриновые основания, сахара, жирные кислоты и т. д. Следовательно, структурные компоненты биомолекул могли возникать спонтанно и в достаточных количествах. Показано также, что в этих условиях могла происходить полимеризация аминокислот в простейшие «заготовки» протеинов — так называемые протеноиды — и полимеризация нуклеотидов в более или менее длинные цепи (*Segal, Körner, Leiterer*, 1982). Под нуклеотидами мы понимаем основные структурные единицы нуклеиновых кислот, состоящих из пуриновых (аденина и гуанина) или пиримидиновых (цитозина, тимина и урацила) оснований, а также пентоз и фосфатных групп. Эти еще примитивные протеноиды и нуклеиновые кислоты образовывались преимущественно в водах, омывавших Праземлю. В свою очередь из них (главным образом из протеноидов) возникали компактные образования — так называемые коацерваты (*Опарин*, 1960; *Fox, Dose*, 1972; *Segal, Körner, Leiterer*, 1982). Решающее значение имело то, что протеноиды обладали простыми каталитическими свойствами, хотя те были выражены в миллионы раз слабее, чем у современных ферментов. При каталитической поддержке протеноидов синтез полинуклеотидов должен был происходить интенсивнее, причем уже имевшиеся цепочки нуклеотидов служили матрицами для синтеза последующих цепочек. Таким или близким к описанному путем протополимеры обрели способность к самовоспроизведению. В результате ошибок при репродуцировании возникали все новые и новые структуры, которые в результате естественного отбора вступали между собой в конкуренцию. Выживали только такие, которые лучше всего приспособлялись к окружающей природной среде, обладали способностью к быстрому и относительно стабильному воспроизведению. На рис. 8.3 представлена формальная схема замещения структур нуклеиновых кислот новыми, все более совершенными. К строгому изложению предложенной Эйгеном концепции ценности мы сможем приступить лишь в последующих разделах. Упомянем пока лишь некоторые качественные аспекты. Способность нуклеиновых кислот к хранению информации (способность, в основе которой лежит мультистабильность первичных структур нуклеиновых кислот), несомненно, играет важную роль при определении «ценности» конкретной первичной структуры для процесса эволюции. «Ценность» с этой точки зрения должна быстро возрастать с увеличением длины ν последовательностей и, следовательно, их информационной емкости $W = \lambda^\nu$. Более длинные нуклеиновые кислоты обладают более развитым аппаратом кодирования протеинов ферментативного характера, необходимых для их стабилизации и самовоспроизводства. Простейшим из существующих ныне бактерий требуется для обеспечения

своего существования, обмена и размножения около 10^3 ферментов; у высокоорганизованных организмов той же работой занято 10^4 – 10^5 ферментов, лишь малую долю которых удалось выделить и исследовать. Тенденция к возрастанию длины и тем самым информационной емкости нуклеиновых кислот противодействует спонтанно происходящим мутациям. Эволюция разворачивается в диалектическом поле напряжений противоречивых тенденций (*Fuchs-Kittowski*, 1976; *Löther*, 1972; *Hörz, Röseberg*, 1981; *Hörz, Wessel*, 1985). Однако роль мутаций в игре эволюции имеет фундаментальное значение: не будь мутаций, эволюция оставалась бы на достигнутом уровне.

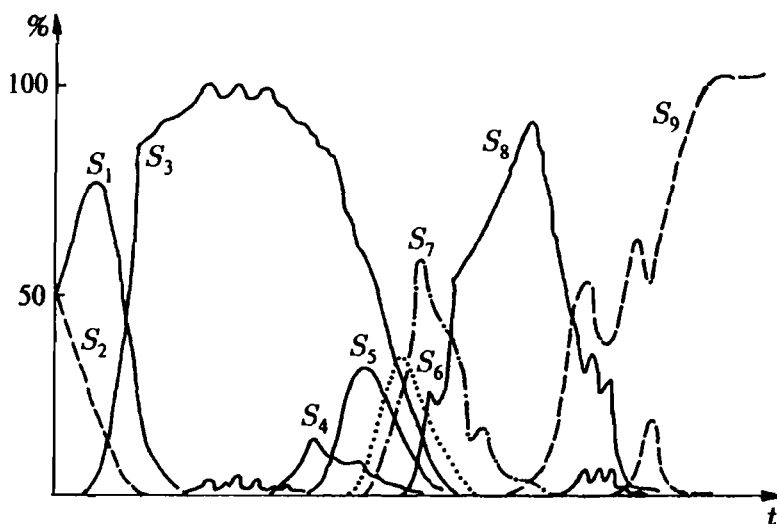


Рис. 8.3. Схема смены структур $S_1 \rightarrow S_3 \rightarrow S_7 \rightarrow S_8 \rightarrow S_9$ и т. д.

Если бы в некоторой нуклеиновой кислоте с определенной структурой воплотился псевдопотенциал безошибочного самовоспроизводства, то для дальнейшей эволюции она не имела бы особого значения: через несколько поколений ее «настигла бы и перегнала» менее совершенная самовоспроизводящаяся структура (*Bresch*, 1977). Таким образом, установление физической концепции ценности должно включать в явном виде способность к мутации (*Küppers*, 1986). То, что самовоспроизведение молекул должно сопровождаться ошибками копирования, физически самоочевидно: выполняющиеся в процессе репликации молекул законы квантовой механики и статистической термодинамики обуславливают неустранимый «внутренний шум», который с необходимостью приводит к ошибкам копирования. В том же направлении оказывают свое действие и неизбежные внешние возмущения — «внешний шум», создаваемый, например, квантами излучения. Как сказываются ошибки в копировании на длине самовоспроизводящихся полимерных молекул? Если обозначить через q вероятность правильного построения фрагментов в копии заданного эталона, и через $1 - q$ — вероятность ошибки, то вероятность получения правильной копии всей молекулы длиной ν определяется величиной

$$Q = q^\nu = [1 - (1 - q)]^\nu \approx 1 - \nu(1 - q). \quad (8.3)$$

При простом самовоспроизведении нуклеиновых кислот (см. разд. 5.1), без участия специализированных ферментов, построение дополнительной копии основано только на образовании относительно сильных водородных связей между дополнительными парами оснований А—У, А—Т и Ц—Г. Однако при нормальной температуре селективность таких взаимодействий не очень велика. В табл. 8.1 представлены константы ассоциации различных пар оснований в неполярном растворителе. Нетрудно видеть, что дополнительные пары оснований отчетливо выделяются

C_6H_6 \ CCl_4	У	А	Ц	Г
У	15 45	550	<50	<1000
А	150	8 22	<50	<1000
Ц	<28	<28	28 50	<1000
Г	<1200	<1200	30000	1200 1000

большими значениями констант ассоциации, однако различие заключено в пределах от 10^1 до 10^2 . Следовательно, вероятность ошибки при образовании дополнительной пары должна лежать в интервале от 0,1 до 0,01; иначе говоря, мы получаем значения q от 0,90 до 0,99. Какова максимальная длина цепи, которая может возникать при столь простом способе самовоспроизведения? Если принять величину $Q_{\min} \approx 1/2$ за нижнюю границу способности к самовоспроизведению, то из формулы (8.3) следует, что максимальная длина цепи определяется величиной

$$\nu_{\max} = \frac{\ln Q_{\min}}{\ln q} \approx \frac{\ln Q_{\min}}{q - 1} \approx \frac{0,7}{1 - q}. \quad (8.4)$$

Для рассмотренного нами выше примера максимальная длина цепи заключена в интервале от 7 до 70. Это позволяет в благоприятном случае осуществить копирование транспортных нуклеиновых кислот (тРНК), содержащих около 80 фрагментов. Более точный анализ должен учитывать, что фрагменты различных сортов могут считываться с различной точностью.

Пусть q_i ($i = 1, \dots, \lambda$) — вероятность правильного построения фрагмента i -го сорта. Вероятность правильного воспроизведения цепи из ν_i фрагментов сорта i определяется выражением

$$Q = \prod_{i=1}^{\lambda} q_i^{\nu_i} \approx 1 - \sum_{i=1}^{\lambda} \nu_i (1 - q_i). \quad (8.5)$$

В предположении о приближенном равномерном распределении компонент $\nu_i \approx \nu/\lambda$ получаем для максимальной длины цепи

$$\nu_{\max} \approx \frac{\lambda \ln Q_{\min}}{\sum_{i=1}^{\lambda} (q_i - 1)}. \quad (8.6)$$

Экспериментально для гуанина и цитозина получен уровень ошибок около 1%, а для аденина и урацила — около 10% (Freese, 1979), т. е.

$$\begin{aligned} q_i &= 0,99 & \text{при } i &= \Gamma, \text{Ц;} \\ q_i &= 0,90 & \text{при } i &= \text{А, У.} \end{aligned}$$

Подставляя полученные значения в формулу (8.6), находим максимальную длину цепи $\nu_{\max} \approx 12$. Из этих оценок следует, что в «протосупе» только наиболее простые и относительно короткие нуклеиновые кислоты имели шанс на самовоспроизведение. Поэтому вычисленный Эйгеном (Eigen, 1979) и Винклер протогеном —

тРНК с 76 нуклеотидами — заведомо является результатом длительного процесса эволюции. Вероятно, что на этой первой фазе цель эволюции состояла главным образом в улучшении механизма воспроизведения. В ходе эволюции этот механизм был доведен до высокой степени совершенства путем создания аппарата ферментов, управляющего воспроизведением. Это относится уже к простейшим организмам — вирусам. В частности, обширный экспериментальный материал накоплен относительно процесса воспроизведения и ошибок у бактериофагов — мельчайших вирусов, нападающих на бактерию как на клетку-хозяина (Küppers, 1979, 1983, 1986).

В частности, такого рода эксперименты проводились с Q -фагами, инфицирующими бактерии *Coli*. Геном этих фагов (совокупность наследственной информации) состоит из одноцепочечной рибонуклеиновой кислоты (РНК). После внедрения вирусного генома в клетку-хозяине продуцируется особый фермент — так называемая РНК-репликаза, делающая возможной репродукцию вирусной РНК. Воспроизведение РНК происходит с фантастической точностью. Вейсман показал, что ошибка происходит только при считывании каждого трехтысячного нуклеотида, т. е. $1 - q = 3 \cdot 10^{-4}$. Эйген (Eigen, 1979) положил для вирусной РНК $Q_{\min} \approx 1/4$ и, подставив в формулу (8.4), получил оценку $\nu_{\max} \approx 4600$. Реальная длина вирусной РНК достигает, по Вейсману, 4500 нуклеотидов, что подтверждает приведенную выше оценку (Frese, 1979). Как мы уже знаем, даже относительно примитивный механизм воспроизведения вирусов обеспечивает устойчивое воспроизведение цепей длиной примерно до 10^4 нуклеотидных единиц. Возникающая на следующей ступени эволюции регулируемая ферментами репликация ДНК у прокариотов функционирует уже с частотой ошибки 10^{-7} на один акт присоединения нуклеотида и поэтому позволяет воспроизводить цепи ДНК длиной до $7 \cdot 10^6$ (по оценке (8.4)). Наконец, развитый у эукариотов механизм рекомбинаторного (полового) воспроизведения достигает непостижимо малой частоты ошибки — около $0,3 \cdot 10^{-9}$ на один акт присоединения нуклеотида. Соответственно, это означает возможность репродуцирования цепи ДНК длиной около 10^9 нуклеотидов. С возрастанием длины генома увеличивается и количество информации, передаваемой следующему поколению. В то время как вирусная РНК может переносить не больше информации, чем обычная книжная страница, ДНК поликариотов содержит уже примерно такую же информацию, как книга, а геном эукариотов по информационной емкости сравним с библиотекой.

8.2. Основы теории Эйгена—Шустер

Манфреду Эйгену принадлежит огромная заслуга: в начале семидесятых годов он заложил основы физико-химически ориентированной теории биологической эволюции (Eigen, 1971, 1976; Эйген, Шустер, 1982). Тем самым его по праву можно считать основоположником молекулярного дарвинизма, в дальнейшее развитие которого весомый вклад был внесен Шустер (Schuster, 1987; Swetina, Schuster, 1982; Fontana, Schuster, 1988). Исходным пунктом теории Эйгена служат следующие вопросы.

1. Почему существуют миллионы видов животных и растений?
2. Почему существует только один универсальный молекулярный механизм жизни: нуклеопротеиновая система, универсальный генетический код, универсальная хиральность макромолекул?
3. Существуют ли механизмы, соответствующие выбору одного вполне определенного сценария эволюции, позволяющего понять исторический путь эволюции, из практически неисчерпаемого множества сценариев эволюции?

Эйген усматривает основополагающий принцип эволюции в дарвиновском принципе естественной эволюции, корни которого уходят в физику и химию фундаментальных процессов. В основе многообразия существующих видов лежит

Таблица 8.2. Аминокислоты, входящие в состав протеинов, и их часто встречающиеся сокращенные обозначения

Символы		Названия аминокислот	
A ₁	S	Ser	Серин
A ₂	L	Leu	Лейцин
A ₃	R	Arg	Аргинин
A ₄	G	Gly	Глицин
A ₅	A	Ala	Аланин
A ₆	V	Val	Валин
A ₇	T	Thr	Треонин
A ₈	P	Pro	Пролин
A ₉	I	Ile	Изолейцин
A ₁₀	K	Lys	Лизин
A ₁₁	E	Gln	Глутаминовая кислота
A ₁₂	D	Asp	Аспарагиновая кислота
A ₁₃	F	Phe	Фенилаланин
A ₁₄	N	Asn	Аспарагин
A ₁₅	Q	Glu	Глютамин
A ₁₆	Y	Tyr	Тирозин
A ₁₇	C	Cys	Цистеин
A ₁₈	H	His	Гистидин
A ₁₉	M	Met	Метионин
A ₂₀	W	Trp	Триптофан

ветвящийся характер процесса эволюции. Подвергнем теперь интересующую нас проблему математическому анализу. Предметом изучения в теории Эйгена служат макромолекулы, которые возникали на самой ранней стадии эволюции жизни в «протосупе» и обладали свойствами самовоспроизведения, отбора и мутации. Представим себе множество линейных гетерополимеров, состоящих из элементов A₁, A₂, ..., A_λ. В случае ДНК и РНК элементарными фрагментами служат, соответственно, нуклеотиды А, Ц, Г, Т и А, Ц, Г, У, а в случае протеинов — 20 аминокислот

$$A_1, A_2, \dots, A_{20},$$

или в другой записи:

Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val (см. табл. 8.2).

В условиях Праземли 3–4 миллиона лет назад могли образоваться как полинуклеотиды, так и полипептиды (см. гл. 12). Проблема состоит в том, что лишь немногие из этих веществ были стабильными в тех условиях и могли самовоспроизводиться. Из астрономического числа $W = \lambda^l$ различных последовательностей (первичных структур) длины l из λ различных элементарных фрагментов лишь очень немногие обладали способностью к образованию высокоорганизованных структур. Согласно тезису Эйгена, «благоприятные» биополимеры образовались лишь в ходе дарвиновского процесса отбора на молекулярном уровне. Предпосылками этого были пять свойств, которыми обладали упомянутые выше макромолекулы.

1. Метаболизм

Макромолекулы постоянно обмениваются энергией и веществом с окружающим раствором, содержащим мономеры с большим запасом энергии. Синтез макромолекул приводится в действие за счет положительного сродства реакции синтеза. Подсистема макромолекул представляет собой открытую термодинамическую систему, способную осуществлять экспорт энтропии, рассмотренный нами в гл. 2 и 3.

2. Самовоспроизведение

Макромолекулы обладают способностью давать инструкции для своего собственного синтеза, а именно: воспроизведение происходит по матричному принципу, т. е. существующие молекулы служат эталоном для синтеза новых однотипных с ними макромолекул (см. гл. 5).

3. *Мультистабильность

Макромолекулы обладают огромным резервуаром потенциальных структур. Каждая из этих структур практически настолько стабильна, что может существовать на протяжении продолжительного времени. Мультистабильность создает возможность хранения информации (см. гл. 5).

4. Отбор

Из-за ограниченного подвода исходных веществ и других ограничивающих факторов между макромолекулами идет непрестанная конкуренция. Благоприятные формы размножаются, неблагоприятные быстро распадаются (см. гл. 6).

5. Изменчивость

Надежность каждого процесса воспроизведения ограничена физическими факторами, например, тепловыми шумами и квантовомеханическими неопределенностями. Ошибки становятся особенно частыми, когда скорость воспроизведения велика, или, иначе говоря, когда энергия активации рассматриваемой реакции синтеза лежит не слишком далеко от уровня энергии флуктуаций.

Ниже мы анализируем роль изменчивости и показываем, что изменчивость не только неизбежно связана с самовоспроизведением, но и может рассматриваться как необходимый фактор эволюции. Ясно, что наличие всех пяти перечисленных выше факторов не только делает эволюцию теоретически возможной, но и действительно необходимой. Поэтому следует ожидать, что и другие физико-химические системы, обладающие всеми пятью свойствами, наделены способностью к эволюции. Экспериментальное доказательство этого утверждения еще предстоит осуществить, так как создать реальную систему, которая обладала бы всеми пятью свойствами, достаточно трудно. Определенные аспекты эволюционных процессов присущи явлениям, происходящим в лазере (Хакен, 1980), росту кристаллов из расплавов и растворов, а также росту металлических нитей («усов») между электродами в кислых растворах солей металлов (фон Фёрстер, Зонф, 1964), процессам в сложных электрических системах (Антер, 1970; Ebeling, Feistel, 1976) и в сетях связанных между собой компьютеров (см. гл. 12).

Обратимся теперь к математическому описанию эволюции макромолекул. Макромолекулу с заданной первичной структурой условимся обозначать буквами:

$$i = (z_1, z_2, \dots, z_\nu), \quad z_k \in \{A_1, \dots, a_\lambda\}. \quad (8.7)$$

В ходе процесса репликации изготавливаются как тождественные копии, так и копии с ошибками:

а) $a + i \rightarrow i + i$ (самовоспроизведение);

б) $a + i \rightarrow i + j$ (мутация).

Поскольку полипептиды не обладают способностью к самовоспроизведению, мы имеем в виду прежде всего полинуклеотиды, т. е. цепи ДНК и РНК, состоящие из четырех сортов элементарных фрагментов (нуклеотидов). Соответственно, a означает молекулу, состоящую из молекул исходных веществ: А, Ц, Г, У или А, Ц, Г, Т.

Ошибка копирования, как правило, означает, что копия отличается от оригинала положением соответствующего основания, например,

$$\begin{aligned} i &= (\text{ГЦГАУЦГА} \dots \text{ГУА}), \\ j &= (\text{ГЦУАУЦГА} \dots \text{ГУА}). \end{aligned} \quad (8.8)$$

Процесс самовоспроизведения и мутацию по Эйгену (*Eigen*, 1973) можно описать следующими дифференциальными уравнениями:

$$\frac{dx_i}{dt} = (A_i - D_i)x_i + \sum_{j \neq i} (A_{ij}x_j - A_{ji}x_i) - k_0x_i. \quad (8.9)$$

Здесь $x_i(t)$ — концентрация сорта i , A_i — полная скорость воспроизведения, включая изготовление копий с ошибками, k_0 — скорость разбавления и A_{ij} — вероятность мутации, превращающей сорт j в сорт i . Уравнение (8.9) можно записать иначе:

$$\frac{dx_i}{dt} = (A_iQ_i - D_i)x_i + \sum_{j \neq i} A_{ij}x_j - k_0x_i, \quad (8.10)$$

где

$$Q_i = 1 - \frac{1}{A_i} \sum_{j \neq i} A_{ji}, \quad (8.11)$$

— так называемый фактор качества, указывающий долю правильных копий. Обычно этот фактор выражается числом, близким к единице: $0,9 \approx Q_i < 1,0$. В модели Эйгена скорость разбавления $k_0(t)$ регулируется так, что полная концентрация остается постоянной:

$$\sum_i x_i(t) = C = \text{const}. \quad (8.12)$$

При суммировании уравнений (8.9) (или (8.10)) по всем сортам левая часть обращается в нуль, и мы получаем возможность вычислить скорость разбавления k_0 :

$$k_0(t) = \langle E \rangle = \frac{1}{C} \sum_i E_i x_i(t), \quad (8.13)$$

$$E_i = A_i - D_i. \quad (8.14)$$

В результате мы приходим к следующей эквивалентной форме уравнений для скоростей:

$$\frac{dx_i}{dt} = (E_i - \langle E \rangle)x_i + \sum_{j \neq i} (A_{ij}x_j - A_{ji}x_i), \quad (8.15)$$

$$\frac{dx_i}{dt} = (W_i - \langle E \rangle)x_i + \sum_{j \neq i} A_{ij}x_j. \quad (8.16)$$

Величину

$$W_i = A_iQ_i - D_i \quad (8.17)$$

Эйген называет селекционной ценностью i -го сорта. Мы будем пользоваться эйгеновскими обозначениями W_i и E_i . Определим матрицу

$$\begin{aligned} E &= \{E_{ij}\}, \\ E_{ij} &= E_i \quad \text{при} \quad i = j, \\ E_{ij} &= A_{ij} \quad \text{при} \quad i \neq j. \end{aligned} \quad (8.18)$$

Кроме того, введем функцию

$$D(t) = \frac{1}{C} \sum_{i,j} E_{ij} x_j(t). \quad (8.19)$$

Если обозначить через $\mathbf{x}(t) = \{x_i(t)\}$ вектор концентраций, то уравнения (8.15) в матричной записи примут вид

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{E}\mathbf{x}(t) - \mathbf{D}(t)\mathbf{x}(t). \quad (8.20)$$

Таким образом, уравнение скоростей полностью аналогично системе уравнений, описывающих реакции с перекрестным катализом, о котором упоминалось в разд. 6.3. Если определить диагональную матрицу $\mathbf{P}$ собственных значений и модальную матрицу $\mathbf{T}$ собственных векторов соотношением

$$\mathbf{E}\mathbf{T} = \mathbf{T}\mathbf{P}, \quad (8.21)$$

то решение уравнений (8.20) можно представить в следующем виде (*Thompson, McBride, 1974; Ebeling, Feistel, 1976, 1977; Jones, Enns, Rangnekar, 1976*):

$$\mathbf{x}(t) = C \frac{\mathbf{T} \exp \{\mathbf{P}t\} \mathbf{T}^{-1} \mathbf{x}(0)}{\mathbf{I}^T \mathbf{T} \exp \{\mathbf{P}t\} \mathbf{T}^{-1} \mathbf{x}(0)}. \quad (8.22)$$

Рассматриваемая нами теперь проблема мутации отличается от проблемы перекрестного катализа тем, что число сортов в случае биополимеров практически бесконечно и поэтому матрицы имеют необычайно большие размеры, хотя большинство недиагональных элементов равны нулю. Разумеется, при таких условиях задача на собственные значения почти неразрешима, и формальное решение (8.22) имеет ограниченную ценность. В предположении, что скорости мутаций A_{ij} достаточно малы, устойчивое стационарное решение может быть получено методом возмущений, которые сходятся к решению (8.22):

$$\begin{aligned} x_i &= \frac{C A_{im}}{E_m - E_i} \quad \text{при } i \neq m, \\ x_m &= C \left[1 - \sum_{j \neq m} \frac{A_{jm}}{E_m - E_j} \right], \end{aligned} \quad (8.23)$$

где m — сорт с наибольшей скоростью воспроизведения, т. е.

$$E_m > E_i \quad \forall i. \quad (8.24)$$

В первом порядке теории возмущений мы получаем для наибольшего собственного значения

$$\lambda_m = E_m - \sum_{j \neq m} A_{jm}. \quad (8.25)$$

Как было показано в разд. 6.3 (соотношение (6.96)), функция $D(t)$ при больших временах сходится к наибольшему собственному значению

$$D(t) = \langle E \rangle \rightarrow \lambda_m. \quad (8.26)$$

Эйген (*Eigen, 1973, 1976*) вывел из этого соотношения общий принцип отбора. Применяя этот принцип, необходимо соблюдать осторожность и учитывать, что соотношение (8.26) надлежит понимать лишь как статический принцип эволюции

в смысле разд. 6.1. Иначе говоря, из соотношения (8.26) в общем случае не обязательно следует монотонное возрастание функции $D(t)$.

Эйген и Шустер (Эйген, Шустер, 1982) существенно развили свою теорию, включив в нее нелинейную связь между конкурирующими сортами. Основная идея так называемой теории гиперциклов состоит в явном учете связей между полипептидами и полинуклеотидами. Как уже было показано, эволюция полинуклеотидов без создания специального аппарата, обеспечивающего надежное воспроизведение длинных цепей, остается на весьма примитивной ступени. Для создания такого аппарата практически подходят только протеины с их потенциально неисчерпаемым разнообразием возможных пространственных структур, позволяющих передавать сложные функции. Еще предшественники протеинов — протеноиды, спонтанно возникавшие в условиях Праземли, — обладали каталитическими свойствами, необходимыми и для воспроизведения нуклеиновых кислот. Простейшие белки, по-видимому, исторически появились задолго до нуклеиновых кислот, так как образование простейших белков может с успехом происходить в самых различных условиях. Многие протеноиды с самого начала обладали каталитическими и прочими функциями современных протеинов. Решающим препятствием стало то обстоятельство, что протеноиды не могли систематически развивать эти свойства дальше и оптимизировать их, поскольку не обладали способностью к самовоспроизведению. Наоборот, нуклеиновые кислоты, появившиеся как исторические преемники протеноидов и представленные в «протосупе» лишь в очень малых концентрациях, обладали способностью к самовоспроизведению и тем самым предпосылкой к эволюции.

Результатом первой фазы эволюции, по Эйгену, явилось возникновение относительно коротких нуклеиновых кислот, напоминавших тРНК. Под воздействием протеноидов эти кислоты размножались в протосупе, претерпевали процесс отбора. Следующая фаза охарактеризовалась установлением более тесной связи между нуклеиновыми кислотами и протеноидами, их объединение в суперструктуру, наделенную новыми функциями. По Эйгену (Eigen, 1973, 1976), а также по Эйгену и Шустер (Эйген, Шустер, 1982), наиболее вероятной формой новой связи является гиперциклическая структура. Согласно этим авторам, гиперциклический порядок служит теоретически обоснованной необходимой предпосылкой к интеграции в структуры, наделенные способностью к самовоспроизведению и большей информационной емкостью. О том, как именно происходит соединение фрагментов в структуру, теория Эйгена—Шустер умалчивает. На рис. 8.4 показан пример гиперциклической структуры, состоящей из 5 нуклеиновых кислот и 5 протеинов.

Кинетические уравнения для гиперциклов могут быть записаны следующим образом (x_i — концентрации нуклеиновых кислот, y_k — концентрации протеинов):

$$\dot{x}_i = E_i x_i + \sum_{j \neq i} (A_{ij} x_j - A_{ji} x_i) + f_i(x_i; y_1, y_2, \dots) - k_0 x_i, \quad (8.27)$$

$$\dot{y}_k = g_k(y_1, y_2, \dots; x_1, x_2, \dots) \quad (8.28)$$

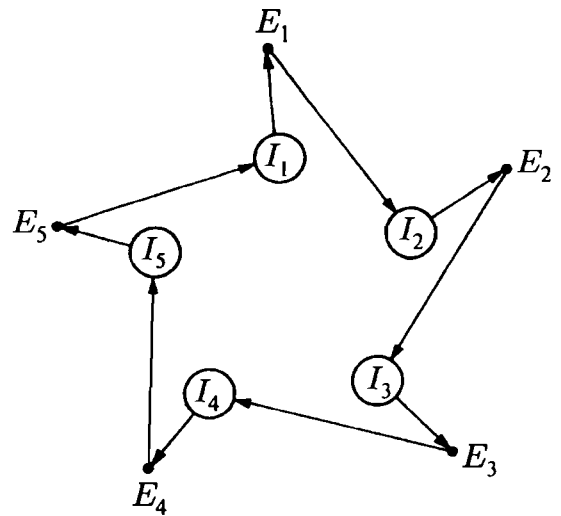


Рис. 8.4. Схема гиперцикла по Эйгену (I_n — нуклеиновые кислоты, E_n — белки)

Здесь функции f_i описывают каталитическое действие протеинов, а g_k — кинетическая функция протеинов, относительно которой мы пока не будем делать уточняющих предположений. Если предположить, что кинетические процессы на уровне белков протекают быстро, то следует положить

$$\dot{y}_k \approx 0, \quad g_k(y_1, y_2, \dots; x_1, x_2, \dots) \approx 0 \quad (8.29)$$

и в рассматриваемом приближении полностью исключить концентрации протеинов. Запишем возникающие при этом уравнения кинетики нуклеиновых кислот с учетом не более чем квадратичных членов в следующем виде:

$$\dot{x}_i = e_i x_i + \sum_{j \neq i} (a_{ij} x_j - a_{ji} x_i) + \sum_{j \neq i} b_{ij} x_i x_j - k_0 x_i. \quad (8.30)$$

Это основное уравнение теории гиперциклов, интенсивно исследуемое в последнее время (Эйген, Шустер, 1982; Schuster, Sigmund, Wolff, 1978, 1979; Jones, 1977; Ebeling et al., 2000). При условии постоянства общего числа частиц (8.12) для константы k_0 в уравнении (8.30) получаем:

$$k_0 = \frac{1}{C} \left[\sum_k e_k x_k + \sum_{k,l} b_{kl} x_k x_l \right]. \quad (8.31)$$

Гиперциклы в узком смысле, структура которых представлена на рис. 8.4, соответствуют особой структуре матрицы $\{b_{ij}\}$, приводящей к характерным кольцам:

$$\{b_{ij}\} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & b_1 \\ b_2 & 0 & 0 & \dots & 0 & 0 \\ 0 & b_3 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & b_n & 0 \end{pmatrix}. \quad (8.32)$$

Матрица (8.32) с ее циклической перестановочной структурой в точности соответствует гиперциклу с n членами. Соответствующее дифференциальное уравнение имеет вид:

$$\dot{x}_i = e_i x_i + \sum_j (a_{ij} x_j - a_{ji} x_i) + b_i x_i x_k - k_0 x_i, \quad (8.33)$$

$$k = i - 1 + \delta_{i|n}.$$

В более общем случае, когда имеется несколько конкурирующих гиперциклов, матрица $\{b_{ij}\}$ состоит из нескольких циклических подматриц. В силу условия постоянства общего числа частиц (8.12) пространство состояний такой системы представляет собой совокупность плоскостей, пересекающих все координатные оси x_i на расстоянии C от начала координат. На этом так называемом симплексе разыгрывается вся динамика системы. Структура такого фазового пространства для гиперциклических систем весьма сложна, поэтому в нем обычно существует несколько аттракторов и областей притяжения, разделенных сепаратрисами. Следовательно, конечное состояние гиперциклической системы всегда зависит от того, в какой области притяжения выбрано начальное состояние, поскольку динамическая система не может пересекать границы (сепаратрисами) между областями притяжения. Эта особенность обуславливает поведение гиперциклической системы типа «все или ничего», т. е. однажды обретенная динамическая структура (область притяжения) фиксирована раз и навсегда. Покажем теперь, что это свойство оказывает сильнейшее влияние на конкурентное поведение гиперциклов. При этом (как

и в случае проблем, рассмотренных в разд. 6.3) оказывается целесообразным переход к так называемой экологической картине (Feistel, Ebeling, 1978). Сосредоточим внимание на графе, порождаемом матрицей $\{b_{ij}\}$, поскольку он описывает самые сильные связи в системе. В простейшем случае этот граф состоит из неприводимых подграфов, не связанных друг с другом. Каждый из подграфов может обладать, например, кольцевой или какой-нибудь аналогичной структурой, которую мы будем, как и прежде (см. разд. 6.3), называть связкой. Таким образом, наша система допускает разложение на гиперциклические связки, которые мы перенумеруем индексом $l = 1, 2, 3, \dots$. По сравнению с введенными в разд. 6.3 «простыми» связками, основанными на линейных связях или мутациях, гиперциклические связки представляют новое качество. Каждая из таких связок ведет себя как единое целое; образующие ее молекулы живут и умирают только все вместе. Введем вместо исходных переменных x_i совокупную переменную S_l связки l , которую можно определять по-разному, например,

$$S_l = \sum_{i \in l} x_i \quad \text{или} \quad S_l = \frac{1}{n_l} \sum_{i \in l} x_i, \quad (8.34)$$

где n_l — число элементов связки. Внутреннюю структуру связки характеризует переменная ρ_{il} , задающая долю элементов i в связке l . Мы снова будем предполагать, что внутренние переменные связок устанавливаются значительно быстрее, чем совокупные переменные. Такое допущение позволяет вывести замкнутые уравнения для совокупных переменных S_l . Эти уравнения всегда содержат квадратичные члены. Например, если пренебречь мутациями, то

$$\dot{S}_l = e_l^* S_l + b_l^* S_l^2 - k_0 S_l. \quad (8.35)$$

Рассмотрим теперь конкуренцию двух гиперциклов. Так как $S_1 + S_2 = C$, одну из переменных можно полностью исключить, и мы получаем

$$\dot{S}_1 = -\frac{b_1^* + b_2^*}{C} S_1 (S_1 - C) \left(S_1 - \frac{C b_2^* - e_1^* - e_2^*}{b_1^* + b_2^*} \right). \quad (8.36)$$

Такая система обладает (по крайней мере при условиях

$$C b_2 > (e_1 - e_2), \quad C b_1 > (e_2 - e_1) \quad (8.37)$$

двумя устойчивыми стационарными решениями $S_1^{(1)}$ и $S_2^{(3)}$, между которыми заключено неустойчивое решение $S_1^{(2)}$:

$$S_1^{(1)} = 0, \quad S_1^{(2)} = \frac{C b_2^* - e_1^* + e_2^*}{b_1^* + b_2^*}, \quad S_1^{(3)} = C. \quad (8.38)$$

С учетом мутаций эта картина претерпевает только качественные изменения. Вместо уравнения (8.35) мы получаем

$$\dot{S}_l = e_l^* S_l + b_l^* S_l^2 + \sum_m (a_{lm}^* S_m - a_{ml}^* S_l) - k_0 S_l. \quad (8.39)$$

В случае конкуренции двух гиперциклов мы приходим к простому дифференциальному уравнению

$$\begin{aligned} \dot{S}_1 &= \gamma - \beta S_1 + \alpha S_1^2 - \zeta S_1^3, \\ \gamma &= a_{12}^* C, \quad \beta = C b_2^* + e_2^* - e_1^* + a_{12}^* + a_{21}^*, \\ \alpha &= \frac{1}{C} [C b_1^* + 2C b_2^* - e_1^* + e_2^*], \quad \zeta = \frac{1}{C} [b_1^* + b_2^*]. \end{aligned} \quad (8.40)$$

Сравнение с результатами, полученными в разд. 5.4, показывает полную аналогию с дифференциальным уравнением бистабильной реакции Шлёгеля. В предположении, что скорости мутаций достаточно малы, получаем для двух устойчивых решений

$$S_1^{(1)} \approx \frac{a_{12}^* C}{C b_2^* + e_2^* - e_1^*}, \quad S_1^{(3)} \approx C - \frac{a_{21}^* C}{C b_1^* + e_1^* - e_2^*}. \quad (8.41)$$

Которое из двух решений описывает поведение системы, т. е. который из двух

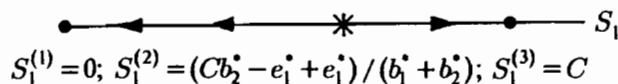


Рис. 8.5. Пространство состояний конкуренции двух гиперциклов

гиперциклов — 1 или 2 — побеждает в конце концов, зависит исключительно от начальных состояний. Резюмируя, можно утверждать следующее: в случае конкуренции двух гиперциклов всегда существуют два устойчивых стационарных решения $S_1^{(1)} \approx 0$ и $S_1^{(3)} \approx C$, а между этими двумя стационарными значениями — неустойчивое стационарное решение, которое одновременно является сепаратрисой в фазовом пространстве (рис. 8.5). Этот пример в точности соответствует рассмотренным в разд. 5.4 бистабильным реакциям. В нашем случае побеждает либо гиперцикл 1, либо гиперцикл 2; какой из них выживает, зависит в конце концов от того, расположена ли задаваемая начальными условиями представляющая точка справа или слева от сепаратрисного значения $S_1^{(2)}$. Представим себе, что 1 — установившийся, а 2 — вновь сформировавшийся гиперцикл, который сначала представлен в весьма малой концентрации, т. е. его начальные условия лежат очень близко от правого устойчивого состояния. Тогда динамика стремится к состоянию $S_1 \approx C$ и $S_2 \approx 0$ независимо от того, насколько хорош второй гиперцикл. Следствием гиперциклической организации является отбор типа «все или ничего». Возникший гиперцикл допускает в определенных пределах усовершенствование своих параметров, но исключает появление любых качественно новых гиперциклических структур. Эйген усматривает в этом теоретическую основу универсальности генетического кода, а также принцип действия клеток с учетом киральности используемых элементарных фрагментов. Выход из временного тупика, в который попадает эволюция из-за фиксации единственного гиперцикла, Эйген (Eigen, 1976) усматривает в компартментации системы. Происходящие в компартментированных системах процессы отбора и оптимизации могли привести к образованию протоклеток, ставших исходным пунктом биологической эволюции видов (см. гл. 2).

8.3. Стохастика процессов эволюции

В процессах мутации и эволюции случайность играет решающую роль. Хотя эволюция в соответствующих физических условиях закономерна и необходима, все же в отдельных случаях она подвержена влиянию случайности. Именно поэтому только стохастическая теория может адекватно отражать сущность процессов эволюции. Существует, однако, и еще один аспект, в силу которого изложенная в разд. 8.2 детерминистическая теория не вполне корректна, — дискретная природа молекулярных единиц процесса эволюции. В нормальных физических, химических и биологических процессах дискретной природой молекул можно пренебречь и перейти к описанию с помощью непрерывных концентраций. Однако в процессах мутации и эволюции один-единственный молекулярный акт — возникновение

единственной молекулы нового сорта, обладающего теми или иными преимуществами, — может стать началом процесса макроскопического усиления, который приводит к полному изменению всей системы. Это основополагающее свойство эволюционирующих систем в литературе часто называется принципом усиления (*Timofeev-Ressovsky, Rompe, 1959; Ratner, 1974; Патнер, 1983, 1986*). Чтобы правильно описывать элементарный молекулярный акт эволюции, теория процессов мутации и эволюции должна была бы, по существу, следовать из квантовой теории систем с переменным числом частиц, например, выводиться из нее методом вторичного квантования. К сожалению, квантово-статистическая теория систем, в которых протекают химические реакции, делает в настоящее время лишь первые шаги (*Климонтович, 1975, 1980, 1986; Klimontovich, Kremp, Kraeft, 1987*). Полностью разработана только стохастическая теория реакций, на которую мы опираемся в дальнейшем.

В рамках стохастической теории реакций элементарным кинетическим актом является увеличение или уменьшение числа N_i молекул i -го сорта на единицу, т. е.

$$N_i \rightarrow N_i \pm 1. \quad (8.42)$$

Состояние системы в момент времени t характеризуется множеством чисел заполнения для всех входящих в систему сортов:

$$N_1(t), N_2(t), N_3(t), \dots$$

Числа заполнения могут принимать только дискретные значения:

$$N_i = 0, 1, 2, 3, \dots$$

В этой главе мы в основном рассматриваем системы, удовлетворяющие условию постоянства общего числа частиц (*Eigen, 1973*)

$$\sum_i N_i = N = \text{const}. \quad (8.43)$$

В пространстве чисел заполнения такие системы могут двигаться только по поверхности, определяемой условием (8.43). В силу сохранения числа частиц в элементарном переходе всегда участвуют два партнера:

$$N_i \rightarrow N_i + 1, \quad N_j \rightarrow N_j - 1. \quad (8.44)$$

Пространство состояний и элементарные переходы в нем представлены на рис. 8.6. Важная особенность рассматриваемых нами процессов эволюции по сравнению с процессами отбора, о которых шла речь в разд. 6.2, состоит в том, что при учете мутаций состояниями $N_i = 0$ можно пренебречь. Тем самым эти состояния утрачивают характер поглощающих состояний в пространстве состояний (см. рис. 8.6 и 8.7). Эта математическая ситуация выражает физически чрезвычайно важное обстоятельство, состоящее в том, что при мутации сорта $N_j > 0$ может возникнуть новый сорт, который до этого вообще не был представлен в системе, т. е. до мутации должно выполняться равенство $N_i = 0$, а после мутации — равенство $N_i = 1$. С физической точки зрения, мутации всегда приводят к возникновению новых сортов молекул, не существовавших прежде в системе; с математической точки зрения, появление новых сортов соответствует расширению пространства состояний за счет включения в него новых размерностей. Как следует из оценок, приведенных в разд. 8.1, мы должны исходить из того, что пространство состояний макромолекулы имеет размерность порядка 10^{600} . Так как во всей «Вселенной» число нуклонов достигает лишь 10^{79} , а из них можно построить не более чем 10^{73} макромолекул, мы заключаем, что по крайней мере для 10^{527} размерностей должны выполняться равенства $N_k = 0$,

или, иначе говоря, что представляющая точка состояния всегда должна лежать на одной из граней определяемой уравнениями (8.41) симплицальной решетке. Таким образом, мутация означает, что представляющая точка удаляется от грани в новой, еще не занятой, размерности, а возникающий процесс отбора соответствует движению по новой грани при условии, что речь идет об инновации, т. е. о благоприятной мутации (рис. 8.7).

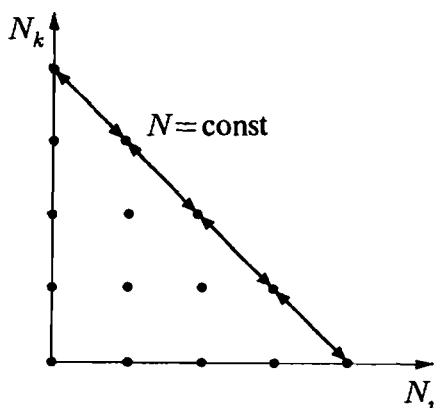


Рис. 8.6. Пространство состояний и элементарные переходы в стохастическом процессе эволюции при постоянном числе частиц (размерность 2)

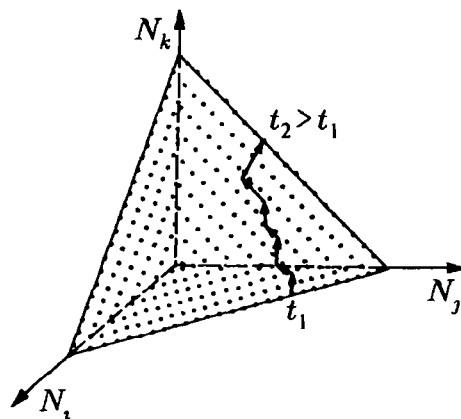


Рис. 8.7. Решетка-симплекс пространства состояний с примером траектории, образовавшейся после мутации

Существует альтернативный вариант графического представления связанных с мутацией и отбором стохастических процессов в пространстве нуклеотидов, или последовательностей. В основе его лежит идея, предложенная Рехенбергом (*Rechenberg*, 1973). Это представление исходит из того, что каждый нуклеотид в последовательности длиной ν может находиться только в четырех состояниях: Г, А, Т, Ц. Для представления состояния такой последовательности можно воспользоваться ν -мерным гиперкубом, состоящим в каждом направлении из трех «поставленных» один на другой элементарных кубов. На рис. 8.8 показан пример для случая $\nu = 2$. Нетрудно видеть, что каждому из 16 состояний соответствует одна из пар-последовательностей ГГ, ГА, ГТ, ГЦ, АГ, АА, АТ, АЦ, ТГ, ТА, ТТ, ТЦ, ЦГ, ЦА, ЦТ, ЦЦ — вершины прилегающих друг к другу квадратов.

В случае последовательностей длины $\nu = 3$ необходимо включить еще третье измерение, в результате чего возникает куб с 64 вершинами, соответствующими 64 различным последовательностям-тройкам. Точковая мутация соответствует сдвигу вдоль ребра, параллельного одной из осей. Длины элементарных ребер не имеют непосредственного значения, так как априори все переходы $\Gamma \rightarrow A$, $\Gamma \rightarrow T$, $\Gamma \rightarrow C$, $A \rightarrow T$, $A \rightarrow C$, $A \rightarrow G$, $C \rightarrow G$, $C \rightarrow A$, $C \rightarrow T$ равнозначны. Эволюционный путь от одной последовательности нуклеотидов к другой, состоящий, вообще говоря, из многих мутаций, т. е. из многих шагов вдоль ребер, соответствует некоторому связному пути по гиперкубу. Примером может служить изображенный на рис. 8.8 эволюционный путь от АГ к ГТ. Если рассматривать популяции, то числу представителей N_i последовательности определенного типа $i = A_1 \dots A_\nu$ может быть сопоставлено скопление точек мощности N_i в непосредственной окрестности соответствующей i вершины (или жирная точка). Эволюция популяции описывается

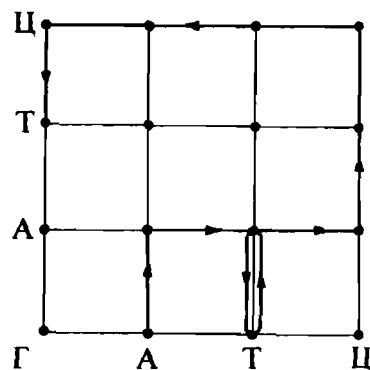


Рис. 8.8. Пространство нуклеотидов для последовательности пар и один из путей эволюции

блужданием скопления точек по пространству нуклеотидов. Графическое представление в пространстве нуклеотидов особенно удобно для наглядного истолкования путей эволюции и при обсуждении стратегий эволюции (см. разд. 12.2 настоящей книги и (*Rechenberg*, 1973)). На языке теории автоматов переходы в пространстве нуклеотидов могут быть описаны как процессы в автоматах Рабина-Скотта (*Ebeling, Feistel, Jiménez-Montaño*, 1977).

Мы подходим теперь к определению вероятностей элементарных переходов (8.44) в единицу времени. При этом мы будем следовать более ранним работам, в которых вероятности переходов были получены сначала для более простых эволюционных процессов Эйгена (*Ebeling, Eigen*, 1977), а позднее уточнены с учетом образования гиперциклов (*Ebeling, Jiménez-Montaño*, 1980; *Ebeling et al.*, 2000). Исходя из формулы (6.69), приведенной в разд. 6.2, получаем при явном учете мутаций

$$W(N_1, \dots, N_i + 1, \dots, N_j - 1, \dots, N_k, \dots \mid N_1, \dots, N_i, \dots, N_j, \dots, N_k) = \\ = a_{ij}N_j + \frac{1}{N}e_iN_iN_j + \frac{1}{NV}b_{ik}N_iN_jN_k. \quad (8.45)$$

Здесь a_{ij} — входящие в выражение (8.45) скорости мутаций, e_i — скорости репродукции нуклеиновых кислот, b_{ij} — коэффициенты гиперциклических связей. Пренебрегая каталитическим действием протеинов, мы положили бы просто $b_{ij} = 0$, $e_i = E_i$ и $a_{ij} = A_{ij}$ и получили бы вероятности перехода, соответствующие формуле (8.15). Приведенные выше вероятности перехода позволяют легко вычислить стохастические траектории в пространстве состояний при заданных начальных условиях. Ценные выводы относительно исследуемого случайного процесса могут быть получены с помощью методов стохастической динамики, т. е. путем систематического исследования большого числа индивидуальных траекторий. Трудности, возникающие при применении этих методов к процессам эволюции, сводятся по существу к тому, что коэффициенты в выражении (8.45) известны лишь в очень грубом приближении. Другой метод исследования случайного процесса эволюции состоит в исследовании распределения вероятности

$$P(N_1, N_2, \dots; t) \quad (8.46)$$

как функции времени. Для этого необходимо предварительно записать основное кинетическое уравнение. Следуя изложенной в разд. 5.2 и 6.2 общей схеме, получаем

$$\frac{\partial}{\partial t}P(N_1, N_2, \dots; t) = \sum_{i,j,k} W(N_i, N_j, N_k \mid N_i - 1, N_j + 1, N_k) \times \\ \times P(N_1, \dots, N_{i-1}, \dots, N_j \mid 1, \dots, N_k, \dots) - \\ - \sum_{i,j,k} W(N_i + 1, N_j - 1, N_k \mid N_i, N_j, N_k) \times \\ \times P(N_1, \dots, N_i, \dots, N_j, \dots, N_k). \quad (8.47)$$

Распределение вероятности представляет собой поверхность, построенную над симплициальной решеткой; описанием деформации этой поверхности во времени служит случайный процесс эволюции. Разумеется, высказать определенное утверждение относительно распределения вероятности в пространстве столь высокой размерности необычайно трудно. К этому добавляется еще принципиальная трудность применения метода основного кинетического уравнения к общему эволюционному процессу: при заданных начальных распределениях, сосредоточенных на относительно узкой области пространства состояний (симплициальной решетки), основное кинетическое уравнение описывает систематическую тенденцию к относительно равномерному

распределению вероятностей на всем пространстве состояний необычайно высокой размерности. Соответственно, возможные утверждения носят достаточно расплывчатый характер, а конечное стационарное распределение, к которому стремится система, по существу лишено определенного физического смысла. С другой стороны, основное кинетическое уравнение особенно хорошо подходит для изучения таких специальных вопросов, как выживание новых сортов и сосуществование новых сортов. Покажем, что основное кинетическое уравнение согласуется с детерминистическим уравнением (8.33). Для этого умножим уравнение (8.47) на N_k и просуммируем по всем числам заполнения. Как показано в разд. 6.2, в результате мы получим уравнение для средних чисел заполнения:

$$\begin{aligned} \frac{d}{dt} \langle N_k \rangle = \sum_l (a_{kl} \langle N_l \rangle - a_{lk} \langle N_k \rangle) + e_k \langle N_k \rangle + \frac{1}{V} \sum_m b_{km} \langle N_k N_m \rangle - \\ - \frac{1}{N} \sum_l e_l \langle N_k N_l \rangle - \frac{1}{NV} \sum_{l,m} b_m b_{lm} \langle N_k N_l N_m \rangle. \end{aligned} \quad (8.48)$$

Факторизуя средние значения и деля на объем, получаем из уравнения (8.48) уравнения для средней плотности $X_k = \langle N_k \rangle / V$, совпадающие с уравнениями (8.30) и (8.31). Тем самым доказано, что стохастическое описание согласуется с детерминистическим. Исследуем теперь вопрос о вымирании или выживании нового сорта, возникающего в системе вследствие мутации. Ограничимся сначала простыми эйгеновскими системами без гиперциклических связей. Предположим, что в системе имеется только один штаммовый сорт со скоростью воспроизведения E_1 и только один представитель мутантов со скоростью воспроизведения E_2 , т. е. в начальном состоянии $t = 0$ выполняются равенства

$$N_1 = N - 1, \quad N_2 = 1.$$

Условимся рассматривать N_2 как независимую переменную и пренебрегать в основном кинетическом уравнении всеми членами выше первой степени по N_2 (что вполне допустимо, если учесть начальные условия). В результате мы получим упрощенное уравнение

$$\begin{aligned} \frac{\partial}{\partial t} P(N_2, t) = E_2(N_2 - 1)P(N_2 - 1) + \\ + E_1(N_2 + 1)P(N_2 + 1) - (E_1 + E_2)N_2 P(N_2). \end{aligned} \quad (8.49)$$

Известно решение этого уравнения при начальном условии

$$P(N_2, 0) = \delta_{1, N_2} \quad (8.50)$$

(Bartholomay, 1958). Если нас интересует вероятность вымирания мутанта по истечении времени t , то

$$P(0, t) = \frac{[\exp \{E_2 t - E_1 t\} - 1]}{[(E_2/E_1) \exp \{E_2 t - E_1 t\} - 1]}. \quad (8.51)$$

Введем для удобства записи фактор преимущества

$$\delta = \frac{E_2 - E_1}{E_1} \quad (8.52)$$

и найдем вероятность выживания нового сорта (Allen, Ebeling, 1983):

$$1 - P(0, t) = P_{\text{выж}}(t) = \frac{\delta}{1 + \delta - \exp \{-E_1 \delta t\}}. \quad (8.53)$$

Если определить $n = E_1 t$ как меру числа поколений, которые выживают мутанты, то получится очень простая формула

$$P_{\text{выж}}(n) = \frac{\delta}{1 + \delta - \exp\{-n\delta\}}. \quad (8.54)$$

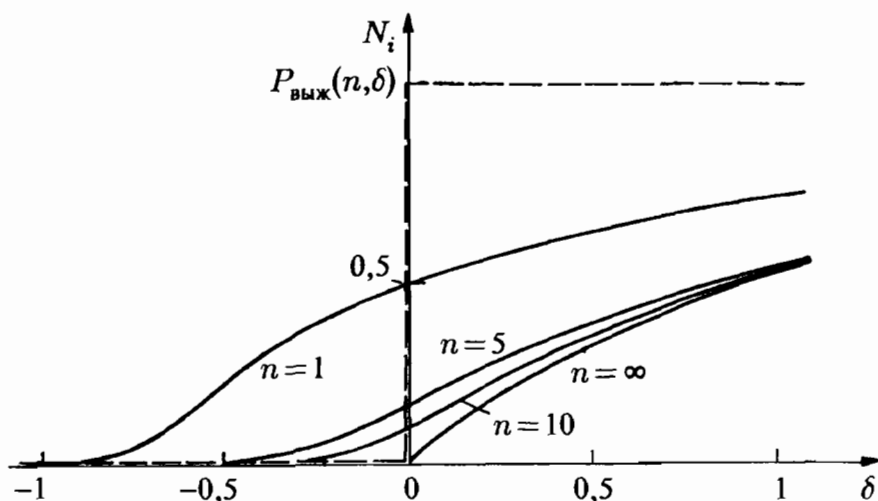


Рис. 8.9. Вероятность выживания мутанта через n поколений как функция фактора выживания δ (Allen, Ebeling, 1983) (сплошные линии — стохастическая теория; штриховые линии — детерминистическая теория)

На рис. 8.9 показана зависимость вероятности выживания от фактора преимущества и числа поколений. В то время как в детерминистической теории лучшие мутанты с $\delta > 0$ заведомо выживают, а худшие с $\delta < 0$ заведомо вымирают, более точная стохастическая теория предсказывает не столь определенное поведение. Как показано на рис. 8.9, вероятность выживания сначала возрастает с увеличением δ , но подъем при $\delta = 0$ довольно плоский, вследствие чего примерно до $n = 10$ мутанты с $\delta > 0$ не очень существенно отличаются от $\delta < 0$. Мутанты, лежащие в определенной области (примерно в полосе $\pm 10\%$ от ценности штаммового сорта), в отношении отбора ведут себя почти нейтрально. Этим объясняется, почему в природе генотипы, представляющие вид, не едины, а охватывают группу сходных генных структур (дикий тип). Как показывает стохастическая теория, естественный отбор действует не очень точно, а допускает целое семейство генотипов, лежащих в области толерантности отбора.

Другое интересное утверждение, следующее из стохастической теории, состоит в том, что поведение гиперциклов типа «все или ничего» нарушается или по крайней мере ослабевает. Как показано в разд. 8.2, конкуренция между установившимся штаммовым гиперциклом и претерпевшим мутацию гиперциклом согласуется с динамическим поведением бистабильной системы. В детерминистической теории спонтанные переходы между двумя стабильными состояниями такой системы исключаются, стохастическая теория оставляет определенную вероятность таких переходов, как было показано в разд. 5.4 для реакции Шлёгеля. Малость реакционного объема заведомо сыграла роль в ходе эволюции «протосупа». Примером тому служат реакции в коацерватах и в малых пылинках скальных пород. Решающий перелом к новому, лучшему гиперциклу в небольшом реакционном объеме мог произойти спонтанно и оттуда уже инфицировать макроскопическую окрестность.

8.4. Игра в имитацию

Одна из принципиальных трудностей при создании теории реального процесса эволюции состоит в том, что относительно значений входящих в теорию параметров, например, скоростей воспроизведения и мутаций, известно очень мало. Можно надеяться, однако, что новое направление экспериментальных исследований — проведение направленных эволюционных экспериментов на микроорганизмах (Küppers, 1979) — позволит получить новые данные. Ныне представляется разумным подвергнуть теорию проверке с помощью компьютерных экспериментов. Численные эксперименты не отражают полностью реальный процесс эволюции, но позволяют имитировать характерные особенности реальных процессов.

Начнем с изложения компьютерной модели Ферстерлинга, Куна и Тьюза (Försterling, Kuhn, Tews, 1972). Речь идет о последовательностях длиной $\nu = 9$ из четырех букв А, Т, Г и Ц, постоянно воспроизводящихся, иногда с ошибками (мутациями), при условии постоянного общего числа частиц $N = 50$. Мы исходим из некоторой начальной случайной последовательности, например,

$$t = 0: \quad N_i = 50, \quad j = (\text{ЦГТАТЦГТЦ}).$$

Таким образом, процесс может начаться лишь с одной из

$$S = 4^9 = 262\,144$$

возможных последовательностей, т. е. процесс начинается с одной из вершин симплекса. Имеющиеся последовательности удваиваются по прошествии определенных интервалов времени. С помощью случайной программы в процессе удвоения с определенной вероятностью W при вписывании буквы совершается ошибка. После каждого удвоения число последовательностей снова понижается до 50. Какая из последовательностей вымирает и какая выживает, решает случайная программа. При определении ценности последовательности авторы исходят из некоторой идеальной произвольно заданной последовательности, например, из последовательности

$$i = (\text{АГТТЦЦГАГ}).$$

Путем сравнения с идеальной последовательностью i мы определяем число ошибочных мест $F = d(k, i)$ в заданной последовательности k . Величина F указывает число несовпадений между последовательностями k и i . Ее можно рассматривать как метрическое расстояние между последовательностями (см. разд. 11.3). Ценность идеальной последовательности назначается произвольно. Для всех последовательностей с одной ошибкой ценность уменьшается в r раз, для всех последовательностей с двумя ошибками — в r^2 раз и т. д. Эти модельные предположения могут быть легко перенесены и в общий формализм, развитый в разд. 8.3. Модель Ферстерлинга, Куна и Тьюза соответствует стохастическому процессу на симплексе с $N = 50$ и вероятностями перехода

$$W(N_1, \dots, N_k + 1, \dots, N_l - 1, \dots, N_S \mid N_1, \dots, N_S) = A_{kl} N_l + \frac{1}{N} E_k N_k N_l, \quad (8.55)$$

где

$$A_{kl} = W^F, \quad F = d(k, l), \quad E_k = \frac{E_i}{r^{d(i, k)}}, \quad (8.56)$$

i — заданная идеальная последовательность. В играх, проведенных Ферстерлингом, Куном и Тьюзом, $r = 1,5$, а частоты ошибок $W = 0,0025; 0,01; 0,04$. Некоторые результаты представлены в табл. 8.3. Во всех случаях система обнаруживала стремление

Таблица 8.3. Компьютерная модель процесса эволюции по Ферстерлингу, Куну и Тьюзу (Försterling, Kuhn, Tews, 1972).(a: $W = 0,0025$, б: $W = 0,01$, в: $W = 0,04$)

Поколение	0			1			2			3			4			5			6		
	a	б	в	a	б	в	a	б	в	a	б	в	a	б	в	a	б	в	a	б	в
0																50	50	50			
10												11		4	29		42	6		4	4
20						1			12		14	33		23	4		12		1		
30						6		5	34		45	9			1						
40			4		3	10		34	23		13	13									
50			19		14	17		36	13			1	47			3					
60			18		23	19		26	12		1	1									
70		1	15		33	25		15	10		1										
80		14	29		31	17		5	3			1									
90		41	30		8	13		1	7												
100		46	33		4	15			1	49		1	1								
110		47	28		3	20			2												
120		41	30		9	16			3			1									
130		49	28		1	16			5			1									
150							49			1											

к идеальной последовательности i , но с различным качеством и с различной скоростью. Благоприятной частотой ошибки оказалась величина $W = 0,01$: при таком значении W после 100 поколений 92% последовательностей достигают идеальной структуры, в то время как, например, при $W = 0,04$, подгонка идет очень плохо, а при $W = 0,0025$ процесс длится необычайно долго.

Аналогичная компьютерная модель была исследована Эйгеном (Eigen, 1976). Эйген рассматривал все последовательности длиной $\nu = 20$, которые можно составить из заглавных букв латинского алфавита, знаков препинания и пробелов и. В качестве идеальной последовательности он выбрал осмысленную фразу «На ошибках учимся»:

L E R N u A U S u D E N u F E H L E R N.

Компьютер работал не непосредственно с буквами алфавита, а предварительно переводил их в последовательности двоичных знаков 0 и 1 (поскольку $2^5 = 32$, для кодирования одной буквы требуется 5 двоичных знаков). В двоичном алфавите идеальная последовательность принимает следующий вид:

10001010101011110011001010011011010110000010101001

01010100110010101011010100110110001010101011110011

В этом пространстве двоичных последовательностей длины $\nu = 100$ разыгрывается игра отбора. При этом рассматриваются только 10 копий, т. е. стохастический процесс протекает на симплексе с $N = 10$. Всего существует $S = 2^{100}$ альтернативных последовательностей. Определим для произвольной последовательности k число $F = d(k, i)$ несовпадений с приведенной выше идеальной последовательностью. Как и в предыдущем примере, ценность реальной последовательности k убывает с увеличением отклонения от идеальной последовательности по закону

$$E_k = E_i r^{-F} = E_i r^{-d(k,i)}. \quad (8.57)$$

В конкретной модели Эйген (Eigen, 1976) выбрал $r = 2,7$, а скорость мутаций определил с помощью соотношения

$$A_{kl} = (1 - q)^{d(k,l)}. \quad (8.58)$$

Таблица 8.4. Компьютерная модель процесса эволюции по Эйгену (*Eigen*, 1976)

Поколение	Частота ошибки	Лучшая последовательность	Частота
0	0,001	KORN AUS DEN FELDERN*)	10
32	0,001	KORN AUS DEN FEDDERN	9
69	0,001	LURN AUS DEN FEXLERN	7
0	0,01	KORN AUS DEN FELDERN	10
1	0,01	KSRN AUS DEN FEL?ERN	1
11	0,01	LERN AUS DEN FEHLERN	1
15	0,01	LERN AUS DEN FEHLERN	3
32	0,01	LERN AUS DEN FEHLERN	4
0	0,02	KORN AUS DEN FELDERN	10
23	0,02	LERN IUS DER FEHLER	1
0	0,03	KORN AUS DEN FELDERN	10
11	0,03	?ARA GUY??!NCVENTUNA	1

*)Зерно с полей

Здесь q — вероятность правильного вписывания одной буквы, и, соответственно, $W = 1 - q$ — вероятность ошибки. В табл. 8.4 представлены результаты проведенной Эйгеном игры. Приведенные данные позволяют прийти к следующим выводам.

- 1. Слишком малые частоты ошибок медленно формирует эволюционный процесс, частота прогресса остается малой.
- 2. Слишком высокие частоты ошибок приводят к разрыву последовательностей. Это происходит, как только превышает некий порог ошибки, определяемый множеством символов.
- 3. Благоприятные условия для эволюции соответствуют уровню, лежащему чуть ниже этого порога.

В той же работе Эйген исследовал интересную модель с гиперциклическими связями между воспроизводящимися последовательностями. Определенный недостаток компьютерных моделей, исследованных Ферстерлингом, Куном и Тьюзом, а также Эйгеном, состоит в том, что должна быть задана «цель эволюции» — идеальная последовательность. Разумеется, для реальных процессов эволюции идеальная последовательность, если она вообще существует, не задана. Существование априорно заданной цели эволюции сделало бы излишним сложный и сопряженный с большими затратами процесс поиска, поскольку был бы известен кратчайший путь к достижению цели. Именно поэтому представляют интерес модели эволюции, которые не исходят из предположения о существовании заданной идеальной последовательности. Такие модели были развиты в более давних работах (*Ebeling, Feistel*, 1974, 1977; *Ebeling, Mahnke*, 1979). Изложим наиболее существенные особенности этих моделей.

Рассмотрим последовательности из букв A, B, C и D длиной до 20 букв, т. е.

$$1 \leq \nu \leq 20, \quad S = 4 + 4^2 + 4^3 + \dots + 4^{20} \approx 10^{12}.$$

Приведем несколько примеров таких последовательностей

A, DA, DAC, DABC, BDADB.

Общее число возможных последовательностей S необычайно велико. Компьютерная игра проводится на симплексе $N = 100$. Стохастические переходы на симплексе происходят по правилам, вытекающим из общего соотношения (8.55). Технически каждая последовательность хранится в сжатом виде в «регистре» длиной в 20 символов. Например, две последние последовательности хранятся в форме

DABC0000000000000000, BDADB0000000000000000.

Каждой из 100 участвующих в игре последовательностей следует поставить в соответствие такой «регистр». Опишем теперь процесс мутации, состоящий либо в перестановке отдельных букв, либо в обрыве какой-то из начинающейся слева последовательности символов. Алгоритм мутации в компьютерной игре состоит из следующих правил.

1. Вписать в одну из ячеек любого из 100 регистров одну из четырех букв А, В, С, D или нуль.

2. Переставить циклически все ячейки, т.е. расположить их в таком порядке, чтобы ячейка 1 шла за ячейкой 2, ячейка 2 — за ячейкой 3,..., ячейка 20 — за ячейкой 1.

3. Вычеркнуть все нули и все части последовательности, стоящие справа от нуля. Возникающая после этих операций последовательность представляет собой результат мутации.

Нетрудно видеть, что в этой игре перестановка буквы в 20-ю ячейку эквивалентна приписыванию этой буквы к левому концу последовательности. Возможны также и перестановки буквы в нулевую позицию справа от последовательности или приписывание буквы к правому концу последовательности. Таким образом, алгоритм мутации допускает удлинение и сокращение цепей с обеих сторон. Для оценки каждой последовательности произвольно устанавливается код из перекрывающихся дублетов. Исходные позиции А, В, С, D оцениваются величиной $w = 1, 2, 3, 4$, а каждая последующая позиция оценивается в зависимости от ее ближайшего соседа по следующей таблице:

AA $w := w$,	CA $w := w/3$,
AB $w := w$,	CB $w := w + q/3$,
AC $w := w + 1$,	CC $w := w$,
AD $w := w/q$,	CD $w := w/3$,
BA $w := w + q/2$,	DA $w := w + 4/q$,
BB $w := w$,	DB $w := w/4$,
BC $w := w/2$,	DC $w := w + q/4$,
BD $w := w + 2$,	DD $w := w$.

Если p — позиция первой буквы дублета, то $q = p - g[(p - 1)/g]$, где квадратные скобки обозначают целую часть числа. Приведем несколько примеров:

$$w(\text{DAC}) = 4 + 4 + 1 = 9,$$

$$w(\text{DACBA}) = 4 + 4 + 1 + 1 + 2 = 12,$$

$$w(\text{BDADBBA}) = (2 + 2 + 4/2) \cdot (1/3) \cdot (1/4) \cdot 1 + 6/2 = 3,5.$$

На рис. 8.10, 8.11 и 8.12 представлены три различных примера компьютерных реализаций эволюционного процесса при различных частотах мутации. Отчетливо видно, как в игру эволюции вступают новые, более приспособленные сорта, доминируют в течение какого-то периода и затем уступают место еще лучшим сортам. При таком процессе длина цепи постоянно растет, и образуются характерные структуры. На рис. 8.13 показано несколько путей эволюции, по которым могло бы пойти развитие системы. Возникает характерное дерево эволюции со все более сильно ветвящейся кроной. На рис. 8.13 представлена лишь часть возможных маршрутов, соответствующих наиболее часто проходимым ветвям. В действительности к последовательностям ведут 262 144 ветвей. Усиливающееся ветвление является характерной особенностью рассматриваемой модели, в то время как модели Ферстерлинга, Куна и Тьюза (*Försterling, Kuhn, Tews, 1972*), а также Эйгена (*Eigen, 1976*), основанные на задании идеальной последовательности (цели эволюции), должны

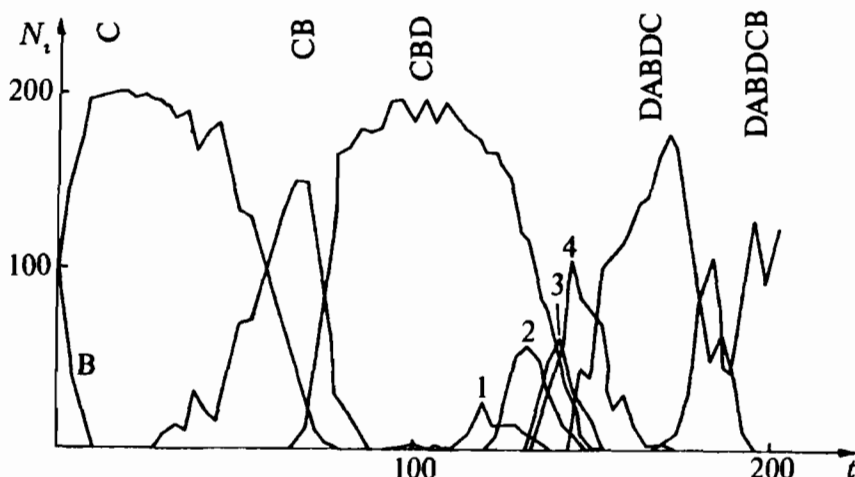


Рис. 8.10. Компьютерная реализация стохастического процесса отбора (тип I) с частотой мутаций 0,01 на репликацию и признак (1 — CBDD; 2 — CBDDC; 3 — CBVDC; 4 — CBDDCB)

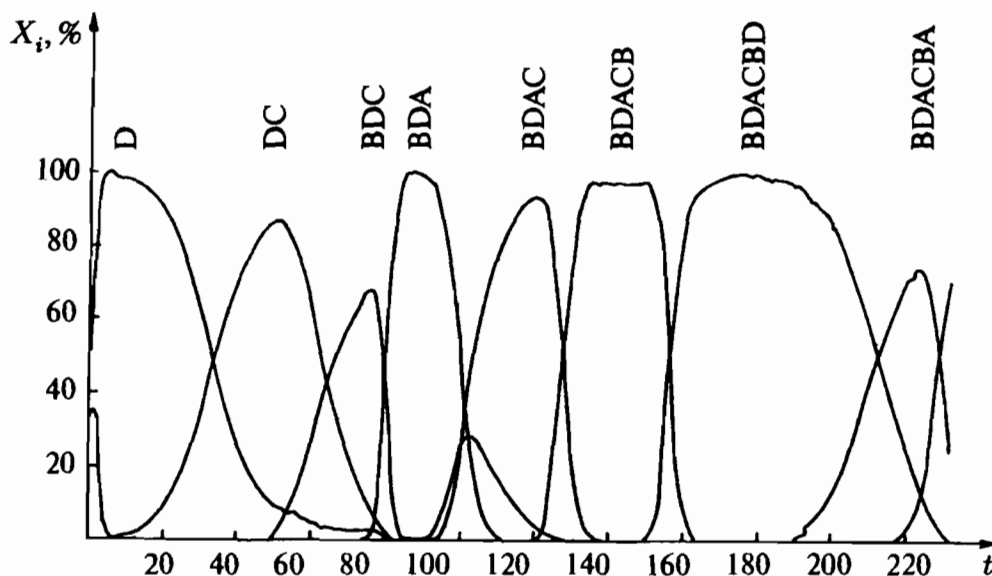


Рис. 8.11. Компьютерная реализация детерминистического процесса отбора со стохастическим появлением мутаций с частотой 0,03 на репликацию и признак (тип I)

сходиться к этой последовательности. В нашей модели процесс протекает на ветви дерева, которая ветвится все чаще по мере подъема от корня дерева. Переход между различными ветвями дерева вследствие глубоких провалов между соответствующими селекционными ценностями относительно маловероятен, но отнюдь не исключаются полностью. Из-за многочисленных ветвей система, как правило, не может достигнуть вершины дерева, так как с высокой вероятностью вершина не лежит на той ветви, по которой происходит эволюция системы. Тенденция направлена на достижение относительного преимущества, эволюция протекает под девизом «все выше и выше», но никакой заранее заданной цели процесса не существует.

Если эволюция последовательностей в моделях Ферстерлинга и др. (Försterling et al., 1972) и Эйгена (Eigen, 1976) происходит при фиксированной длине цепей, то «высота» на основе нашей системы оценок связана с увеличением длины последовательностей. Поэтому в нашей модели изменяется процентное отношение (доля) ошибочных копий, т. е. как E_i (см. соотношение (8.14)), так и Q_i (см. соотношение (8.11)) изменяют свое значение от последовательности к последовательности. Это отчетливо видно на рис. 8.14: с увеличением длины возрастает как селекционная ценность E_i , так и число производимых мутантов соответствующей штаммовой

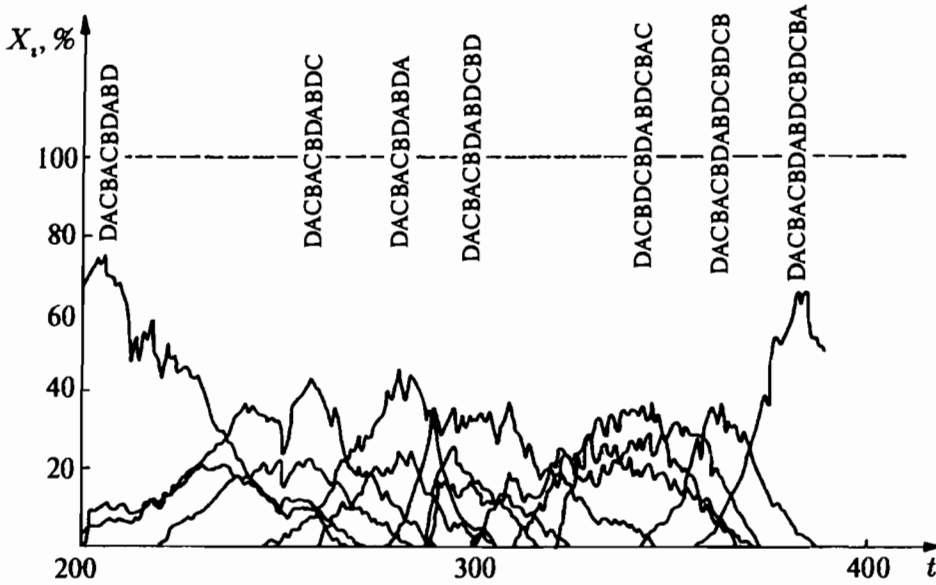


Рис. 8.14. Мутационная катастрофа: продолжение процесса, начало которого представлено на рис. 8.12, приводит к хаотическому распределению последовательностей

процесс завершается «мутационной катастрофой». Возникает спектр последовательностей, ни одна из которых не образуется с устойчивой частотой. Нестабильность частот связана с приближением к эйгеновской предельной длине (8.4). Чтобы достичь надежной репликации длинных последовательностей, частота ошибки на одну букву должна убывать обратно пропорционально длине последовательности. Известно, что в пребиологической эволюции повышение точности репликации, например, в результате появления репликаз, должно было иметь центральное значение.

В нашей игре в имитацию типа II мы учитываем необходимое повышение точности репликации и вместо фиксированной частоты мутации *MR* используем переменную частоту мутации

$$m = \frac{MR}{1 + \nu/8}.$$

Затем в этой игре в имитацию исследованию подлежит новый «фрустрированный» тип вычисления значений. Очень простое по сравнению с типом I правило имеет следующий вид:

$$E_i = w_i, \tag{8.60}$$

где *w* = 1, 2, 3, 4 надлежит подставить вместо букв A,B,C,D на первом месте по следующему алгоритму:

IF *a*\$(*p* TO *p* + 1) = AB, BC, CD, DA LET *w* := *w* + 1,
 IF *a*\$(*p*) = *a*\$(*p* + *g*) LET *w* := *w* + *b*.

Этот «рецепт» ставит каждой паре букв, расположенных в алфавитном порядке, приращение текущего значения на одну единицу (пара букв DA считается расположенной в алфавитном порядке), а периодичности длиной *g* — приращение *b*. В то время как в игре в имитацию типа I сложность последовательности в конечном счете следует из сложности правил вычисления значения, при игре в имитацию типа II она следует из сложности правил вычисления только в том случае, если два простых частичных правила противоречат друг другу (фрустрация).

Первое из двух частичных правил отдает предпочтение последовательностям типа

DABCDABCDABCDABCDABC . . . ,

в то время как второе частичное правило отдает предпочтение последовательностям с периодом g . При $g \neq 4$ может быть выполнено лишь одно из двух требований, предъявляемых к оптимальности. Например, если положить $g = 5$, то после 4-й позиции образуется ошибочный порядок, например,

DABCCDABCCDABCCDABCC

Если длина равна 20, то период 4 имеет значение $w = 23$, а период 5 с ошибочным порядком — значение $w = 19 + 15b$. Критическое значение для доминирования одного или другого правила определяется величиной

$$b_{\text{крит}} = \frac{l}{g(l - g)} \quad \text{при } l > g,$$

или в случае длинных последовательностей — величиной $b_{\text{крит}} = 1/g$. Для генерирования более сложных последовательностей параметр b полезно поддерживать на критическом значении: $b = b_{\text{крит}}$. Это позволяет достигать максимального эффекта фрустрации. Более строгие соображения относительно сложного эффекта фрустрации, играющего важную роль в процессе эволюции, приведены в гл. 10.

В табл. 8.5 приведены некоторые результаты численного моделирования при $g = 5$ и $b = 1/5$. Интересно отметить, что сложность полученных последовательностей, несмотря на очень простые правила, не уступает сложности последовательностей, образующихся в случае игры в имитацию.

В третьем варианте имитации (или в игре в имитацию типа III) правила игры смешанные: они представляют собой комбинацию правил I и II типа. Выбранное значение равно

$$E_i = w_i,$$

где $w = 1, 2, 3, 4$ для начальных («заглавных») букв A, B, C, D. Правила вычисления значений последовательностей букв сводятся к следующим.

IF	$a\$(p \text{ TO } p + 1) = AB$	LET	$w := w + 1,$
IF	$a\$(p \text{ TO } p + 1) = AD$	LET	$w := w - 1/(g + 1 - q),$
IF	$a\$(p \text{ TO } p + 1) = BA$	LET	$w := w - 1/3,$
IF	$a\$(p \text{ TO } p + 1) = BC$	LET	$w := w + q/3,$
IF	$a\$(p \text{ TO } p + 1) = BD$	LET	$w := w - 1/(g + 1 - q),$
IF	$a\$(p \text{ TO } p + 1) = CA$	LET	$w := w + q/4,$
IF	$a\$(p \text{ TO } p + 1) = CB$	LET	$w := w - 1/2,$
IF	$a\$(p \text{ TO } p + 1) = CD$	LET	$w := w + 2,$
IF	$a\$(p \text{ TO } p + 1) = DA$	LET	$w := w + 1,$
IF	$a\$(p \text{ TO } p + 1) = DB$	LET	$w := w/4,$
IF	$a\$(p \text{ TO } p + 1) = DC$	LET	$w := w + q/4,$
IF	$a\$(p) = a\$(p + g)$	LET	$w := w + b.$

В табл. 8.6 представлены результаты нескольких вариантов численного моделирования при $g = 5$ и $b = q = p - g [(p - 1)/g]$ при постоянных частотах мутации.

Имитация эволюции последовательностей с помощью численных экспериментов сопряжена с относительно большой затратой компьютерного времени, но на первых этапах численное моделирование широко использовалось для достижения понимания наиболее существенных особенностей реальной эволюции (Swetina, Schuster, 1982; Anderson, Stein, 1984; Schuster, 1987; Fontana, Schuster, 1988; Kaufmann, 1993; Редько, 1986). Наконец, для более глубокого понимания эволюционных моделей был предпринят также ряд аналитических исследований (McCaskill 1984; Demetrius, 1983; Demetrius et al., 1985; Rumschitzky, 1986).

Таблица 8.5. Игра в имитацию с фрустрацией (тип II) для 100 последовательностей и 3 зависящих от длины частот мутаций (для максимальной длины $\nu = 32$) В каждом случае приведены 5 «наиболее ценных» последовательностей с их селекционными ценностями (в скобках) и число их представителей

$t = 0$	$100 \times \text{ABCCDABC}$	$(E = 7, 6)$	$100 \times \text{ABCCDABC}$	$(E = 7, 6)$	$100 \times \text{ABCCDABC}$	$(E = 7, 6)$
$t = 1$	$1 \times \text{CBCCDABC}$ $79 \times \text{ABCCDABC}$ $8 \times \text{ABCDDABC}$ $2 \times \text{BBCCDABC}$ $3 \times \text{BBCCBABC}$	$(8, 4)$ $(7, 6)$ $(7, 6)$ $(7, 4)$ $(5, 4)$	$13 \times \text{DABCCDABCC}$ $46 \times \text{DABCCDABC}$ $7 \times \text{DABCCDABCB}$ $4 \times \text{DABCCDABCA}$ $1 \times \text{DABCCDABAC}$	$(12, 0)$ $(11, 8)$ $(11, 8)$ $(11, 8)$ $(10, 8)$	$4 \times \text{CABCCDABC}$ $1 \times \text{CBCCDABC}$ $3 \times \text{ABCCDABCC}$ $5 \times \text{ABCCDABCA}$ $1 \times \text{BBCCDABC}$	$(9, 6)$ $(8, 4)$ $(7, 8)$ $(7, 6)$ $(7, 4)$
$t = 2$	$19 \times \text{DBCCDABC}$ $1 \times \text{CBCCDABC}$ $57 \times \text{ABCCDABC}$ $11 \times \text{ABCDDABC}$ $2 \times \text{ABCCDABC}$	$(9, 4)$ $(8, 4)$ $(7, 6)$ $(7, 6)$ $(7, 4)$	$1 \times \text{DCDABCCDABCCC}$ $2 \times \text{CDABCCDABCCC}$ $39 \times \text{DABCCDABC}$ $1 \times \text{DABCCDABCA}$ $5 \times \text{DABCCDABC}$	$(13, 2)$ $(12, 2)$ $(11, 8)$ $(11, 8)$ $(10, 8)$	$4 \times \text{DABCCDABCA}$ $5 \times \text{CABCCDABCC}$ $5 \times \text{DABCCDDDBCA}$ $3 \times \text{CABCCDABCB}$ $3 \times \text{DABCCDDDBCA}$	$(11, 8)$ $(9, 8)$ $(9, 6)$ $(9, 6)$ $(9, 6)$
$t = 3$	$42 \times \text{DBCCDABC}$ $5 \times \text{BDCBCCDABC}$ $5 \times \text{DBCCDABCD}$ $1 \times \text{DBCCDABCB}$ $25 \times \text{ABCCDAABC}$	$(9, 4)$ $(9, 4)$ $(9, 4)$ $(9, 4)$ $(7, 6)$	$2 \times \text{DABCCDABCB}$ $2 \times \text{CDABCCDABC}$ $27 \times \text{DABCCDABC}$ $16 \times \text{DDABCCDABC}$ $5 \times \text{DDABCCDABCB}$	$(12, 2)$ $(12, 0)$ $(11, 8)$ $(11, 8)$ $(11, 8)$	$1 \times \text{DABCCDABCD}$ $17 \times \text{DABCCDABCA}$ $1 \times \text{DABCCDABDA}$ $2 \times \text{DABCCDABCA}$ $1 \times \text{DADACDABCA}$	$(12, 8)$ $(11, 8)$ $(11, 6)$ $(10, 8)$ $(10, 4)$
$t = 4$	$1 \times \text{DBCCDABCD}$ $1 \times \text{DBCCDABCC}$ $58 \times \text{DBCCDABC}$ $23 \times \text{BCDDBCCDABC}$ $6 \times \text{ABCCDABCD}$	$(10, 4)$ $(9, 6)$ $(9, 4)$ $(9, 4)$ $(8, 8)$	$4 \times \text{DDABCCDABCD}$ $2 \times \text{DABCCDABCD}$ $43 \times \text{DABCCDABC}$ $2 \times \text{DABCCDABCB}$ $1 \times \text{DABCCDABC}$	$(12, 8)$ $(12, 8)$ $(11, 8)$ $(11, 8)$ $(11, 8)$	$1 \times \text{DABCCDABCCB}$ $1 \times \text{DABCCDABCC}$ $37 \times \text{DABCCDABCA}$ $3 \times \text{DABCCDABCA}$ $1 \times \text{DABCCDABCAD}$	$(12, 0)$ $(12, 0)$ $(11, 8)$ $(11, 8)$ $(11, 8)$
$t = 5$	$11 \times \text{DBCCDABCD}$ $1 \times \text{DBCCDABCC}$ $49 \times \text{DBCCDABC}$ $17 \times \text{BCDDBCCDABC}$ $2 \times \text{DBCCDABCC}$	$(10, 4)$ $(9, 6)$ $(9, 4)$ $(9, 6)$ $(9, 4)$	$1 \times \text{CDABCCDABCCD}$ $2 \times \text{DABCCDABCD}$ $22 \times \text{DABCCDABC}$ $3 \times \text{DDABCCDABC}$ $13 \times \text{DABCADABC}$	$(13, 4)$ $(12, 6)$ $(11, 8)$ $(11, 8)$ $(10, 8)$	$5 \times \text{DABCCDABCCD}$ $1 \times \text{DABCCDABCB}$ $1 \times \text{DABCCDABCB}$ $1 \times \text{DABCCDABCAD}$ $7 \times \text{DABCCDABCA}$	$(12, 8)$ $(12, 8)$ $(12, 6)$ $(12, 0)$ $(11, 8)$

Таблица 8.6. Игра в имитацию типа III (100 последовательностей, 3 различные частоты мутаций MR при общей длине последовательностей $\nu = 32$)

	$(MR) = 0,03$	$(MR) = 0,1$	$(MR) = 0,3$
$t = 0$	100 × ABCDABCD	100 × ABCDABCD	100 × ABCDABCD
$t = 1$	1 × CBCDABCD 98 × ABCDABCD 1 × ABCBABCD	1 × BABCDABCD 84 × ABCDABCD 5 × ABCDABCA	1 × CDABCDABCD 21 × DABCDABCD 6 × DABCDADCD
$t = 2$	1 × DABCDABCD 89 × CBCDABCD 10 × ABCDABCD	2 × CBCDABCD 98 × ABCDABCD	1 × DAABCDABCD 35 × DAABCDABCD 6 × DAABCDABCD
$t = 4$	81 × DABCDABC 6 × ABCDABCD 6 × CBCDABCD	1 × DABCDABCD 7 × AABCDABCD 78 × ABCDABCD	1 × CDACBCDABCD 18 × CDACBCDABCD 22 × DDAABCDABCD
$t = 7$	99 × DABCDABCD 1 × ABCDABCD	1 × DAAABCDABCD 61 × CABCDABCD 26 × ABCDABCD	1 × DCDACDCDABCD 29 × DCDACDCDABCD 40 × DCDACACDABCD
$t = 10$	99 × DABCDABCD 1 × ABCDABCD	1 × DABCDABCDAD 45 × CCABCDABCD 53 × CABCDABCD	1 × DACDACDCDABCD 47 × DACDACDCDABCD 4 × DACDACDCCABCD
$t = 13$	99 × DABCDABCD 1 × DABCDACCD	9 × DABBCDABCD 38 × DABCDABCD 23 × CABCDABCD	2 × DACDACDCDABCD 34 × DACDACDCDABCD 18 × DACDACBCDABCD

8.5. Инновации в сложных системах

Механизмы мутации играют важную роль не только в процессах биологической эволюции, но и всюду, где только происходят эволюционные процессы — в технической эволюции, развитии языков, поведения, общественных отношений, процессах, протекающих в нашем сознании, и т. д. Правда, «мутации», происходящие в сложных системах, объединяет с мутациями полимеров только общий принцип: новые структуры возникают в системах стохастически и подвергаются проверке, причем из новых свойств отбираются только такие, которые соответствуют «прогрессу». Еще Гегель, Маркс и Дарвин видели в возникновении новых свойств самую суть эволюции. И в наше время этот аспект занимает центральное место в любой научной теории эволюции. Существование достаточно обширного поля еще неиспользованных возможностей является неременной предпосылкой каждого истинно эволюционного процесса. Имеющиеся возможности не обязательно должны реализоваться в материальной форме, и это доказано, например, в случае биологической эволюции; они вполне могут возникать и подвергаться отбору как чисто мысленные возможности. Именно такие возможности обусловили необычайно высокий темп эволюции после появления на Земле гомо сапиенса. В частности, в основе эволюции наук по существу лежит такая «мутационная стратегия», хотя имеются и другие факторы. «Теоретическая мутация» выступает как новый, причем весьма эффективный, метод наряду с «реальной мутацией», хотя и не может полностью заменить последнюю. Примерами, подтверждающими это, могут служить экспериментальные проверки новых теорий, испытания новых изобретений, товаров, методов обучения, планов преподавания различных дисциплин, проверка на практике организационных структур и т. д. На всех ступенях организации способность системы порождать инновации, позволяющие отбирать и распространять благоприятные варианты, является решающей предпосылкой ее способности эволю-

ционировать. Системы, участвующие в эволюционном процессе, должны порождать инновации ценой своего существования, хотя новые благоприятные возможности открываются гораздо реже, чем неблагоприятные возможности. Отклонение инноваций лишь на первый взгляд кажется оптимальной тактикой, но как стратегия оно губительно, о чем отчетливо свидетельствует биологическая эволюция.

Как будет показано в дальнейшем, развитый в разд. 5.2 и 8.3 формализм чисел заполнения является удобным теоретическим базисом для описания процессов эволюции в сложных системах. Правда, для этого теорию необходимо определенным образом обобщить. Как и прежде, мы будем исходить из основополагающего предположения о том, что множество структур, возникающих в ходе подлежащего описанию процесса эволюции, счетно. В этом случае каждой структуре можно поставить в соответствие натуральное число

$$i = 1, 2, 3, \dots$$

В случае биологической эволюции различные i соответствуют генотипам, т. е. определенным последовательностям нуклеотидов, которым в качестве номеров (натуральных чисел) можно присвоить гёделевские номера (см. разд. 11.3). Предположение о перечислимости (счетности) множества возникающих структур, несомненно, является ограничением, тем не менее счетные множества охватывают все структуры, допускающие описание с помощью последовательности букв некоторого алфавита, например, алфавита компьютера — 0 и 1 или латинского алфавита $a, b, c, \dots, z$. Класс таких структур весьма широк. Мы предполагаем также, что в рассматриваемой эволюционной системе имеется N_i тождественных подсистем структуры i . Величины N_i называются числами заполнения системы. Полный набор чисел заполнения $N_1, N_2, \dots$ характеризует состояние эволюционной системы. Если величины N_i отложить по осям координат, то состояние системы можно представить как точку в некотором многомерном пространстве — пространстве чисел заполнения. Поскольку числа заполнения $N_i(t)$ зависят от времени, точка, соответствующая состоянию системы, движется в пространстве чисел заполнения, или, точнее говоря, совершает скачки вследствие дискретного характера движения. Предположим, что в ходе элементарного процесса одновременно может изменяться только одно число заполнения, а в ходе элементарного процесса перехода — самое большее два числа заполнения, т. е. что могут представиться только следующие случаи:

$$N_i \rightarrow N_i + 1, \quad N_i \rightarrow N_i - 1, \quad N_i, N_j \rightarrow N_i + 1, N_j - 1.$$

Следует также предположить, что рассматриваемый процесс эволюции имеет марковский характер (относительно пространства чисел заполнения) и что вероятности переходов в единицу времени для указанных выше элементарных процессов могут быть представлены следующим образом.

1. Самоорганизация

$$\begin{aligned} W(\dots, N_i + 1, \dots, N_j, \dots, N_k, \dots \mid \dots, N_i, \dots, N_j, \dots, N_k, \dots) = \\ = A_i^{(1)} N_i + A_{ij}^{(1)} N_i N_j + B_{ij}^{(1)} N_i N_j + C_{ijk}^{(1)} N_i N_j N_k. \end{aligned} \quad (8.61)$$

2. Распад

$$\begin{aligned} W(\dots, N_i - 1, \dots, N_j, \dots, N_k, \dots \mid \dots, N_i, \dots, N_j, \dots, N_k, \dots) = \\ = A_i^{(2)} N_i + B_{ij}^{(2)} N_i N_j + B_{ii}^{(2)} N_i (N_i - 1) + B_{ik}^{(2)} N_i N_k. \end{aligned} \quad (8.62)$$

3. Превращение

$$W(\dots, N_i + 1, \dots, N_j - 1, \dots, N_k, \dots \mid \dots, N_i, \dots, N_j, \dots, N_k, \dots) = \\ = A_{ij}^{(3)} N_j + B_{ij}^{(3)} N_i N_j + C_{ijk}^{(3)} N_i N_j N_k. \quad (8.63)$$

Эти простые полиномиальные выражения получаются из формул для вероятностей переходов, выведенных в разд. 5.2 и 6.2. Напрашиваются дальнейшие обобщения путем учета членов более высокого порядка. Коэффициенты в правых частях соотношений (8.61)–(8.63) следует рассматривать как функции общего числа частиц, объема и других параметров системы. Отдельные члены в правой части равенства (8.61) соответствуют спонтанному самовоспроизведению, воспроизведению с ошибками и каталитическому воспроизведению, в то время как члены в правой части равенства (8.62) описывают самопроизвольный распад, распад из-за внешних помех и самопомех, а члены в правой части равенства (8.63) — мутационные переходы, а также «нацеленные» переходы. Разумеется, точный смысл членов зависит от специфики эволюционной проблемы.

Следующие примеры показывают общность введенных моделей. Подчеркнем, что речь идет лишь о чисто формальных рамках, которые могут позволить плодотворно использовать те или иные конкретные аналогии.

1. Самовоспроизведение биополимеров в модели Эйгена

Если положить

$$A_{ij}^{(3)} = a_{ij}, \quad B_{ij}^{(3)} = \frac{1}{N} e_i, \quad C_{ijk}^{(3)} = \frac{1}{NV} b_{ik}, \quad (8.64)$$

а все остальные коэффициенты равными нулю, то получится в точности та модель, которая была рассмотрена в разд. 8.3.

2. Экономическое производство при капитализме в условиях свободной конкуренции

В некоторых работах предпринималась попытка использовать изложенную выше модель для формализации теории Маркса эволюции системы капиталистических производителей в условиях свободной конкуренции (Feistel, 1977; Ebeling, 1977, 1978). Основные идеи такого подхода сводились к следующему. Перенумеруем производителей определенного продукта последовательными натуральными числами $i = 1, 2, \dots$ и обозначим через N_i число товаров, которые может купить i -й производитель. Тогда движение товаров может быть описано следующим выбором коэффициентов:

$$A_i^{(1)} = \alpha M_i, \quad B_{ij}^{(3)} = \beta \kappa_i, \quad A_{ij}^{(3)} = A_{ij}. \quad (8.65)$$

Все остальные коэффициенты равны нулю. Величина M_i зависит от индивидуальной прибыли i -го производителя, а κ_i — от стоимости производимого им товара. При этом стоимость понимается в смысле теории Маркса — как мера заключенного в продукте абстрактного труда. Коэффициент A_{ij} , задающий частоту мутаций, описывает при таком подходе возникновение производителей, работающих по новым технологиям. Эта модель приводит к среднему значению, задаваемому формулой (6.145).

3. Возникновение и разработка технологических новшеств (инноваций)

Как показано в работе Хименеса-Монтаньо и Эбелинга (Jiménez-Montaño, Ebeling, 1980), процесс технологической эволюции при определенных предположениях

может быть описан точно таким же образом. В этом случае индекс i соответствовал бы определенной технологии, а N_i — числу предприятий, использующих эту новую технологию. Коэффициенты выбираются следующим образом:

$$A_i^{(1)} = A_i, \quad A_i^{(2)} = D_i, \quad A_{ij}^{(3)} = A_{ij}; \quad B_{ij}^{(2)} = (IM)A_i. \quad (8.66)$$

Все остальные коэффициенты равны нулю. Кроме того, A_i — коэффициент воспроизводства; (IM) — коэффициент имитации; D_i — постоянная распада и A_{ij} — мера возникновения новых технологий в результате научных исследований и конструкторских разработок.

4. Миграция ученых из одной области науки в другую

Если разделить сферу научных исследований на области и перенумеровать последние натуральными числами $i = 1, 2, 3, \dots$, то процесс перемещения ученых из одной области в другую протекает так же, как описанная выше динамика инноваций (*Bruckner et al.*, 1989). Пусть N_i — число ученых, работающих в i -й области. Тогда коэффициент A_i выражает скорость самовоспроизведения области, т. е. скорость получения образования молодыми учеными, собирающимися посвятить себя работе в соответствующей области. Коэффициент D_i описывает уход от научной деятельности. Коэффициенты $A_{ij}^{(3)}$ характеризуют новые научные открытия (освоение ранее не известных областей), а $B_{ij}^{(2)}$ — соответствующие наметившейся тенденции переходы в «современные» области науки.

Изложенные выше в общих чертах модели сложных процессов носят лишь формальный характер, а их биологическое, экономическое или науковедческое содержание оказывается запрятанным в значениях коэффициентов. Авторы убеждены, что такие модели позволят обнаружить аналогии между различными дисциплинами и тем самым установить между ними соответствие.

Глава 9

Эволюция в пространстве фенотипов

Всякий раз бывает приятно по-новому взглянуть на старые вещи.

Ричард П. Фейнман

9.1. Адаптивный ландшафт

Объектом ранних фаз биологической эволюции были выполнявшие роль носителей информации макромолекулы, отличавшиеся последовательностью входивших в них оснований. Мутации приводили к ошибочной репликации этой последовательности, т. е. к перестановке различных оснований. Приспособленность таких полинуклеотидных цепей определяется их химическими свойствами, а значит, непосредственно последовательностью мономеров и взаимодействием между ними (см. разд. 8.1 и 10.3). Таким образом, отбор и мутация действуют на одном и том же — молекулярном — уровне.

По мере продвижения по ступеням эволюции на ходе процесса репродукции начинает сказываться каталитический белок (механизм влияния белка может описываться, например, гиперциклом, о котором мы упоминали выше). Свойства такого белка имеют первостепенное значение для приспособленности вида, поэтому характерные особенности, скрытые в полинуклеотидах, в этом случае уже не определяются непосредственно их химическими свойствами. Однако мутационные процессы продолжают протекать на уровне последовательности оснований в нуклеотидах, так как случайные ошибки при синтезе ферментов не исключаются. В ходе эволюции такая тенденция развития проявляется все более отчетливо и с возникновением биологических видов приводит к полному отделению уровня функционирования от уровня отбора и мутации. Приспособленность вида лишь в весьма узких пределах определяется химической стабильностью последовательности оснований в ДНК. Способность особей к выживанию в ходе отбора в гораздо большей степени зависит от так называемых фенотипических свойств: размеров, веса, остроты зрения, быстроты и т. д. Как случайные ошибки при репликации последовательности оснований в ДНК мутации по существу не зависят от фенотипических свойств.

В этой главе мы, исходя из рассмотренных выше моделей ранней фазы эволюции и с учетом всех отмеченных особенностей, построим простую математическую модель последующих, более поздних, стадий биологической эволюции и проанализируем ход решений, допускаемых нашей моделью.

Поскольку фенотипические свойства вида в общем случае изменяются непрерывно, разумно ввести некое гипотетическое пространство (так называемое пространство фенотипа), по осям координат которого отложены все фенотипические свойства $q_1, \dots, q_d$, необходимые для однозначного задания вида. Например, в качестве первой координаты q_1 можно выбрать вес, в качестве второй — характерный линейный размер и т. д. Каждая точка $q = (q_1, \dots, q_d)$ пространства фенотипа соответствует вполне определенной комбинации всех указанных свойств. Большинство

таких комбинаций лишены смысла с биологической точки зрения, соответствующая им приспособленность равна нулю. Однако существуют особые комбинации фенотипических свойств, т. е. точки или небольшие области в пространстве фенотипа, гарантирующие приспособленность, достаточную для выживания. Такие «биологически осмысленные» фенотипы мы ставим в соответствие биологическим видам. Таким образом, центральной величиной в рассматриваемой нами модели является функция плотности $n(q, t)$, показывающая, с какой частотой фенотип q («вид» q) встречается в момент времени t . Эволюционные процессы отражаются в изменении функции плотности $n(q, t)$ во времени. Сравнивая такой подход с моделями, рассмотренными в гл. 5 и 6, мы видим, что функция плотности $n(q, t)$ представляет собой не что иное, как континуальное обобщение вектора с компонентами $n_i(t)$, $i = 1, \dots, N$, — концентрациями макромолекул различных сортов i . Вместо дискретного индекса i в нашей модели используется непрерывная переменная q .

Мы рассматриваем только гомогенные системы и, следовательно, пренебрегаем зависимостью плотности $n(q, t)$ от координат реального трехмерного пространства. Начнем с модели, не содержащей мутационных членов. В этом случае временная эволюция функции плотности $n(q, t)$ определяется только процессами самовоспроизведения, которые в самом общем случае могут быть описаны уравнением

$$\partial_t n(q, t) = w(q, t; n(q, t)) \cdot n(q, t). \quad (9.1)$$

При этом частота воспроизведения w изменяется не от вида к виду и не от одного

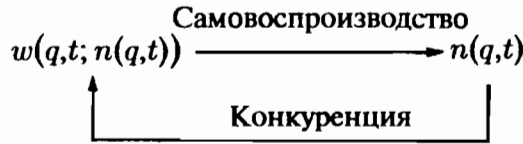


Рис. 9.1. Взаимозависимость темпа самовоспроизведения w и функции плотности n в моделях пространства фенотипов

момента времени к другому, т. е. зависит не только от каждой из переменных q и t в отдельности, но и от самой плотности $n(q, t)$. Именно последняя зависимость вводит в игру процессы конкуренции, так как дальнейшее развитие вида q зависит от наличия в системе другого вида q' . Мы видим, что зависимость между w и n имеет вид петли обратной связи (рис. 9.1). В качестве простых примеров выберем

$$w(q, t; n(q, t)) = E(q) - \frac{\int dq n(q, t) E(q)}{\int dq n(q, t)} \quad (9.2)$$

и

$$w(q, t; n(q, t)) = a(q) - \int dq' b(q, q') n(q', t). \quad (9.3)$$

Формула (9.2) соответствует континуальному обобщению модели Эйгена (6.21); при этом зависимость частоты самовоспроизведения $w(q, t; n(q, t))$ от распределения $n(q, t)$ ограничена аддитивной зависящей только от времени константой

$$\langle E \rangle = \frac{\int dq n(q, t) E(q)}{\int dq n(q, t)}. \quad (9.4)$$

Формула (9.3) представляет собой обобщение дискретной модели Вольтерры—Лотки (6.121). Обратное воздействие плотности $n(q, t)$ на $w(q, t; n(q, t))$ определяется ядром интеграла $b(q, q')$ и уже в силу одного этого носит нелокальный характер.

Как и в дискретном случае, ядро $b(q, q')$ описывает взаимодействие между видами q и q' (Ebeling, Feistel, 1985; Ebeling, Schimansky-Geier, 1988). Качественно ход решения уравнения (9.1) может быть наглядно прослежен в частном случае (9.2). В областях пространства фенотипа, соответствующих большим значениям $E(q)$, плотность $n(q, t)$ возрастает быстрее, чем в других областях. Однако с другой стороны, как и в дискретном случае, выполняется соотношение

$$\partial_t N = \partial_t \int dq n(q, t) = 0 \quad (9.5)$$

(см. равенство (9.4)), содержащее динамику, а также общее число особей, поэтому распределение $n(q, t)$ все больше стягивается к максимуму функции приспособленности $E(q)$. Поэтому при $t \rightarrow \infty$ мы получаем:

$$n(q, t) \rightarrow N \cdot \delta(q - q_{\max}), \quad (9.6)$$

причем

$$E(q_{\max}) = \max_{q, n(q, 0) > 0} E(q). \quad (9.7)$$

Следовательно, выживает тот вид, который обладает наибольшей приспособленностью среди всех видов, представленных в начальном условии $n(q, 0)$. Такое «восхождение на пики в горном ландшафте приспособленности» локально можно описать эволюцией во времени положения «центра тяжести» распределения плотности $n(q, t)$, который определяется величиной

$$\langle q \rangle = \frac{1}{N} \int dq q n(q, t). \quad (9.8)$$

Уравнение (9.1) и соотношение (9.2) позволяют записать следующее уравнение:

$$\partial_t \langle q \rangle = \frac{1}{N} \int dq q E(q) n(q, t) - \langle q \rangle \langle E \rangle.$$

В предположении, что $E(q)$ изменяется на больших масштабах, чем те, в которых локализовано распределение $n(q, t)$, функцию $E(q)$, стоящую под интегралом, можно разложить в ряд относительно $\langle q \rangle$ и получить уравнение

$$\partial_t \langle q \rangle = \frac{\partial E}{\partial q}(\langle q \rangle) [\langle q \cdot q \rangle - \langle q \rangle^2]. \quad (9.9)$$

Как видно из уравнения (9.9), скорость дрейфа центра $\langle q \rangle$ локально пропорциональна градиентам функции приспособленности и разброса фенотипических свойств конкурирующих видов. Ускорение эволюции вследствие разнообразия фенотипов составляет содержание так называемого закона Фишера и давно известно (Fisher, 1930; Wright, 1932).

Включение мутаций в модель типа уравнения (9.1) требует принятия предположений о взаимосвязи между замещениями оснований в ДНК и возникающими в результате таких замещений фенотипическими изменениями. В совершенно общем случае мутации могут приводить к возникновению вида q из уже существующего вида q' и поэтому могут быть описаны путем включения в уравнение (9.1) члена вида

$$\int dq' [a(q, q') n(q', t) - a(q, q') n(q, t)]. \quad (9.10)$$

Функция $a(q, q')$ задает частоту перехода от вида q' к виду q . Хотя замещение даже одного основания в ДНК при определенных обстоятельствах может иметь далеко

идущие последствия для индивидуума и даже поставить под угрозу его существование, мы примем в качестве простого приближения предположение о том, что отношения наследственности для генотипа и фенотипа совпадают. Тогда небольшие мутации не соответствуют никаким изменениям в фенотипе, и функция $a(q, q')$ существенно отлична от нуля только на малых расстояниях $|q - q'|$. При этих условиях член (9.10) уравнения можно заменить диффузионным членом вида

$$\partial_t D(q) \partial_q n(q, t), \quad (9.11)$$

где коэффициент диффузии определяется выражением

$$D(q) = \frac{1}{2} \int dq' (q - q') a(q, q') (q - q') \quad (9.12)$$

(Feistel, Ebeling, 1982). Из уравнения (9.1) с учетом члена (9.11) получаем полную простую модель эволюции в пространстве фенотипов:

$$\partial_t n(q, t) = w(q, t; n(q, t)) \cdot n(q, t) + \partial_q D(q) \partial_q n(q, t). \quad (9.13)$$

Дополнительно введенный диффузионный член приводит к двум принципиально важным следствиям. Согласно одному из них, при $t \rightarrow \infty$ распределение не сходится к δ -функции, как в случае (9.6), поскольку у представленных в модели видов всегда имеется хвост мутантов. Как и в дискретном случае, вместо видов в конкурентную борьбу вступают определенные распределения видов — введенные еще в разд. 6.5 квазивиды. Во-вторых, мутации могут приводить к образованию новых видов, поэтому искать требуется не локальный максимум функции $E(q)$ по области пространства фенотипов, занимаемой при $t = 0$, как в (9.7), а ее глобальный максимум, к которому она стремится в ходе эволюционного процесса в результате последовательного отбора и ненаправленной диффузии. Это означает также, что центр $\langle q \rangle$ распределения $n(q, t)$ при наличии диффузионного члена не всегда сдвигается вдоль градиента функции приспособленности, как в случае (9.9), но может иногда перемещаться и против градиента. Тем самым возникает возможность появления локальных максимумов функции приспособленности $E(q)$, более благоприятных (соответствующих большему значению $E(q)$), чем другие максимумы. Этот аспект представляет основной интерес при применении стратегий эволюции в ходе процесса оптимизации и будет подробно рассмотрен нами позднее.

Существенной проблемой в рассматриваемой модели остается задание явного вида функций $w(q, t; n(q, t))$, $E(q)$, $a(q)$ и $b(q, q')$. Точная зависимость функции приспособленности $E(q)$ или взаимодействия $b(q, q')$ даже для простейших биологических видов зависит от фенотипических свойств q и q' очень сложным образом и известна лишь в общих чертах. Во всяком случае можно ожидать, что эти функции обладают разнообразной структурой и имеют много минимумов и максимумов, соединенных между собой сложной весьма высокой размерности сетью седловых точек. Вопрос о том, как надлежит в общем случае характеризовать такого рода ландшафт, мы будем рассматривать и в следующей главе. В качестве первой попытки преодолеть такого рода трудности можно воспользоваться методом, давно и успешно используемым в физике твердого тела для исследования неупорядоченных систем. В некотором смысле ситуация там сравнима с нашей. Для вычисления плотности электронных состояний легированного примесью или аморфного твердого тела в принципе необходимо знать явный вид электронного потенциала $U(\mathbf{r})$. Однако такого рода информацией физики не располагают, так как порождающая электронный потенциал неупорядоченная мозаика из большого числа твердых тел приводит к тому, что этот потенциал как функция от $\mathbf{r}$ имеет очень сложный вид. При попытке зондирования детальной структуры электронного потенциала результат каждого

измерения оказывается иным. В качестве возможного выхода в физике твердого тела вместо неизвестного потенциала $U(\mathbf{r})$ используют реализацию $V(\mathbf{r})$ по ансамблю стохастических потенциалов $\{V(\mathbf{r})\}$, характеризующемуся распределением вероятности $P[V(\mathbf{r})]$. Предполагается, что такие важные физические свойства, как, например, плотность состояний, одинаковы для любой выбранной реализации $V(\mathbf{r})$, т. е. что они зависят не от детальной структуры потенциала, а лишь от статистических свойств ансамбля $\{V(\mathbf{r})\}$, например, от корреляционных функций и средних. Если эти параметры удастся связать со статистическими характеристиками реального потенциала $U(\mathbf{r})$, то задачу можно решить, не зная явный вид функции $U(\mathbf{r})$.

Аналогичный подход напрашивается и в нашей задаче: рассматривать функции $E(q)$, $a(q)$ и $b(q, q')$ как представителей ансамбля стохастических функций, наделенного определенными статистическими функциями. Такой подход имеет смысл в том случае, если важные характеристики протекающего процесса эволюции зависят только от соответствующих статистических характеристик ансамбля. Как будет показано в следующем разделе, подобный подход применим по крайней мере в случае (9.2) (Ebeling et al., 1981).

При использовании столь сложных и неизвестных во многих деталях функций, как функции приспособленности, возникает вопрос, обладают ли описываемые ими процессы каким-либо сходством с процессами эволюции или сильная структурированность ландшафта в конце концов приводит к хаотическому развитию. Для ответа на этот вопрос полезно выяснить, существуют ли для рассматриваемой системы функции Ляпунова (см. Haken, 1983). Поскольку функции Ляпунова монотонно изменяются со временем, они независимо от сложности динамики позволяют правильно оценивать общую тенденцию. Нетрудно убедиться в том, что в случае (9.2) определяемое соотношением (9.4) среднее от функции приспособленности по ансамблю $\langle E \rangle$ при $D = 0$ монотонно возрастает. Действительно, дифференцируя выражение (9.4), получаем (Feistel, Ebeling, 1982)

$$\begin{aligned} \partial_t \langle E \rangle &= \frac{1}{N} \int dq E(q) \partial_t n(q, t) = \\ &= \frac{1}{N} \int dq E^2(q) n(q, t) - \langle E \rangle^2 = \langle (E(q) - \langle E \rangle)^2 \rangle \geq 0. \end{aligned} \quad (9.14)$$

Тем самым обеспечивается независимость глобального направленного процесса от конкретного вида функции $E(q)$. Неравенство (9.13) может быть также выведено из критерия эволюции, предложенного Гленсдорфом и Пригожиным (см. разд. 3.4) (Feistel, Ebeling, 1982). Для непрерывной модели типа Лотки—Вольтерры (9.3) в случае положительной матрицы $b(q, q')$ можно построить функционал Ляпунова

$$L = - \int dq [n(q, t) - n^0(q) \ln n(q, t)],$$

где $n^0(q)$ — стационарное распределение ($n^0(q) = n(q, t \rightarrow \infty)$) (Feistel, Ebeling, 1984). В случае неположительной матрицы взаимодействия $b(q, q')$ могут возникать известные колебания, поэтому функционал Ляпунова не существует.

При включении диффузионного члена ситуация несколько усложняется, однако в простом случае (9.2) по-прежнему существует функционал Ляпунова, имеющий структуру функционала Гинзбурга—Ландау (см. соотношение (4.13)) (Feistel, Ebeling, 1982).

9.2. Скачкообразный характер эволюции в простой модели

Этот раздел посвящен исследованию модели эволюции

$$\partial_t n(q, t) = (E(q) - \langle E \rangle) n(q, t) + D \partial_q^2 n(q, t). \quad (9.15)$$

Эта модель возникает из соотношений (9.2) и (9.4) как частный случай рассмотренной в предыдущем разделе модели пространства фенотипов (9.13), причем дополнительно мы можем предположить, что коэффициент диффузии постоянен: $D(q) = D$. На основе простой нелинейности, форма которой соответствует дискретной модели Эйгена, мы можем высказать некоторые утверждения относительно характера поведения решений уравнения (9.15). В частности, нас интересует, какую информацию можно извлечь относительно стохастической функции приспособленности $E(q)$.

Для более осмысленной постановки вопроса попытаемся прежде всего представить наглядно, как должна себя вести функция $E(q)$. На рис. 9.2 показан фрагмент возможного графика функции $E(q)$ и начальное условие

$$n(q, t = 0) = n_0(q) \quad (9.16)$$

для уравнения (9.15). В течение короткого времени в уравнении (9.15) главным является член, описывающий отбор, а распределение сначала локализовано в окрестности ближайшего максимума функции приспособленности. Затем возникающие мутации начинают «зондировать» более широкую окрестность, и как только обнаруживается новый, более высокий, максимум функции $E(q)$, среднее значение $\langle E \rangle$ возрастает. Это означает, что доминировавшие прежде виды вымирают, и распределение $n(q, t)$ сосредоточивается в окрестности следующего локального максимума ландшафта приспособленности. Требуется выяснить, распространяются ли эти качественные представления на все функции $E(q)$ и как зависят от коэффициента диффузии D и размерности d пространства фенотипов средние времена переходов максимум—минимум.

Для ответа на эти вопросы целесообразно ввести преобразование

$$y(q, t) = n(q, t) \exp \left\{ \int_0^t \langle E \rangle(t') dt' \right\}, \quad (9.17)$$

позволяющее получить из (9.15) линейное уравнение для новой функции плотности $y(q, t)$:

$$\partial_t y(q, t) = E(q) y(q, t) + D \partial_q^2 y(q, t). \quad (9.18)$$

Формальное решение уравнения (9.18) представимо в виде разложения по собственным функциям эллиптического оператора

$$H = D \partial_q^2 + E(q),$$

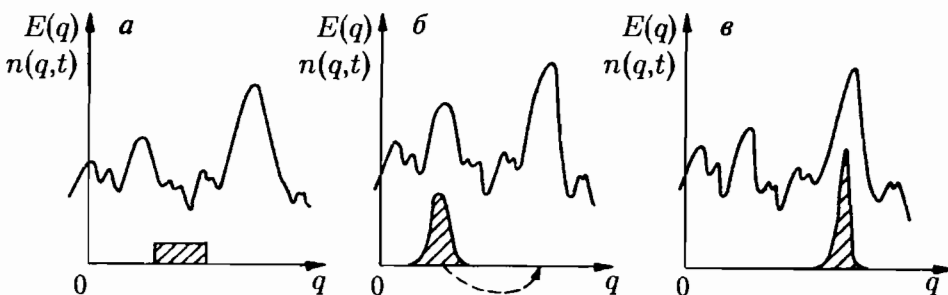


Рис. 9.2. Наглядные этапы процесса эволюции стохастического ландшафта

т. е. в виде ряда

$$y(q, t) = \sum_n c_n \exp \{E_n t\} \psi_n(q), \quad (9.19)$$

где

$$c_n = \int dq \psi_n(q) n_0(q), \quad (9.20)$$

а E_n и $\psi_n(q)$ определяются путем решения задачи на собственные значения

$$D \partial_q^2 \psi_n(q) + E(q) \psi_n(q) = E_n \psi_n(q). \quad (9.21)$$

Ясно, что $\psi_n(q)$ представляют собой континуальное обобщение квазивидов Эйгена и Шустер (Эйген, Шустер, 1982), E_n — их приспособленность. С точностью до знака «потенциала» $E(q)$ и собственных значений E_n уравнение (9.21) соответствует стационарному уравнению Шрёдингера (Feistel, Ebeling, 1982), что позволяет использовать различные результаты квантовомеханических исследований. В качестве простого примера рассмотрим вопрос о том, какой из двух существующих максимумов функции $E(q)$ может быть реализован видом, обладающим более высокой приспособленностью, и, следовательно, в каком из максимумов сосредоточено распределение при $t \rightarrow \infty$. На рис. 9.3 показан при $d = 1$ один максимум высотой $E^{(1)}$ и шириной $\Delta q^{(1)}$ и второй более плоский ($E^{(2)} < E^{(1)}$), но более широкий максимум ($\Delta q^{(2)} > \Delta q^{(1)}$). Нетрудно видеть, что при $D = 0$ без учета мутаций решающим является только локальное значение приспособленности $E(q)$, поэтому при $t \rightarrow \infty$ распределение сосредоточивается в высоком и узком максимуме. При $D \neq 0$ из квантовой механики для наиболее глубокого собственного значения E_0 такой потенциальной ямы известно следующее выражение:

$$E_0^{(i)} = E^{(i)} - \left(\frac{\pi}{\Delta q^{(i)}} \right)^2 \cdot D \quad (9.22)$$

(Ландау, Лифшиц, 1989). Таким образом, при достаточно большом коэффициенте D более плоский, но более широкий максимум функции приспособленности реализует приспособленность квазивида (подчеркнем еще раз различие в знаках между моделью эволюции и квантовой механикой). Такой вывод согласуется и с биологическими представлениями, поскольку с возникновением мутаций каждый вид q тащит за собой «хвост» мутантов, что требует учета не только значения функции $E(q)$ в самой точке q , но и в некоторой окрестности этой точки. Поэтому узкие максимумы, хотя они и относительно высоки, остаются «незаселенными». Подобный эффект становится еще более заметным при переходе к высоким размерностям, так как обобщение соотношения (9.22) на случай размерности d имеет вид

$$E_0^{(i)} = E^{(i)} - d \left(\frac{\pi}{\Delta q^{(i)}} \right)^2 \cdot D.$$

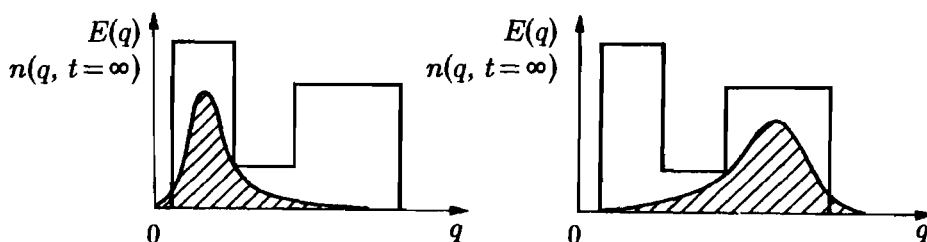


Рис. 9.3. Конечное состояние эволюции для простой функции приспособленности с двумя различными локальными максимумами при малой (слева) и большой (справа) интенсивности мутаций

Так как пространство фенотипов имеет очень высокую размерность $d \gtrsim 100$, определение даже асимптотического распределения $n(q, t \rightarrow \infty)$ становится нетривиальной задачей.

Приведенное выше преобразование к уравнению Шрёдингера имеет особое значение в случае стохастической функции приспособленности. Так как функция $E(q)$ играет в уравнениях (9.18) и (9.21) роль потенциала, наша эволюционная задача с математической точки зрения допускает далеко идущую аналогию с задачей о поведении электрона в случайном потенциале. Это позволяет перенести на интересующий нас случай результаты, полученные при решении того круга задач, которые занимают центральное место в теории неупорядоченных тел и поэтому исследовались и продолжают исследоваться весьма основательно (Thouless, 1974; Anderson, 1978; Lee, Ramakrishnan, 1985). Особый интерес представляют следующие три результата.

1. Плотность состояний собственного значения E_n уравнения (9.21) имеет характерный вид, представленный на рис. 9.4. Здесь важно обратить внимание на существование так называемой границы подвижности $E_{\text{непр}}$, отделяющей непрерывную часть спектра $E < E_{\text{непр}}$ с собственными значениями, соответствующими нелокализованным собственным функциям, от дискретной части $E > E_{\text{непр}}$ с собственными значениями E_n , соответствующими локализованным функциям (Mott, 1967; Kimball, 1978). При $E \rightarrow \infty$ асимптотические выражения для $\rho(E)$ существуют для различных статистик потенциала (Лифшиц и др., 1982).

2. Соответствующие локализованным собственным значениям $E_n > E_{\text{непр}}$ собственные функции $\psi_n(q)$ локализованы, т. е. обладают центром локализации q_n , и на больших расстояниях от этого центра их асимптотика имеет вид

$$\psi_n(q) \sim \exp \left\{ -\frac{|q - q_n|}{l_n} \right\}, \quad (9.23)$$

где l_n — так называемая длина локализации (Anderson, 1958).

3. Для гауссовского случайного поля $E(q)$ с нулевой корреляционной длиной и, следовательно, с корреляционной функцией

$$\langle E(q)E(q') \rangle = \text{const } \delta(q - q') \quad (9.24)$$

при $d \geq 4$ локализованные состояния не существуют, т. е. $E_{\text{непр}} \rightarrow \infty$, в то время как в случае стохастического потенциала с отличной от нуля корреляционной длиной локализованные состояния могут образоваться при любой размерности d (Thouless, 1976; Cardy, 1978; Садовский, 1979).

Выясним прежде всего, к каким следствиям приводит для нашей задачи п. 1. Из соотношений (9.13) и (9.17) для $n(q, t)$ получается следующее представление в виде отношения двух рядов по собственным функциям:

$$n(q, t) = N \frac{\sum_n c_n \exp \{E_n t\} \psi_n(q)}{\sum_n c_n \exp \{E_n t\} a_n}, \quad (9.25)$$

где $a_n = \int dq \psi_n(q)$. Представление (9.25) показывает, что при возрастании t в плотности $n(q, t)$ начинают все сильнее доминировать собственные функции $\psi_n(q)$

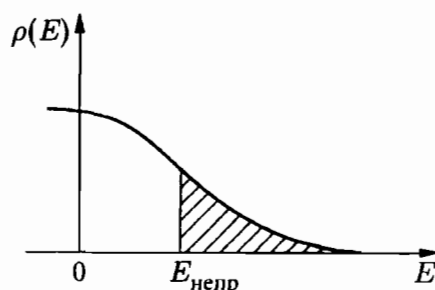


Рис. 9.4. Типичный ход плотности состояний для оператора вида $\partial^2 + E(q)$ со стохастической функцией $E(q)$. Заштрихованная область содержит собственные значения, соответствующие локализованным собственным функциям

с большими собственными значениями E_n . Если мы исходим из постоянного начального условия $n_0(q) = \text{const}$, то при $t \gtrsim 1/E_{\text{непр}}$ существенный вклад в $n(q, t)$ вносят только локализованные собственные функции. Такое «возникновение островков» можно наглядно интерпретировать, как возникновение видов; по истечении определенного времени из первоначально недифференцированной массы индивидов под действием все возрастающего давления отбора жизнеспособными остаются только особые, четко различимые между собой комбинации фенотипических черт. Аналогичный процесс структурирования при $D = 0$ допускает описание в рамках теории перколяции, или протекания (Зельдович, 1983). В общем случае можно утверждать, что возникновения разумного процесса эволюции (появления все более приспособленных видов) в нашей модели можно ожидать только тогда, когда имеется локализованная собственная функция. В силу приведенного выше п. 3 ландшафт приспособленности $E(q)$ характеризуется не короткодействующими корреляциями — он должен быть достаточно гладким (что заведомо выполняется при $d > 4$ для модели пространства фенотипов). Это так называемое требование гладкости, которому должна удовлетворять функция приспособленности (Conrad, 1978), допускает, таким образом, вполне наглядную интерпретацию. Что же касается типичных реализаций ансамбля случайных функций с корреляционной функцией (9.24), то они хотя и непрерывны, но не дифференцируемы, и поэтому при переходе от точки к точке могут изменяться весьма заметно и резко. Поскольку в этом случае система в среднем не может по известному значению функции приспособленности в точке q строить какие-либо заключения о значениях, принимаемых функцией приспособленности в ближайшей окрестности точки q , не может быть выработана стратегия, и процесс отбора не может компенсировать разрушительное действие мутаций. В результате каждое локализованное начальное условие $n_0(q)$ претерпевает расплывание типа диффузионного по всему пространству фенотипов, т. е. мы имеем дело с поведением, ничем не напоминающим реальные процессы эволюции. То, что наша модель приводит к такому режиму только при нулевой длине корреляции $l = 0$, обусловлено диффузионным приближением (9.11) для мутаций. В действительности мутации в силу их дискретного характера всегда порождают изменения фенотипа на некоторой минимальной длине $l_{\min}$, а описанный выше режим наблюдается при $l \ll l_{\min}$. Диффузионное приближение соответствует $l_{\min} \rightarrow 0$, почему и был выделен случай $l = 0$.

Если существуют локализованные собственные значения, то они соответствуют большим собственным значениям E_n , и эволюция плотности $n(q, t)$ при больших временах характеризуется переходами между различными локализованными состояниями $\psi_n(q)$ (рис. 9.2). Для более точного описания этих переходов рассмотрим начальное распределение $n_0(q)$, общий вид которого представлен на рис. 9.2. Такое распределение соответствует суперпозиции (9.19) вида

$$y(q, t = 0) = \sum_n c_n \psi_n(q),$$

где $c_0 \approx 1$ и $c_n \ll 1$ при $n \neq 0$, т. е. в распределении доминирует локализованное состояние ψ_0 . Временная эволюция такого начального условия определяется разложением (9.19). Поскольку по прошествии времени

$$t_n = \frac{\ln(c_0/c_n)}{(E_n - E_0)} \quad (9.26)$$

справедливо именно равенство

$$c_n \exp \{E_n t_n\} = c_0 \exp \{E_0 t_n\},$$

при $t = t_n$ собственные функции ψ_0 и ψ_n представлены в суперпозиции (9.19) с одинаковыми амплитудами, что соответствует переходу к новому максимуму функции приспособленности $E(q)$, образующему состояние ψ_n (рис. 9.2 в). Используя локализованность собственной функции ψ_n (см. формулу (9.23)), получаем оценку

$$\frac{c_n}{c_0} \approx \exp \left\{ -\frac{|q_0 - q_n|}{l_n} \right\}$$

и с учетом соотношения (9.26) находим

$$t_n \approx \frac{|q_0 - q_n|}{l_n(E_n - E_0)}. \quad (9.27)$$

Нас интересует, когда происходит переход от «квазивида» ψ_0 с приспособленностью E_0 к другому «квазивиду» ψ_n с приспособленностью E_n . Следовательно, нам необходимо рассмотреть состояние ψ_n , самое близкое к исходному состоянию ψ_0 из всех состояний с собственными значениями, принадлежащими интервалу $(E_n, E_n + dE)$, т. е. состояние ψ_n с минимальным значением $|q_0 - q_n|$. При больших E_n корреляции между различными локализованными состояниями очень слабы, и поэтому вполне допустимо предположить, что q_n с плотностью $\rho(E_n)$ равномерно распределены в пространстве фенотипов. При этих условиях можно определить среднее кратчайшее расстояние до центра локализации состояния ψ_n с собственным значением из интервала $(E_n, E_n + dE)$ от q_0 (Engel, 1983). Подставляя в (9.27), получаем

$$t_n \approx \frac{\Gamma(\frac{d}{d+1})}{l_n(E_n - E_0)} \left[\frac{d}{S_d \rho(E_n)} \right]^{1/d}, \quad (9.28)$$

где $\rho(E_n)$ — плотность состояний, S_d — площадь поверхности d -мерного единичного шара и $\Gamma(x)$ — гамма-функция Эйлера. При определенных статистиках потенциала можно воспользоваться асимптотическими выражениями для $l_n = l(E_n)$ и $\rho(E_n)$ и получить время перехода $t_n = t(E_n)$ как функцию приспособленности E_n конечного состояния и параметров задачи $(E_0, D, d, P[E(q)])$. Общий вид этой функции представлен на рис. 9.5. Он не зависит от конкретной статистики $P[E(q)]$ функции приспособленности и претерпевает лишь незначительные изменения при учете корреляции между состояниями, которыми первоначально мы пренебрегали (Engel, 1985). Важно существование минимума $E_{\min}$, так как он соответствует наиболее

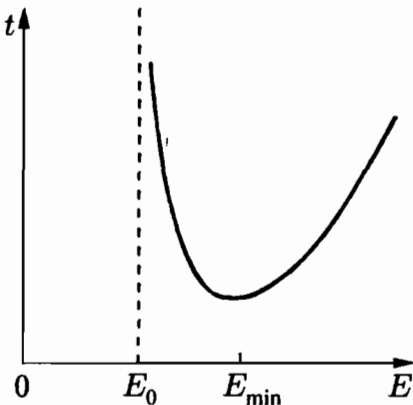


Рис. 9.5. Качественный вид зависимости времени перехода t от приспособленности E конечного состояния



Рис. 9.6. Типичная временная зависимость приспособленности ансамбля $\langle E \rangle$ для процесса эволюции в стохастическом ландшафте приспособленности

быстрым и, следовательно, наиболее часто реализующимся процессам. Возникновение уровня нетрудно обосновать с помощью простых и наглядных соображений. Переходы со значительным улучшением приспособленности $(E - E_0) \gg (E_{\min} - E_0)$ связаны с одновременным действием нескольких благоприятных мутаций, поскольку конечное состояние, как правило, расположено в пространстве фенотипов довольно далеко. Столь «большие прыжки», необходимые для преодоления расстояния до конечного состояния, весьма маловероятны, что находит отражение в больших временах перехода $t(E)$ при $E \gg E_{\min}$. Наоборот, очень малые улучшения приспособленности $(E - E_0) \ll (E_{\min} - E_0)$ могут весьма быстро порождаться мутациями, но соответствующий процесс отбора в силу слабого превосходства нового рода тянется весьма долго. При $E \rightarrow E_0$ возникает особенность $t(E) \rightarrow \infty$. Значение $E_{\min}$ реализует благоприятный компромисс между этими двумя предельными случаями. Из графика на рис. 9.5 следует, что мы в праве ожидать, что средняя по ансамблю приспособленность $\langle E \rangle$ будет изменяться так, как показано на рис. 9.6. Если доминирует плотность $n(q, t)$ какого-нибудь одного локализованного состояния, то $\langle E(t) \rangle \approx E_n = \text{const}$, причем малые флуктуации среднего $\langle E \rangle$ обусловлены постоянным действием мутаций. По истечении определенного времени происходит переход в новое состояние ψ_m , который приводит к скачкообразному изменению среднего $\langle E \rangle$ на величину $E_m > E_n$. Времена пребывания в одном состоянии (между переходами) являются случайными величинами со средним значением $t(E_{\min})$; типичная величина прироста («улучшения») функции приспособленности за один переход $\delta E = E_{\min} - E_0$ фиксирована, поэтому δE играет роль «кванта эволюции». Соответственно, становится дискретной и динамика «центра тяжести плотности» $\langle q \rangle$ (см. формулу (9.8)) в пространстве фенотипов. Качественно такая динамика представлена на рис. 9.7.

Замечательно, что скачкообразный характер эволюционной динамики обнаружен в модели, предложенной для описания именно непрерывных изменений фенотипа. Таким образом, обнаруженные палеонтологами разрывы в ходе эволюции жизни на Земле (Gould, 1977) не обязательно обусловлены только дискретным действием мутаций, а являются гораздо более общей характеристикой процессов эволюции. Таким образом, наши выводы позволяют по-новому взглянуть на старый спор между сторонниками непрерывного характера видообразования, или градуализма, и сторонниками прерывистого равновесия, или пунктуализма (Волькенштейн, 1984; Sander, 1988; Engel, Ebeling, 1987).

Взятые вместе, приведенные в этом разделе исследования показывают, что неизвестные функции $w(q, t, n(q, t))$ моделей эволюции в пространстве фенотипов разумно представлять с помощью стохастических функций. Многие интересные особенности эволюционных процессов определяются только статистическими свойствами этих функций и не зависят от их детальной структуры. Вопрос о реалистических статистиках таких функций находится в центре внимания в гл. 10.

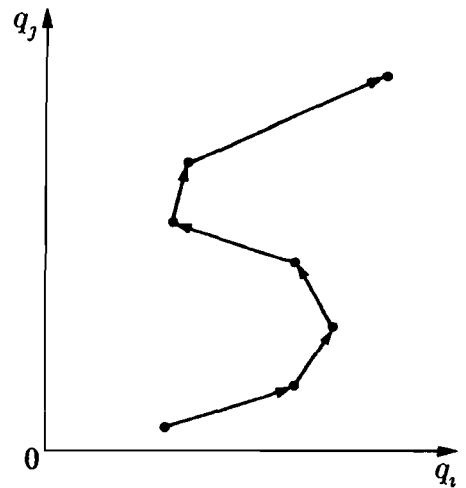


Рис. 9.7. Скачки центров локализации распределения плотности $n(q, t)$ в пространстве фенотипов на двумерной проекции

9.3. Индивидуальное развитие фенотипа

До сих пор в рамках этой главы мы исходили из того, что для самовоспроизводства не требуется никакого времени. Копии входят в пространство фенотипов

мгновенно вместе с оригиналом. Разумеется, такое предположение нереалистично, и поэтому мы хотим усовершенствовать модель и учесть то, что нам известно из гл. 7 относительно индивидуального развития. С формальной точки зрения это означает прежде всего, что множество фенотипических свойств необходимо пополнить еще одним элементом — возрастом τ . Соответственно, пространство фенотипов получает дополнительную координату $q_{d+1} = \tau$ и обретает структуру

$$\bar{q} = (q_1, \dots, q_d, \tau).$$

В обобщенном пространстве фенотипов основное уравнение динамики представимо в следующем виде:

$$\partial_t x(\bar{q}, t) + \partial_{\bar{q}} [f(\bar{q}) x(\bar{q}, t)] = \int d\bar{q}' A(\bar{q}, \bar{q}') x(\bar{q}', t) - [d(\bar{q}) + \Phi(t)] x(\bar{q}, t). \quad (9.29)$$

Обобщение (по сравнению с уравнением (9.13)) состоит в том, что введен дополнительный дрейфовый член, а вместо членов, описывающих самовоспроизведение и мутации, в правую часть входит интегральная форма весьма общего вида. Ядро интеграла A позволяет учитывать конечные переходы в пространстве фенотипов.

Рассмотрим прежде всего несколько подробнее дрейфовый член. Его назначение состоит в том, чтобы описывать движение фенотипа в ходе индивидуального развития. Предположим, что

$$\frac{d}{dt} q_i(\tau) = f_i(q_1, q_2, \dots, q_d, \tau), \quad i = 1, 2, \dots, d. \quad (9.30)$$

Здесь f_i — функции роста фенотипа. Они описывают, например, увеличение веса и изменение других свойств с увеличением возраста τ . Решение уравнения (9.30) задает траекторию индивидуального развития (рис. 9.8). С учетом соотношения (9.30), получаем для дрейфового вектора в уравнении (9.29)

$$f(q) = \begin{Bmatrix} f_1 \\ f_d \\ \vdots \\ 1 \end{Bmatrix}.$$

Рассмотрим теперь член, описывающий переходы. Так как новорожденный вид всегда должен иметь возраст $\tau = 0$, ядро интеграла должно иметь специальный вид:

$$A(q, \bar{q}') = \delta(\tau) B(q(\tau'), q', \tau').$$

С динамической точки зрения самовоспроизведение (рождение) представляет собой скачок на плоскость $t = 0$. Другая возможность динамической интерпретации состоит в предположении существования источников на плоскости $\tau = 0$. В уравнении (9.29) под функциями $q(\tau')$ понимаются те траектории (решения уравнения (9.30)), которые удовлетворяют начальному условию $q(0) = q$. С точки зрения динамики полный жизненный цикл выглядит как блуждание вдоль траектории роста в полупространстве $\tau > 0$. Смерть особи соответствует исчезновению точки на этой траектории, рождение потомка (с формальной точки зрения) — появлению новой точки в основании той же самой траектории (в случае тождественного самовоспроизведения) или в основании близкой траектории (в случае самовоспроизведения

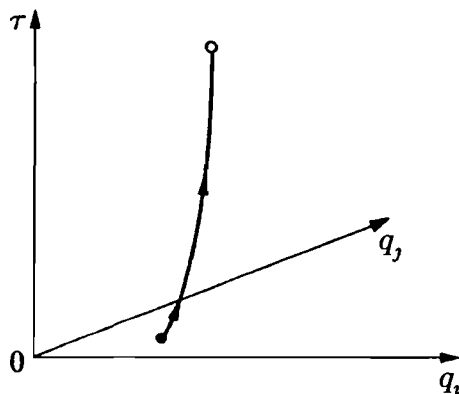


Рис. 9.8. Траектория индивидуального развития в пространстве фенотипов, начало на плоскости $t = 0$ соответствует рождению индивидуума, а конец — его выживанию из популяции

с ошибкой). На рис. 9.9 показан пучок траекторий, состоящий из центральной (родительской) траектории и соседних траекторий потомков.

Основное уравнение популяционной динамики при положительном возрасте $\tau > 0$ имеет следующий вид:

$$\begin{aligned} \partial_t x(q, \tau, t) + \partial_\tau x(q, \tau, t) + \sum_{i=1}^d \partial_{q_i} [f_i(q) x(q, \tau, t)] = \\ = -[d(q, \tau) + \Phi(t)] x(q, \tau, t). \end{aligned} \quad (9.31)$$

К уравнению (9.31) следует также присоединить начальное условие на плоскости $\tau = 0$:

$$x(q, 0, t) = \int_0^\infty d\tau' \int dq' B(q(\tau'), q', \tau') x(q', \tau', t). \quad (9.32)$$

В случае тождественного воспроизведения ядро интеграла и начальное условие имеют вид

$$B = b(q(\tau'), \tau') \prod_{i=1}^d \delta[q_i(\tau') - q'_i].$$

Интегрирование в этом случае проводится в точности вдоль центральной траектории роста. Воспроизведение с ошибкой может быть описано с помощью гауссовских распределений $g(x)$, центрированных на траектории родителей. Начальное условие при этом имеет следующий вид:

$$x(q, 0, t) = \int b(q(\tau'), \tau') \prod_{i=1}^d g_i[q_i(\tau') - q'_i] x(q', \tau', t) dq' d\tau'. \quad (9.33)$$

Ценность видов определяется функциями $b(q, \tau)$ и $d(q, \tau)$.

В заключение еще несколько замечаний о необходимых обобщениях теории. Первое замечание относится к включению в уравнение взаимодействий типа Лотки—Вольтерры. Для жертв распадный член следует заменить интегралом:

$$\begin{aligned} \partial_t x(q, \tau, t) + \partial_\tau x(q, \tau, t) + \sum \partial_{q_i} [f_i(q) x(q, \tau, t)] = \\ = - \left[d(q, \tau) + \int D(q, \tau; q^*, \tau^*) x(q^*, \tau^*, t) dq^* d\tau^* \right] x(q, \tau, t). \end{aligned} \quad (9.34)$$

Ядро интеграла D описывает взаимодействие рассматриваемой жертвы с различными хищниками. Для хищных видов дрейфовый член, описывающий рост, следует дополнить интегральным членом

$$f_i(q, \tau) = f_i^{(0)}(q, \tau) + \int F_i(q, \tau; q^*, \tau^*) x(q^*, \tau^*, t) dq^* d\tau^*.$$

Для учета половой репликации в теорию необходимо ввести две различные функции плотности

$$x^\sigma(q, \tau, t) \quad \text{и} \quad x^\varphi(q, \tau, t).$$

Математическая теория таких процессов необычайно сложна и до сих пор разработана весьма слабо.

Глава 10

Фрустрация и иерархическое упорядочение

Нельзя желать того, что не происходит.

Петер Хакс

10.1. Дарвиновская эволюция как задача фрустрированной оптимизации

Дарвиновский принцип эволюции — выживание наиболее приспособленного вида — может быть интерпретирован как задача оптимизации: при заданных экологических краевых условиях определить тот вид (или то семейство видов), который реализует максимальную индивидуальную или коллективную частоту самовоспроизведения. Как показано в последней главе нашей книги, решение задачи оптимизации может быть получено с помощью эволюционного процесса, в ходе которого в игру попеременно вступают отбор и мутация. С сильно упрощенными примерами задач оптимизации, решаемых с помощью эволюционной динамики (эволюционных процессов), мы познакомились еще в разд. 6.4. В рамках имитационных игр на компьютере там было показано, что в результате процессов самовоспроизведения, отбора и мутации последовательностей букв возникают определенные характерные последовательности букв. Мы видели также, какую важную роль играет при этом выбранный алгоритм оценивания приспособленности. Самые интересные и разнообразные структуры возникали в наших численных экспериментах в тех случаях, когда алгоритм оценивания был внутренне противоречив. Это явление мы называли фрустрацией. Типичным для этих экспериментов по численному моделированию было порождение процессом, стремящимся удовлетворить одновременно нескольким противоречивым требованиям оптимальности, весьма сложных и иерархически упорядоченных последовательностей символов. Нам представляется, что тем самым затронут весьма общий и глубокий аспект эволюции, заслуживающий более глубокого рассмотрения.

В этой главе мы займемся исследованием специфических свойств таких экстремальных задач, которые возникают в ходе эволюции, а также характерные особенности их решений. При этом выясняется принципиальное сходство между процессами эволюции, сложными техническими проблемами оптимизации, которая будет рассмотрена более подробно позднее, и поведением магнетиков, так называемых спиновых стекол, при низких температурах, которое было предметом подробных исследований в физике твердого тела в последние 10–15 лет. Успехи в теории названных выше систем привели к глубокому пониманию некоторых свойств процессов эволюции и оптимизации, поэтому мы также уделяем внимание этим системам. На уровне наглядных представлений отличительная особенность сложных задач оптимизации, в том числе и задач, возникающих в биологической эволюции, состоит в многочисленности и сложной взаимозависимости тех требований, которые должны быть выполнены. В простых ситуациях оптимизации подлежат только немногие,

главным образом слабо коррелированные между собой свойства. В результате в такого рода случаях существует некое идеальное решение, обладающее в любом аспекте лучшими свойствами, чем любое другое решение. Алгоритм решения задач на оптимизацию в простых случаях, вообще говоря, достаточно прост (*Rechenberg*, 1973). В случае сложных задач на оптимизацию ситуация совершенно иная, поскольку многие требования сильно взаимосвязаны, а некоторые просто противоречивы.

На рис. 10.1 схематически изображен сильно упрощенный пример из области биологической эволюции. Для конкуренции двух видов предпочтительно, чтобы каждое из подлежащих сравнению фенотипических свойств — сила, вес, быстрота — были реализованы со всей возможной полнотой. Тем самым предполагается, что из двух в остальном совершенно эквивалентных видов выживает тот, который в большей степени наделен этими свойствами. Но эти свойства не являются независимыми, а качественно коррелированы по своему характеру и интенсивности, что на схеме условно обозначено знаками плюс и минус. Сила и быстрота, а также вес и сила благоприятствуют друг другу, быстрота и вес — свойства взаимно дополнительные.

Идеальное решение задачи невозможно из-за внутренней противоречивости требований; правда, вполне мыслимы различные компромиссы, позволяющие посредством подходящей комбинации преимуществ и недостатков в одной области достигать локально оптимальные решения в другой области. По аналогии с образованием понятий в теории спиновых стекол такого рода задачи на оптимизацию принято называть фрустрированными. Нет ни малейшего сомнения в том, что существует множество интересных задач на оптимизацию такого типа и что нефрустрированные ситуации следует считать скорее исключением, чем правилом. Из своего повседневного опыта мы хорошо знаем, и многие пословицы, поговорки и крылатые выражения убеждают нас в том («нет в мире совершенства», «все знать никому не дано» и т. д.), как глубоко укоренилось понятие фрустрации в нашем сознании. Столкнувшись со сложной задачей, мы никогда не ожидаем идеального решения, а обычно интересуемся сильными и слабыми сторонами какого-то решения. Несмотря на столь явную приверженность понятию фрустрации, точная математическая формулировка этого понятия и некоторых проистекающих из него следствий стала возможна лишь недавно в результате исследований по статистической механике спиновых стекол.

Хотя исследования, о которых идет речь, находятся в начальной стадии, будет уместно кратко изложить существующую теорию, избранные результаты и указать те следствия, к которым они приводят для теории биологической эволюции. Поэтому в следующем разделе мы кратко излагаем наиболее существенные для нас результаты теории спиновых стекол и обсуждаем основанную на этой теории недавно предложенную модель (*Anderson*, 1983) эволюции пребиологических макромолекул.

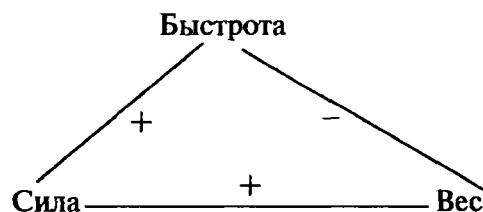


Рис. 10.1. Фрустрированное соотношение треугольника для сильно упрощенной эволюционной проблемы. Знаки плюс и минус соответствуют положительным и отрицательным корреляциям свойств

10.2. Некоторые результаты теории спиновых стекол

Под спиновыми стеклами мы понимаем неупорядоченные твердые тела, в которых магнитные ионы расположены в немагнитной решетке случайным образом (например, ионы Fe в Au или ионы Mn в Cu). Электроны проводимости решетки-«хозяина» осуществляют взаимодействие $J(R)$ между ионами, знак которого изменяется в зависимости от расстояния R между ионами. Точнее говоря, при больших R

взаимодействие имеет вид

$$J(R) = J_0 \frac{\cos(2k_F R + \xi_0)}{(k_F R)^3}, \quad (10.1)$$

где J_0, k_F и ξ_0 — некоторые константы (Binder, Young, 1986). Выражение (10.1) описывает дальное действие, так как интенсивность такого взаимодействия спадает лишь степенным образом. Но знак его изменяется периодически как функция R (рис. 10.2). Соответственно, между одними спинами (как мы условимся для краткости называть магнитные ионы) связь ферромагнитная, т. е. для них энергетически выгодна параллельная ориентация, а между другими спинами связь антиферромагнитная, т. е. энергетически выгодна антипараллельная ориентация. При температурах ниже так называемой температуры замерзания $T_{\text{зам}}$, лежащей в большинстве случаев в интервале от 10 до 50 Кельвин, спиновые стекла обнаруживают различные экспериментальные особенности. Резкий пик восприимчивости переменного поля указывает на переход беспорядок—порядок, но теплоемкость не обнаруживает аналогичных аномалий, которыми обычно сопровождаются фазовые переходы. Остаточное намагничение низкотемпературной фазы очень слабо и затухает чрезвычайно медленно: как показывают экспериментальные измерения, релаксация происходит на временах масштаба от 10^{-12} с до нескольких часов. Кроме того, для спиновых стекол при наложении медленно изменяющегося магнитного поля характерны явления гистерезиса различного вида.

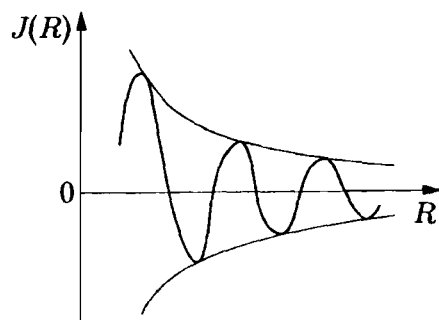


Рис. 10.2. Схематический ход взаимодействия между магнитными ионами металлического спинового стекла

В качестве простого исходного пункта при теоретическом описании может служить так называемая модель Изинга. Рассмотрим правильную пространственную решетку и каждому ее узлу $i = 1, \dots, N$ поставим в соответствие спин Изинга $S_i = \pm 1$. Соседние спины взаимодействуют через коэффициенты связи J_{ij} , поэтому полная энергия системы определяется функцией Гамильтона

$$H = - \sum_{i,j}^N J_{ij} S_i S_j. \quad (10.2)$$

Положительные J_{ij} соответствуют ферромагнитным связям, так как при $S_i = S_j$, т. е. при параллельных спинах, энергия отводится. Антиферромагнитные взаимодействия $J_{ij} < 0$ требуют, наоборот, $S_i = -S_j$ для конфигураций меньшей энергии. С помощью стандартных методов статистической механики (Ландау, Лифшиц, 1995) из функции Гамильтона (10.2) находят по формуле

$$F = -T \ln \sum_{S_i} \exp \left\{ - \frac{H(S_i)}{T} \right\} \quad (10.3)$$

свободную энергию и определяют термодинамические свойства системы. В выражении (10.3) T означает абсолютную температуру, а сумма берется по всем возможным конфигурациям спинов. Если связи не зависят от узлов решетки, т. е. если $J_{ij} = J$, то результат такой программы хорошо известен: при $J > 0$ ($J < 0$) система претерпевает фазовый переход второго рода в ферромагнитную (антиферромагнитную) низкотемпературную фазу. В смысле нашей постановки задачи это соответствует решению

нефрустрированной задачи на оптимизацию, так как при низких температурах система реализует состояния с минимальной энергией. В случаях, о которых мы только что говорили, существуют ровно два таких состояния: когда все спины параллельны, т. е. $S_i = +1$ (или $S_i = -1$), — в случае ферромагнетиков, и спины с чередующейся ориентацией — в случае антиферромагнетиков. Для нас важно, что существует состояние системы, которое удовлетворяет всем условиям взаимодействия. Учитывая статистическую природу спиновых стекол, следует ожидать сложной вариации коэффициентов связи J_{ij} при переходе от узла решетки к узлу, если простая модель (10.2) позволяет получать результаты, которые хотя бы качественно согласуются с экспериментом. Особенно интересной представляется возможность рассматривать обменные связи как случайную (стохастическую) величину, т. е. рассматривать J_{ij} как независимые случайные величины, имеющие одно общее распределение вероятностей $P(J_{ij})$. В этом случае при заданной конфигурации коэффициентов связи J_{ij} можно вычислить свободную энергию и получить распределение вероятностей P_{ij} по различным возможным конфигурациям. Осуществить такие вычисления аналитически пока удастся лишь с помощью плохо контролируемых математических трюков и в относительно простых приближениях. Поэтому было проведено много численных экспериментов, результаты которых обладают убедительным сходством с результатами натуральных экспериментов, что свидетельствует об адекватности модели Эдвардса—Андерсона (ЭА) задачи о спиновых стеклах (Binder, Young, 1986).

Чем физически отличается эта модель от простой модели Изинга для ферро- и антиферромагнетиков, как она объясняет принципиально различные макроскопические свойства?

На рис. 10.3 представлены две возможные элементарные ячейки плоской модели ЭА с возможными конфигурациями минимальной энергии. Слева, хотя спины беспорядочно ориентированы то вверх, то вниз, ситуация в целом обладает далеко идущей аналогией с упорядоченными системами, так как изображенная конфигурация удовлетворяет всем связям. Это принципиально невозможно во втором случае: в каком бы направлении ни был ориентирован спин, обозначенный вопросительным знаком, всегда одно из четырех взаимодействий оказывается нарушенным; следовательно, не существует конфигурации, в которой все вклады в энергию (10.2) одновременно минимальны. По предложению П. У. Андерсона, системы с такого рода конкурирующими взаимодействиями называются фрустрированными. Поиск состояния с минимальной энергией в модели ЭА представляет тем самым наглядный пример фрустрированной задачи на оптимизацию. Обобщая пример, приведенный на рис. 10.3, нетрудно показать, что все конфигурации спинов, при которых произведение связей отрицательно (положительно), фрустрированы (нефрустрированы) (Toulouse, 1977). Если распределение $P(J_{ij})$ обменных связей симметрично относительно $J_{ij} = 0$, то у макроскопической системы существует примерно столько фрустрированных, сколько и нефрустрированных конфигураций спинов. Столь сложное смешение разрешимых и принципиально неразрешимых задач при минимизации энергии является микроскопической причиной необычных термодинамических свойств спиновых стекол.

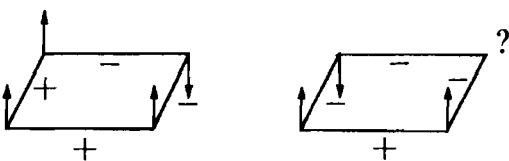


Рис. 10.3. Конфигурация с минимальной энергией для нефрустрированной (слева) и фрустрированной (справа) элементарной ячейки плоской модели Эдвардса—Андерсона. Знаки определяются обменными связями

Из рис. 10.3 мы получаем также первое указание на число спиновых конфигураций, реализующих возможный минимум энергии. Для нефрустрированной конфигурации, изображенной на рис. 10.3 слева, помимо нее самой существует еще только одна конфигурация с такой же энергией, возникающая при инверсии всех спинов и соответствующая преобразованию симметрии $S_i \rightarrow -S_i$ функции Гамильтона (10.2). В случае фрустрированной конфигурации взаимодействие нарушается и в состоянии с минимальной энергией, причем фрустрирована может быть каждая из четырех имеющихся связей. При дополнительном учете вырождения энергии $S_i \rightarrow -S_i$ мы получаем даже для такой системы всего лишь четырех спинов восемь возможных конфигураций с минимальной энергией. Столь сильное вырождение основного состояния обуславливает сложные явления гистерезиса, наблюдаемые экспериментально в спиновых стеклах. Совершенно ясно, что «идеальные» решения нефрустрированных задач на оптимизацию позволяют получить представление лишь о вырождении симметрии избранной постановки задачи, тогда как в случае фрустрированных задач мы принципиально можем получить только компромиссные решения различного рода.

Важные для нас наиболее существенные результаты теории спиновых стекол, кратко резюмированные и рассмотренные здесь, получены с помощью методов статистической механики и поэтому имеют под собой прочную основу. Значение проблемы спиновых стекол для теории сложных систем в том и состоит, что спиновые стекла, с одной стороны, достаточно сложны, чтобы в них могли наблюдаться такие явления, как фрустрация и компромисс, но, с другой стороны, достаточно просты и поэтому допускают применение математических методов. Подробное введение в физику спиновых стекол можно найти в различных обзорах (*Binder, Young, 1986; Rammal et al., 1986*).

Большинство аналитических результатов в теории спиновых стекол соответствуют так называемому приближению молекулярного поля. Это приближение отвечает исследованию модели Шеррингтона—Киркпатрика (ШК), в которой каждый спин взаимодействует с каждым. Модель ШК также описывается функцией Гамильтона (10.2), в которой суммирование проводится по всем парам $i \neq j$, а не только, как в модели ЭА, по ближайшим соседям. Физически это соответствует подавлению корреляций между флуктуациями различных спинов, что сильно упрощает рассмотрение задачи. Тем не менее свойства этого простейшего приближения еще весьма многообразны.

Прежде всего при критической температуре $T_{\text{зам}}$ происходит резкий переход от эргодического поведения к неэргодическому. При высоких температурах доминирует тепловая энергия, и спиновое стекло обладает парамагнитной фазой со свойствами, аналогичными тем, которыми обладает парамагнитная высокотемпературная фаза ферромагнетиков. В этой фазе система эргодична, т. е. все микроскопические конфигурации достижимы из любой начальной конфигурации. По этой причине локальное намагничение при всех i составляет величину $m_i = \langle S_i \rangle = 0$. Здесь и в дальнейшем угловые скобки означают тепловое среднее значение, т. е. математическое ожидание соответствующей величины при реализации связей J_{ij} , а черта над какой-то величиной означает среднее по конфигурациям с распределением вероятностей $P(J_{ij})$. Равенство $\langle S_i \rangle = 0$ означает просто, что тепловые флуктуации достаточно сильны, вследствие чего спин в i -м узле решетки одинаково часто ориентирован как вверх, так и вниз.

При понижении температуры до величины порядка характерного взаимодействия $J = (\overline{J_{ij}^2})^{1/2}$ между спинами эта эргодичность нарушается: из данного начального состояния теперь достижима только часть фазового пространства. Остальные

области лежат за барьерами свободной энергии, высота которых в термодинамическом пределе $N \rightarrow \infty$ расходится, в силу чего преодоление их требует бесконечно большого времени. Фазовое пространство разделяется на несколько несвязных областей, называемых его компонентами, или чистыми фазами. Внутри одной фазы система остается эргодической, но переходы между фазами невозможны. Если рассматривать все конфигурации одной фазы, то какой-нибудь спин, например S_i , может быть чаще ориентирован вверх, чем вниз, вследствие чего локальное намагничение может быть отлично от нуля. Точнее говоря, каждая чистая фаза α характеризуется набором локальных намагничений m_i^α . Переходы от эргодического к неэргодическому поведению не представляют собой ничего необычного для систем с кооперативными взаимодействиями. Фазовое пространство ферромагнитной модели Изинга также распадается при низких температурах на две области, разделенные барьером, который расходится при $N \rightarrow \infty$ (N — число спинов). Две фазы отличаются и по значению локального намагничения $m_i^\alpha = \langle S_i \rangle_\alpha = m^\alpha$, которое в данном случае не зависит от узла решетки. Кроме того, $m^1 = -m^2$, т. е. одну фазу можно отображать в другую с помощью преобразования симметрии $S_i \rightarrow -S_i$.

Качественно новая особенность переходов в спиновых стеклах появляется уже в числе возникающих фаз. Можно показать, что ниже температуры замерзания $T_{\text{зам}}$ существует около 2^{aN} чистых фаз, где $a \simeq 0,2$ при $T = 0$ и $a \rightarrow 0$ при $T \rightarrow T_{\text{зам}}$ (Tanaka, Edwards, 1980; Bray, Moore, 1980; De Dominicis et al., 1980). Это возрастающее, очень быстро стремящееся к бесконечности с увеличением N число несвязных областей в фазовом пространстве показывает, сколько различных микроскопических разновидностей спинового стекла может возникнуть в результате замерзания при низких температурах. При $T = 0$ каждая фаза соответствует локальному минимуму энергии, т. е. число фаз соответствует качественно определяемым возможностям нахождения компромисса при минимизации энергии. Второе существенное отличие от фазового перехода парамагнетик—ферромагнетик состоит в том, что не существует преобразований симметрии, которые отображали бы одну фазу в другую. (Преобразование $S_i \rightarrow -S_i$, изменяющее число фаз только в 2 раза, во внимание не принимается; оно не характерно для спиновых стекол и может быть исключено, например, с помощью внешнего поля.) Столь сложной структурой фазового пространства обусловлены как новые свойства спиновых стекол, так и математические трудности, возникающие при их теоретическом описании. Появление большого числа чистых фаз ставит вопрос о вероятности их реализации. Если образец спинового стекла очень часто нагревать до температур выше $T_{\text{зам}}$, а затем охлаждать до температур ниже $T_{\text{зам}}$, то сколько раз образец окажется в одной и в другой фазе? Оказывается, что вероятность P_α реализации фазы α , если $F\{m_i^\alpha\}$ — свободная энергия фазы α , определяется выражением

$$P_\alpha = \frac{\exp \{ - F\{m_i^\alpha\}/T \}}{\sum_\gamma \exp \{ - F\{m_i^\gamma\}/T \}}. \quad (10.4)$$

Величину $F\{m_i^\alpha\}$ мы находим по формуле (10.3), причем вклад в сумму дают только такие конфигурации, которые принадлежат области фазового пространства фазы α . Среднее по конфигурациям свободной энергии в этом случае удастся вычислить, и мы получаем

$$\overline{F_\alpha} = Nf + f_\alpha + o(1). \quad (10.5)$$

Здесь f и f_α не зависят от N , и f_α — независимые, одинаково распределенные случайные величины (распределение которых известно). Соотношение (10.5) показывает, что при $N \rightarrow \infty$ главным становится первый, не зависящий от α , член, т. е.

в термодинамическом пределе все фазы имеют одну и ту же свободную энергию на степень свободы. Это — не что иное, как математическое выражение интуитивно ожидаемой эквивалентности различных компромиссных решений при минимизации энергии. С другой стороны, поправки f_α определяют вероятности появления различных фаз, так как из соотношений (10.5) и (10.4) при $N \rightarrow \infty$ следует

$$P_\alpha = \frac{\exp \{-f_\alpha/T\}}{\sum_\gamma \exp \{-f_\gamma/T\}}. \quad (10.6)$$

Поэтому можно ожидать, что при низких температурах с заметно отличными от нуля вероятностями реализуется лишь часть возможных фаз. Для понимания термодинамических свойств фазовых стекол необходимо более детальное знание соотношения между многочисленными чистыми фазами, возникающими при низких температурах, когда происходит замерзание спинового стекла. В частности, если ввести меру «подобия» или «расстояния» между фазами α и β , то это позволяет рассматривать на количественном уровне сложную топологию пространства состояний. Изящным решением задачи о мере в пространстве состояний может служить так называемое перекрытие $q^{\alpha\beta}$ (Parisi, 1983):

$$q^{\alpha\beta} = \frac{1}{N} \sum_i m_i^\alpha m_i^\beta, \quad -1 \leq q^{\alpha\beta} \leq 1. \quad (10.7)$$

Если локальные намагничивания m_i^α и m_i^β в очень многих узлах решетки равны, то $q^{\alpha\beta}$ близки к единице, и обе фазы очень похожи. Наоборот, если m_i^α совершенно не зависит от m_i^β , то $q^{\alpha\beta} = 0$, т. е. между фазами нет никакой корреляции. На эвристическом уровне это означает, что две фазы в этом случае находятся в пространстве на большом расстоянии друг от друга, поскольку перевести одну фазу в другую мы можем, только опрокинув большое число спинов.

Если температура T в точности равна нулю, то тепловые флуктуации исчезают, и $m_i^\alpha = \pm 1$. В этом случае перекрытие просто зависит от расстояния Хэмминга $d^{\alpha\beta}$, которое показывает долю несовпадающих знаков двух N -значных двоичных чисел. Нетрудно убедиться в том, что

$$d^{\alpha\beta} = \frac{1}{2}(1 - q^{\alpha\beta}). \quad (10.8)$$

Здесь $d^{\alpha\beta}$ — метрика (мера расстояния), характеризующая различие конфигураций, а параметр $q^{\alpha\beta}$ служит количественной мерой сходства между двумя фазами. Соответственно, при $\alpha = \beta$ мы получаем $d^{\alpha\alpha} = 0$ и $q^{\alpha\alpha} = 1$. При $T = 0$ каждая фаза представлена конфигурацией спинов, которая в свою очередь отвечает возможному компромиссному решению задачи на минимизацию энергии. Перекрытие и расстояние Хэмминга позволяют дать наглядное и количественное описание сходства различных компромиссов. Эта проблематика играет важную роль в дальнейшем, когда речь идет об отыскании подходящих численных процедур решения сложных задач оптимизации при использовании свойств ландшафта ценностей. Простая зависимость (10.8) между перекрытием и расстоянием Хэмминга нарушается, когда температура становится отличной от нуля, так как тогда локальные намагничивания m_i^α могут принимать произвольные значения между -1 и 1 , в силу чего возможны, например, случаи, когда $q^{\alpha\alpha} < 1$. Для описания термодинамических свойств спинового стекла заведомо не нужно точно знать перекрытия $q^{\alpha\beta}$ всех возможных пар чистых фаз. Более того, можно ожидать, что вполне достаточно знать, мало или много

существует фаз с большим или малым перекрытием, т. е. каково характерное соотношение между двумя случайно выбранными фазами. Эта информация содержится в распределении $P(q)$, которое показывает, с какой вероятностью система при двух независимых измерениях находится в двух фазах с перекрытием q :

$$P(q) = \sum_{\alpha, \beta} P_{\alpha} P_{\beta} \delta(q - q^{\alpha\beta}). \quad (10.9)$$

Действительно, можно показать (Paris, 1980), что свободная энергия и, следовательно, все термодинамические свойства функционально зависят от распределения вероятностей $P(q)$, причем в качестве параметра порядка вместо функции $q(x)$, $0 \leq x \leq 1$, обратной относительно $P(q)$, формально удобнее воспользоваться первообразной распределения $P(q)$, т. е. функцией $x(q) = \int_{-1}^q P(q') dq'$. В эргодической

фазе функция $x(q)$ есть константа, для ферромагнетиков ниже критической температуры $T_{\text{крит}}$ она имеет два плато, соответствующих двум чистым фазам. В случае спиновых стекол функция $x(q)$ непрерывно изменяется между $-q_{\text{ЭА}}$ и $+q_{\text{ЭА}}$, где параметр порядка Эдвардса—Андерсона $q_{\text{ЭА}} = q^{\alpha\alpha}$ задает самоперекрытие и тем самым локальный порядок. На рис. 10.4–10.6 представлены графики трех функций $P(q)$, $x(q)$ и $q(x)$ для трех случаев: парамагнетика (а), ферромагнетика ниже $T_{\text{крит}}$ (б) и спинового стекла ниже $T_{\text{зам}}$ (в).

Хотя термодинамика полностью определяется функцией $P(q)$, введение перекрытия открывает возможность более подробного описания топологии пространства состояний на уровне, задаваемом функцией $P(q)$. Например, интересны соотношения между перекрытиями $q^{\alpha\beta}$, $q^{\beta\gamma}$ и $q^{\alpha\gamma}$ трех произвольно выбранных чистых фаз α , β и γ . Для вероятности $P(q_1, q_2, q_3)$ того, что $q^{\alpha\beta} = q_1$, $q^{\alpha\gamma} = q_2$ и $q^{\beta\gamma} = q_3$,

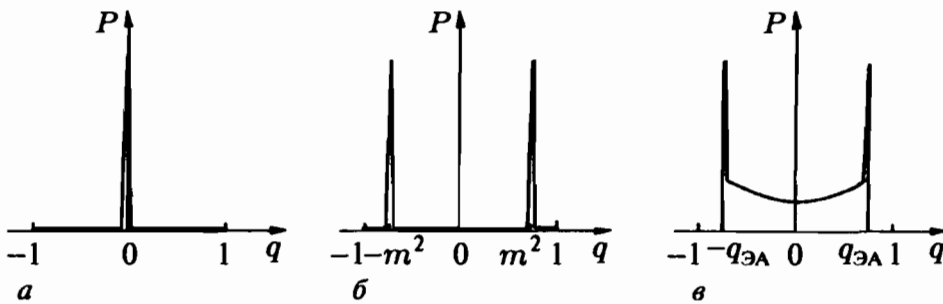


Рис. 10.4. Распределение вероятностей $P(q)$ для наложения двух чистых фаз: (а) парамагнетик $P(q) = \delta(q)$; (б) ферромагнетик $P(q) = \frac{1}{2}\delta(q + m^2) + \frac{1}{2}\delta(q - m^2)$; (в) спиновое стекло

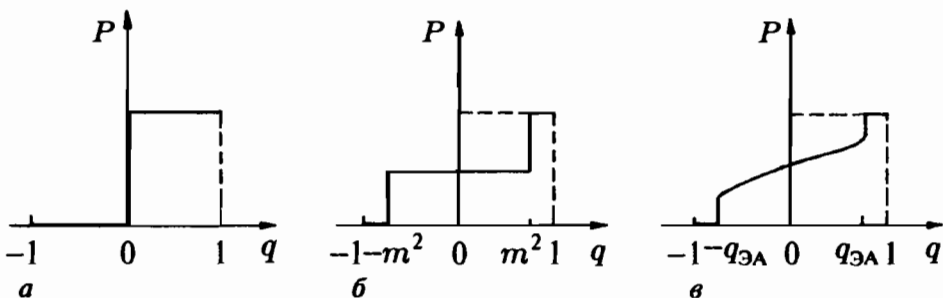


Рис. 10.5. Кумулятивное распределение вероятностей $x(q) = \int_{-1}^q P(q') dq'$ для тех же случаев, что и на рис. 10.4

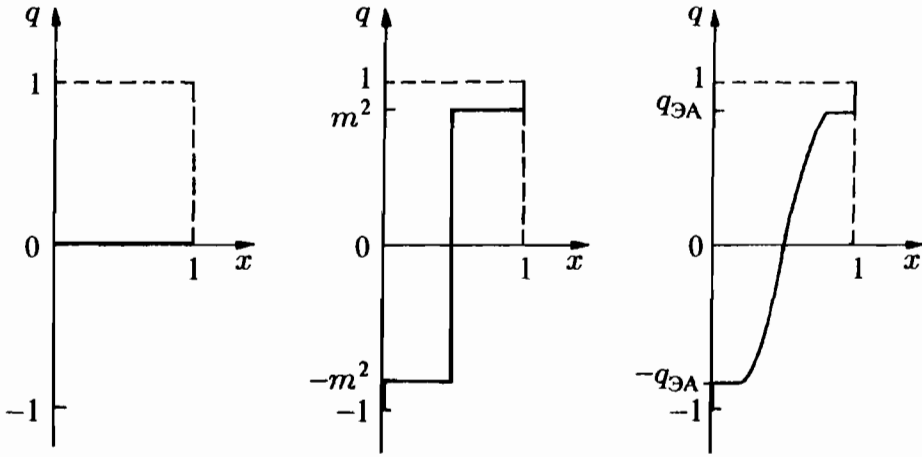


Рис. 10.6. Функция параметра порядка $q(x)$ для тех же случаев, что и на рис. 10.4

получается следующий поразительный результат (Mezard et al., 1984):

$$P(q_1, q_2, q_3) = \frac{1}{2} P(q_1) x(q_1) \delta(q_1 - q_2) \delta(q_1 - q_3) + \\ + \frac{1}{2} [P(q_1) P(q_2) \delta(q_1 - q_2) \delta(q_2 - q_3) + \text{перестановки}]. \quad (10.10)$$

Соотношение (10.10) означает, что для произвольных фаз α, β, γ либо все перекрытия попарно равны, либо по крайней мере два перекрытия равны и больше третьего. Если фазы интегрировать как точки некоторого пространства, а перекрытие как меру расстояния от этих точек, то все треугольники в таком пространстве должны быть либо равносторонними, либо равнобедренными с более коротким основанием. Пространства с такими топологическими свойствами называются ультраметрическими (Bourbaki, 1966). До начала интенсивных исследований по проблематике спиновых стекол ультраметрические пространства не играли никакой роли в физике, хотя в математической литературе они обсуждались с конца XIX столетия (см. Rammal et al., 1986). В результате проведенных исследований было установлено, что любое ультраметрическое пространство может быть

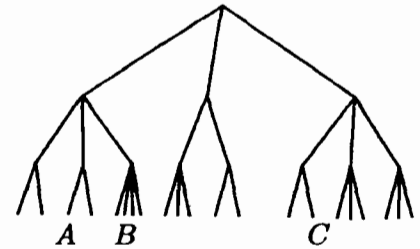


Рис. 10.7. Генеалогическое дерево в качестве примера ультраметрического пространства. Точки пространства представлены свободными узлами («концами ветвей») дерева, которые соответствуют также самому нижнему уровню иерархии

отображено на граф, имеющий вид перевернутого дерева (рис. 10.7). Точкам пространства соответствуют концы ветвей; расстояние между двумя точками A и B определяется как число «поколений», которые нужно отсчитать назад, чтобы дойти до общего «предка» точек A и B . Например, на рис. 10.7 $d_{AB} = 2$, $d_{AC} = d_{BC} = 3$. Как нетрудно убедиться на примере графа, изображенного на рис. 10.7, приведенное выше определение расстояния гарантирует, что по крайней мере две стороны любого треугольника всегда имеют одинаковую длину. На математическом языке это означает, что справедливо сильное неравенство треугольника

$$d^{AB} \leq \max(d^{BC}, d^{AC}), \quad (10.11)$$

которое с учетом соотношения между перекрытием и расстоянием представимо в виде

$$q^{AB} \geq \min(q^{BC}, q^{AC}) \quad (10.12)$$

и в этой форме эквивалентно соотношению (10.10).

Какие следствия имеет неравенство (10.12) для соотношения между многочисленными чистыми фазами в пространстве конфигураций спинового стекла? Рассмотрим отношение R_q , которое должно быть определено следующим образом:

$$\begin{aligned} \alpha R_q \beta &\leftrightarrow q^{\alpha\beta} \geq q, \\ -q_{ЭА} &\leq q \leq q_{ЭА}. \end{aligned} \quad (10.13)$$

Отношение R_q может означать, например, «подобная q », так как оно возникает только между двумя фазами, перекрытие которых не меньше q . Ясно, что отношение R_q обладает следующими свойствами:

$$\alpha R_q \alpha, \quad (10.14)$$

так как $q^{\alpha\alpha} = q_{ЭА}$, и

$$\alpha R_q \beta \rightarrow \beta R_q \alpha, \quad (10.15)$$

так как $q^{\alpha\beta} = q^{\beta\alpha}$ (см. формулу (10.7)).

Наконец, из неравенства (10.12) следуют свойства

$$\alpha R_q \beta \text{ и } \beta R_q \gamma \rightarrow \alpha R_q \gamma, \quad (10.16)$$

так как при $q^{\alpha\beta} \geq q$ и $q^{\beta\gamma} \geq q$ мы получаем

$$q^{\alpha\gamma} \geq \min(q^{\alpha\beta}, q^{\beta\gamma}) \geq q.$$

Таким образом, отношение R_q рефлексивно (10.14), симметрично (10.15) и транзитивно (10.16), т. е. представляет собой отношение эквивалентности. В качестве такого отношения R_q порождает разбиение чистых фаз на непересекающиеся классы, причем каждая фаза принадлежит ровно одному классу эквивалентности. Тем самым фазовое пространство спинового стекла распадается на множество фаз, которые могут быть разделены на классы эквивалентности, любые два из которых не имеют общих элементов (или полностью совпадают) (рис. 10.8). Все фазы, принадлежащие одному классу, «подобны q ». Если определить с помощью соответствия (10.13) еще одно отношение $R_{q'}$ с $q < q' \leq q_{ЭА}$, то все сказанное выше об отношении R_q переносится и на $R_{q'}$. В результате каждый из классов эквивалентности, порожденных отношением R_q , в свою очередь распадается на непересекающиеся классы фаз, имеющие друг с другом перекрытие большее, чем $q' > q$ (рис. 10.8). Аналогичным образом можно продолжать все более тонкое разбиение на подклассы, пока не будет достигнуто равенство $q^{(n)} = q_{ЭА}$ и каждый класс эквивалентности не будет состоять только из одной фазы. Таким образом, из соотношения (10.10) следует вполне определенная иерархическая организация фаз спинового стекла ниже температуры замерзания $T_{зам}$.



Рис. 10.8. Фазовое пространство спинового стекла распадается на классы эквивалентности фаз с попарным перекрытием, превышающим q . Каждый из этих классов в свою очередь распадается на классы фаз с попарным перекрытием, превышающим $q' > q$

Такая иерархическая структура, наглядно изображенная на рис. 10.7, является существенным результатом теории спиновых стекол. Значение иерархической организации для динамики сложных систем обсуждается давно (Simon, 1962), однако

тесная взаимосвязь фрустрации и иерархического упорядочения была обнаружена впервые только при обсуждении результатов теории спиновых стекол. Компромиссы служат решениями фрустрированных задач на оптимизацию, — решениями, которые могут как походить друг на друга, так и отличаться одно от другого. Если мысленно представить себе решения задачи, порождаемые иерархией принятых решений (волевых актов), причем сначала решаются принципиальные, а затем частные вопросы, то подобие, или сходство, двух решений существенно определяется тем, в какой плоскости иерархии впервые принимаются различные решения (волевые акты) (см. рис. 10.7). В результате мы приходим к множеству возможных компромиссных решений, которые вполне очевидным образом могут быть упорядочены в соответствии с иерархией принятия решений (волевых актов). Непосредственное отношение всего сказанного к биологической эволюции очевидно, а следствия для решения технических задач оптимизации мы обсудим позднее.

Спиновые стекла, по-видимому, являются простейшими фрустрированными системами, их рассмотрение возможно в рамках статистической механики, т. е. с помощью математических методов. Поскольку спиновые стекла обнаруживают многие свойства, обладающие качественным сходством с наиболее существенными особенностями нефизических комплексных систем, можно ожидать, что понятие и методы, разработанные в ходе исследования таких систем, переносятся и на проблематику спиновых стекол.

10.3. Модель Андерсона пребиологической эволюции

В этом разделе мы еще раз вернемся к проблеме эволюции пребиологических макромолекул, которая была рассмотрена в гл. 6. Там было показано, каким образом может быть решен в принципе центральный вопрос этого периода эволюции — возникновение сложных, стабильно самовоспроизводящихся на протяжении многих поколений полинуклеотидных цепей из запаса недифференцированных мономеров с богатым содержанием энергии. Установлены существенно важные механизмы: самовоспроизведение, отбор и мутация. Их совместное действие, согласно модели Эйгена (*Eigen*, 1973) описывается уравнением, которое мы уже рассматривали в гл. 8:

$$\partial_t x_i(t) = [E_i - \langle E \rangle] x_i(t) + \sum_{j \neq i} [a_{ij} x_j(t) - a_{ji} x_i(t)]. \quad (10.17)$$

Введенная «руками» приспособленность E_i как мера «доброкачественности» полинуклеотида в зависимости от его структуры при этом более подробно не характеризуется. В этом разделе мы исследуем, какими общими свойствами должна обладать функция приспособленности, чтобы уравнение (10.17) действительно допускало решения, имеющие непосредственное отношение к упомянутой выше проблеме, и попытаемся выяснить, как эти свойства следуют из определенных физических и химических особенностей полинуклеотидов.

Начнем с рассмотрения мысленного эволюционного эксперимента, в результате которого образуются полинуклеотидные цепи длиной N . Обозначим через $M_{\text{возм}}$ число возможных стабильных цепей, а через $M_{\text{эксп}}$ — число экспериментально реализуемых цепей. Для упрощения предположим, что возникновение всех $M_{\text{возм}}$ стабильных макромолекул равновероятно. В соответствии с гл. 11, это (в случае дискретного хода эксперимента) позволяет вычислять образующуюся информацию по формуле

$$I = \log_2 \left(\frac{M_{\text{возм}}}{M_{\text{эксп}}} \right). \quad (10.18)$$

Если величина E_i устроена так, что могут встречаться только некоторые немногочисленные стабильные полинуклеотидные цепи, то образующаяся информация очень мала: $I = o(1)$. Простым примером был бы учет при вычислении E_i энергии связи только соседних мономеров. Так как связь гуанин—цитазин существенно сильнее, чем связь аденин—тимин, в рассматриваемом случае всегда доминируют цепи вида Ц—Г—Ц—Г—Ц—... . Таким образом, в данном случае не существует потенциально лучшей цепи, которая могла бы быть найдена с помощью мутации и отбора, т. е. эволюционный эксперимент завершается установлением «идеального» вида. Кроме того, исход эксперимента предreshен с самого начала и воспроизводится при повторении эксперимента. Следовательно, никакой информации эксперимент не производит, поскольку $M_{\text{возм}} = 1$, и мы получаем $I = 0$. Такие случаи жесткого выхода на конечную последовательность моделируются в разд. 6.4.

Рождение многих различных полинуклеотидов можно обеспечить, если слегка варьировать E_i в зависимости от i , т. е. если сделать график функции приспособленности очень плоским. В этом случае возможно образование почти всех мыслимых цепей, и можно ожидать, что $M_{\text{возм}} \sim 4^N$. Однако в ходе каждого эксперимента в результате ошибочного или сильно подавленного отбора получается недифференцированная масса полинуклеотидов, поэтому $M_{\text{эсп}} \sim e^N$ и мы снова получаем $I = o(1)$. При дальнейшей эволюции доминируют те мутации, которые приводят к напоминающей диффузию смене распределения частот, и виды получают возможность нестабильного воспроизведения на протяжении многих поколений.

На этих двух примерах мы отчетливо видим, каким требованиям должна удовлетворять функция приспособленности E_i . Только в том случае, если имеется экспоненциально много стабильных решений, из которых в каждом случае реализуются лишь немногие, мы получаем экстенсивно производимую информацию $I = o(N)$, что благоприятствует образованию все более длинных цепей. Повторное проведение экспериментов приводит в этом случае к различным результатам. В пространстве последовательностей полинуклеотидов мы имеем подлинное нарушение симметрии, исход эксперимента в этом случае уже не предreshен с самого начала. Наконец, множество потенциальных, не реализуемых в ходе повторных экспериментов решений создают достаточно возможностей для дальнейшей эволюции.

Неясно, каким образом может возникать функция приспособленности со столь специфической структурой. Аналогии между желательными свойствами функции приспособленности E_i и обсуждаемыми в разд. 10.2 характерными свойствами функции энергии спиновых стекол наводят на мысль о том, что специфическая структура функции приспособленности может быть обусловлена фрустрацией и беспорядком.

Рассмотрим более подробно одну модель, предложенную П. У. Андерсоном в 1983 г. Чтобы подчеркнуть формальные аналогии со спиновым стеклом Изинга, начнем с анализа цепей, построенных лишь из двух различных мономеров A и B и предположим, что действительно реализуемый случай с четырьмя основаниями обнаруживает качественное сходство с нашей упрощенной моделью. В нашем случае полинуклеотид может быть однозначно отображен на некоторую одномерную конфигурацию спинов S_i , причем $S_i = 1$ (-1) означает, что на i -м месте последовательности мы найдем основание A (B). Требуется найти функцию $E(\{S_i\})$, которая заданной последовательности $\{S_i\}$ ставит в соответствие определенную приспособленность E_i . Так как $S_i^2 = 1$, эту функцию можно разложить в ряд по произведениям спинов в различных узлах решетки:

$$E(\{S_i\}) = \sum_i h_i S_i + \sum_{i \neq j} J_{ij} S_i S_j + \sum_{i \neq j \neq k} C_{ijk} S_i S_j S_k + \dots \quad (10.19)$$

Воспроизведение цепей происходит по механизму Уотсона—Крика, в силу чего «наследником» цепи $\{S_i\}$ является цепь $\{-S_i\}$. Поэтому представляется заманчивым потребовать, чтобы $E(\{S_i\}) = E(\{-S_i\})$, отчего в разложении (10.19) пропадут все нечетные члены. Ограничиваясь первым неисчезающим членом, получаем:

$$E(\{S_i\}) \sim \sum_{i \neq j} J_{ij} S_i S_j. \quad (10.20)$$

Несмотря на формальное сходство с формулой (10.2), это выражение для $E(\{S_i\})$ само по себе не обеспечивает свойств спиновых стекол: необходимо выбрать такие J_{ij} , чтобы могли действовать беспорядок и фрустрация.

Существенной предпосылкой фрустрации являются конкурирующие взаимодействия, которые в одномерной модели могут иметь место только в том случае, если связи J_{ij} дальнедействующие (рис. 10.9). В действительности, в полинуклеотидных цепях существуют



Рис. 10.9. В одномерных системах возникают только дальние взаимодействия эффектов фрустрации

различные механизмы, способные осуществлять такие взаимодействия на больших расстояниях. На рис. 10.10 показаны два простых примера. В первом случае на расположении 12 оснований сказывается винтовая («спиральная») структура цепи — шаг винтовой линии. Такая структура может оказаться эффективной, например, когда цепь растет на субстрате. Вторая возможность создает эффективные дальнедействующие ограничения: если в двух местах цепи встречаются длинные взаимно дополнительные фрагменты, в результате чего цепь складывается и возможность дальнейшего воспроизведения утрачивается. Такой механизм устанавливает столь жесткие ограничения упоминавшейся выше предпочтительности связи Ц—Ц, что казавшаяся сначала оптимальной молекула Г—Ц—Г—Ц—... утрачивает способность к самовоспроизведению.

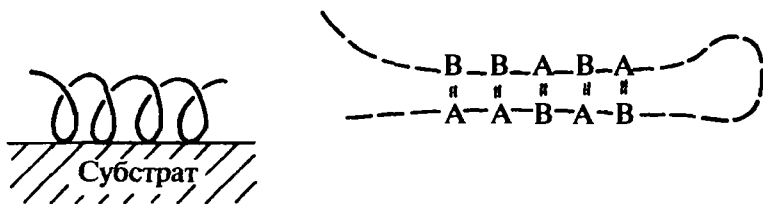


Рис. 10.10. Винтовая структура и складывание как примеры дальних взаимодействий между основаниями полинуклеотида

Ясно, что взаимодействия такого рода и различного происхождения в некоторых случаях вступают в противоречие между собой. К тому же существуют многочисленные причины для случайного или хаотически действующего влияния окружающей среды, простым примером которого могут служить аморфные или несоизмеримые субстратные материалы. Поэтому функция приспособленности $E(\{S_i\})$, задаваемая формулой (10.20), со случайными и дальнедействующими связями J_{ij} является правдоподобной моделью эволюции пребиологических макромолекул.

Ясно, что взаимодействия такого рода и различного происхождения в некоторых случаях вступают в противоречие между собой. К тому же существуют многочисленные причины для случайного или хаотически действующего влияния окружающей среды, простым примером которого могут служить аморфные или несоизмеримые субстратные материалы. Поэтому функция приспособленности $E(\{S_i\})$, задаваемая формулой (10.20), со случайными и дальнедействующими связями J_{ij} является правдоподобной моделью эволюции пребиологических макромолекул.

Для исследования свойств таких молекул были проведены обширные численные эксперименты на компьютерах (Rokhsar et al., 1983). Как показали работы Блума (Blum, 1962), при этом возникает цикл с двумя фазами различной температуры, воспроизводящие колебания день—ночь на Праземле. К системе за один цикл подводится постоянное количество мономеров, при этом по аналогии с моделью Эйгена возникает давление отбора на образующиеся полинуклеотиды. В фазе с более низкой температурой мономеры и уже имеющиеся цепи выстраиваются последовательно, и при соединении примыкающих друг к другу соседних частей могут образовываться

более длинные цепи, которые в завершающей фазе с более высокой температурой снова отделяются от своей «матрицы» (рис. 10.11). В результате образуются все более длинные полинуклеотиды. В высокотемпературной фазе для всех имеющихся последовательностей приспособленность вычисляется по формуле (10.20):

$$\bar{E}(\{S_i\}) = \frac{E(\{S_i\})}{[N(N-1)/2]^{1/2}} \quad (10.21)$$

(величина $\bar{E}$ нормирована на длину цепи) и по формуле

$$P(\{S_i\}) = \frac{\exp \{ \beta(\bar{E} - \mu) \}}{1 + \exp \{ \beta(\bar{E} - \mu) \}} \quad (10.22)$$

определяет вероятность P выживания цепи. Параметр μ вводит масштаб приспособленности; например, $\mu = E$ приводит к $P = 1/2$, а параметр β регулирует резкость перехода от $P = 0$ к $P = 1$ (рис. 10.12). В зависимости от величины P последовательность либо изымается из системы, либо выживает и принимает участие в следующем цикле. Наконец, время от времени происходят стохастические перевороты спина, соответствующие изменениям последовательности оснований в результате мутаций. Результаты численного моделирования мы рассмотрели лишь кратко и на качественном уровне, более подробный их анализ читатель может найти в литературе (Rokhsar *et al.*, 1983). При надлежаще выбранных значениях параметров выживают около 10–15% всех полимеров, из которых, однако, лишь немногие ($\lesssim 10$) встречаются в большом числе. Эти доминирующие цепи образуются во всех экспериментах после примерно 100 поколений (температурных циклов) и до конца эксперимента (500–600 поколений) остаются стабильными. В различных экспериментах доминируют различные полинуклеотиды. Это свидетельствует о том, что результат не предопределен начальными и граничными условиями. Однако всегда образуется небольшая группа макроскопических стабильных последовательностей. Это свидетельствует о том, что модель правильно отражает существенные моменты эволюции полинуклеотидов, обсуждаются и ее обобщения на более поздние фазы молекулярной эволюции (Stein, Anderson, 1984).

Изложенные в этом разделе соображения показывают, что эффекты фрустрации позволяют правдоподобно объяснить важные предпосылки процессов эволюции,

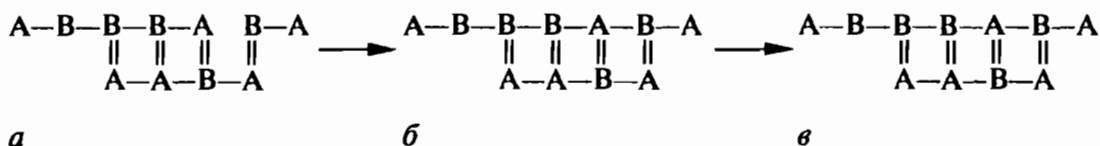


Рис. 10.11. Дополнительные участки цепей при более низких температурах располагаются рядами — (а), в результате чего возникает возможность связи двух цепей — (б). В высокотемпературной фазе молекулы диссоциируют и разделяются друг от друга — (в)

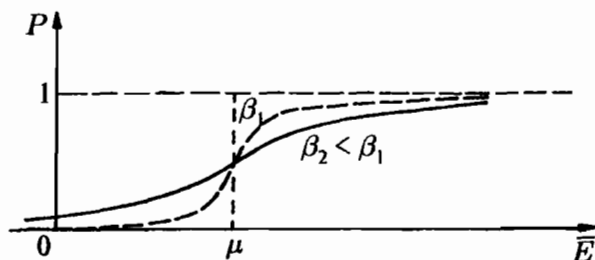


Рис. 10.12. Вероятность выживания P полинуклеотидной цепи как функция ее приспособленности E , задаваемая формулой (10.22), при двух различных значениях параметра β

например, специфическую структуру функции приспособленности. Конкурирующие взаимодействия являются универсальным свойством сложных систем, поэтому следует ожидать, что по мере развития эволюции роль явлений фрустрации будет все больше нарастать. Впрочем, до сих пор надежно охарактеризовать наиболее важные свойства фрустрированной системы удавалось лишь в простейших случаях. Общей теории фрустрированных случайных функций пока не существует, и большинство приложений, как и рассмотренная нами модель, основаны на прямых формальных аналогиях с задачей о спиновых стеклах. Успехи в этой еще только зарождающейся области математики и теоретической физики заведомо будут иметь интересные последствия для моделирования биологической эволюции.

Эволюция информации

О честь, что ты не понимаешь!

Фридрих Гельдерлин

11.1. Переработка информации как процесс самоорганизации

Информация неразрывно связана с физическими объектами и процессами. Немалое влияние было оказано физиками и при создании основ теории информации. Решающее влияние на формирование идей Шеннона оказали Максвелл, Больцманн, Гиббс, Сциллард и фон Нейман. Особенно выдающееся научное значение имели статья Сцилларда (1929) «Об уменьшении энтропии в термодинамической системе при вмешательстве разумного существа» и монография Бриллюэна (1956) «Наука и теория информации». Физики хорошо осознали важный факт: перенос информации между системами всегда связан с переносом энергии и энтропии. Вместе с тем информация не тождественна энергии или энтропии. Она означает некое другое качество. Другой факт, сознавать который также важно, состоит в том, что информация может встречаться только в системах с несколькими возможными состояниями, в которой относительно реального состояния системы имеется некоторая неопределенность. Мы имеем в виду макросостояния, так как информационные системы по существу носят макроскопический характер. Наконец, следует пояснить, что информационные системы могут действовать только в неравновесных условиях (*Волькенштейн, Чернавский, 1979; Küppers, 1986; Ebeling, 1987*). Для информационных систем необходима накачка высокоценной энергией. Тем самым устанавливается связь между переработкой информации и самоорганизацией. Сознательно упрощая ситуацию, физик понимает под информацией обменную величину, тесно связанную с энергией и энтропией и понижающую неопределенность состояния системы. Мы исходим из того, что обмен информацией может происходить между двумя системами, условно называемыми источником и приемником.

Существуют две основные ситуации.

1. Источник приготавливает приемник с помощью сигнала, вызывая в приемнике переход из одного состояния в другое.

2. Приемник является средством наблюдения, измерительным прибором: прием сигнала указывает на переход источника в определенное состояние.

Приготовление и измерение представляют собой необратимые акты, их невозможно обратить. Сложные информационные системы, например, компьютер, используют последовательности изменений состояний; эти изменения также необратимы, но в принципе могут протекать и обратимо или почти обратимо. Этому направлению посвящены проведенные в последнее время исследования Беннета, Ландауэра и Фейнмана. Только начальное приготовление и заключительное наблюдение должны быть безусловно необратимыми, так как в противном случае было бы нарушено второе начало термодинамики.

Изложенные выше понятия и в частности физическое определение информации как уменьшения неопределенности состояния позволяют физике ввести в своей системе количество (величину) информации. Семантическая сторона информации не поддается физическому описанию; поэтому мы вынуждены полностью исключить ее из рассмотрения. Обмен информацией между двумя системами неизбежно связан с обменом энергией и энтропией. Хотя обмен энергией необходим, интенсивность обмена несущественна: даже наименьшие количества энергии, например, квант света, могут уменьшить неопределенность состояния системы. С другой стороны, как показывают фундаментальные исследования Сцилларда и Бриллюэна, уменьшение неопределенности количественно связано с обменом энтропией. Количество передаваемой информации определяется уменьшением неопределенности:

$$\Delta I = -\Delta H, \quad (11.1)$$

при этом неопределенность для системы с n состояниями по Шеннону задается формулой

$$H = \sum_{i=1}^n p_i \log p_i^{-1}. \quad (11.2)$$

Шенноновская H -функция есть среднее значение неопределенностей отдельных состояний $\log p_i^{-1}$. Сравнивая выражение (11.2) с энтропией Больцмана—Гиббса, мы обнаруживаем, что каждое изменение неопределенности связано с изменением энтропии системы формулой

$$\Delta S = k_B \Delta H. \quad (11.3)$$

Величина H вычисляется с помощью натуральных логарифмов, а k_B — постоянная Больцмана:

$$k_B = 1,381 \cdot 10^{-23} \text{ Дж/К}. \quad (11.4)$$

Таким образом, получение информации о системе осуществляется ценой передачи энтропии. Иначе говоря, поток информации всегда сопровождается пропорциональным ему потоком энтропии. Взаимосвязь между энтропией и мерой информации носит качественный характер и весьма существенна, это — фундаментальный закон, связывающий физику и теорию информации. Правда, не каждая передача энтропии связана с передачей информации, поэтому в более общем случае справедливо неравенство

$$\Delta I \leq k_B^{-1} |\Delta S|. \quad (11.5)$$

Оно означает, что передача информации всегда меньше (в единицах k_B) передачи энтропии. В реальных процессах энтропия может быть передана многими различными способами, например, путем теплопередачи; передача информации представляет собой лишь частный случай переноса энтропии между двумя системами. В гл. 3 мы показали, что энтропия удовлетворяет не закону сохранения, а второму началу термодинамики. Это начало утверждает, что энтропия, хотя и может быть произведена, никогда не уничтожается. Поэтому энтропия и тем самым неопределенность состояния системы может уменьшаться только в системах с накачкой, отдающих энтропию окружающей среде. Сама по себе энтропия и тем самым неопределенность состояния системы может только возрасти, т. е. информация сама по себе может только утрачиваться. Только в открытых системах (с накачкой), отдающих энтропию, информация может приобретаться. Так как в термодинамическом равновесии энтропия и тем самым неопределенность состояния максимальны, для информационных процессов необходимы неравновесные условия. Поэтому переработку информации можно рассматривать как некую частную разновидность самоорганизации.

Количественное соотношение между физической энтропией и неопределенностью состояния позволяет высказать ряд важных физических утверждений об информационных процессах. Соотношение (11.5) Сциллара—Бриллюэна между энтропией и шенноновской мерой неопределенности можно в определенном смысле сравнить с соотношением Эйнштейна между массой и энергией

$$E = mc^2. \quad (11.6)$$

Оба соотношения эквивалентности означают не тождественность величин, стоящих в их правых и левых частях, и не превращаемость одной величины в другую. Подобно тому, как масса не может переходить в энергию, физическая энтропия не может переходить в неопределенность состояния. Однако одна форма массы (соответственно, энтропии) может превращаться в другую, и параллельно с этим превращением соотношение эквивалентности вызывает соответствующее превращение одной формы энергии (соответственно, неопределенности состояния) в другую. Именно в этом смысле мы можем говорить об информационной энтропии. Речь идет о форме энтропии, непосредственно связанной с информационными процессами. Из подобной точки зрения следует, что не каждая передача энтропии связана с передачей информации, а лишь такая, которая сопряжена с частью, наделенной информационными свойствами, т. е. с информационной энтропией. Полная передача энтропии при информационном процессе, разумеется, может быть больше, чем доля информационной энтропии, о чем и свидетельствует неравенство (11.5).

Как правило, изменение энтропии так же, как и изменение энергии, в информационных процессах чрезвычайно малы по сравнению с полным запасом энтропии и энергии. Часто из этого делается вывод о том, что энтропийный аспект информационных систем якобы несуществен. Так, известный биофизик Л. А. Блюменфельд считает, что в термодинамическом смысле биологическая система упорядочена ничуть не больше, чем равный ей по весу камень, и что соответствующая энтропия не превышает ту, которая приводит к испарению нескольких сотен граммов воды. Такие рассуждения не отражают сути дела. По сравнению с энергией покоя вещества все химические процессы соответствуют лишь небольшой «завитушке» на поверхности энергии, однако делать из этого вывод, будто энергия несущественна для химических процессов, означало бы немедленно вызвать противоречия. Для информационных процессов поток энтропии является существенной составляющей, хотя доля информационной энтропии от общей энтропии системы, как правило, чрезвычайно мала; однако именно эта доля существенна для обмена информации. Значение имеет не отношение информационной энтропии к полной энтропии, а абсолютная величина информационной энтропии, так как именно она соответствует уменьшению энтропии по сравнению с максимальным значением в состоянии равновесия. Подобно тому, как энергия при обмене между двумя системами выступает в различной форме (механическая, тепловая, электрическая энергия и т. д.), энтропия при обмене также принимает различные формы (термо-калорическая энтропия или энтропия смешивания), и одной из этих форм является информационная энтропия. Но подобно тому, как мы не приписываем связанной энергии ту или иную форму, не следует придавать какую-нибудь форму и связанной энтропии. Старый спор о том, можно ли вообще считать информацией связанную, покоящуюся информацию, с этой точки зрения беспредметен. Связанная энтропия (соответственно, неопределенность положения) является не информационной энтропией, а простой энтропией, она не имеет никакой формы и обретает форму только в процессе обмена.

Какие физические ограничения существуют для функционирования информационных систем? Одно фундаментальное физическое ограничение для всех информационных процессов следует из соотношения эквивалентности: за получение

информации нужно чем-то платить, и это «что-то» и есть энтропия. Один бит информации «стоит» от $k_B \ln 2$ и больше. Прием и передача информации связаны с необратимыми затратами энтропии, не зависящими от температуры. Ни одна самая хитроумная конструкция не может позволить нам обойти второе начало термодинамики. Если информационную систему подпитывать данными, то ее энтропия всегда будет возрастать, поскольку энтропия является мерой внутреннего беспорядка. Следовательно, в стационарном режиме энтропию необходимо отводить от системы, например, путем теплопроводности или излучения, что может создать определенные проблемы. Информационные системы должны экспортировать энтропию, и поэтому она соответствует процессам самоорганизации. К счастью, внутренние переходы в информационной системе могут протекать обратимо или почти обратимо, поэтому никаких принципиальных ограничений здесь не возникает. Понятие почти обратимого компьютера на протяжении нескольких лет интенсивно обсуждается физиками (*Bennet, Benioff, Fredkin, Toffoli, Landauer, Feynman*).

Физические ограничения на информационные системы представляют не только технический, но и принципиальный интерес. Из современной математической теории сложности известно, что затраты на решение многих задач растут экспоненциально с увеличением числа элементов и поэтому быстро превосходят возможности имеющейся вычислительной техники. Возможно, когда-нибудь в результате совместных математико-кибернетических, физических и биологических исследований станет ясно, какие задачи действительно непреодолимо сложны и не допускают сведения к более простым задачам и где проходит граница анализа сложных систем. На пути к всеобъемлющей междисциплинарной теории информации главную роль играет исследование переработки информации в биологических системах и ее эволюции.

11.2. Информация и жизнь

Способность хранить и перерабатывать информацию принадлежит к числу наиболее важных свойств живых систем. В ходе биологической эволюции эта способность необычайно усилилась, качественно и количественно. В живых системах используются главным образом три формы информации:

- 1) генетическая информация;
- 2) информация в нервной системе и в головном мозге;
- 3) экстрасоматическая информация, хранящаяся вне организма в записях, книгах и т. д.

Эти три формы информации образовались последовательно, одна за другой, в ходе эволюции. Хранение генетической информации началось с возникновением жизни на Земле 3–4 миллиарда лет назад и быстро развивалось, пока примерно через 10 миллионов лет после возникновения земноводных, пресмыкающихся и млекопитающих не вышло на плато на уровне около 10^{10} битов генетической информации (рис. 11.1). Количество информации, хранящейся в нервной системе и в головном мозге прежде всего на много порядков меньше. Лишь с появлением земноводных количество этой информации становится сравнимым с количеством генетической информации, а с появлением млекопитающих и особенно человека количество информации, хранящейся в нервной системе и в головном мозге, превосходит количество генетической информации на много порядков. Человек располагает около 10^{10} битами генетической информации и более 10^{13} битами информации в головном мозге и нервной системе, причем обладает еще доступом к еще большему количеству экстрасоматической информации в библиотеках и других хранилищах (*Sagan, 1978*). История эволюции хранения информации в живых существах наглядно представлена на рис. 11.1.

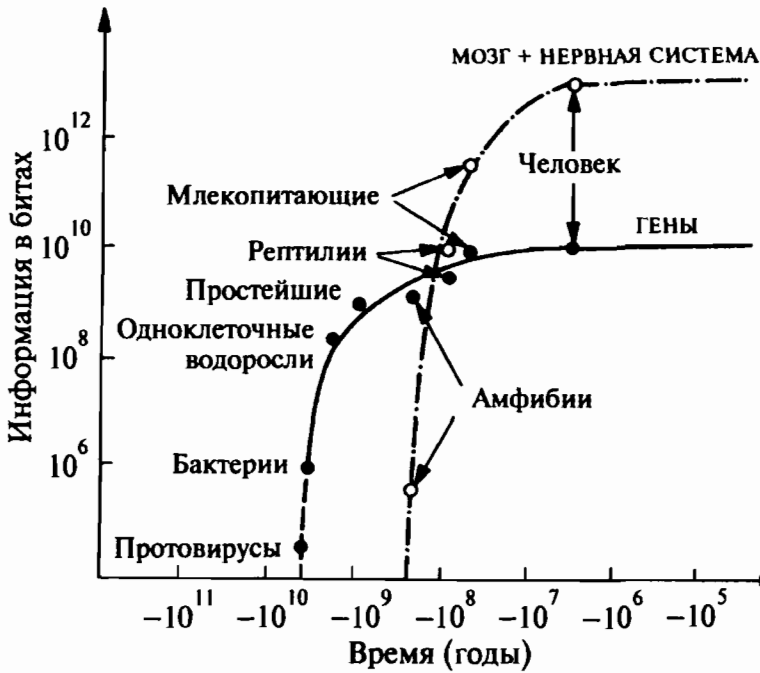


Рис. 11.1. Развитие генетической и нейронной информационной емкости в ходе биологической эволюции (по Сагану (*Sagan*, 1978))

Экстрасоматическое хранение информации (экстрасоматическое — вне тела) развивается на протяжении примерно 10^3 лет. Количество экстрасоматической информации уже существенно превзошло количество информации, хранимой в мозге и в нервной системе, и заведомо приведет к образованию третьего плато. Нас будет интересовать генетическая информация и физические аспекты ее хранения.

Ясно, что биологические системы в состоянии не только хранить и перерабатывать огромные количества информации, но и производить информацию. Чтобы понять, как функционируют биологические системы, проанализируем физические условия, которым должны удовлетворять информационные системы. В своем изложении мы будем следовать Волькенштейну и Чернавскому (*Волькенштейн, Чернавский*, 1979; *Волькенштейн*, 1979).

1. Система должна быть мультистационарной. В простейшем случае должны были бы существовать два стабильных стационарных состояния, а между ними одно нестабильное стационарное состояние. Такая система с двумя стабильными состояниями может хранить один бит информации. Разумеется, в системах с многими стационарными состояниями может храниться большее число битов информации. В мультистабильных системах в отличие от бистабильных систем существует много аттракторов, к которым притягиваются все траектории, принадлежащие некоторой определенной области. Число состояний-аттракторов всегда дискретно, и в конечной информационной системе оно, разумеется, конечно. Если N — число аттракторов, то информационная емкость системы, измеренная в битах, составляет величину

$$I = \log_2 N. \quad (11.7)$$

2. Система должна быть диссипативной. Вследствие требуемой устойчивости аттракторов должны существовать диссипативные механизмы, препятствующие покиданию траекториями занятой ими области аттрактора. Если система оказывается в области определенного аттрактора, то перейти в область другого аттрактора она может только в результате внешнего воздействия. Именно в этом и заключается механизм памяти информационных систем (*Landauer*, 1976, 1978).

3. Обусловленные флуктуациями переходы, принципиально всегда возможные, должны происходить настолько редко, что за характерное время функционирования системы они практически не встречаются. Из этого условия возникают определенные ограничения на размеры фазового пространства (чтобы исключить квантовомеханические эффекты), на число частиц (чтобы флуктуации относительно среднего числа частиц можно было считать малыми), а также на температуру и уровень внешнего шума.

4. В пространстве состояний должна существовать область, из которой достижимы все N аттракторов. К области какого аттрактора направится траектория из этой нейтральной области зависит от флуктуаций или внешних воздействий. Это и есть механизм производства и хранения информации.

Перечисленные физические условия показывают, что отнюдь не любая система способна к производству, хранению и переработке информации. Например, сразу же отпадают обширные классы консервативных гамильтоновых систем. К системам, удовлетворяющим условиям 1–4, относятся переключательные схемы электроники и микроэлектроники, а также мультистабильные реакции. В ходе естественной эволюции последний класс информационных систем выдвинулся на передний план. Идеальными информационными системами естественной эволюции являются реакции синтеза биополимеров. Биомолекулы представляют собой гетерополимеры с длиной цепи от 10^2 до 10^6 элементарных звеньев. Последние в случае полинуклеотидов состоят из 4 сортов нуклеотидов (А, Ц, Г и Т или У), а в случае полипептидов — из 20 сортов аминокислот (А, С, D, Е, ..., V, W, Y). Нейтральное исходное состояние реакции синтеза представляет собой смесь свободных элементарных звеньев необходимого брутто-состава. Состояния-аттракторы этой информационной системы соответствуют химически устойчивым конфигурациям биополимеров. Для линейного полимера, состоящего из λ элементарных звеньев, существуют

$$W = \lambda^\nu \quad (11.8)$$

вариантов полимера, где ν — его длина. В общем случае число W невообразимо велико. Например, если мы рассмотрим полинуклеотид длины $\nu = 1000$, то получим

$$W = 4^{1000} \sim 10^{600} \quad (11.9)$$

способов расположения нуклеотидов в цепи. Чтобы как-то упорядочить это огромное множество и сделать его более обозримым, целесообразно представить множество последовательностей геометрически в пространстве последовательностей. Такое представление позволяет наглядно упорядочить первичные структуры биополимеров. Вторичные и третичные структуры мы рассматривать не будем, а интересующихся читателей отошлем к литературе (Лифшиц, 1973; Волькенштейн, 1975; Damaschun, Buder, 1978; Blumenfeld, 1977). В дальнейшем мы рассмотрим взаимосвязь между физической структурой генетических носителей информации, т. е. первичной структурой полинуклеотидов и полипептидов, и некоторыми количественными аспектами хранящихся сообщений. Для этого нам понадобятся элементы теории передачи сообщений, теории сложности и теории языков, однако мы не ставим перед собой задачи построения биологически ориентированной теории информации (см. Tembrock, 1971, 1977; Hassenstein, 1970; Atlan, 1972; Küppers, 1986; Волькенштейн, 1986). Нас в первую очередь будут интересовать физические аспекты проблемы, однако мы не считаем разумным проведение строгих границ.

Открытие молекулярных основ и принципа кодирования наследственной информации по праву может считаться одним из важнейших результатов естествознания за последние 50 лет. При решении этой загадки существенную помощь современной биологии оказали такие физики, как Дельбрюк, Шрёдингер и Гамов.

Хранение, переработку и накопление информации можно рассматривать как характерные особенности живого. Живые организмы и образуемые ими структуры (биоценозы, трофические системы, общественные, социальные, экономические структуры и т. д.) отличаются столь высокой сложностью, что их построение, поддержание и дальнейшее развитие было бы немыслимо без хранения и передачи информации (накопленного опыта, чертежей конструкций, памяти). С другой стороны, не известна ни одна система, перерабатывающая информацию (за исключением тривиальных случаев), которая существовала бы вне и независимо от живого. Разумеется, компьютер может осуществлять некоторые информационные процессы независимо от людей, но в конечном счете эта информация имеет значение только в рамках технических и научных систем человека. Вопрос о том, где в процессе возникновения жизни лежит начало живого, согласно современным представлениям часто связывают с накоплением информации в химических соединениях.

Отличительная особенность физики состоит в исследовании величин, допускающих количественное описание. Соответственно, физика эволюции должна интересоваться измеримостью информации. Мерой информации служит информационная емкость запоминающего устройства (памяти), задаваемая логарифмом числа различных возможных состояний памяти. Однако эта величина ничего не говорит о степени загруженности памяти информацией. Такие меры, как информационная (шенноновская) энтропия или сложность последовательности символов, позволяют судить о доле емкости запоминающего устройства сверх минимально необходимой для хранения информации по отношению к доле минимально необходимой емкости (избыточность).

Наконец, гораздо труднее ответить на вопрос, какой ценностью обладает определенная информация для перерабатывающей ее системы, например, для живого существа. Этот вопрос затрагивает семантику и поэтому выходит за рамки чисто синтаксических исследований. Но в качестве первого приближения для оценки содержания информации можно воспользоваться синтаксическими мерами, поскольку информационное содержание высокоизбыточного сообщения заведомо не может быть максимальным.

Таким образом, мы в праве ожидать, что генетические последовательности символов должны обладать относительно высокой сложностью. С другой стороны, генотипическая информация представляет собой программу построения живых существ, длина которой была оптимизирована в процессе отбора. Из опыта работы с компьютерами нам известно, что, используя языки программирования, мы можем сократить программу и записать ее в более удобном для хранения в запоминающем устройстве виде. Если мы хотим применить те же методы к генетическому программированию, то прежде всего необходимо предположить существование определенных грамматических правил. Но наличие каких-то правил или закономерностей в последовательности символов означает уменьшение сложности; по-видимому, генетическая информация обладает не максимальной, а «почти» максимальной сложностью.

Альтернативная концепция «ценности информации» была предложена Волькенштейном (1978, 1979, 1984, 1986). По Волькенштейну, ценность информации определяется биологическим значением замены имеющихся носителей информации (последовательности) любыми другими носителями. В качестве конкретной меры Волькенштейн предлагает, например, для кодона сумму средних различий гидрофобий кодированных аминокислот, которые возникали бы при замене отдельных нуклеотидов в кодоне. Под гидрофобией здесь понимается изменение свободной энергии при переводе аминокислоты из спирта в воду. В отличие от сложности эта мера информации ориентирована на семантику и поэтому определена лишь для специальных (в данном случае генетических) носителей информации.

11.3. Биопоследовательности и дерево эволюции

Рассмотрим произвольную цепь молекул сортов $\{A_1, A_2, \dots, A_\lambda\}$ суммарной длины ν . Пусть

$$X^\nu := \{p_1 p_2 \dots p_\nu \mid p_k \in X = \{A_1, A_2, \dots, A_\lambda\}\} \quad (11.10)$$

— множество различных последовательностей. Число элементов этого множества (кардинальное число) равно λ^ν . Множество всех последовательностей длины $\nu \leq m$, которое мы обозначим $X^{[m]}$, содержит

$$S = \sum_{\nu=0}^m \lambda^\nu = \frac{\lambda^{m+1} - 1}{\lambda - 1} \quad (11.11)$$

элементов. Ясно, что множество всех последовательностей произвольной длины ($m \rightarrow \infty$), которое мы обозначим X^* , счетно и бесконечно. Для исследования биопоследовательностей с успехом применяются методы теории кодирования, используемые при передаче сообщений. Несколько примеров эффективности такого подхода уже существует (*Иваницкий и др.*, 1978; *Ebeling, Feistel, Jiménez-Montaño*, 1977; *Ebeling, Mahnke*, 1979). В смысле теории кодирования мы будем называть элементы последовательности p_k буквами, множество $X = \{A_1, A_2, \dots, A_\lambda\}$ — алфавитом и последовательности $\{p_1 p_2 \dots p_\nu\}$ — словами, или молекулярными словами. Слово в дальнейшем мы будем обозначать одной буквой, например, $p = \{p_1 p_2 \dots p_\nu\}$. Пустое слово условимся обозначать буквой e . Каждой последовательности p различными способами можно поставить в соответствие некоторое натуральное число $g(p)$. Оно называется гёделевским числом (последовательности p).

Гёделевские числа порождают взаимнооднозначное отображение множества последовательностей (слов) на множество натуральных чисел, т. е. для каждой последовательности p существует ровно одно натуральное число $g(p)$, и для каждого натурального числа n существует ровно одно слово $p = g^{-1}(n)$. Для данного слова длины ν над алфавитом $\{A_1, \dots, A_\lambda\}$

$$\{z_{i_1} z_{i_2} \dots z_{i_\nu}\}$$

гёделевское число равно

$$g(z_{i_1} z_{i_2} \dots z_{i_\nu}) = \sum_{k=1}^{\nu} i_k \lambda^{\nu-k}, \quad (11.12)$$

например,

$$g(\text{АГТ}) = 1 \cdot 4^2 + 3 \cdot 4^1 + 4 \cdot 4^0 = 32.$$

Множество всех слов X^* над алфавитом $X = \{A_1 \dots A_\lambda\}$ с операцией умножения

$$pq = \{p_1 \dots p_\nu q_1 \dots q_\mu\}$$

и единичным элементом e

$$ep = pe = \{p_1 \dots p_\nu\}$$

образует полугруппу — так называемую полугруппу слов.

Наконец, важно подчеркнуть, что множество всех слов X^* метризуемо. Проблема метрики на кодированных молекулярных последовательностях была впервые рассмотрена Бейером, Штейном, Смитом и Уламом (*Beyer, Stein, Smith, Ulam*, 1974), а позднее и рядом других авторов (*Sankoff*, 1975; *Ebeling, Feistel, Jiménez-Montaño*, 1977). В рамках теории кодирования расстояние между последовательностями равной

длины впервые было введено Хэммингом (1950). Расстояние (метрика) Хэмминга определяется как

$$d(p, q) = \sum_{k=1}^{\nu} (1 - \delta_{z_k^p z_k^q}). \quad (11.13)$$

Здесь k — положение элемента. Рассмотрим пример:

$$\begin{aligned} p &= \text{АТЦАГЦ}, \\ d(p, q) &= 2. \\ q &= \text{АГЦГГЦ}, \end{aligned} \quad (11.14)$$

Расстояние Хемминга между двумя последовательностями равной длины соответствует числу позиций, занятых несовпадающими элементами. В случае последовательностей различной длины расстояние Хэмминга определяется как минимальное число позиций, занятых несовпадающими элементами при всех возможных способах наложения одной последовательности на другую. Например, последовательности

$$p = \text{ТАЦГ}, \quad q = \text{ГТАТГГ}$$

мы накладываем одну на другую следующими способами:

$$\begin{array}{ccc} \text{ТАЦГ} & \text{еТАЦГ} & \text{ееТАЦГ} \\ \text{ГТАТГГ} & \text{ГТАТГГ} & \text{ГТАТГГ} \end{array}$$

При максимальном числе совпадений величина $d(p, q)$ оказывается равной 3. Аналитически такое расстояние можно записать следующим образом:

$$d(p, q) = \min_i \left\{ \sum_{k=-\infty}^{+\infty} (1 - \delta_{z_k^p z_{k+i}^q}) \right\}. \quad (11.15)$$

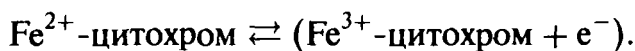
Можно показать, что выражение (11.15) обладает свойствами расстояния, или метрики:

$$\begin{aligned} d(p, q) &> 0 \quad \text{при } p \neq q; \\ d(p, q) &= 0 \quad \text{только при } p = q; \\ d(p, q) &= d(q, p) \quad \text{при всех } p, q; \\ d(p, q) &\leq (d(p, r) + d(r, q)) \quad \text{при всех } p, q, r. \end{aligned} \quad (11.16)$$

Расстояние между двумя молекулами ДНК p и q называется также генетическим расстоянием. Существует наименьшее число мутаций, или локальных изменений последовательностей, переводящих молекулу p в молекулу q и наоборот. Селлерс назвал это расстояние эволюционным расстоянием. Вследствие фундаментального характера кодирования информации в биопоследовательностях перечисленные выше математические свойства множества биопоследовательностей имеют некоторое значение и для описания биологической эволюции. Например, понятие генетического расстояния было с большим успехом использовано при построении дерева эволюции (Dayhoff, 1969; Dickerson, 1972; Beyer, Stein, Smith, Ulam, 1974; Sankoff, 1975; Муркин, Родин, 1977).

Большинство выполненных к настоящему времени исследований проведено на белке цитохром c , для которого точная последовательность элементарных звеньев известна более чем для 30 организмов (Dayhoff, 1969). Этот белок играет роль в цепи дыхания и поэтому содержится почти во всех организмах. В состав цитохрома входит Fe-порфирин; это кольцевая система, образующаяся при замыкании в кольцо

четырёх азотсодержащих пирольных колец с атомом железа в центре. Центральный атом железа участвует в переносе электронов в цепи дыхания, и в этом процессе определенную роль играет реакция



Вся молекула в целом представляет собой клетку из полипептидных цепей, внутри которой заключена активная группа (гем-группа). Связь с гем-группой осуществляется через положения аминокислот 14 (Cys), 17 (Cys), 18 (His) и 80 (Met). Эти центральные позиции совпадают, как показывает сравнение, в молекулах цитохрома с у всех организмов. Кроме того, совпадает примерно треть из 104 позиций в пептидных цепях, а в остальных позициях имеются более или менее сильные расхождения.

Построение дерева эволюции с помощью последовательностей цитохрома с основано на гипотезе, согласно которой все последовательности, встречающиеся в настоящее время в организмах, возникли из одной-единственной протеиновой прапоследовательности p_0 в ходе естественной эволюции. Прапоследовательность p_0 соответствует нуклеотидная прапоследовательность n_0 . Первичный процесс в ходе эволюции состоит в мутациях нуклеотидной последовательности, отражающихся в изменениях протеиновой последовательности. Число необходимых мутаций в точности соответствует генетическому расстоянию. Эволюционное дерево следует строить с двоичными ветвлениями. Оно должно «расти» из прапоследовательности, приводить к имеющейся ныне последовательности кратчайшим путем (с наименьшим числом промежуточных ветвлений) и не должно содержать замкнутых петель. Каждая ветвь дерева соответствует эволюционному пути от прапоследовательности к существующей ныне последовательности. Проблема состоит в том, что для N существующих ныне сортов имеется $(N - 1)$ возможных последних ветвлений, $(2N - 2)$ ветвей, а общее число топологически различных расположений промежуточных вершин графа (ветвлений) достигает

$$\frac{(2N - 2)!}{(N - 1)!2^{N-1}}. \quad (11.17)$$

Таким образом, для 33 финальных последовательностей мы получаем около 10^{44} , т. е. чудовищно большое число, возможных топологий эволюции.

Понятие генетического расстояния позволяет ввести критерий для отбора «лучших» деревьев эволюции. Фитч и Марголиаш (*Fitch, Margoliash*, 1967) первыми предложили связывать длину ветви с расстоянием между видами. При заданной топологии дерева они потребовали минимизировать выраженное в процентах стандартное отклонение между вычисленным расстоянием, разделяющим рассматриваемые виды, и фактическим расстоянием. «Наилучшая» топология дерева соответствует наименьшему выраженному в процентах стандартному отклонению от существующих ныне фактических расстояний. С учетом чрезвычайно большого числа возможных топологий дерева эффективный поиск наилучшей топологии возможен только с помощью мощных компьютеров. Составленная группой Дэйхоффа программа действует следующим образом (*Dayhoff*, 1969).

Прежде всего угадывается прапоследовательность, из которой выделяются известные последовательности цитохрома с существующих ныне видов. Особенно простая возможность отыскания прапоследовательности открывается, если воспользоваться следующим рецептом: выписать одну под другой все известные последовательности цитохрома с и под каждой позицией поставить ту аминокислоту, которая встречается чаще всего. Эйген (*Eigen*, 1979) воспользовался этим рецептом для поиска

прапоследовательности тРНК. Применительно к известным ныне последовательностям цитохрома с этот простой метод приводит к следующей прапоследовательности: GDVEKGKKIFVQKCAQCSTVEKGGKHKKTGPNLHGLFGRKTGQAPGFSYTDANKN KGITWGEDTLMEYLENPKKYIPGTKMIFAGIKKKGERALDIAYLK KATNE

Полужирные буквы обозначают здесь позиции, в которых все последовательности совпадают. Ясно, что именно эти позиции имеют жизненно важное значение для выполнения функции цитохрома с. Исходя из этой гипотетической прапоследовательности, мы строим дерево. Концы его ветвей соответствуют существующим ныне последовательностям. Затем компьютер вычисляет полное число замен аминокислот, ведущих от исходной последовательности к существующим теперь последовательностям. Топология дерева считается тем лучше, чем с меньшим числом замен аминокислот оно сопряжено. Компьютер сравнивает различные топологии и различные исходные последовательности и находит «лучшее (с точки зрения приведенного выше критерия) дерево эволюции». Полученный Дэйхоффом результат такой оптимизации представлен на рис. 11.2. Сильное сходство построенного дерева эволюции с результатами таксономических исследований свидетельствует об эффективности анализа эволюции, основанного на использовании метрических расстояний. Строгие математические исследования построения деревьев эволюции были предприняты Миркиным и Родиным (Миркин, Родин, 1977) и Дрессом (*Dress et al.*, 1982).



Рис. 11.2. Дерево эволюции, выведенное на основе последовательностного анализа цитохрома с по Дикерсону

Для обеспечения структуры и функционирования живых организмов требуется огромное количество информации, например, для вируса 10^4 битов, для бактерии 10^6 битов, для одноклеточных 10^8 битов и для млекопитающих 10^{12} битов. У млекопитающих подавляющая часть этого количества информации хранится в молекулах ДНК хромосом. Наследственность соответствует передаче сообщений от родителей детям. В цикле биосинтеза клеток переносчиками этой информации служат полипептидные цепи (белков, ферментов и т. д.), выступающие в роли носителей определенных функций. Механизм такого переноса информации хорошо известен (Волькенштейн, 1978, 1981).

Прежде всего информация переписывается на простую нуклеотидную нить — информационную, или матричную, РНК (мРНК) — с помощью специального фермента (РНК-полимеразы). Распознавание различных аминокислот, из которых состоит молекула соответствующего белка, производится молекулами транспортной РНК (тРНК), длина которых составляет около 70–80 нуклеотидных единиц. Молекулы тРНК позволяют распознавать различные аминокислоты и осуществлять их транспорт. Синтез белка происходит в рибосоме — «химической фабрике», в которой по образцу информационной РНК аминокислоты, транспортируемые тРНК, выстраиваются в правильной последовательности (рис. 11.3).

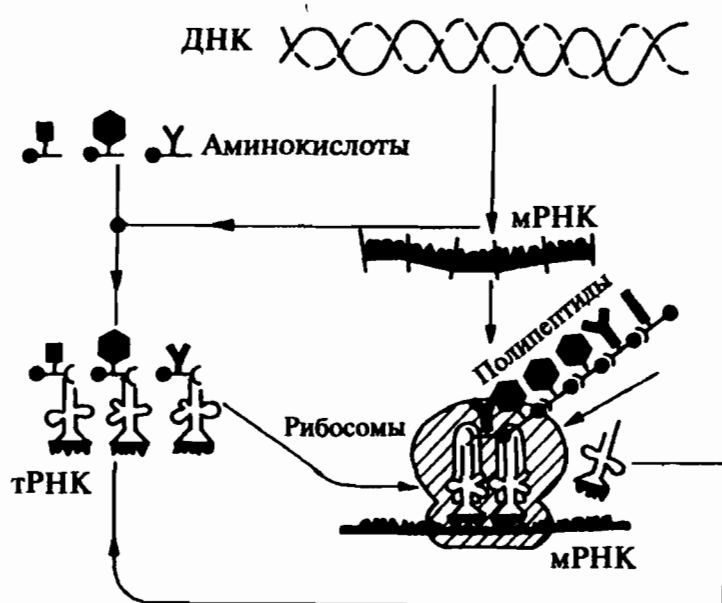
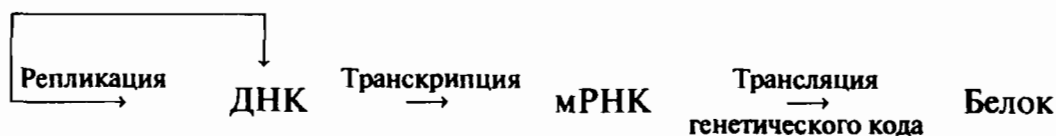


Рис. 11.3. Схема генетического аппарата синтеза белков

Знанием генетического кода мы по существу обязаны работам Хорана, Маттеи, Ниренберга, Очоа и их сотрудников. Каждая аминокислота кодируется триплетом из нуклеотидных оснований А, Ц, Г, У в случае РНК и А, Ц, Г, Т в случае ДНК. Ключ для декодирования («перевода») приведен в табл. 11.1.

Поток генетической информации можно наглядно представить с помощью следующей схемы:



Информационный поток исходит из ДНК и через РНК поступает к белку.

Таблица 11.1. Генетический код

		Вторая позиция				
		У	Ц	А	Г	
Первая позиция	У	Phe	Ser	Tyr	Cys	У
		Phe	Ser	Tyr	Cys	Ц
		Leu	Ser	Term	Term	А
		Leu	Ser	Term	Term	Г
	Ц	Leu	Pro	His	Arg	У
		Leu	Pro	His	Arg	Ц
		Leu	Pro	Gln	Arg	А
		Leu	Pro	Gln	Arg	Г
	А	Ile	Thr	Asn	Ser	У
		Ile	Thr	Asn	Ser	А
		Ile	Thr	Lys	Arg	Ц
		Met	Thr	Lys	Arg	Г
	Г	Val	Ala	Asp	Gly	У
		Val	Ala	Asp	Gly	Ц
		Val	Ala	Asp	Gly	А
		Val	Ala	Asp	Gly	Г

11.4. Энтропия и сложность

Количественная оценка хранимых в биополимерах и переносимых ими сообщений чрезвычайно затруднительна. Она должна исходить прежде всего из того, что обмен информацией всегда представляет собой взаимное отношение между ее отправителем и получателем и затрагивает следующие стороны (*Klix*, 1974).

1) Количественная и структурная сторона, т. е. определение количества информации, соподчинений и отношений между элементами сообщения. Это — метрический и синтаксический аспект информации.

2) Содержательная сторона, т. е. определение того, что означают входящие в сообщение символы для принимающей системы. Два сообщения одинакового объема и одинаковой структуры могут обладать различной значимостью. Это — семантический аспект информации.

3) Оценочная сторона, т. е. полезность и значимость принятого сообщения. При одной и той же значимости важность и польза сообщения могут зависеть от условий, в которых находится получатель. Это — прагматический аспект информации.

Все эти вопросы являются предметом изучения теории информации (*Abramson*, 1963; *Яглом, Яглом*, 1973; *Ebeling et al.*, 1998; *Tembrock*, 1971; *Стратонович*, 1975; *Klix*, 1974; *Völz*, 1982, 1983). Нас особенно интересует один специальный аспект информации, а именно ее метрический аспект. Творцом важнейших из мер информации — информационной энтропии — является американский инженер Клод Шеннон, опубликовавший в 1948 г. на эту тему фундаментальную монографию. Шеннон связал определение энтропии системы с состояниями $i = 1, 2, \dots, W$, предложенное Гиббсом и Эйнштейном, и соответствующие вероятности p_i , введя соотношение

$$S = -k \sum_{i=1}^W p_i \ln p_i. \quad (11.18)$$

В случае равновероятных событий $p_i = W^{-1}$ соотношение (11.18) сводится к формуле Больцмана—Планка—Хартли

$$S = k \ln W. \quad (11.19)$$

Основная идея Шеннона состояла в том, чтобы построение последовательности символов из алфавита $\{A_i; i = 1, 2, \dots, \lambda\}$ рассматривать как случайный процесс, в котором каждый символ генерируется с вероятностью p_i . В качестве меры неопределенности, устраняемой при реализации буквы A_i , Шеннон выбрал величину $\log p_i^{-1}$. В среднем каждый символ устраняет неопределенность

$$H = \sum_{i=1}^{\lambda} p_i \log \left(\frac{1}{p_i} \right) = - \sum_{i=1}^{\lambda} p_i \log p_i, \quad (11.20)$$

(основание логарифмов может быть выбрано произвольно). Величина H называется информационной энтропией (иногда просто информацией или энтропией). Словесно рецепт Шеннона можно сформулировать следующим образом:

Информация на символ = средняя величина неопределенности на символ.

Аналогия шенноновской информационной энтропии (10.20) с физической энтропией системы с W состояниями Гиббса—Эйнштейна очевидна. Шенноновская информационная энтропия зависит только от вероятностей, с которыми стохастический источник порождает символы A_i , поэтому она является свойством источника сообщений и характеризует не отдельное сообщение, состоящее из конкретной последовательности молекул или букв. Последнее утверждение имеет большое значение для нашей проблемы; другие меры информации, характеризующие отдельное сообщение, мы рассмотрим в конце этого раздела.

Во многих приложениях целесообразно наряду с шенноновской энтропией ввести высшие энтропии, использующие в качестве элементов не отдельные буквы, а r -слова. Например, если алфавит содержит λ букв, то над ним можно образовать λ^2 пар букв. Источник, порождающий пары $(A_i A_k)$ с вероятностями $p(A_i A_k)$, обладает, по определению, информационной энтропией второго порядка (Яглом, 1973)

$$H_2 = - \sum_{i,k=1}^{\lambda} p(A_i A_k) \log p(A_i A_k). \quad (11.21)$$

Если $p(A_i A_j \dots A_k)$ — вероятность порождения r -слов, то информационная энтропия r -го порядка определяется выражением

$$H_r = - \sum_{i,j,k=1}^{\lambda} p(\overbrace{A_i A_j \dots A_k}^r) \log p(\overbrace{A_i A_j \dots A_k}^r). \quad (11.22)$$

Информационную энтропию r -го порядка можно снова связать с отдельными буквами, положив, по определению (Gatlin, 1972; Ebeling, Feistel, Herzel, 1987),

$$H^{(r)} = \frac{1}{r} H_r, \quad h_r = H_{r+1} - H_r. \quad (11.23)$$

Для информационной энтропии, разумеется, справедливо равенство $H^{(1)} = H_1$. В предположении, что источник информации стационарен и генерирует эргодическую последовательность Маркова, согласно Хинчину (Khinchin, 1957) справедливо неравенство $H^{(r+1)} \leq H^{(r)}$ и существует предел

$$H = \lim_{r \rightarrow \infty} H^{(r)} = \lim_{r \rightarrow \infty} \frac{H_r}{r} = \lim_{r \rightarrow \infty} h_r. \quad (11.24)$$

Величина H называется энтропией источника информации. В конкретных случаях вычислить предел (11.24) часто бывает очень трудно, так как требуется анализировать весьма длинные последовательности. Задача определения энтропии источника H сильно упрощается, если источник информации обладает лишь конечной памятью. Мы говорим об источнике без памяти ($m = 0$), если каждый порождаемый символ полностью не зависит от предыдущего. Память первого порядка ($m = 1$) означает, что генерация символа A_j зависит только от символа, стоящего непосредственно перед A_j , т. е. если вероятности удовлетворяет соотношениям

$$\begin{aligned} p(A_i A_j) &\neq p(A_i) p(A_j), \\ p(A_i A_j A_k) &= \frac{p(A_i A_j) p(A_j A_k)}{p(A_j)}. \end{aligned} \quad (11.25)$$

Соответствующие разложения допускают и высшие вероятности. Подставляя выражения для информационной энтропии источника в соотношение (11.22), получаем

$$H_r = H_1 + (r - 1)(H_2 - H_1), \quad (11.26)$$

т. е. все высшие энтропии сводятся к H_1 и H_2 . Разделив на r и совершив предельный переход, получим для марковского источника первого порядка

$$H = \lim_{r \rightarrow \infty} \frac{H_r}{r} = H_2 - H_1. \quad (11.27)$$

Обобщение на случай марковского источника m -го порядка (Gatlin, 1972) имеет следующий вид:

$$H_r = H_{r-1} + (r - m)(H_{m+1} - H_m). \quad (11.28)$$

После деления на r и предельного перехода $r \rightarrow \infty$ получаем

$$H = \lim_{r \rightarrow \infty} \frac{H_r}{r} = H_{m+1} - H_m. \quad (11.29)$$

Мы видим, что для анализа марковского источника m -го порядка достаточно вычислить энтропии до $(m + 1)$ -го порядка включительно. К сожалению, порядок памяти биополимеров точно не известен, так как мы не располагаем еще достаточным количеством точных данных. В простейшем случае мы принимаем $m = 0$ для белков и $m = 1$ для полинуклеотидов (Galtin, 1972). В новых работах постулируют $m = 1$ для полипептидов и $m = 2$ для полинуклеотидов. В дальнейших исследованиях мы опираемся главным образом на последние предположения. Как видно из примеров, последовательность энтропий $H^{(r)}$, $r = 1, 2, 3, \dots$, позволяет сделать важные заключения о внутренней структуре соответствующего языка. Выбор основания логарифмов в соотношениях (11.20) и (11.21) до сих пор мы оставляли без внимания. Фиксируя основание, мы тем самым задаем единицу измерения энтропий. В теории информации чаще всего выбирают основание 2; возникающую при этом единицу принято называть двоичной (бинарной) единицей, или сокращенно битом. Если алфавит содержит λ символов, то в качестве основания логарифмов разумно выбрать число λ и назвать соответствующую единицу λ -единицей. Если выбрать натуральные логарифмы, то мы приходим к описанию энтропий без единицы меры.

Рассмотрим теперь некоторые применения введенных выше энтропий. Основная трудность состоит в том, что в случае передачи биологических или общественных сообщений мы никогда не имеем дела с определенными источниками, обладающими известными вероятностями испускания сигнала. Известны более или менее длинные

ANuKONNTEuIGHuNICHTuDIEuKLEINSTEuBEWEGUNGMEHRu
 AUSFUEHREN.ICHuSTAND,ODERuHINGuVIELMEHR,IMuLUFTBETT,MITu
 DERuMETALLHUELLEuINuEINSuVERFUGT.ALSuICHuDENuBLICKuHOB,SAHu
 ICHuDURCHuDIEuVORGEWOELBTEuSCHEIBEuDIEuWAENDEuDESu
 SCHACHTESuUNDuWEITERuOBENuMODDARDSuDARUEBERGENEIGTESu
 GESICHT.ESuVERSCHWANDuSOFORT,UNDuFINSTERNISuBRACHuHEREIN,
 DENNuVONuOBENuWURDEuDERuSCHWEREU SCHUTZKEGELuAUFGESETZT.
 ICHuHOERTEuDENuACHTMALuWIEDERHOLTENuPFIFFuDERu
 ELEKTROMOTOREN,DIEuDIEuSCHRAUBENuFESTZOGEN.DANN,DASu
 ZISCHENuDERuLUFT,DIEuINuDIEuAMORTISATORENuEINGELASSENu
 WURDE.DASuAUGEuGEWOEHNTEuSICHuANuDIEuFINSTERNIS.SCHONu
 SAHuICHuDENuBLASSGRUENENuUMRISSuDESuEINZIGENu
 KONTROLLANZEIGERS.FERTIG,KELVIN?ERTOENTEuESuINuDENu
 KOPFHOERERN.FERTIG,MODD.

Относительно частоты оснований и пар оснований в последовательностях ДНК имеется необычайно богатый материал. В качестве примера в табл. 11.2 приведены относительные частоты для ДНК *Micrococcus phlei*. Информационный анализ соответствующих последовательностей оснований с помощью шенноновской энтропии проведен в книге Гатлина (Gatlin, 1972).

В дальнейшем мы будем заниматься анализом четырех приведенных выше стандартных последовательностей. Мы предполагаем, что для белков можно ввести память первого порядка ($m = 1$), а для остальных последовательностей — память второго порядка ($m = 2$). Последовательность информационных энтропий для рассматриваемых примеров приведена в табл. 11.2 и 11.3.

Таблица 11.2. Относительные частоты для оснований и пар оснований ДНК *Micrococcus phlei*

	А	Т	Ц	Г
$h^{(1)}(\cdot)$	0,164	0,162	0,337	0,337
$h^{(2)}(A)$	0,024	0,031	0,064	0,045
$h^{(2)}(T)$	0,012	0,026	0,061	0,063
$h^{(2)}(C)$	0,063	0,045	0,090	0,139
$h^{(2)}(G)$	0,065	0,060	0,122	0,090

Таблица 11.3. Информационные энтропии естественных последовательностей в алфавите из λ символов

	H_1	H_2	H_3	H_4	H_5	H
ДНК	0,988	1,94	2,75	3,17	3,38	1,23
РНК	0,983	1,94	2,79	3,24	3,42	1,20
Белок	0,918	1,47	1,53	1,54	1,54	1,64
Солярис	0,840	1,45	1,77	1,94	1,98	1,79

Эти исследования показывают, что между последовательными символами естественных языков существуют определенные корреляции. Такие корреляции уменьшают количество передаваемой информации, поскольку уменьшается степень устранения неопределенности. Уменьшение количества информации отражает взаимное влияние символов в тексте. Свойство текстов содержать излишнюю информацию называется избыточностью. Слово избыточность, по латыни *redundantia* — переизбыток, означает «велеречивость» или «избыток слов». Передача избыточной информации в естественных языках обеспечивает надежность, защищая от возможной

ошибки. Можно поэтому утверждать, что избыточность означает повторную информацию и поэтому одновременно потенциальную информацию на случай искаженной передачи (Hassenstein, 1970). С помощью информационных энтропий избыточность можно определить следующим образом (Klix, 1974):

$$R_r = 1 - \frac{H_r}{H_r^{\max}}, \quad R = 1 - \frac{H}{H^{\max}}. \tag{11.31}$$

Здесь $H^{\max}$ — максимальное значение информационной энтропии, задаваемой с помощью равномерного распределения λ^r возможных частичных слов, т. е.

$$H_r^{\max} = \log \lambda^r = r \log \lambda, \quad H^{\max} = \log \lambda. \tag{11.32}$$

Вычисленные по этим формулам избыточности для случаев, приведенных в табл. 11.3 и 11.4, см. в табл. 11.5.

Таблица 11.4. Информация в битах на одну букву

	$H^{(1)}$	$H^{(2)}$	$H^{(3)}$	$H^{(4)}$	$H^{(5)}$
ДНК	1,98	1,94	1,88	1,58	1,35
РНК	1,97	1,94	1,86	1,62	1,37
Белок	3,96	3,17	2,20	1,67	1,33
Солярис	4,20	3,61	2,95	2,43	1,98
Немецкий	4,10	3,50	2,80	—	—
Английский	4,03	3,32	3,10	—	—
Русский	4,35	3,53	3,01	—	—
Самоа	3,40	2,68	—	—	—

Таблица 11.5. Избыточность различных естественных текстов и языков

	R_1	R_2	R_3
ДНК	0,012	0,028	0,083
РНК	0,017	0,030	0,070
Белок	0,082	0,275	0,490
Солярис	0,160	0,323	0,410
Немецкий	0,14	0,26	0,33
Английский	0,15	0,30	0,35
Русский	0,14	0,29	0,40
Самоа	0,17	0,34	—

Как видно из табл. 11.5, языки биополимеров обладают относительно малой избыточностью по сравнению с языками человеческого общения. Это свидетельствует о высокой надежности передачи биологической информации. В языке ДНК заметно выделены трехбуквенные слова ГГГ = Gly и ТЦГ = Ser, в языке РНК — трехбуквенные слова ГГГ = Gly, ЦУГ = Leu, АЦЦ = Thr, и ГГА = Val (они соответствуют наиболее часто встречающимся аминокислотам). Этим объясняется сильное повышение избыточности групп из трех букв. В случае аминокислот мы наблюдаем особенно заметное повышение избыточности при переходе к группам из двух букв. В этом отражается предпочтительность таких пар, как KG, TG, VE, EK, GK, IF, KT, PG, NK, GI, YI, означающая, что соответствующие аминокислоты образуют между собой связи особенно часто.

Для структуры языка наряду с абсолютными частотами знаков особое значение имеют бинарные отношения, наглядно представимые на диаграмме — языковом графе. Обозначим буквы алфавита кружками, а наиболее частые корреляции — штриховыми стрелками. На рис. 11.4, 11.5 и 11.6 показаны языковые графы для языка нуклеиновых кислот, языка протеинов (белков) и немецкого языка. Узлы графа

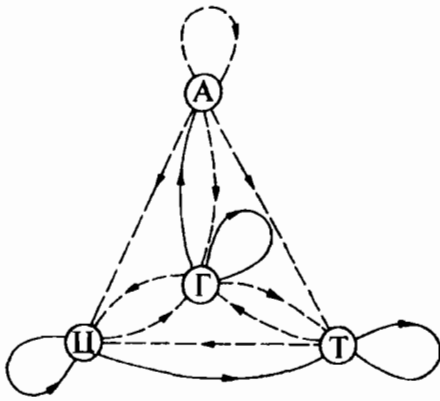


Рис. 11.4. Граф языка ДНК (сплошные стрелки соответствуют частоте последования от 8 до 13%, штриховые стрелки — частоте последования от 4 до 8%)

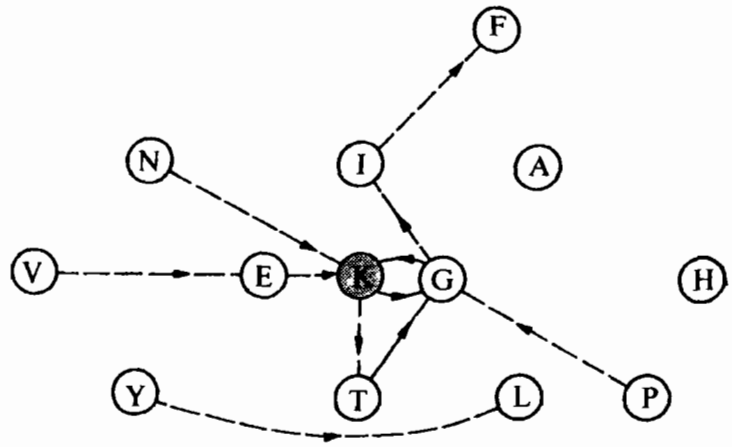


Рис. 11.5. Граф языка белков (сплошные стрелки соответствуют частоте последования от 2 до 3%, штриховые стрелки — частоте последования от 1 до 2%)

изображены с таким расчетом, что их расстояние от центра языка (наиболее часто встречающегося знака) служит мерой убывающей частоты.

Приведенные примеры относительно коротких последовательностей отнюдь не всегда являются представительными для классов последовательностей ДНК, РНК и белков. Выдающиеся исследования Гатлина (*Gatlin, 1972*) последовательностей ДНК и белка обнаруживают относительно большое рассеяние. Гатлин предположил, что биополимеры можно рассматривать как марковские источники первого порядка ($m = 1$), и вычислял энтропию этих источников по формуле

$$H = H_2 - H_1.$$

Соответственно, избыточность определяется выражением

$$R = \frac{\log \lambda - H}{\log \lambda}.$$

Для некоторых последовательностей ДНК Гатлин (*Gatlin, 1972*) вычислил также высокие значения шенноновской энтропии (рис. 11.7). Некоторые более новые результаты вычислений для ДНК вируса саркомы Ру представлены на рис. 11.8. Для этого был выбран фрагмент последовательности ДНК длиной $l = 9\,304$, причем особое внимание было уделено учету поправок на конечность цепи (*Ebeling, Feistel, Herzel, 1987*). Существуют указания о том, что наблюдаемое плато при $r > 5$ не соответствует истинному пределу и что при увеличении значений r оно сменяется спадом. В упоминавшихся выше работах сформулирована гипотеза, согласно которой энтропия биопоследовательностей и других текстов на естественных языках равна нулю. Для таких последовательностей характерно существование дальнедействующих корреляций. Вследствие этих корреляций информации, носителями которой служат

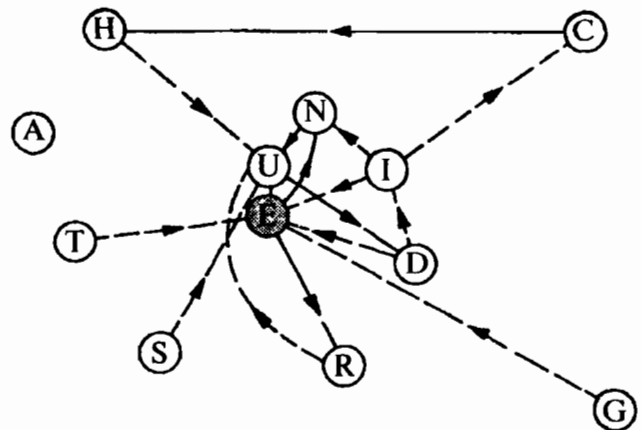


Рис. 11.6. Граф немецкого языка на основе приведенного в тексте отрывка из «Соляриса» Станислава Лема (сплошные стрелки соответствуют частоте последования от 2,5 до 3%, штриховые стрелки — частоте последования от 1,2 до 2,5%)

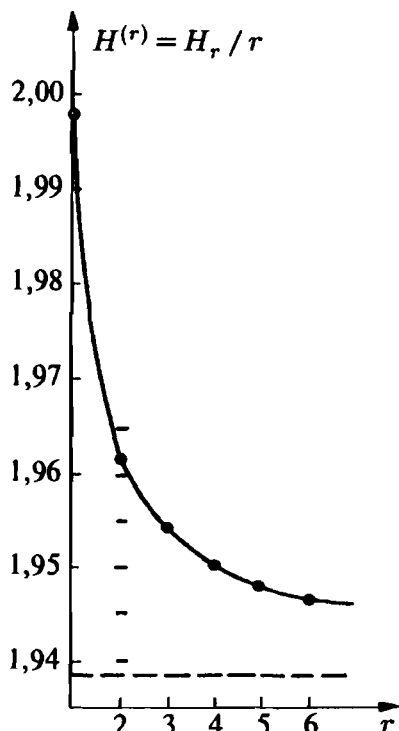


Рис. 11.7. Высшие энтропии Шеннона для ДНК из печени собаки по Гатлину (Gatlin, 1972)

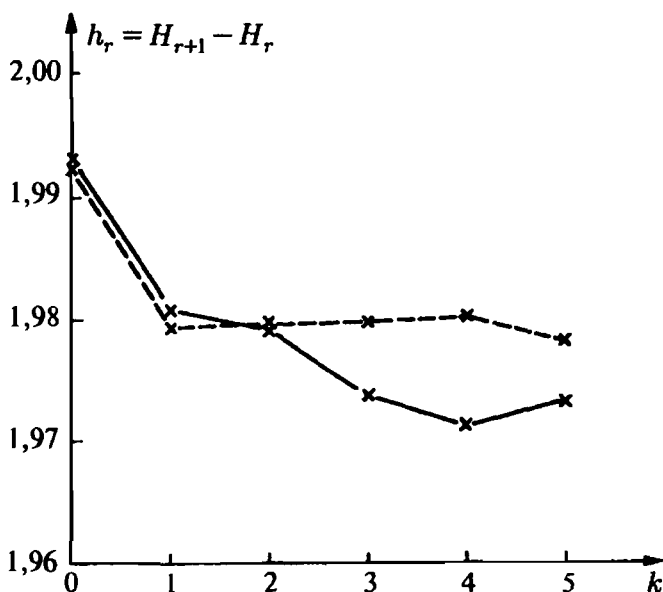


Рис. 11.8. Энтропия на букву для ДНК вируса саркомы Ру; для сравнения штриховой ломаной показаны результаты в случае цепи Маркова первого порядка (Ebeling, Feistel, Herzel, 1987)

новые буквы, становится все меньше. В пределе длинных последовательностей новая добавляемая буква в значительной мере «предопределена» правилами. Другой способ прочтения обсуждавшейся выше гипотезы состоит в том, что число «имеющих биологический смысл» последовательностей возрастает не экспоненциально

$$N^* \sim \exp \{ \nu H \}, \quad (11.33)$$

а по степенному закону

$$N^{**} \sim \nu^G = \exp \{ G \ln \nu \}. \quad (11.34)$$

Здесь G — константа, зависящая от «языка», но не от длины последовательностей. Мы будем называть G характеристическим индексом языка. Несмотря на имеющийся обширный эмпирический материал по биопоследовательностям, какие-либо надежные статистические утверждения пока отсутствуют.

Одна из основных трудностей при вычислении высших энтропий состоит в том, что число возможных частичных слов необычайно быстро возрастает с увеличением длины, поэтому хорошая статистика достижима только вместе с весьма обширным материалом по последовательностям. Как показано в одной более поздней работе, вычисление трансинформации или распределений расстояний Хэмминга значительно менее трудоемко (Ebeling, Feistel, Herzel, 1987). Понятие трансинформации определяется следующим образом:

$$I(k) = \sum_{i,j} p^k(A_i A_j) \ln \left[\frac{p^k(A_i A_j)}{p(A_i) p(A_j)} \right]. \quad (11.35)$$

Здесь $p^k(A_i A_j)$ — вероятность найти в последовательности буквы A_i и A_j на расстоянии k одну от другой. При $k \geq 3$ трансинформация $I(k)$ вычисляется значительно легче, чем $H^{(k)}$ или h_k . Например, при анализе ДНК для оценки рассматриваемой

последовательности необходимо знать только 16 вероятностей. На рис. 11.9 показан результат анализа ДНК вируса саркомы Ру. Пики распределения соответствуют корреляциям. Вплоть до максимума $k = 3$ все наблюдаемые максимумы кратны трем, что, возможно, является следствием принципа кодирования.

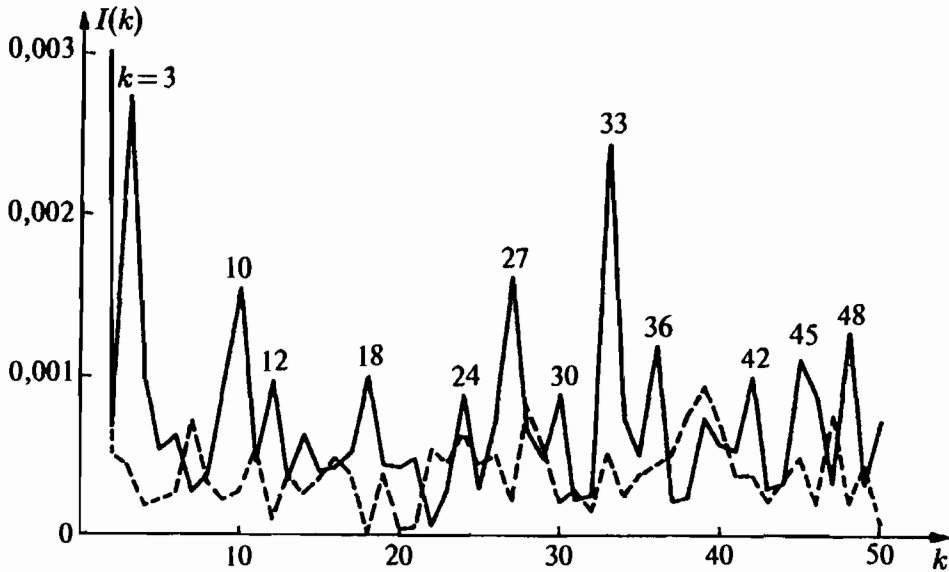


Рис. 11.9. Трансинформация для ДНК вируса саркомы Ру. Штриховая линия соответствует цепи Маркова первого порядка (Ebeling et al., 1987)

Для обеспечения структуры и функционирования живых организмов, как оказалось, необходимо огромное количество информации, хранимой наследственностью. Еще Шрёдингер, которому не была известна молекулярная основа наследственности, постулировал существование нерегулярной молекулярной структуры — «аперiodического кристалла» — как единственно мыслимой возможности для принципа кодирования. Для полинуклеотидов длины ν существуют 4^ν различных вариантов расположения элементарных звеньев, откуда для информационной емкости получаем:

$$I = \frac{\ln 4^\nu}{\ln 2} = 2\nu \text{ в битах (двоичных единицах),} \quad (11.36)$$

$$I = \nu \text{ в четверичных единицах.}$$

Для полипептида из 20 элементарных звеньев для информационной емкости получаем:

$$I = \frac{\ln 20^\nu}{\ln 2} = \nu \cdot 4,3219 \text{ в битах,} \quad (11.36)$$

$$I = \nu \text{ в двадцатиричных единицах.}$$

Эти соотношения носят общий характер, и при выводе их не используются никакие предположения об информационном содержании отдельного конкретного полинуклеотида или белка. Интуитивно ясно, что упорядоченная структура с определенными симметриями может переносить меньшее количество информации, чем аперiodическая структура. Например, последовательности

$$p_1 = \text{ААААЦЦЦЦ} \quad \text{или} \quad p_2 = \text{АЦАЦАЦАЦ}$$

содержат меньше информации, чем последовательности

$$p_3 = \text{АЦТГГАТЦ} \quad \text{или} \quad p_4 = \text{ГЦЦАГААЦГТ},$$

поскольку последовательности p_1 и p_2 могут быть построены по простому и краткому правилу, в то время как в структуре последовательностей p_3 и p_4 не обнаруживается никакой регулярности.

Информационная емкость, измеренная в некоторых выбранных единицах (обычно в битах), играет в каждом случае важную роль в качестве верхней границы подлежащего измерению информационного содержания носителя, каким бы оно ни было. Если носитель информации, обладающий информационной емкостью I , полностью и без избыточности загружен сообщениями, то сообщение может быть передано без потерь другому носителю информации (причем независимо от способа кодирования — с помощью частотной или амплитудной модуляции, цифрового или аналогового представления в виде функции или интегрального преобразования) только в том случае, если другой носитель обладает такой же информационной емкостью I . Так как хромосомы содержат около 10^9 нуклеотидов, можно исходить из того, что их емкость как запоминающего устройства согласно формуле (11.35) соответствует удвоенной длине и составляет около $2 \cdot 10^9$ битов. Представим себе наглядно величину емкости такой памяти. Для этого подсчитаем, сколько книг объемом в 500 страниц потребуется, чтобы вместить такое количество информации. Для каждого из 32 знаков нашего алфавита требуется $\ln 32 / \ln 2 = 5$ битов. Это означает, что с помощью последовательности ДНК можно закодировать около $4 \cdot 10^8$ букв или при средней длине слова в 6 букв — около $6 \cdot 10^7$ слов. Поскольку обычная печатная страница содержит около 300 слов, емкость последовательности ДНК как запоминающего устройства составляет примерно $2 \cdot 10^5$ печатных знаков, т. е. соответствует небольшой библиотеке в 400 томов. Такой же емкостью памяти обладали бы 100 магнитофонов, каждый из которых может хранить на магнитной ленте $512 \cdot 1024 \cdot 48$ битов.

Обсуждаемая нами сейчас мера информации — информационная емкость, или емкость запоминающего устройства, не зависит от формы хранения информации и, как показывают примеры, легко вычисляется для любого конкретного запоминающего устройства.

Попытаемся теперь придать более точный смысл интуитивному понятию информационного содержания данной конкретной структуры. Прежде всего ясно, что реальное информационное содержание $I(p)$ данного конкретного запоминающего устройства или последовательности p не может быть больше информационной емкости I , поскольку последняя является верхним пределом для $I(p)$:

$$\text{информационное содержание } I(p) \leq \text{информационная емкость } I. \quad (11.37)$$

Напрашивается предположение о том, что информационное содержание последовательности находится в тесной связи с информационной энтропией источника сообщений, определяемой как средняя неопределенность, устраняемая на один символ. Так как информационная энтропия определяется свойствами источника, а с другой стороны информационное содержание возрастает с увеличением длины последовательности, однако эти меры не тождественны. Итак, мы отчетливо видим, что информационное содержание является мерой информации, отличной от информационной энтропии. В то время как информационная емкость книги объемом в 500 страниц, согласно приведенным выше расчетам, составляет около $5 \cdot 10^6$ битов, реальное информационное содержание такой книги, как правило, существенно меньше. Например, если наша книга состоит только из повторений высказывания Фауста «В начале было дело», то ее информационное содержание, несомненно, не идет ни в какое сравнение с информационным содержанием научного труда того же объема. Вместе с тем вполне возможно, что какой-нибудь первоклашка извлечет из такой «Фаустовской книги» больше информации, чем от чтения

монографии по современной алгебре. Однако эта проблема уводит нас в область семантики, занимающейся изучением отношений между знаком и его значением, и мы не будем вдаваться в этот круг проблем более подробно. Анализ со стороны физики ограничивается элементарными, «наивными» аспектами проблем. Прежде всего для физика важно, если это возможно, рассматривать информационное содержание как измеримую величину, соответствующую некоторому действительному числу. Затем физика интересуют свойства такой информационной меры и ее связь с информационной энтропией источника. В духе физического мышления и ради сознательного упрощения проблемы в целом мы сосредоточим свои усилия на поиске величин, измеримых или вычислимых по результатам измерений (например, последовательностей молекул), которые коррелированы с величиной, интуитивно воспринимаемой как информационное содержание. Начнем с вопроса о том, может ли мера информационного содержания быть независимой от способа кодирования и считывания информации. Довольно очевидно, что невозможно наряду с информационной емкостью найти вторую, не зависящую от нее, инвариантную относительно способа кодирования меру информации.

При поиске подходящей меры мы следуем идее Колмогорова, согласно которой информационное содержание структуры убывает с возрастанием степени ее «регулярности» и увеличивается с возрастанием степени ее «нерегулярности». Проблема состоит в том, чтобы найти меру регулярности или случайности (апериодичности) последовательности. Математическое определение случайности, или апериодичности, последовательности символов было предложено независимо в работах академика А. Н. Колмогорова из Москвы и бывшего тогда студентом Нью-Йоркского университета Чейтина в 1965 г. Аналогичная идея была высказана еще в 1960 г. Соломоновым при попытке квантификации простоты научных теорий. Соломонов, бывший тогда сотрудником американской «Zator Company», рассматривал научные наблюдения, как серию двоичных знаков. Задача теории состоит в том, чтобы объяснять имеющиеся наблюдения и предсказывать новые. Соломонов определяет теорию как алгоритм, воспроизводящий серию наблюдений. Если существуют два таких алгоритма, то предпочтение отдается более простому. Поскольку данные наблюдений совершенно нерегулярны, нет никакой программы, которая была бы короче, чем ряд данных, и, следовательно, никакая теория сформулирована быть не может.

Рассмотрим в качестве примера следующие двоичные последовательности

$$p_5 = 1111111111111111,$$

$$p_6 = 0101010101010101,$$

$$p_7 = 0110110011011110.$$

Ясно, что каждая из последовательностей p_5 и p_6 представима с помощью простой короткой программы:

$$p_5 = 16 \text{ раз повторить } 1,$$

$$p_6 = 8 \text{ раз повторить } 01.$$

В последовательности p_7 никакой закономерности нет. По Чейтину и Колмогорову, последовательность случайна, если кратчайший алгоритм, позволяющий воспроизвести последовательность, требует для своей записи примерно такого же числа битов, какое необходимо для записи исходной последовательности. Иначе говоря, последовательность случайна, если она несжимаема, т. е. если не существует более короткого алгоритма, способного воспроизвести ее. Разумеется, любую заданную последовательность всегда можно произвести с помощью бесконечно большого числа алгоритмов. Кратчайший из них назовем минимальной программой. Каждая минимальная

программа с необходимостью случайна, как это следует из определения случайных последовательностей. Понятие сложности определяется следующим образом:

сложность последовательности = длина минимальной программы,
производящей эту последовательность, в битах.

Рассмотрим в качестве примера последовательности нулей и единиц длиной 16. Существует 2^{16} различных последовательностей такого типа, из которых три приведены выше (см. последовательности p_5 , p_6 и p_7). Последовательность p_5 может быть воспроизведена программой длиной около $\log_2 16$ битов. Для воспроизводства последовательности p_6 требуется программа длиной не менее 5 битов, а для последовательности p_7 — программа длиной не менее 15 или 16 битов. Подавляющее большинство рассматриваемых последовательностей из общего числа их 2^{16} обладают очень большой сложностью, близкой к 16 битам, и только немногие исключительные последовательности обладают малой сложностью (например, две последовательности, составляющие 2^{-15} от всех последовательностей, обладают сложностью, равной 4 битам).

Попытаемся уточнить введенное выше понятие сложности или случайности. В 1965 г. Колмогоров предложил для описания этих свойств новое понятие энтропии — так называемую алгоритмическую энтропию. Это новое понятие имеет большое значение для исследования последовательностей (слов), поскольку позволяет различать случайные последовательности от регулярных. Общая идея Колмогорова связывает сложность слова p с длиной кратчайшей программы q , позволяющей представить p в рамках некоторого алгоритма, языка и машины (автомата) F , т. е. $F(q) = p$. Последовательность называется случайной, если не существует более короткой программы, т. е. если $l(q) \geq l(p)$. С другой стороны, для регулярных последовательностей существует более короткое представление $F(q) = p$ с $l(q) < l(p)$, причем сокращение длины отражает существующие в последовательности закономерности («регулярности»). Правда, Колмогоров использовал вместо понятия программы или алгоритма понятие частично рекурсивных функций, исходя из предположения о том, что класс алгоритмически вычислимых функций эквивалентен классу частично рекурсивных функций. Понятие функции применимо к нашей проблеме, если мы исходим из существования взаимно однозначных отображений последовательностей на натуральные числа. Если $g(p)$ и $g(q)$ — гёделевские числа, соответствующие p и q , то представимость p по q с помощью функции F означает, что

$$g(p) = F(g(q)).$$

Иначе говоря, существует отображение гёделевского числа, соответствующего последовательности q , в гёделевское число, соответствующее последовательности p .

По Колмогорову, сложность двоичного слова $p \in X$ относительно алгоритма (или рекурсивной функции) $F(q)$, $q \in X$, определяется следующим образом (Звонкин, Левин, 1970):

$$K_F(p) := \min l(q): F(q) = p. \quad (11.38)$$

По теореме Колмогорова—Соломонова, существует оптимальный алгоритм F_0 в том смысле, что

$$K_{F_0}(p) \leq K_F(p) + C \quad \forall p \in X^*.$$

Сложность относительно этого оптимального алгоритма обозначается $K(p)$ и называется сложностью последовательности p . По теореме Колмогорова, $K(p)$ является верхней гранью сложности последовательности символов, задаваемой длиной последовательностей, т. е.

$$K(p) \leq [l(p) + C], \quad (11.39)$$

где C — не зависящая от p константа. Понятие алгоритмической энтропии является обобщением шенноновской энтропии. Шенноновскую энтропию последовательности мы определяем по формуле (11.22) как величину

$$H_1(p) = - \sum_{k=1}^{\lambda} h_k \log h_k, \quad (11.40)$$

где $h_k = \nu_k/\nu$ — относительная частота появления символа A_ν , ν_k — абсолютная частота символа A_ν в слове p . По Звонкину и Левину (Звонкин, Левин, 1970), для двоичной последовательности справедлива теорема

$$K(p) \leq [\nu H_1(p) + C_0 + C_1 \ln \nu], \quad (11.41)$$

где ν — длина последовательности, C_0 и C_1 — константы, не зависящие от p .

Исследования, о которых мы упоминали выше, естественно приводят к вопросу о том, как связаны между собой сложность и шенноновская энтропия. Связь между этими величинами устанавливает важная теорема Звонкина и Левина (Звонкин, Левин, 1970) и более поздняя работа (Leung-Yan-Cheong, Cover, 1978).

Теорема. Для каждого эргодического процесса с конечным алфавитом соотношение

$$\frac{K(p)}{l(p)} \rightarrow H(p) \quad (11.42)$$

выполняется в пределе при $l(p) \rightarrow \infty$ с вероятностью единица. Таким образом, сложность в смысле Колмогорова—Чейтина с вероятностью единица сходится к шенноновской энтропии источника сообщений.

Учитывая это, мы определяем алгоритмическую энтропию на символ (букву) следующим образом:

$$H_{\text{алг}}(p) = \frac{K(p)}{l(p)}. \quad (11.43)$$

Как нетрудно убедиться, величина

$$K_H(p) = l(p)H(p) \quad (11.44)$$

удовлетворяет аксиомам Тиле и поэтому является сложностью в смысле Тиле. Мы вычислили H -сложности для некоторых биологических последовательностей в предположении, что биополимеры в первом приближении можно рассматривать как марковские процессы первого порядка, т. е. что

$$H(p) = H_2(p) - H_1(p).$$

Избыточность вычислялась по формуле

$$R_H = 1 - \frac{K_H(p)}{K_H^{\max}(p)} = 1 - \frac{H(p)}{H^{\max}(p)}. \quad (11.45)$$

Мы видим, что H -избыточность совпадает с шенноновской избыточностью с точностью до выбора $H^{\max}(p)$. При вычислении шенноновской избыточности, вообще говоря, вводят предположение о том, что $H^{\max}(p) = \log \lambda$ (Gatlin, 1972).

Более подробный анализ показывает, что формула $H^{\max}(p) = \log \lambda$ неприменима к относительно коротким последовательностям. В частности, это следует иметь в виду при вычислении избыточности белков. Рассмотрим в качестве примера

последовательности цитохрома с длиной 104. В такой последовательности могут реализоваться самое большее 103 различные пары, т. е.

$$H_2^{\max} = \log [l(p) - 1] = \log 103 = 1,547.$$

Чтобы построить 103 различные пары, нам необходимо самое большее 11 различных символов, из которых 5 символов должны встречаться 10 раз и 6 символов 9 раз, т. е.

$$H_1^{\min} = \log 104 - \frac{1}{104} [50 \log 10 + 54 \log 9] = 0,800.$$

Окончательно получаем:

$$H^{\max} = H_2^{\max} - H_1^{\min} = 0,747 \quad (20 \text{ знаков}).$$

Например, для цитохрома с человека вычисленное нами значение $R_H = 0,267$ оказывается значительно ниже значения $R = 0,447$, полученного Гатлином. Другие исправленные значения для различных белков приведены в табл. 11.6. У последовательностей ДНК этот эффект играет не столь заметную роль, поскольку для них $l(p) \gg 16$ (Ebeling, Feistel, Herzel, 1987).

Таблица 11.6. Сложности, избыточности и длины программ, вычисленные на основе порождающих грамматик (K_G, R_G, L_G), анализа частичных слов (K_T, R_T) и шенноновских энтропий (K_H, R_H) (Ebeling, Jiménez-Montaño, 1980).

Сорт	K_G	K_T	K_H	R_G	R_T	R_H	l	L_G
ДНК <i>E. coli</i> , ген тирозина тРНК	85	120,37	119,70	0,105	0,0298	0,045	126	37
ДНК вируса X 174, ген I	78	—	99,91	0,125	—	0,100	111	32
ДНК человека, 55 г KB карцинома	84	—	112,56	0,086	—	0,062	120	46
Цитохром с человека	99	102	57,41	0,048	0,0038	0,267	104	8
Цитохром с макаки-резус	99	102	57,41	0,048	0,0038	0,267	104	8
Цитохром с свиньи, овцы	99	101,96	56,99	0,048	0,0041	0,266	104	8
Цитохром с собаки	99	101,92	56,99	0,048	0,0045	0,266	104	8
Цитохром с пекинской утки	100	102,02	57,10	0,038	0,0036	0,265	104	7
Цитохром с курицы, индейки	101	102,08	57,41	0,029	0,0030	0,261	104	7
Цитохром с лягушки-быка	102	102,08	56,78	0,019	0,0030	0,269	104	5
Цитохром с тунца	102	—	—	0,010	—	—	103	3
Цитохром с пекарских дрожжей	105	—	—	0,028	—	—	108	5
Эволюционный узел 1	99	—	—	0,048	—	—	104	8
Эволюционный узел 2	99	—	—	0,048	—	—	104	8
Эволюционный узел 5	100	—	—	0,038	—	—	104	7
Эволюционный узел 9	101	—	—	0,029	—	—	104	5
Эволюционный узел 10	102	—	—	0,010	—	—	103	3
Эволюционный узел 11	104	—	—	0,009	—	—	105	3
Эволюционный узел 13	105	—	—	0,000	—	—	105	0
Эволюционный узел 14	105	—	—	0,009	—	—	106	3
Эволюционный узел 15	105	—	—	0,009	—	—	106	3

Из определения K_H следует, что хотя K_H можно считать мерой сложности в смысле Тиле (Thiele, 1974), поскольку эта величина удовлетворяет аксиомам 1–6, однако она не является мерой сложности в смысле Колмогорова—Чейтина, т. е. длиной минимальной программы. Сложность Колмогорова—Чейтина будет введена нами в следующем разделе и применена к анализу биопоследовательностей.

11.5. Генетический язык и грамматика биопоследовательностей

Генетическое сообщение, записанное в виде последовательности ДНК и РНК и выражаемое в функциональной плоскости как белковая последовательность, — основной объект исследования молекулярной биологии. Наряду с методами теории информации высокую оценку микробиологов получили и другие эффективные методы исследования биопоследовательностей, позволяющие делать количественные и — потенциально — качественные выводы, в последнее время — методы теоретического языкознания (*Иваницкий и др.*, 1978; *Ebeling, Jiménez-Montaño*, 1979, 1980; *Jiménez-Montaño, Zamora-Cortino*, 1981). Ныне не подлежит сколько-нибудь обоснованному сомнению, что генетический язык действительно существует, хотя его особенности и грамматика известны далеко не полностью. Согласно Ратнеру (*Ratner*, 1974), мы называем генетическим языком множество всех кодовых состояний генетической памяти. Генетический язык существует в двух формах — языка нуклеиновых кислот (языка НК) и языка полипептидов (языка ПП). Язык аминокислот состоит из «осмысленных» линейных последовательностей над алфавитом А, Ц, Г, Т (У), а язык полипептидов — из «осмысленных» линейных последовательностей над алфавитом из 20 аминокислот. Кроме того, существуют знаки препинания и грамматические правила, а также иерархия плоскостей генетических языков. Следуя работам Ратнера (*Ratner*, 1974; *Ратнер*, 1983), мы изложим сначала обе формы генетических языков на чисто качественном уровне.

Язык НК. Все наделенные смыслом конструкции этого языка, а также все знаки препинания состоят из последовательностей букв А, Ц, Г и Т (или У), означающих различные *нуклеотиды*. В силу отношения дополнительности А—Т, А—У, Ц—Г для каждой конструкции языка НК существует дополнительная конструкция.

Если нуклеотиды, представленные буквами, или символами, образуют самый нижний уровень языка НК, то следующий его уровень образуют *кодоны*. Под кодоном мы понимаем триплет из 3 нуклеотидов, ориентированных в направлении 5' — 3'. Известно 61 кодонов, которые кодируют 20 аминокислот. Триплеты УАА, УАГ и УГА указывают концы конструкций. Отдельные кодоны не разделяются знаками препинания; их смысл устанавливается в рамках перевода. Начала конструкций помечены кодонами АУГ и ГУГ. В результате возникает некоторая неоднозначность, поскольку эти же кодоны могут обозначать аминокислоты *met* и *val*. Однако в процессе перевода из-за некоторых особенностей указанных аминокислот эта неоднозначность не возникает. Тем самым в сумме возникает однозначное отображение 64 возможных кодонов на 20 аминокислот, операции «старт», и «стоп». Однозначный обратный перевод, естественно, невозможен, так как вместо одной аминокислоты могут стоять два, три или шесть кодонов. По оценкам Йоки (*Yockey*, 1978), существует около 10^{52} последовательностей ДНК, соответствующих последовательности аминокислот цитохрома с дрожжей *Candida*. Столь высокое вырождение генетического кода, естественно, приводит к невозможности перевода с языка ПП на язык НК; это обстоятельство служит обоснованием так называемой центральной догмы молекулярной биологии, согласно которой информационный поток всегда (за исключением отдельных весьма специфических случаев) направлен от нуклеиновых кислот к белкам.

Следующий уровень в языке НК образуют *цистроны*. Цистрон представляет собой ориентированную последовательность кодонов. Каждый цистрон начинается с открывающего (начального) символа АУГ или ГУГ и заканчивается завершающим символом УАА, УАГ или УГА. Цистроны служат единицами перевода; они соответствуют фрагментам мРНК и полипептидов.

Следующий уровень образуют *скриптоны*, служащие единицами транскрипции. Скриптон всегда начинается с промотора и завершается терминатором транскрип-

ции. Скриптоны, как правило, состоят из нескольких (до десяти) цистронов, знаков препинания и неактивных элементов.

Следующий уровень образуют *репликоны*. Репликон — единица репликации. Каждый репликон представляет собой последовательность нуклеотидов, ограниченную старт- и стоп-символом, или промотором и терминатором, репликации. Как правило, репликоны состоят из нескольких (до 1000) скриптонов, а также знаков препинания и других функциональных единиц.

У организмов, размножающихся половым путем, существует еще уровень *сегрегонов*. Сегрегон представляет собой нить нуклеотидов, служащих единицами при редукционном делении — так называемом мейозе.

Наконец, высший уровень языка НК образуют *геномы*. Геном представляет собой совокупность наследственной информации клетки, хранящейся в последовательностях нуклеотидов.

Язык ПП. Все осмысленные конструкции этого языка построены над алфавитом из 20 аминокислот. В языке ПП нет знаков препинания, поскольку не существует информационного процесса, в котором материальные носители языка ПП служили бы матрицами. Особенность языка ПП состоит в том, что его материальные носители не только кодируют сообщение, но и доставляют его «адресату». Значение осмысленных конструкций языка ПП задается ПП-кодом. Под ПП-кодом мы понимаем набор правил, выражающих взаимосвязь между последовательностью аминокислот и функцией соответствующего белка. Правила эти, если не считать редких исключений, пока не известны. Однако не приходится сомневаться в том, что соответствие между последовательностью аминокислот и биологической функцией белка действительно существует. В целом о языке ПП известно гораздо меньше, чем о языке НК.

Фундаментальные математические основы формальной лингвистики были заложены в работах Хомского (Хомский, 1962; Chomsky, 1966, 1972). Хомский исходит из понятия грамматики. Под грамматикой он понимает систему правил, позволяющих породить из данного алфавита некоторое множество предложений. Хомский устанавливает иерархию грамматик. Простейший класс охватывает регулярные грамматики, в свою очередь порождающие регулярные языки. В основе регулярных грамматик лежат правила, порождающие предложения в результате последовательности принятия решений о выборе элементов алфавита по порядку слева направо. Иначе говоря, сначала мы выбираем первый, самый левый элемент (символ), а каждый последующий элемент определяется характером предшествующего элемента. Простым примером служит правила: за А следует Ц, за Ц следует Г, за Г следует Т, за Т следует А. Они порождают предложения

ГТАЦГТАЦГТАЦГТАЦ

Регулярные языки могут быть определены с помощью конечных автоматов (Gössel, 1972). Хомский показал, что некоторые языки нерегулярны, например, предложения с зеркальной симметрией типа

АГГА, АТЦТТЦТА

не могут быть порождены регулярными грамматиками. На этом основании Хомский ввел более обширные классы структурных грамматик, при этом различая контекстносвободные и контекстносенситивные грамматики.

Уточним введенные выше понятия с помощью следующих определений (Lindemayer, 1971).

Определение 1. Алфавитом X называется конечное множество символов: $\{A_1, A_2, \dots, A_\lambda\}$. В случае ДНК $X = \{A, Ц, Г, Т\}$, в случае РНК $X = \{A, Ц, Г, У\}$ и в случае

белка $X = \{A, C, D, \dots, Y, W\}$ ($\lambda = 20$). Множество всех конечных последовательностей (слов), которые можно построить из алфавита, обозначается X^* (свободная полугруппа слов над X). Пустая последовательность ϵ также считается принадлежащей X^* . Множество последовательностей без ϵ обозначается X^+ .

Определение 2. Языком над алфавитом X называется подмножество свободной полугруппы X^* .

Определение 3. Если p и q из X^* , то их композиция pq (конкатенация) есть снова слово из X^* .

Определение 4. Структурной грамматикой называется упорядочная четверка $G = \{N, T, P, S\}$, где

- 1) N — конечное множество переменных (вспомогательный алфавит), называемых нетерминальными; к их числу принадлежит старт-символ S ;
- 2) T — конечное множество, $T \subseteq X$, элементов, называемых терминальными;
- 3) P — конечное множество порождающих правил $p \rightarrow q$.

Определение 5. Контекстносенситивной грамматикой называется система правил вида $p\sigma q \rightarrow p\sigma'q$, $p, q \in T^*$, $\sigma \in N$, $\sigma' \in (N \cup T)^+$.

Контекстносвободной грамматикой называется система правил вида $\sigma \rightarrow p$, $\sigma \in N$ и $p \in (N \cup T)^*$.

Регулярной грамматикой называется система правил вида $\sigma \rightarrow x$ или $\sigma \rightarrow xq$, $\sigma, x \in N$ и $q \in T$ или $q = \epsilon$.

Определение 6. Языком, порожденным грамматикой G , называется множество всех последовательностей, состоящих только из терминальных символов и порожденных из старт-символа в результате поочередного применения конечного числа порождающих правил. В зависимости от характера порождающей грамматики язык называется регулярным, контекстносвободным или контекстносенситивным.

С остальными понятиями и результатами математической лингвистики читатель может познакомиться в работе Хомского (Хомский, 1962). Математическая лингвистика нашла ряд весьма интересных приложений в биологии.

Но вернемся к проблеме сложности. Среди различных реализаций функции сложности, рассмотренных Тиле (Thiele, 1974), мы находим сложность (на основе порождающих грамматик) контекстносвободных грамматик в смысле Хомского (Хомский, 1962). Исходя из решеточной модели, развитой Бёмом и Кликсом для психологических исследований по распознаванию и переработки образов, Гейсслер и Шейдереитер предложили меру сложности слов над некоторым алфавитом. Тиле (Thiele, 1974) и Шейдереитер (Scheidereiter, 1974) развили и усовершенствовали понятие «грамматической сложности». Для этого они воспользовались так называемой праволинейной порождающей грамматикой, задаваемой следующей упорядоченной четверкой:

$$G = \{N, X, P, S\}.$$

Здесь X — алфавит, играющий роль терминальных символов грамматики, N — множество нетерминальных символов, P — множество правил и S — старт-символ ($S \in N$). Правила считаются контекстносвободными, причем допускаются только правила вида

$$\sigma \rightarrow q, \quad \text{где } \sigma \in N, \quad q \in (X \cup N)^*.$$

Речь идет о контекстносвободной грамматике в смысле математической лингвистики. По Шейдереитеру (Scheidereiter, 1974), в качестве сложности правила, по определению, принимается длина слова, стоящего в правой части:

$$K(\sigma \rightarrow q) = l(q).$$

В качестве сложности слова с учетом специального представления с помощью множества вспомогательных символов N принимается сумма сложностей всех правил:

$$K_N(p) = \sum_{x \in N} K(\sigma \rightarrow q). \quad (11.46)$$

Наконец, грамматическая сложность определяется как минимум сложности по всем представлениям:

$$K_G(p) = \min_N K_N(p). \quad (11.47)$$

Работы Шейдерейтера (*Sheidereiter*, 1974) и Тиле (*Thiele*, 1974) не содержат каких-либо утверждений относительно сложности слов, состоящих из ν -кратного повторения одного-единственного символа. Чейтин (*Chaitin*, 1975) предложил приписывать слову, состоящему из ν одинаковых символов двоичного алфавита, например, 000...0 (ν раз), сложность $\log_2 \nu$. Поэтому мы, по определению, полагаем (*Ebeling, Jiménez-Montaño*, 1980)

$$K(\sigma \rightarrow A_1 A_2 \dots A_\nu) = (1 + \text{Mod } \log_2 \nu) \quad (11.48)$$

при

$$A_1 = A_2 = \dots = A_\nu.$$

При этом речь идет об уточнении 4-й аксиомы Тиле.

С помощью грамматической сложности мы определяем грамматическую избыточность по формуле

$$R_G = 1 - \frac{K_G(p)}{K_G^{\max}(p)}, \quad K_G^{\max}(p) \leq l(p). \quad (11.49)$$

В качестве еще одной величины, характеризующей слово, мы определяем длину правил $L_G(p)$ как число правил, использованных при построении слова. Каждое правило $\sigma \rightarrow q$, за исключением стартового правила, подсчитывается столько раз, сколько оно используется в минимальной программе. Введение длины правил, называемой также длиной вывода или сложностью вывода, восходит к работам Гладкого и Игараши.

Мы вычислим значение введенных выше величин для некоторых полимеров (*Ebeling, Jiménez-Montaño*, 1980); результаты представлены в табл. 11.6. Ход вычислений мы показываем на трех примерах. Для рассмотренного выше гена ДНК *E. coli* мы обнаружили в качестве короткой производящей программы следующую производящую грамматику:

$$\begin{aligned} S &\rightarrow 333\gamma G\alpha\gamma A3\zeta 3G6A0AC\gamma CAGA4G4AG715\zeta BC \\ &2\zeta\zeta A\zeta AC\gamma GT\alpha 0CT4C5\zeta 1A2GC \\ (K_1 &= 55) \end{aligned}$$

$$\begin{array}{llll} \alpha \rightarrow AA & (K = 2), & 3 \rightarrow T6 & (K_3 = 3), \\ \zeta \rightarrow CC & (K = 2), & 4 \rightarrow TTTA & (K_4 = 4), \\ \gamma \rightarrow GG & (K = 2), & 5 \rightarrow 0G & (K_5 = 2), \\ 0 \rightarrow TG & (K = 2), & 6 \rightarrow TG & (K_6 = 2), \\ 1 \rightarrow 0C7\zeta TT & (K_1 = 6), & 7 \rightarrow TC & (K_7 = 2), \\ 2 \rightarrow 6\gamma\alpha & (K_2 = 3). & & \end{array} \quad (11.50)$$

Для простоты мы используем в качестве вспомогательного алфавита греческий алфавит и однозначные натуральные числа. Если произвести все замены в соответствии

с приведенными 12 правилами, то возникает первоначальное молекулярное слово — последовательность ДНК гена *E. coli*. Таким образом, перед нами действительно производящая грамматика для слова ДНК. Сложность этой грамматики, по определению, равна сумме длин всех правил, т. е.

$$K_G \leq 55 + 2 + 2 + 2 + 2 + 6 + 3 + 3 + 4 + 2 + 2 + 2 = 85.$$

Для вычисления избыточности мы используем максимальное значение сложности $K_G^{\max}$. При этом мы рассуждаем следующим образом. Каждая произвольная последовательность длины 126 над четырехбуквенным алфавитом представима в виде последовательности 63 пар. Но поскольку существует только 16 различных пар, непременно должны быть повторения, и возникает необходимость воспользоваться вспомогательным алфавитом из 16 символов. Для такого вспомогательного алфавита мы получили бы

$$K_N(p) = 63 + 2 \cdot 16 = 95.$$

Последовательности с большей сложностью невозможны, и поэтому мы с полным основанием интерпретируем значение 95 как максимальную сложность. В общем случае последовательности над алфавитом из λ символов длиной $l(p) > 4\lambda^2$ обладают универсальным представлением в виде последовательности пар, обладающим сложностью

$$K_N(p) = \frac{1}{2} [l(p) + 1] + 2\lambda^2. \quad (11.51)$$

Это значение, справедливое для произвольных последовательностей, мы рассматриваем как максимальную сложность. В случае очень длинных последовательностей было бы естественно прежде всего доказать, не приводит ли универсальное представление в виде последовательности троек, четверок и т. д. символов к еще меньшему значению. Для рассмотренной последовательности ДНК мы получили избыточность

$$R_G = 1 - \frac{85}{95} = 0,105.$$

Наконец, мы приходим к длине правил

$$L_G = 2 + 2 + 2 + 2 + 5 + 3 + 3 + 3 + 4 + 4 + 4 = 37.$$

Для рассмотренной ранее последовательности РНК 55r KB карциномы человека существует наиболее короткое представление

$S \rightarrow 92\beta G 5\alpha 334A \beta 75C 89286A 7\gamma A 7A 19C 15$

$UG 9\gamma 9A 2U 6UG 6G \beta C 746\alpha 319G 4\gamma 12U$

$1 \rightarrow GG$	$(K_1 = 2),$	$7 \rightarrow GC$	$(K_7 = 2),$
$2 \rightarrow CU$	$(K_2 = 2),$	$8 \rightarrow G\alpha 2C$	$(K_8 = 4),$
$3 \rightarrow \beta C$	$(K_3 = 2),$	$9 \rightarrow GU$	$(K_9 = 2),$
$4 \rightarrow 2G$	$(K_4 = 2),$	$\alpha \rightarrow AU$	$(K_{10} = 2),$
$5 \rightarrow 7G$	$(K_5 = 2),$	$\beta \rightarrow AC$	$(K_{11} = 2),$
$6 \rightarrow 1A$	$(K_6 = 2),$	$\gamma \rightarrow UA$	$(K_{12} = 2).$

Характеристические величины этой последовательности имеют следующие значения:

$$K_G = 84, \quad K_G^{\max} = 92, \quad R_G = 0,086, \quad L_G = 46.$$

Наконец, для последовательности цитохрома с человека в качестве минимальной программы мы находим производящую грамматику

$$S \rightarrow \text{GD13IFIMKCSQCHT1GKH2PNLHGLFGR2QAPGYSYTAANKNK} \\ \text{GIIWGEDTLMEYLENP3YIPGTKMIFVGI3KEERADLIAYL3ATNE} \\ 1 \rightarrow \text{VEKG}, \quad 2 \rightarrow \text{KTG}; \quad 3 \rightarrow \text{KK}.$$

Характеристические величины этой грамматики имеют следующие значения

$$K_G = 99, \quad K_G^{\max} = 104, \quad R_G = 0,048, \quad L_G = 9.$$

В свою очередь эти примеры показывают характерное различие между избыточностями полинуклеотидов и полипептидов. Последние упорядочены лишь весьма слабо, и порождающая их грамматика весьма скудна правилами. В качестве еще одной иллюстрации этой зависимости рассмотрим неупорядоченный белок, в котором каждая аминокислота встречается только один раз:

$$S \rightarrow \text{FLIMVSP TAYHQNKDECWRG}.$$

Так как эта последовательность не допускает дальнейшего сжатия, справедливы следующие равенства (5-я аксиома Тиле):

$$K_G(p) = 20, \quad R_G = 0, \quad L_G = 0.$$

Из многочисленных последовательностей РНК длиной 60, кодирующих этот белок, мы выбираем одну особую последовательность. Соответствующая ей минимальная последовательность имеет вид

$$S \rightarrow \text{U661UAU544ACC3AGCA3CA4AGAACAAA12UG65C25} \\ 1 \rightarrow \text{GAU}, \quad 4 \rightarrow \text{UC}, \\ 2 \rightarrow \text{GAG}, \quad 5 \rightarrow \text{GG}, \\ 3 \rightarrow \text{UAC}, \quad 5 \rightarrow \text{UU}.$$

Ее производящая грамматика обладает характеристическими значениями:

$$K_G = 54, \quad R_G = 0,1, \quad L_G = 15.$$

Минимальная программа производящей программы биополимеров и полученные из нее характеристические значения величин K_G , R_G и L_G несомненно дают важную информацию об их структуре. Для изучения эволюции этих величин в табл. 11.6. представлены и соответствующие числа для производящей грамматики эволюционных узлов по Дэйхоффу (*Dayhoff*, 1969). Напомним, что узлы Дэйхоффа последовательностей цитохрома с соответствуют точкам ветвления реконструированного филогенетического дерева эволюции. Например, узел 1 соответствует предкам приматов, узел 2 — предкам млекопитающих и узел 10 — предкам позвоночных (см. рис. 11.10). Филогенетическое дерево эволюции позволяет установить некоторые эволюционные пути. Приведем соответствующие длины правил, так как эти величины обнаруживают заметный рост в ходе эволюции:

- а) узел 13 ($L_G = 0$) → узел 11 ($L_G = 3$) → узел 10 ($L_G = 3$) → узел 9 ($L_G = 5$) → узел 5 ($L_G = 7$) → узел 2 ($L_G = 8$) → узел 1 ($L_G = 8$) → человек ($L_G = 8$);
- б) узел 13 ($L_G = 0$) → узел 10 ($L_G = 3$) → узел 9 ($L_G = 5$) → узел 6 ($L_G = 7$) → курица ($L_G = 7$);
- в) узел 13 ($L_G = 0$) → узел 11 ($L_G = 3$) → узел 10 ($L_G = 3$) → тунец ($L_G = 3$);
- г) узел 13 ($L_G = 0$) → узел 14 ($L_G = 3$) → узел 15 ($L_G = 3$) → пекарские дрожжи ($L_G = 6$).

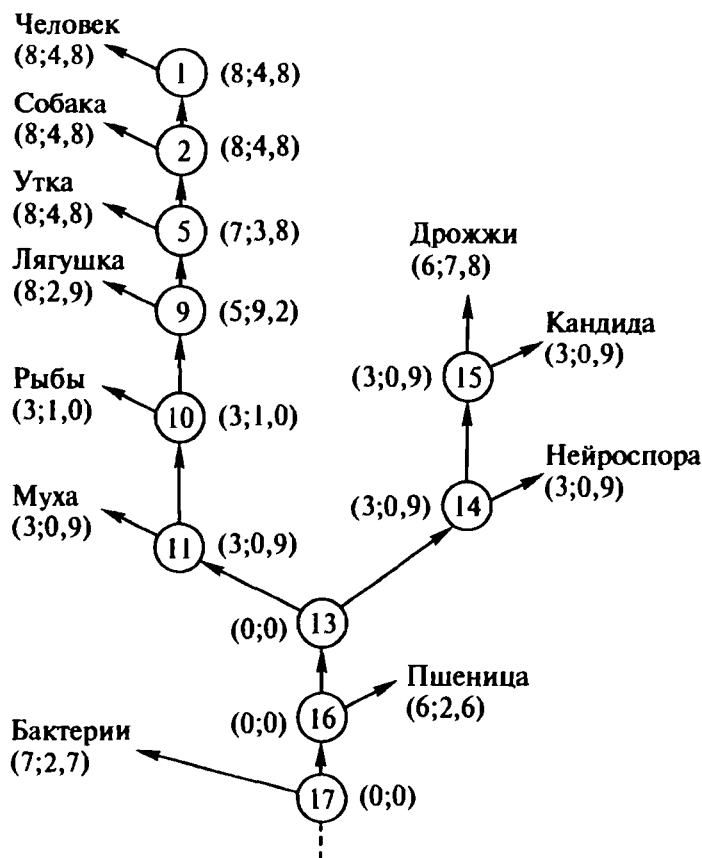


Рис. 11.10. Растущая длина правил и убывающая сложность ветвей дерева эволюции цитохрома *c*. Числа в скобках означают L_G и R_G в %

По мере того как прогрессирует эволюция, сложность медленно убывает. Это обстоятельство мы интерпретируем как взаимодействие двух противоположных тенденций в ходе эволюции. С одной стороны мы имеем тенденцию к максимализации хранимой информации и, соответственно, максимализации сложности. Эта тенденция, по-видимому, должна была доминировать на первой стадии эволюции, т. е. до образования узла 13, соответствующего высокоразвитым клеточноподобным предшественникам синих водорослей с одной стороны и бактерий — с другой. На более поздних фазах эволюции биополимеры образуют грамматическую структуру, соответствующую биологическим осмысленным последовательностям. Образование грамматических структур неизбежно связано с понижением сложности.

Рассмотренные нами контекстносвободные грамматики отражают лишь часть грамматических структур, например, дальний порядок белков, наблюдаемый у цитохрома *c* в фиксированных позициях для всех сравнимых последовательностях, передан неполно. Для описания белковых структур необходимо вводить контекстносенситивные грамматики (Ebeling, Jiménez-Montaño, 1980). Если рассмотреть дерево эволюции цитохрома *c* по Дэйхоффу (Dayhoff, 1969), то выяснится, что на очень ранней стадии эволюции, т. е. еще до образования узла 13, возникает следующая грамматическая структура цитохрома *c*:

$$S \rightarrow \chi_n G \chi_4 G \chi_7 C \chi_2 C H \chi_8 K \chi_1 G P \chi_1 L \chi_1 G \chi_{13} Y \chi_{10} W \chi_7 Y \chi_2 N P K \chi_6 K M \chi_1 F \chi_4 K \chi_m; \quad (11.52)$$

$$\begin{aligned} \chi_k &\rightarrow q, \quad q \in A_1 A_2 \dots A_k, \quad l(q) = k; \\ A_i &\in X; \quad n \leq 8, \quad m \leq 17. \end{aligned}$$

Иначе говоря, белок-цитохром состоит из фиксированной решетки, образованной 20 аминокислотами, которые занимают определенные и неизменные положения, а между этими кислотами вставлено множество последовательностей фиксированной длины, но переменного состава. Описанная нами структура напоминает аperiодический кристалл, о котором говорил Шрёдингер. У всех белков-цитохромов с в указанных 20 положениях находятся одни и те же аминокислоты. Ясно, что такая структура необходима для выполнения функции цитохрома с. Тем самым завершается важнейшая стадия эволюции цитохрома с, а именно обеспечение его функции в обмене веществ и особенно в цепи дыхания. В частности, тем самым фиксируется существенная для функции цитохрома третичная структура белка. К сожалению, мы не располагаем никакими данными об этой самой ранней стадии эволюции, завершившейся образованием узла 13, — данными, которые можно было бы учесть при анализе. Гипотеза о возрастании сложности не получает поэтому подтверждения. Стадии эволюции после образования узла 13 связаны с относительно несущественными свободными позициями белка. При такой тонкой оптимизации определенную роль играет тенденция к стандартизации, т. е. к образованию «строительных белков, или модулей», особенно часто используемых при построении последовательности. Аналогичные тенденции наблюдаются и в технике. Когда инженеры создают новую машину, например, разрабатывают новую конструкцию автомобиля, они направляют свои усилия прежде всего на обеспечение главной функции, или основного предназначения, машины, и лишь позднее обращают больше внимания на стандартизацию и упрощение технологии. Но вернемся к эволюции цитохрома. Выше эволюционного узла 11, представляющего предков всех животных (см. рис. 11.10), бывший до того свободным выбор промежуточных слов χ_k в грамматической структуре (11.52) становится сильно ограниченным. Эти условия можно выразить с помощью следующей контекстносенситивной грамматики:

$K\chi_1 G \rightarrow KA_i G;$	$A_i \rightarrow \{T, V\};$
$P\chi_1 L \rightarrow PNL;$	
$L\chi_1 G \rightarrow LA_i G;$	$A_i \rightarrow \{H, N, Y, W\}$
$M\chi_1 F \rightarrow MA_i F;$	$A_i \rightarrow \{I, V\};$
$C\chi_2 C \rightarrow CqC;$	$q \rightarrow \{SQ, AQ\};$
$Y\chi_2 N \rightarrow YLEN;$	
$G\chi_4 G \rightarrow GqG;$	$q \rightarrow \{DVEK, DIEK, DVAK, NAEN\}$
$F\chi_4 K \rightarrow FqK;$	$q \rightarrow \{VGIK, AGIK, AGLK\};$
$K\chi_6 K \rightarrow KKYIPGTK;$	
$G\chi_7 K \rightarrow GKKqC;$	$q \rightarrow \{IFIMK, IFVQK, TFVQK, TFVQR\};$
$W\chi_7 Y \rightarrow WtTLuE;$	$t \rightarrow \{GED, KEE, GEE, NND, QDD, GDD\};$
	$u \rightarrow \{M, F\};$
$H\chi_8 K \rightarrow HTvGGKH;$	$v \rightarrow \{VEK, CEK, VEN, VEA\};$
$Y\chi_{10} K \rightarrow YwANKxW;$	$w \rightarrow \{TA, TD, TE, TN, SN\};$
	$x \rightarrow \{NKGII, NKGIT, SKGIV, AKGIT\}$
$G\chi_{13} Y \rightarrow GaGRKTGQAbY;$	$a \rightarrow \{LF, IF, LI, LY\};$
	$b \rightarrow \{PGYS, PGFT, PGFS, VGFS, EGTS, AGFS, EGYS, AGFA\};$

$$e\chi_n G \rightarrow eqG;$$

$$\chi_m e \rightarrow cERdLfAYLKge;$$

$$q \rightarrow \{e, PA, GVPA\};$$

$$c \rightarrow \{KE, KTKG, TG, KD, KS, KA, PN, AN\};$$

$$d \rightarrow \{AD, ED, VD, QD, GG\};$$

$$f \rightarrow \{I, V\};$$

$$g \rightarrow \{KATNE, KATKE, DATSK, DATAK, SACKS, SATS, SATK, ESTK\}.$$

Эти грамматические правила приводят к резкому ограничению числа возможных вариантов. До того как в протоокеане возникла грамматика белков, число возможных вариантов образования белка длиной 104 достигало 2^{104} . После возникновения грамматики, соответствующей узлу 13, число возможных вариантов сократилось до

$$20^{(8+4+7+2+8+1+1+1+13+10+7+2+6+1+4+17)} = 20^{91}.$$

По миновании эволюцией узла 11 большинство их этих вариантов оказываются запрещенными новыми правилами, и мы находим только

$$2 \cdot 4 \cdot 2 \cdot 2 \cdot 4 \cdot 4 \cdot 3 \cdot 4 \cdot 6 \cdot 2 \cdot 4 \cdot 5 \cdot 5 \cdot 4 \cdot 8 \cdot 9 \cdot 4 \cdot 2 \cdot 7 \cdot 3 \approx 20^9$$

вариантов. Таким образом, эволюцию можно рассматривать как процесс введения новых правил в грамматику белков, являющихся носителями функций живой системы. Эта тенденция эволюции действует против тенденции к максимизации сложности. Эволюция разыгрывается на почве, допускающей компромисс между двумя противоположными тенденциями.

Подведем итоги сказанному в этом разделе. Прежде всего, следуя работам Ратнера (*Ратнер*, 1974), мы показали, что генетически запрограммированные нуклеиновые кислоты и белки образуют два языка — язык НК и язык ПП. Были рассмотрены алфавит, иерархическая структура и некоторые элементы грамматики и сопоставлены с формализмом математической лингвистики. В качестве возможного приложения лингвистических понятий была предложена новая мера сложности и избыточности биополимеров, опирающаяся с одной стороны на контекстносвободные грамматики, а с другой — на алгоритмическую сложность Колмогорова—Чейтина. На основе математических работ Тиле и Шнейдерейтера по грамматической сложности и Гладкова и Игараши по длине правил эти величины были предложены в качестве количественных мер информации, хранящейся в биопоследовательностях, и грамматики, с соблюдением которой «записано» генетическое сообщение. Обе меры были вычислены для различных последовательностей ДНК, РНК и белков, а результаты сравнены с мерами сложности, введенными в разд. 11.4. Сравнение показало определенные преимущества грамматических величин. Например, в грамматической мере избыточности белков, если они лежат ниже 5%, оказываются меньше, чем избыточности нуклеиновых кислот, лежащих в интервале от 8% до 13% (см. табл. 11.6). Следовательно, вполне возможно, что примененные нами порождающие грамматики наряду с их формальным значением для определения меры сложности имеют и непосредственное молекулярно-биологическое значение. На интуитивном уровне можно утверждать, что порождающая грамматика описывает грамматику построения последовательности из фрагментов, причем все встречающиеся повторения использованы для того, чтобы представить последовательность с помощью как можно меньшего числа относительно коротких подпоследовательностей и их повторений. Недавно появились различные указания на то, что необходимая для синтеза белков мРНК

также составлена из нескольких стандартных фрагментов. Это означает, что методы построения грамматических конструкций удастся успешно сочетать с фрагментами, считанными из ДНК. Если воспользоваться инженерной терминологией, то можно сказать, что использованные выше производящие грамматики служат мерой стандартизации конечного продукта. С этой точки зрения мы стремимся к тому, чтобы использовать «экономические» преимущества максимальной стандартизации конечного продукта и при синтезе биополимеров. Мы предполагаем, что максимальная стандартизация является общим принципом оптимизации в эволюции. Наше предположение подтверждается эволюционными изменениями введенной выше длины правил. Следуя филогенетическому дереву Дэйхоффа (рис. 11.10), мы обнаружили, что длина правил L_G в ходе эволюции изменяется от значения $L_G = 0$ для предков животных и грибов (узел 13), живших примерно 1,2–1,5 миллиардов лет назад, до значения $L_G = 3$ у предков всех высших животных (узел 11), живших около 1 миллиарда лет назад, и значений $L_G = 7$ и $L_G = 8$ для птиц и млекопитающих, а также $L_G = 6$ для пекарских дрожжей. Аналогичным образом вычислялась и грамматическая избыточность. Опираясь на эти данные, мы (Ebeling, Jiménez-Montaño, 1980) сформулировали гипотезу о том, что эволюция представляет собой компромисс в смысле теории игр между тенденцией к увеличению сложности и тенденцией к возрастанию грамматической структуры. Возможно, оптимизация сложности и тем самым количества информации была доминирующей тенденцией на ранних стадиях эволюции биополимеров. На более поздних стадиях доминирующей стратегией, по-видимому, стало введение все более сложных грамматических структур.

Глава 12

Перспективы использования

Всякий раз, когда мне случается узнать нечто новое, я тотчас же принимаюсь размышлять над тем, нельзя ли сотворить из этого что-нибудь для жизни.

Готтфрид Вильгельм Лейбниц

12.1. Проблемы энергии, материалов и физических элементов

Развитие производительных сил происходит в настоящее время с необычайной быстротой, и многое указывает на то, что в ближайшие годы и десятилетия темпы научно-технического прогресса ускорятся еще больше. Существенный вклад в этот процесс принадлежит науке: многие отрасли современной техники и технологии, например, ядерная техника, микроэлектроника или генная инженерия, являются результатами научных поисковых разработок, проводившихся на протяжении десятилетий. Процесс инновации начинается с нового знания, получаемого в результате фундаментального научного исследования, например, открытия расщепления атомного ядра Отто Ганом или объяснения молекулярной структуры наследственности Уотсоном и Криком. Часто в момент совершения ключевого открытия у его авторов существуют самые смутные представления относительно далеко идущих последствий, которые оно может оказать на экономику и политику, на характер производства и образ жизни людей. Еще каких-нибудь сто лет назад электромагнитные волны были чисто гипотетическими, математическими следствиями из уравнений Максвелла, а сегодня наша жизнь немыслима без радио и телевидения.

Разумеется, лишь небольшое число открытий приводят к фундаментальным инновациям; результаты многих исследований в дальнейшем оказываются либо технически неприменимыми, либо имеют лишь ограниченную полезность. Поэтому столь важно внимательно отслеживать тенденции в фундаментальных исследованиях, чтобы правильно оценивать их значение для будущих приложений в технике.

Именно с этой точки зрения следует подходить к приводимым далее соображениям относительно использования принципов самоорганизации и эволюционных стратегий. Хотя ни одна из обсуждаемых ниже идей не привела к глубоким инновациям в технике, тем не менее некоторые из них кажутся весьма многообещающими. Мы надеемся, что краткое изложение некоторых новых идей будет способствовать их критическому, но непредвзятому обсуждению.

Особое значение имеет проверка новых решений проблем энергии, материалов и конструкционных элементов. Диссипативные структуры в том виде, как они возникают в процессе самоорганизации в физических, химических или биологических системах, обнаруживают определенные характеристики, связанные с подводом и хранением энергии, а также тепловыделением. Здесь открываются разнообразные возможности технических приложений, которые до сих пор реализованы лишь весьма незначительно. Так как хранение энергии, как правило, соответствует подводу

другой формы энергии, диссипативные структуры в принципе открывают возможность создания преобразователя энергии на новой технической основе, например, преобразователя химической энергии в электрическую (рис. 12.1). Такой преобразователь энергии может существовать в различных вариантах.

1. Источники энергии. Требуется, чтобы новые источники по своим характеристикам не уступали существующим батареям и аккумуляторам. Если учесть высокий уровень современной аккумуляторно-батарейной техники, то станет ясно, что удовлетворить этому требованию очень трудно. Как показали теоретические оценки, КПД диссипативных структур как преобразователей энергии составляет лишь несколько процентов, и поэтому они вряд ли могут составить конкуренцию батареям и аккумуляторам, работающим вблизи равновесия (*Feistel, Ebeling, 1984, 1988; Malchow, Schimansky-Geier, 1986*). С другой стороны, появились новые источники энергии, например, батареи переменного тока, для которых могут существовать свои особые возможности замены.

2. Сенсоры. Во многих случаях КПД преобразования энергии диссипативными структурами относительно низок, и развиваемые мощности малы. Но для использования в сенсорах этих КПД и мощностей вполне достаточно. Разнообразие возможных диссипативных структур позволяет в принципе разработать различные типы сенсоров для различных областей применения (*Malchow, Schimansky-Geier, 1986*).

3. Методы измерения. Диссипативные структуры часто весьма чувствительно реагируют на изменения краевых условий и при этом могут претерпевать кинетические фазовые переходы различного рода. Фазовые переходы допускают создание точных методов измерения; например, химические диссипативные структуры служат весьма чувствительным средством обнаружения слабых электрических полей.

4. Транспорт энергии. Диссипативные структуры образуются только в открытых системах, постоянно подпитываемых внешним потоком энергии и/или вещества. Возникновение структур приводит к эффективному внутреннему транспорту, который может быть использован для переноса энергии или информации, или к быстрому, энергетически экономичному отводу тепла, или к созданию перепадов температур (*Wheatley, Hoffler, 1983*).

5. Методы разделения веществ. Диссипативные структуры представляют собой относительно простую форму энергии, которая, однако, нередко трудно реализуема с помощью существующих ныне традиционных технологий: градиентов, или перепадов, концентраций химических веществ (*Feistel, Ebeling, 1984*). На этом принципе могут быть созданы экономичные (требующие малых затрат энергии), «холодные» методы разделения веществ (рис. 12.2).

Широкое поле для приложений методов теории самоорганизации открывается в области двигателей внутреннего сгорания. Например, здесь уже достигнуты многообещающие результаты в ходе исследования разброса момента зажигания горючей смеси (*Rigos, Deutch, 1982; Baras et al., 1983*). Большой интерес представляет также

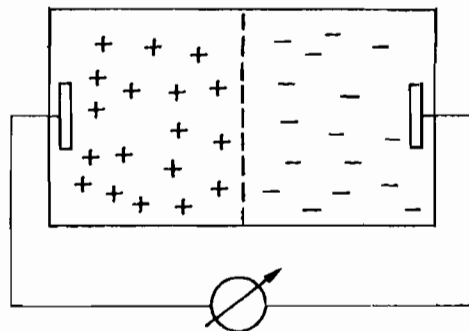


Рис. 12.1. Схема батареи на основе электрических диссипативных структур

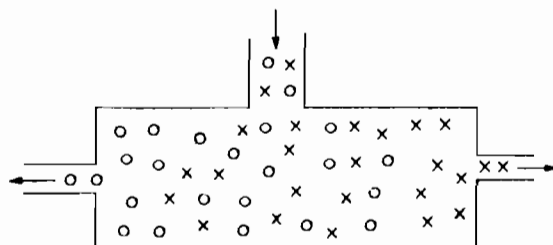


Рис. 12.2. Схема установки для разделения веществ на основе химических диссипативных структур

такая область приложения, как новый подход к расчету КПД энергетических установок с помощью новейших теоретических методов (*Richter et al.*, 1981). Классическая теория машин, преобразующих энергию, исходит из предположения об обратимом характере процессов. В действительности же процессы в энергетических установках протекают вдали от равновесия и поэтому связаны с производством энтропии. Теоретический анализ таких процессов оказался чрезвычайно трудным (*Andresen et al.*, 1977). Преодоление встретившихся трудностей открывает возможности дальнейшей оптимизации традиционных энергетических установок (*Richter, Ross*, 1978).

Интересные возможности использования самоорганизации открываются в такой области техники, как лакокрасочные покрытия для защиты поверхностей от агрессивной среды и повышение износоустойчивости трущихся поверхностей. Важная область применения процессов самоорганизации — антикоррозионные покрытия. Самоорганизация может как способствовать, так и препятствовать нанесению антикоррозионных слоев. Например, при нанесении лака на металлическую поверхность в автомобильной промышленности образование ячеек в результате неустойчивостей граничных поверхностей при испарении летучего растворителя относится к числу нежелательных побочных эффектов (*Linde*, 1984). Однородность лакового покрытия нарушается, и возникают «тонкие места» — центры будущего коррозионного разрушения материала (рис. 12.3). Вторая важная область приложений связана с процессами разрушения. Материал, находящийся под напряжением, можно рассматривать как активную среду (с подкачкой энергии), к каждой пространственной точке которой подводится свободная энергия. При надкритических нагружениях (темпах накачки энергии) материал организуется заново: в нем возникают трещины, сколы и другие новые структуры. Третья область приложений самоорганизации — повышение износоустойчивости трущихся поверхностей. Показано, что использование механизмов самоорганизации может значительно уменьшить износ подшипников. При таком подходе подшипник рассматривается как открытая система с постоянной подкачкой извне высокоценной механической энергии (рис. 12.4). Механизм селективного трения (нулевой износ), открытый в 50-х годах Гаркуновым и Кругельским, основан на том, что часть накачиваемой энергии отводится физико-химическим процессам, уменьшающим износ или сводящим его к нулю (*Гаркунов*, 1982).



Рис. 12.3. Образование утоньшений в сохнущем лаковом покрытии на основе эффекта Марангони

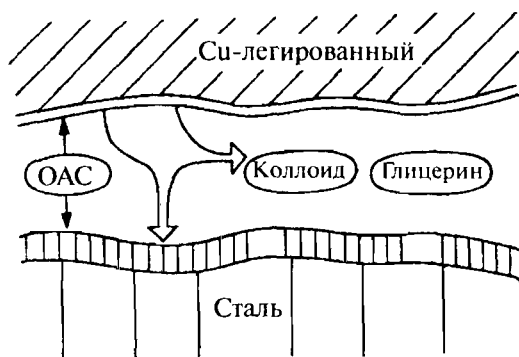


Рис. 12.4. Схема селективного трения по Гаркунову и Кругельскому при образовании диссипативных структур в слое трения

Некоторые возможности открываются и в области развития конструктивных элементов для микроэлектроники и оптоэлектроники. Применение принципов самоорганизации и эволюции ведется весьма интенсивно в различных направлениях усилиями международного содружества исследователей. Перечислим основные направления проводимых работ.

1. Продление срока службы элементов за счет более глубокого понимания процессов их разрушения. По Кернеру и Осипову (*Кернер, Осипов*, 1974, 1977, 1981), многие случаи отказа элементов обусловлены шнурованием тока и пробоем. На эту тему имеется обширная информация (*Кернер, Осипов*, 1981; *Балкарей, Элинсон*, 1983; *Балкарей и др.*, 1979, 1980, 1981).

2. Генерация автоволн в полупроводниках. Ряд идей, часть которых получила экспериментальное подтверждение, был высказан Балкареем и его сотрудниками (*Балкарей, Элинсон, 1977, 1979, 1983*). Следуя этим идеям, можно построить передатчик на основе спиральных волн.

3. Голографическая обработка изображений (*Букатова и др., 1978, Балкарей и др. 1979, 1980, 1981*) в активных полупроводниках.

4. Накачка светопроводов по образцу нервной проводимости (рис. 12.5).

5. Создание динамической кратковременной памяти, работающей по аналогии с головным мозгом человека (*Евтихов и др., 1978; Кринский, Михайлов, 1984*).

6. Развитие функциональной электроники на основе самоорганизующихся сред (*Калнин, Лухинин, 1981, 1983; Калнин, 1982, 1984*).

Правда, нельзя не учитывать, что новые принципы и технологии постепенно вытесняют существующие ныне физические элементы. По-видимому, будет создано новое поколение физических элементов, которое придет на помощь существующим ныне элементам.

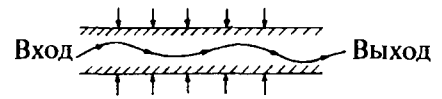


Рис. 12.5. Световод, действующий как насос по принципу нервной проводимости

12.2. Проблемы автоматизации и оптимизации

Особенно интересные приложения принципы самоорганизации и эволюции могут найти в робототехнике. Новые поколения роботов отличаются не только повышенной точностью, повышенной скоростью и гибкостью производимых манипуляций, но и более глубоким восприятием и лучшей переработкой информации, поступающей от окружающего мира, например, с помощью сенсоров и электронной связи с роботами-«коллегами». Человек предоставляет автоматам все бóльшую свободу принятия решений, машины обретают все бóльшую «свободу рук»; тем самым система производства получает больше внутренних степеней свободы (в физическом смысле или в смысле системного подхода). Как показывает теория образования структур, при достижении числа степеней свободы самоорганизации некоторого критического уровня одни «моды» подчиняют себе другие. В результате возникает новая структура (система переходит в новый рабочий режим), желательная и нежелательная, причем возникновение ее может быть иногда весьма неожиданным.

Поясним сказанное на простом примере из области транспортных автоматов — грузоподъемников или лифтов. Эти автоматы обладают «свободой» переходить в рабочий режим только после поступления вызова, а в случае одновременного поступления нескольких вызовов самостоятельно «решать», в какой последовательности их выполнять. При надкритической частоте поступления вызовов внутренние степени свободы в результате самоорганизации выходят на новый режим, в котором система минимизирует или максимизирует свою пропускную способность или быстродействие. Система лифтов в крупном отеле с несколькими лифтовыми шахтами представляет собой типичный пример такой самоорганизации. Если от постояльцев отеля одновременно поступает много вызовов, все лифты автоматически синхронизируются и поднимаются или спускаются одновременно. В результате такой синхронизации среднее время ожидания для постояльца отеля становится максимальным. Этот нежелательный эффект заведомо не был предусмотрен конструкторами лифтов. Самоорганизация системы лифтов приводит к отрицательным последствиям для постояльцев отеля.

Аналогичный эффект наблюдается и в обычном общественном транспорте. Если автобус или трамвай немного отстает от расписания, то на остановках скапливается

много ожидающих, и из-за задержек при посадке опоздание еще больше увеличивается. Зато выигрывает следующий автобус: он прибывает вовремя, чтобы забрать умеренное скопление пассажиров. Профессиональные аналитики высказывают также опасения по поводу того, что огромные сети автоматов, связанные с созданием современного ракетного оружия, могут приводить к возникновению опасной тенденции к самоорганизации. Известны эффекты самоорганизации и в сложных системах программ, и в компьютерных сетях; доскональное знание этих эффектов и умение предотвращать их нежелательные последствия приобретают все большее значение.

Еще одним примером нежелательных процессов самоорганизации могут служить самопроизвольные колебания рельсов. Значительная часть энергии, развиваемой тепловозом или электровозом, превращается в колебания рельсов и насыпи, на которую они уложены, и эти колебания могут приводить к разрушению железнодорожного полотна. Более тщательно изучение этого крайне нежелательного процесса может способствовать увеличению продолжительности интервалов между дорогостоящими работами по замене изношенных участков рельсового пути новыми.

Приведенные примеры показывают, что знание процессов самоорганизации играет важную роль при разработке автоматизированных производственных установок, ибо позволяет исключить возможность возникновения нежелательной самоорганизации и, если обстоятельства позволяют, использовать самоорганизацию для оптимизации технических, экономических, экологических и др. процессов. Для этого чрезвычайно важны дальнейшие исследования. Одно из интересных направлений таких исследований — самоорганизация моделей прогнозирования (*Ivachnenko, Müller, 1984*).

Обширной областью приложений принципов эволюции является актуальная проблема автоматизации процессов проектирования и оптимизации проектов (*Rechenberg, 1973; Schwefel, 1977*). Оптимизация несомненно принадлежит к числу основных проблем современной экономики и технологии. Современному обществу требуются простые, надежные и недорогие решения. Примерами использования принципов самоорганизации и эволюции для решения задач оптимизации мы особенно обязаны работам Холланда, Рехенберга и Швefeldя, успешно решивших большое число интересных технических проблем. Мы рассмотрим в этой главе лишь один специальный класс трудных задач на оптимизацию, частные случаи которых встречаются в самых различных областях экономики и техники.

Пример 1. В микроэлектронике стоит задача соединить проводящими путями по заданной схеме относительно большое количество заданных переключательных схем ($n = 10^2 - 10^7$). Как наилучшим образом расположить элементы на чипе, чтобы суммарная длина проводящих путей была возможно меньше? Разумеется, на практике приходится соблюдать множество дополнительных условий, например, следить за тем, чтобы проводящие пути не пересекались и т. д. (*Iwainsky, 1984, 1985; Voigt, 1985*).

Пример 2. Основная проблема современного двигателестроения состоит в том, чтобы построить оптимальный профиль дюз для многофазных и реактивных течений. Полный теоретический расчет протекающих при этом сложных процессов пока невозможен. Рехенберг (*Rechenberg, 1973*) и Швefeldь (*Schwefel, 1977*) показали, каким образом эволюционные стратегии позволяют найти оптимальные решения. Задача состояла в том, чтобы с помощью горячей смеси из паров и капель калия создать сопло с максимальной тягой. Для экспериментальной адаптивной оптимизации было подготовлено 330 конических сопел различного продольного сечения. Оптимальная форма сопла (рис. 12.6) была найдена методом проб и ошибок.

Пример 3. Торговые и транспортные проблемы нередко требуют расположить потребителей в такой последовательности, чтобы их можно было объехать по крат-

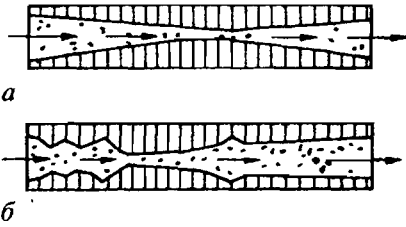


Рис. 12.6. Задача Рехенберга и Швевеля состоит в том, чтобы найти оптимальный профиль дюзы, обеспечивающий максимальную тягу двухфазного течения: (а) до; (б) после применения эволюционной стратегии

чайшему пути или снизить дорожные расходы до минимума. Эту проблему, известную под названием «задачи коммивояжера», пытались решить многие исследователи (Kirkpatrick et al., 1983, 1985).

Пример 4. Важной проблемой для эффективности производства является рациональное использование станочного парка. Речь идет о таком использовании имеющихся станков, при котором сводятся до минимума простои или расходы на переналадку. Аблай (Ablay, 1987) показал, что и такого рода задачи могут решаться с помощью эволюционных стратегий.

Трудность решения задач бегло очерченного выше класса связана с тем, что число производимых операций растет экспоненциально с увеличением числа элементов (неполиномиальная, или НП-, полнота). Это приводит к чудовищным временам счета и обуславливает непригодность классических стандартных методов оптимизации. Эффективность новых методов поиска решений мы продемонстрируем на примере задачи об обходе всех ребер графа (рис. 12.7). Требуется найти маршрут обхода всех ребер, при котором длина пути была бы минимальной в заданной метрике. Метод решения состоит в том, чтобы, исходя из произвольно выбранного начального маршрута, с помощью стохастического перебора ребер приблизиться, насколько это возможно, к неизвестному оптимуму. Существует огромное число математических работ, в которых наряду с классической градиентной стратегией предлагается ряд более практичных способов решения. Совершенно новый физически мотивированный подход был предложен в работе (Kirkpatrick, Gelatt, Vecchi, 1983). Авторы назвали свой метод стратегией моделируемого отжига. Идея метода состоит в следующем. Мы моделируем процесс стохастического перебора ребер замкнутого маршрута с вероятностями перехода, удовлетворяющими распределению Больцмана с псевдотемпературой T (метод Монте—Карло, алгоритм Метрополиса). Начальную температуру мы выбираем высокой, а затем постепенно ее понижаем. В качестве функции Гамильтона служит суммарная длина пути

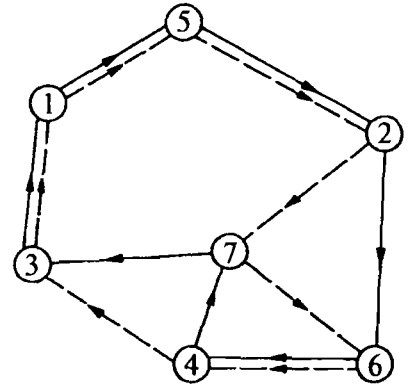


Рис. 12.7. Задача коммивояжера состоит в том, чтобы найти кратчайший маршрут, проходящий через n городов

$$H = L = \sum L_{ij}.$$

Иначе говоря, мы подходим к решению задачи так же, как доменщик, стремящийся выплавить высококачественный металл. Метод моделируемого охлаждения, известный также под названием больцмановской стратегии, применялся во многих научных исследованиях и неоднократно усовершенствовался и видоизменялся. Один из частных вариантов этого метода был предложен и авторами настоящей книги (Ebeling, Engel, 1986; Boseniuk, Ebeling, Engel, 1987). Основная идея состоит в сочетании больцмановской стратегии с дарвиновской стратегией. Мы исходили из следующего простого соображения: пример эволюции учит нас, что для возникновения качественно новой и обладающей какими-то преимуществами структуры из ансамбля

взаимодействующих атомов необходимо определенное сочетание больцмановской и дарвиновской стратегий (Ebeling, 1986). Затем удалось получить теоретическое доказательство возможных преимуществ смешанных стратегий (Ebeling, Engel, 1986) и, наконец, подтвердить преимущества нового образования численным моделированием на персональном компьютере (Boseniuk et al., 1987). В ходе наших исследований мы систематически подвергали сравнению три стратегии.

1. *Стохастическая градиентная стратегия.* Замкнутые маршруты подвергаются стохастическим вариациям (быстрая динамика). Положительные результаты, т. е. такие изменения, которые приводят к сокращению длины маршрута, принимаются, отрицательные результаты отвергаются.

2. *Моделируемое охлаждение.* Положительные результаты мутаций всегда принимаются, отрицательные результаты вводятся с малой вероятностью

$$\exp \left\{ - \frac{\Delta L}{T} \right\}. \quad (12.2)$$

Это предположение позволяет системе покидать неблагоприятные побочные минимумы и перемещаться в направлении лучших минимумов. Температура медленно понижается (медленная динамика).

3. *Смешанная стратегия Больцмана—Дарвина.* Рассматривается ансамбль из n коммивояжеров, которые попарно производят сравнения выбранных маршрутов через определенные стохастически распределенные временные интервалы. Менее удачливый коммивояжер после сравнения следует по лучшему маршруту. В ходе дальнейшего развития событий маршруты двух коммивояжеров, как правило, снова расходятся из-за процессов стохастической мутации. На рис. 12.8 приведены для сравнения лучшие результаты, достигнутые в численном моделировании десятью коммивояжерами при использовании трех указанных выше стратегий. На рис. 12.9 показан лучший маршрут. Моделирование проводилось в 50 прогонах. Нетрудно

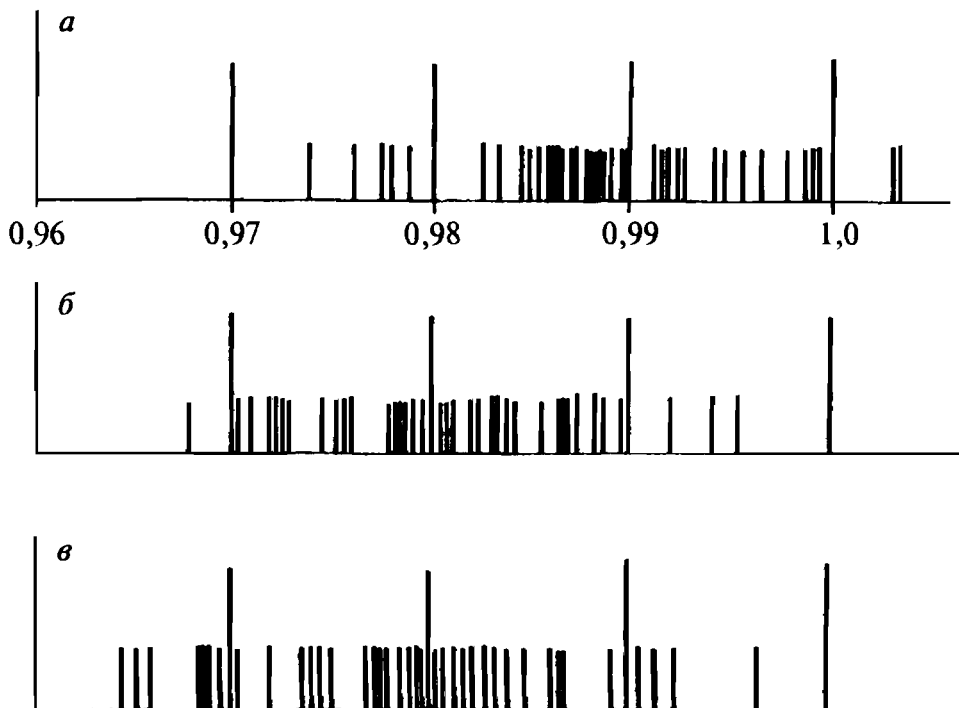


Рис. 12.8. Распределение лучших результатов 50 оборотов коллектива, состоящего из 10 торговцев: (а) независимая стохастическая градиентная стратегия; (б) независимая термодинамическая стратегия; (в) введение соревнования между торговцами (смешанная стратегия Больцмана—Дарвина по Босенюку и др. (Boseniuk et al., 1987)

видеть, что смешанная стратегия обладает заметными преимуществами перед двумя другими стратегиями.

Приведенные выше стохастические замкнутые сетевые графики можно рассматривать как примеры работы с эволюционными стратегиями. Особый интерес замкнутые сетевые графики представляют для задач оптимизации. Такие сети часто используют в компьютерных моделях и даже в конструкциях компьютеров. Сотрудники фирмы «Bell Laboratories» разработали в последнее время различные проекты кремниевых чипов, позволяющих с помощью измерения токов на чипе из $n \times n$ элементов решать задачу коммивояжера. Технологии VLSI позволяют достигать большой поверхностной плотности элементов, т. е. работать с достаточно большим числом городов. В ближайшем будущем следует ожидать появления специальных чипов для серийного решения задач на оптимизацию типа задачи коммивояжера (Hopfield, Tank, 1985, 1986).

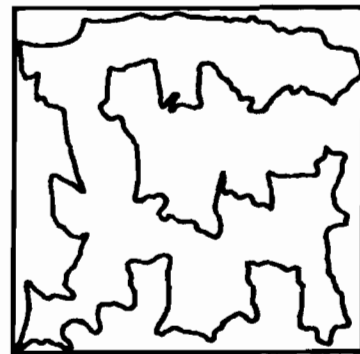


Рис. 12.9. Оптимальный маршрут, найденный для задачи о круговом объезде коммивояжером 100 городов с помощью смешанной стратегии Больцмана—Дарвина (Boseniuk et al., 1987)

12.3. Проблемы биотехнологии

Ныне нет никаких сомнений в том, что биотехнология наряду с микроэлектроникой принадлежит к числу наиболее важных технологий, которые необходимо всемерно развивать. «Революция в биохимии, достигшая своей кульминации в начале 60-х годов с открытием генетического механизма и связи между нуклеиновыми кислотами и синтезом белка, — это прорыв в знании, быть может, не столь глубокий, как тот, который был вызван физико-химическими открытиями в первое десятилетие нашего столетия, но гораздо более сложный и влекущий за собой гораздо дальше идущие последствия» (Bernal, 1967). В конце концов речь идет о том, что человек по образу и подобию природы становится конструктором биотической эволюции в рамках законов природы. Особенно далеко идущих последствий для экономики следует ожидать от геной инженерии и энзимотехники, ставших ныне главным направлением научно-технической революции. Для научного анализа процессов биотехнологии знание теории самоорганизации имеет определенное значение.

Один из вопросов, которые ставит теория самоорганизации и эволюции, состоит в следующем. Если система состоит из большого числа взаимодействующих, относительно однотипных и относительно самостоятельных элементов (например, популяция микробов) и если эти элементы обладают способностью медленно изменять свои свойства (мутации, процессы обучения), то как изменяются свойства элементов? Как изменяются свойства системы в целом? Как реагирует («отвечает») система на изменение краевых условий?

За этими внешне простыми вопросами, несомненно, кроются чрезвычайно сложные проблемы. Пока мы имеем лишь самые общие представления о том, как надлежит описывать и решать их (Feistel, Ebeling, 1982, 1989). Попытаемся в общих чертах описать сферу применимости соответствующих результатов. В микробиологических реакторах живут и размножаются одноклеточные, в которых искусственно введен генетический материал — информация, необходимая для производства определенных ферментов, например, инсулина. Поскольку производство инсулина не требуется для размножения клеток (и даже замедляет последнее), эта способность постепенно понижается в результате мутаций, и в дальнейшем клеточная структура перестает быть необходимой для производства инсулина. Возникает вопрос, какие

условия надлежит создать в реакторе, чтобы естественная эволюция популяции клеток приводила не к уменьшению выработки инсулина, а к ее повышению.

Другой пример — поиск заменителей антибиотиков как средства борьбы с возбудителями болезней. Естественная эволюция бактерий приводит к иммунитету против медикаментов, поэтому уже сейчас известны, например, штаммы малярии, устойчивые к действию хинина или венерических болезней и действию пенициллина. Проблема эволюционной теории состоит в том, можно ли создать условия, при которых естественная эволюция действовала бы не против интересов человека, а в интересах последнего. С аналогичной ситуацией мы встречаемся и в сельском хозяйстве, когда речь заходит о борьбе с вредителями или об удобрениях, а также использовании фармацевтических препаратов в животноводстве. Нетрудно предвидеть, что значительные изменения в химической промышленности будут сопровождаться все более широким использованием ферментов. Правда, сегодня эти изменения выступают лишь как теоретическая возможность, но успехи генной инженерии и микроэлектроники могут быстро изменить всю картину. Почему так происходит? Дело в том, что современная химическая промышленность часто осуществляет превращение одних веществ в другие при высоких температурах, т. е. затрачивая большие количества энергии. При всех успехах химии синтетических веществ следует иметь в виду, что спектр химических соединений, производство которых требует экономически оправданных затрат, относительно узок. Что же касается живых клеток, то они способны осуществлять весьма сложный химический синтез с большой быстротой, высокой точностью и с минимальными затратами энергии. Основную функцию в химии клеток выполняют цепи нуклеиновых кислот («программа») и ферменты («автомат»). Промышленное копирование химии клеток пока невозможно, с одной стороны, потому что функция ферментов недостаточно известна, а с другой — из-за высокой стоимости синтеза ферментов. Главную роль в биохимической регуляции играют так называемые аллостерические ферменты. Скорость, с которой они превращают одно вещество в другое, определяется концентрацией третьего вещества. Возникает далеко идущая аналогия с функцией транзистора в электронных схемах. Математически эта аналогия между электрическими и химическими цепями проявляется особенно отчетливо в термодинамике сетей (*Oster et al.*, 1973; *Schnakenberg*, 1976). В этом смысле клетку можно сравнить с высоко интегрированной электронной схемой, а технические применения современных электронных схем позволяют судить о потенциальных возможностях «химических схем» будущего (*Conrad*, 1985). Идея технического применения системы химических реакций по образу и подобию электроники разрабатывалась несколькими авторами, в особенности Конрадом и Е. Е. Сельковым. В этой связи следует упомянуть также работы (*Haddon, Lomola*, 1983; *Conrad*, 1985, 1986; *Nöldechen*, 1986), привлечшие внимание специалистов.

В отличие от электронных схем, реализуемых как твердые тела, биохимические сети работают в жидкой фазе. Пространственное упорядочение элементов в последнем случае возможно лишь весьма условно, так как в жидкости диффузионные и конвективные процессы существенно влияют на свойства всей системы. Теория самоорганизации здесь может прийти на помощь, указав, при каких условиях устанавливается, самоорганизуется та или иная желательная функция системы. Ответы на важные фундаментальные вопросы теории самоорганизации дают современные исследования свойств систем реакций с диффузией, проводимые с большой интенсивностью во всем мире (*Романовский и др.*, 1984; *Cartianu*, 1986; *Васильев и др.*, 1987). Первым шагом на пути к химии ферментов можно считать генную инженерию: во-первых, она способствовала развитию методов получения ферментов, во-вторых, весьма настойчиво развивала и развивает экспериментальную технику. Последующий

прогресс может приобретать сильнейшее самоускорение, так как ферменты могут возникать и самопроизвольно в ходе производства, подобно тому, как это с высокой вероятностью происходило при возникновении жизни на Земле (*Ebeling, Feistel, 1982*).

12.4. Ассоциативная память

Хранение информации играет центральную роль в обработке информации живыми системами. Если в гл. 11 мы подробно рассмотрели на элементарном уровне хранение генетической информации в последовательностях ДНК, то в данном случае речь идет о более сложных формах хранения биологической информации, представляющих большой интерес и с точки зрения их технического использования (*Hinton, Anderson, 1981; Amari, Arbib, 1982; Kohonen, 1984*). В этом отношении особый интерес представляет ассоциативная память. Хранимая в такой памяти информация опрашивается путем введения в память неполной или частично ложной информации. Если вводимый сигнал достаточно сильно напоминает содержимое памяти, то из памяти будет извлечена полная или исправленная информация. Такого рода ассоциативность характерна для центральной нервной системы многих высокоорганизованных существ и в особенно совершенной форме — для головного мозга человека. Что же касается современных цифровых запоминающих устройств, то они работают с адресами, т. е. неассоциативно.

С физической точки зрения ассоциативная память представляет собой динамическую систему с большим числом аттракторов, представляющих хранимую в памяти информацию. Любое распределение информации — содержимое памяти — окружено некоторой областью притяжения, определяемой конкретной динамикой системы (рис. 12.10), в силу чего начальная конфигурация, лежащая в окрестности того или иного аттрактора, под действием динамики системы асимптотически стремится к аттрактору. Тем самым частичная информация выступает в роли стимула, инициирующего выдачу полной информации, хранящейся в памяти (*Hogg, Hubermann, 1985; Николис, 1987*). Если мы хотим, исходя из этих совершенно общих представлений, провести физические исследования ассоциативной памяти, то нам необходима какая-то конкретная модель, которая с одной стороны допускала бы подробный математический анализ, а с другой стороны позволяла бы воспроизводить как можно больше известных явлений из области биологии и психологии. Одна из таких моделей была предложена в 1982 г. Дж. Дж. Хопфилдом (*Hopfield, 1982; Hopfield et al., 1983*). Модель Хопфилда существенно использует некоторые аналогии с проблематикой спиновых стекол, рассмотрению которой посвящен разд. 10.2. Проведенные за последние годы исследования модели Хопфилда позволили получить ряд интересных результатов. К изложению некоторых из них мы сейчас перейдем.

Мостик к спиновым стеклам возникает прежде всего в силу мультимодальности функции Гамильтона (10.2) спинового стекла и четко выраженному бистабильному характеру нейронов. Нейроны могут находиться в двух различных состояниях — в активном состоянии и в состоянии покоя, которые могут быть сопоставлены спиновой переменной Изинга $S_i = \pm 1$. В том же грубом приближении динамика может быть описана с помощью простой пороговой динамики. Каждый нейрон

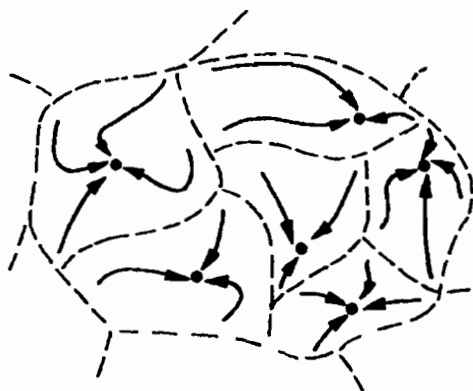


Рис. 12.10. Области притяжения аттракторов динамической системы, действующей как ассоциативная память

$i = 1, \dots, M$ получает через синапсы J_{ij} от других нейронов $j + i$ в зависимости от их состояния S_j потенциал

$$E_i = \sum_{j \neq i} J_{ij} S_j \quad (12.1)$$

и сравнивает его с пороговым потенциалом E_i^0 , который мы для простоты положили равным нулю для всех нейронов ($E_i^0 = 0$). При $E_i > 0$ нейрон переходит в активное состояние $S_i = 1$, при $E_i < 0$ — в состояние покоя. При дополнительном предположении о симметрии синапсов

$$J_{ij} = J_{ji} \quad (12.2)$$

такая динамика соответствует минимизации функции

$$H(\{S_i\}) = - \sum_i E_i S_i = - \sum_{i \neq j} J_{ij} S_i S_j. \quad (12.3)$$

Поскольку существуют как возбуждающие ($J_{ij} > 0$), так и тормозящие ($J_{ij} < 0$) синапсы, функция Гамильтона (10.3) сильно напоминает функцию Гамильтона (10.2) спинового стекла и можно ожидать, что и в случае нейронов образуется сложная структура фазового пространства с большим числом локальных минимумов. Так как эти минимумы соответствуют именно таким аттракторам в нашей динамике, нейронная система подходит в качестве ассоциативной памяти. Существенной проблемой остается задание четкого алгоритма, или «рецепта», для синаптических связей J_{ij} , который гарантировал бы, что именно заранее заданный набор из p конфигураций $\{\xi_i^\mu\}$, $\xi_i^\mu = \pm 1$, $i = 1, \dots, N$; $\mu = 1, \dots, p$, будет минимумами функции (12.3). Эти конфигурации представляют собой содержание памяти, которое может быть опрошено на ассоциативной основе.

Существует большое число так называемых правил обучения $J_{ij} = J_{ij}(\{\xi_i^\mu\})$, из которых мы назовем лишь несколько. Простейшее из них восходит к Хеббу и первоначально было также введено Хопфилдом. Это правило гласит:

$$J_{ij} = \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu. \quad (12.4)$$

Алгоритм (12.4) позволяет хранить в памяти до $p \approx 0,15 \cdot N$ ортогональных конфигураций $\{\xi_i^\mu\}$ и вызывать их на ассоциативной основе, если ошибка достаточна мала ($\lesssim 1,5\%$). До $p = N$ неортогональных конфигураций позволяет безошибочно записать в памяти и вызывать из нее на ассоциативной основе алгоритм

$$J_{ij} = \sum_{\nu, \mu=1}^p \xi_i^\mu (C^{-1})_{\mu\nu} \xi_j^\nu, \quad (12.5)$$

где C^{-1} — матрица, обратная матрице перекрытий (см. гл. 10)

$$C_{\mu\nu} = \frac{1}{N} \sum_i \xi_i^\mu \xi_i^\nu \quad (12.6)$$

(Kanter, Sompolinsky, 1987). Для ортогональной конфигурации $C_{\mu\nu} = \delta_{\mu\nu}$, и алгоритм (12.5) упрощается до (12.4).

Первые исследования такой модели спинового стекла для ассоциативной памяти носили численный характер (Hopfield, 1982; Kinzel, 1985). Существенный прорыв

в этой области наметился, когда модель Хопфилда удалось исследовать аналитическими методами теории спиновых стекол (*Amit et al.*, 1985, 1987). Многие свойства модели Хопфилда стало возможным охарактеризовать точно. В общем случае удается записать в память информацию $O(N)$, т. е. число p (объем памяти) возрастает пропорционально числу нейронов N . Для отношения $\alpha = p/N$ существует критическое значение $\alpha_{\text{крит}}$, отделяющее режим хранения и ассоциативного вызова информации из памяти ($\alpha < \alpha_{\text{крит}}$) от так называемого состояния конфузии, или путаницы ($\alpha > \alpha_{\text{крит}}$), в котором вся записанная в памяти информация утрачивается. Для правил обучения (12.4) и (12.5) мы в соответствии с численными результатами получаем $\alpha_{\text{крит}} = 0,14$ и $\alpha_{\text{крит}} = 1$. Обсуждались различные предложения относительно того, как избежать катастрофы перегрузки — «путаницы», или конфузии, — заменив ее направленным забыванием (*Веденов, Левченко*, 1985; *Parisi*, 1986; *Nadal et al.*, 1986; *Mezard et al.*, 1986; *Hoffmann et al.*, 1986; *Personnaz et al.*, 1985; *Personnaz et al.*, 1985). Исследовалась и (возмущающая) роль «перепутанной» информации. Если в памяти хранятся одновременно несколько информационных, то возникают побочные минимумы функции (12.3) в конфигурациях, представляющих собой комбинации фрагментов хранящихся в памяти информации. Побочные минимумы нарушают функционирование запоминающего устройства и могут приводить к «ложной» памяти (ошибочным интерпретациям). Их влияние может быть в значительной мере подавлено весьма интересным способом — путем введения шумов в нейродинамику (*Amit et al.*, 1985, 1987). Еще весьма многообещающая возможность по крайней мере частичного исключения ложной памяти — так называемое «обратное обучение», которое оказалось связанным с определенной фазой сна (*Hopfield et al.*, 1983; *Crick, Mitchison*, 1983). Изучались реакция системы на отказ определенной части (в процентах) синапсов или на стохастические ошибки в процессе обучения, а также нейронные сети с синапсами, которые могут находиться только в двух состояниях (± 1), что представляет большой интерес для технических приложений. Первоначально наложенные ограничения одно за другим удалось снять. Так, исследовались ситуации, в которых пороговое значение E_i^0 варьируется от нейрона к нейрону или синапсы несимметричны, т. е. когда соотношение (12.2) более не выполняется. Во всех случаях способность к ассоциативной памяти сохраняется, поэтому можно исходить из того, что речь идет о кооперативном эффекте, спонтанно возникающем в большой системе взаимодействующих элементов и согласно гипотезе универсальности статистической механики не зависящем от деталей подсистем. Тем самым в дальнейшем получает обоснование грубое описание нейронов как двоичных переменных. Хороший обзор современного состояния исследований в этой интересной области читатель найдет в трудах конференции (*van Hemmen, Morgenstern*, 1987).

В заключении упомянем о взаимосвязи ассоциативной памяти с важной и сложной проблемой распознавания образов. Если расположить N спиновых переменных на плоскости и сопоставить значениям $S_i = 1$ черное, а значениям $S_i = -1$ — белое (пустое) поле, то каждой конфигурации спинов $\{S_i\}$ соответствует некий двумерный образ. Конфигурации $\{\xi_i^\mu\}$ соответствуют p записанных в памяти образов, которые могут быть запрошены ассоциативно с помощью задания искаженного помехами



Рис. 12.11. Временное проявление эталонного изображения на сети бистабильных узлов как ассоциативная память о букве А (*Kinzel*, 1985)

(«зашумленного») начального образа. На рис. 12.11 представлен пример, в котором в качестве образа выбрана заглавная буква А. Нетрудно видеть, что при сохранении соответствующей емкости памяти уже после нескольких релаксационных шагов из сильно искаженных шумами букв возникает четкий образ, изначально записанный в памяти. Именно в этом и заключается взаимосвязь между ассоциативной памятью изложенного выше типа и проблемой распознавания образа. Следует, однако, отметить, что способность к ассоциативной связи в конфигурационном пространстве ограничена отношением близости, или соседства, т. е. наблюдается только на расстоянии Хэмминга. В настоящее время не существует точных представлений относительно того, каким образом можно было бы обобщить и расширить изложенную выше модель ассоциативной памяти на основе спиновых стекол, чтобы она была применима и к сложной проблеме распознавания повернутых или сдвинутых образов. Как показывает пример изображения на рис. 12.11, буква А, перевернутая «вверх тормашками», не была бы распознана, так как даже смещение неискаженного образа на несколько узлов решетки (растра) превращает его распознавание в проблему, решение которой выходит за рамки ассоциативных возможностей модели Хопфилда. Будущее покажет, удастся ли решить и этот класс проблем с помощью ассоциативных запоминающих устройств.

12.5. Сетевые автоматы и нейрокомпьютеры

В основе существующих ныне методов обработки информации лежит принципиальная схема компьютера, предложенная гениальным венгерским физиком и математиком Джоном (Яношем) фон Нейманом. Процесс обработки информации разлагается в последовательность двоичных логических операций. К числу наиболее явных преимуществ автоматов (компьютеров) фон Неймана относятся их программируемость и универсальность. Автоматы фон Неймана позволяют (по крайней мере в принципе) решать почти любые задачи. Практически, т. е. «в железе», компьютеры фон Неймана реализуются на основе электронных процессов, протекающих в схемах на кремниевых чипах. О необычайно большом значении этого поколения современных компьютеров вряд ли требуется что-нибудь говорить: оно общеизвестно. Однако специалисты не без основания неоднократно указывали на качественные ограничения современного поколения компьютеров. Так, американский специалист по компьютерам Конрад из Детройта приводит в качестве примера проблему распознавания структур и образов, проблему проектирования сложных систем, например, микроэлектронных чипов, и проблему принятия решений в сложных ситуациях (Conrad, 1986). Современные автоматы фон Неймана умеют (в принципе) делать все, что программируемо, но справляются с определенными задачами только ценой огромных затрат. В частности, сказанное относится к программному обеспечению. Рассмотрим пример. Маленькие дети безошибочно узнают голос своей матери среди тысячи других голосов, они без труда распознают животных, марки автомашин, оценивают возраст людей, выносят здравые суждения о характере той или иной ситуации, в том числе распознают опасность и умеют ее избегать. Конрад усматривает основные недостатки современного поколения машин, занимающихся обработкой информации, в их неспособности к обучению, негибкости, неэффективности и подверженности влиянию ошибок и возмущений. В ходе вычислений активно функционируют только 10% элементов современного компьютера, а остальные 90% пребывают в состоянии покоя. Способность современных компьютеров быстро переключаться с одного режима на другой означает малую помехоустойчивость: каждое внешнее воздействие возмущает систему.

Такого рода проблемы переработки информации неустанно решают и живые организмы. С проблемами распознавания образов и принятия решений великолепно

справляются даже простейшие организмы, у них хорошо развиты приспособляемость и обучаемость, а за счет параллельной обработки информации достигается высокая помехоустойчивость. Вместе с тем к недостаткам биологической обработки информации следует отнести ее медленность (скорость распространения сигналов 10–100 м/с), а также отсутствие программируемости и универсальности.

В обоснование своих выводов Конрад привел некоторые весьма весомые аргументы и сформулировал так называемый принцип стоимости. Этот принцип гласит, что не все желательные свойства компьютера, например, программируемость, универсальность, эффективность, приспособляемость, обучаемость и помехоустойчивость (надежность), могут быть реализованы одновременно. Компьютер (автомат) фон Неймана оптимизирован в отношении программируемости и универсальности, информационные системы живых организмов оптимизированы относительно эффективности, обучаемости и помехоустойчивости.

Исходя из этих соображений, в высокоразвитых промышленных странах была поставлена задача разработать информационные системы нового типа, которые соответствующим образом дополняли бы компьютер фон Неймана. Подчеркиваем: дополняли бы, а не заменяли бы. Сетевые компьютеры, о которых пойдет речь в этом разделе, построены по образу и подобию нейронных сетей, но не повторяют в деталях их структуру. Из-за чудовищной сложности реальных нейронных сетей сетевые компьютеры могут в лучшем случае (по крайней мере сейчас) воспроизводить лишь некоторые их аспекты. Дальний прицел исследований по созданию сетевых компьютеров состоит в разработке нового типа компьютеров, получивших в международной литературе название «нейрокомпьютеров» (*Ackley et al.*, 1985). Схематически устройство сетевого компьютера можно представить следующим образом: он состоит из мультистабильных элементов, которые могут быть соединены между собой стохастически. Система состоит из входа, выхода и внутренней части, элементы которой обладают стохастической динамикой. Каждой сети можно поставить в соответствие хроматический (цветной) граф; матрица, соответствующая ребрам такого рода, называется синаптической матрицей. Цвет узлов обозначает внутреннее состояние, например, «внутри» или «наружу». Связи между узлами переменные и подчиняются определенной динамике. Как правило, система изменяется в соответствии с быстрой детерминистической или стохастической динамикой и медленной адаптивной динамикой (принцип двойной динамики Конрада). Дальнейшие связи заведомо допустимы. Системы такого типа традиционно встречаются в кибернетике и физике, теории электрических сетей и линейных фильтров, а также в теории нейронных систем.

Отцом теории электрических сетей, как известно, по праву считается Густав Роберт Кирхгоф, возглавлявший с 1876 по 1889 г. первую кафедру теоретической физики в Берлинском университете. Для сетей, рассматриваемых нами в этом разделе, особый интерес представляют работы (*McCullosh, Pitts*, 1943; *Amari, Arbib*, 1982; *Rumelhart, McClelland*, 1986; *Hoppensteadt*, 1986). Упомянем также о связи такого рода сетей с теорией клеточных автоматов, основы которой были заложены фон Нейманом в его лекции «Общая и логическая теория автоматов» (сентябрь 1948 г.) Несмотря на то, что рассматриваемые нами сети имеют много общего с классическими хорошо исследованными сетями, они обладают также и некоторыми особенностями, состоящими в следующем.

1. Динамика изменения состояний нелинейна и кооперативна.
2. Состояния узлов би- или мультистабильны, состояния всей системы «фрустрированы», и число альтернативных состояний возрастает экспоненциально с увеличением числа узлов.
3. Узлы сети иерархически упорядочены на различных уровнях.

Первые основополагающие исследования сетевых автоматов восходят к Розенблатту (*Розенблатт*, 1965), который предложил и название «перцептрон», а также Минскому и Паперту (*Minsky, Papert*, 1969). С точки зрения приложений в последнее время особое внимание привлекают два типа автоматов:

- автоматы Больцмана (*Askley, Hinton, and Sejnowski*, 1985);
- атоматы распространения ошибки (*Sejnowski, Rosenberg* 1987).

Теория этих неоднократно обсуждавшихся в литературе обучающихся машин делает только первые шаги. Тем поразительнее достигнутые ею отдельные успехи. Под автоматом Больцмана мы будем понимать в дальнейшем стохастическую сеть, обладающую следующими основными элементами (рис. 12.12).

1. Сеть имеет иерархическую структуру и состоит из входного и выходного уровня, между которыми заключен внутренний уровень.

2. Между узлами, расположенными на этих трех уровнях, возникают стохастические связи; существует функция Гамильтона.

3. Состояния узлов (спины или цвет) подчиняются флип-динамике, соответствующей больцмановским переходам.

4. Система попеременно переходит в фазу обучения с жестко заданными отношениями вход—выход, фазу тестирования с фиксированным входом со свободным выходом, а также корректировочную фазу.

5. Поправка, вносимая в матрицу связей в корректировочной фазе, пропорциональна отклонению между требуемым и реальным отношением вход—выход.

6. Машина тренируется (меняет три фазы) до тех пор, пока при заданном входе она не достигает требуемого выхода (при свободной игре динамики). Достижение требуемого выхода означает, что машина «научилась» при заданном входе заведомо находить требуемый выход с помощью своей внутренней динамики.

Машина Больцмана — это тренируемая сеть. Ей необходим тренер, знающий требуемые отношения вход—выход и, исходя из этого, вводящий соответствующие поправки на отклонения в целеположении (задании цели). Динамика корректировки медленная и адаптивная; наоборот, динамика «опрокидывания» («флипа») быстрая и соответствует больцмановскому переходу. «Способности» автоматов Больцмана поистине поразительны. Например, после соответствующей тренировки эти автоматы обучаются классифицировать фигуры по типам симметрии, причем при произвольной ориентации фигур. Еще более впечатляющи достижения автоматов Больцмана, связанные с решением проблемы кодировки (*Ackley et al.*, 1985).

Еще более выдающимися способностями обладают автоматы распространения ошибки, разработанные Сейновским и Розенбергом (*Sejnowski, Rosenberg*, 1987) (рис. 12.13). В отличие от автоматов Больцмана, эти обучающиеся машины обладают непрерывной областью внутренних состояний с аттракторами в 0 и 1.

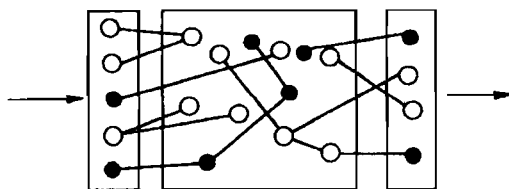


Рис. 12.12. Принцип машины Больцмана по Эккли, Хинтону и Сейновскому (*Ackley, Hinton, Sejnowski*, 1985). Система, обладающая ассоциативной памятью, обучается после нескольких циклов преобразовывать входное изображение в требуемое выходное изображение

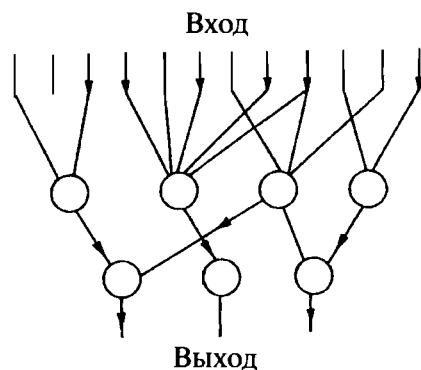


Рис. 12.13. Принцип машины распространения ошибок по Сейновскому и Розенбергу (*Sejnowski, Rosenberg*, 1987)

При заданном входе ошибка на выходном уровне составляет величину

$$E(z) = \sum_{k \in \text{выход}} (z_k - z_k^*)^2$$

(z_k — реальное выходное состояние, z_k^* — требуемое выходное состояние). В рамках медленной адаптивной динамики матричные элементы связей T_{ij} изменяются за один акт обучения при следовании градиентной стратегии на величину

$$\dot{T}_{ij} \sim -\text{grad } E(z).$$

После большого числа фаз обучения (10^2 – 10^4) система, как правило, научается с высокой достоверностью сопоставлять заданным состояниям на входном уровне требуемые состояния на выходном уровне. Наиболее впечатляющим из известных ныне примеров обучения является разработанная Сейновским и Розенбергом программа NETalk (Sejnowski, Rosenberg, 1987): после обучения эта система обретает способность правильно прочесть любой заданный (неизвестный) текст. Задача состоит в том, чтобы данному напечатанному слову поставить в соответствие оптический сигнал, свободный от помех и искажений. Автоматы Больцмана после длительной тренировки демонстрируют такую же способность читать, какой обладают школьники 2-го или 3-го класса. При этом автоматы Больцмана совершают те же ошибки, что и дети, например, при чтении редких слов и исключений. Другой эксперимент относится к эффекту повторений, перерывов в обучении и т. п. При возмущении 65% связей в сети способность к чтению утрачивается, но скорость повторного обучения оказывается более высокой, чем при первичном обучении. Как показывают имеющиеся эксперименты, автоматы Больцмана удобны для проверки психологических гипотез. Область применения сетевых автоматов далеко не исчерпана.

Рассмотрим теперь возможные или уже существующие реализации обсуждаемых моделей в аппаратном оформлении компьютеров. Цель исследований, проводимых в этом направлении, состоит в том, чтобы найти наиболее простую и изящную физическую реализацию сетей, доказавших свои преимущества на уровне моделирования, разумеется, с учетом возможностей современной техники (VLSI-технологии). Мы уже упоминали об одном таком специальном элементе — чипе, решающем задачу коммивояжера, т. е. чипе, связанном с нелинейной RC -цепью (Alek-sander, 1965, 1988). Частный вариант представляет собой схема, составленная из RC -цепей (рис. 12.14) (Hopfield, Tank, 1986). Каждый элемент имеет несколько входов и один выход, напоминая тем самым структуру нейрона. Правила Кирхгофа для i -узла цепи позволяют записать уравнение

$$C_i \frac{du_i}{dt} + \frac{u_i}{R_i} - \sum_j T_{ij} V_j = I_i,$$

где $T_{ij} V_j$ — токи на входе синапса, I_i — ток на выходе. Зависимость между напряжением u_i в RC -контуре и выходным напряжением определяется отображением с нелинейной характеристикой

$$V_i = g(u_i),$$

где g — функция типа сигма-функции. Тем самым RC -контур обретает способность быть активным (возбуждаться) или блокировать прохождение импульса. И в этом

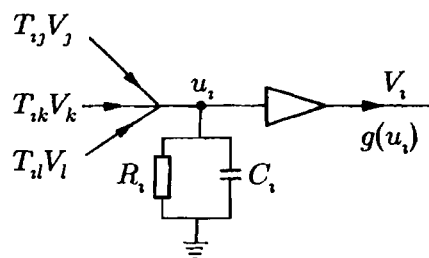


Рис. 12.14. Нелинейный RC -контур, из которого можно построить сеть, действующую аналогично нейронной сети

случае аналогия с нейроном вполне очевидна. Можно показать, что цепь, представленная на рис. 12.14, обладает функцией Гамильтона

$$H = -\frac{1}{2} \sum_{ij} T_{ij} V_i V_j - \sum_i I_i V_i + \sum_i \frac{1}{R_i} \int_0^{V_i} g^{-1}(V) dV.$$

При фиксированной схеме эта система позволяет легко реализовать градиентную стратегию, так как выполняется неравенство:

$$\frac{dH}{dt} \leq 0$$

(H — функция Ляпунова системы при $T_{ij} = T_{ji}$). Стратегия разработки физических элементов состоит в следующем: сначала с помощью компьютерной модели следует найти «подходящую» синаптическую матрицу, а затем нанести проводящие дорожки на кремневый чип. В силу неизменяемой конфигурации проводников такая RC -цепь может решать только одну задачу, например, распознавание одного класса объектов. Потенциальные приложения такой подход, по-видимому, найдет в массовом контроле дешевых продуктов и, возможно, в военной области.

Недостатком сетей, реализуемых с помощью кремневой технологии, следует считать отсутствие возможности варьировать конфигурацию проводников; синаптическая матрица определяется после фазы апробирования. Существенно большей гибкостью обладают оптические реализации, которые уже имеются по крайней мере на стадии испытания несовершенных образцов. При этом роль синаптической матрицы переходит к объемной или плоскостной голограмме. Тем самым достигается высокая гибкость и вариабельность. В настоящее время проводятся испытания систем двух основных типов (Gibbs, 1986).

1. В качестве узлов используются прозрачные круглые окна в экране, за которым могут располагаться нелинейные оптические элементы. С помощью полупрозрачного зеркала лучи снова направляются на экран, причем с помощью синаптической матрицы нацеливая их на другие окна. После прохождения через экран лучи рассеиваются на объемной голограмме. Принципиальная схема установок изображена на рис. 12.15.

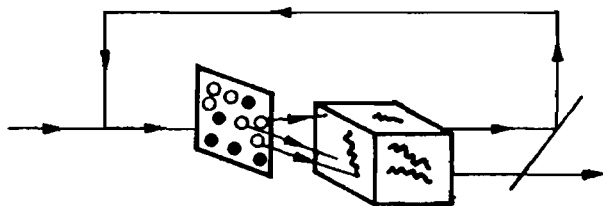


Рис. 12.15. Схема оптической нелинейной сети, состоящей из бистабильных шаблонов и объемной голограммы, моделирующей синаптическую матрицу

2. В качестве узлов служат LED из арсенида галлия и детекторы, наносимые на чип с помощью VLSI-технологии. Синаптическая матрица реализована с помощью расположенного внизу зеркала и плоской голограммы (рис. 12.16).

Вряд ли приходится сомневаться в том, что оптические «нейронные» сети представляют собой особенно изящное решение проблемы. Однако первоначальная эйфория по поводу реализуемости цифрового оптического компьютера сменилась с 70-х годов весьма трезвой оценкой. Однако не следует упускать из виду, что в последние несколько лет появились весьма эффективные оптические элементы (Gibbs, 1986). Ведущие специалисты в области оптической обработки информации

предполагают, что большие возможности открываются перед оптическими компьютерами при параллельной обработке информации. В литературе обсуждаются прототипы оптических сетевых компьютеров.

3. Третье направление реализации сетей связано с предложенными Конрадом, Кирби и Хонгом ферментативными нейронами (Kirby, Conrad, 1984, 1986). Основная идея этого подхода состоит в том, чтобы нанести на искусственную мембрану светочувствительные молекулы (родопсин). В тесной связи с этим подходом находятся оптохимические сети, экспериментально исследованные Кунертом и сотрудниками. Если отвлечься от совершенно иной химической структуры, используемой в этих сетях, то принцип оптохимического подхода в большей или меньшей степени соответствует ферментативным нейронам. В основе оптохимических сетей лежит открытый Кунертом (Kuhnert, 1986) светочувствительный вариант реакции Белоусова—Жаботинского. Светочувствительность используемого нового катализатора (рутения) позволяет управлять локальным состоянием реакции непосредственно с помощью оптического входного сигнала (рис. 12.17). Внутренняя структура оптохимической сети чрезвычайно проста, так как соседние ячейки сообщаются между собой только с помощью диффузионной связи. Установление связи может производиться, например, с помощью нелинейных электродов, положение которых варьируется адаптивно. Если возможность оптического управления ходом реакции была убедительно доказана Кунертом (Kuhnert, 1986), относительно адаптивной связи (рис. 12.18) пока имеются только теоретические исследования (Ebeling, 1986).

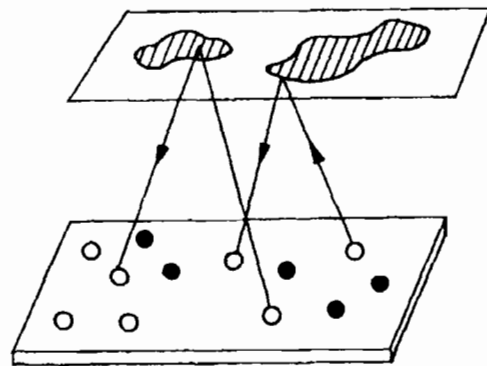


Рис. 12.16. Схема нелинейной оптической сети, состоящей из би-стабильных диодов на кремниевой пластине и голограммы с зеркальной поверхностью, моделирующей синаптическую матрицу

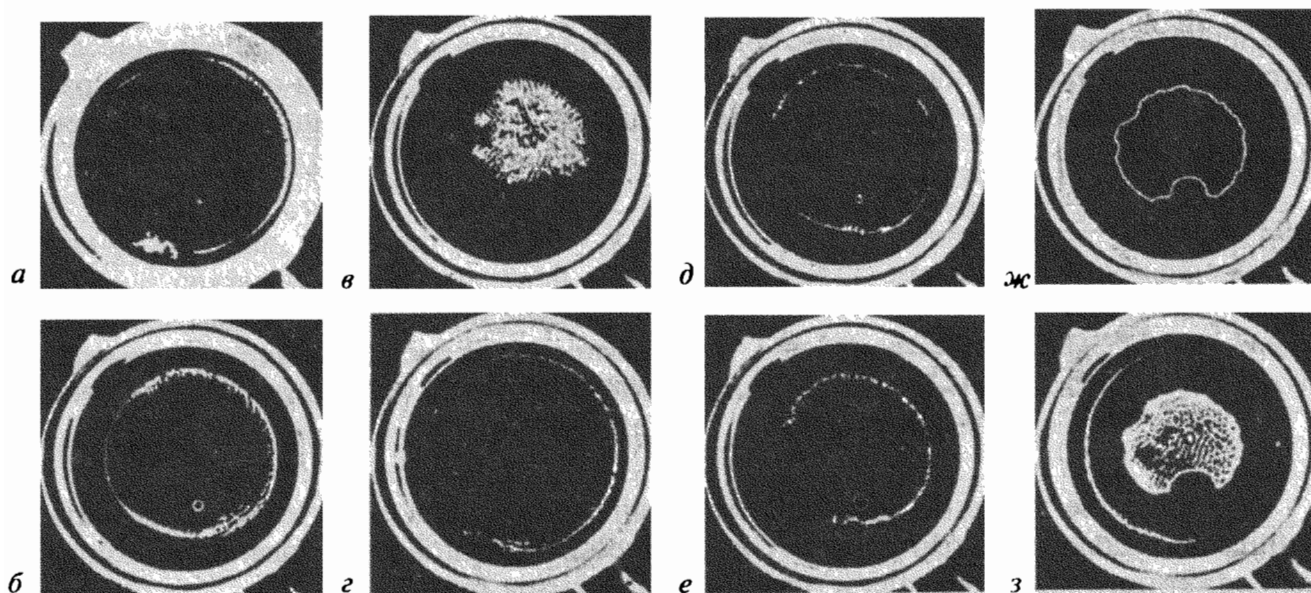


Рис. 12.17. Оптическая манипуляция светочувствительной реакцией Белоусова—Жаботинского—Кунерта с рутениевым катализатором: (а)—(в) первый цикл темновой реакции. Возникающая у края триггерная волна превращается в фазовую волну с осциллирующей средой вокруг нее; (г)—(е) повторение цикла темновой реакции; (ж) включение света переводит превращение в фазовую волну. Триггерная волна стабилизируется; (з) после выключения света превращение в фазовую волну продолжается. (С разрешения автора — Kuhnert, 1986.)



Рис. 12.18. Принцип оптической манипуляции реакции Белоусова—Жаботинского—Кунерта. Крестиками обозначены выходные электроды

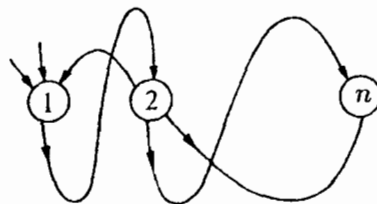


Рис. 12.19. Схема сети, состоящей из n связанных процессоров

Нельзя не упомянуть и о возможности реализации сложных сетей с помощью материалов, напоминающих по своим свойствам цеолиты (*Dress et al.*, 1982). В заключение скажем несколько слов и о многообещающем моделировании сетей с помощью процессов. Так, калифорнийская фирма TRW предлагает так называемый нейронный процессор. С 1987 г. на рынок выпущен процессор типа Марк IV, состоящий из восьми специальных процессоров (рис. 12.19), которые позволяют моделировать сеть с $1/4$ миллионов узлов и 5 миллионами соединений; 85% узлов и соединений приходится в большой RAM-памяти на долю синаптической матрицы. Современные нейронные процессоры позволяют моделировать все известные типы нейронных сетей от Хопфилда до Хинтона. Специализированный компьютер для моделирования сетей обходится дешевле (при прочих равных параметрах), чем моделирование на большом компьютере, и требует меньших затрат времени. Нейрокомпьютеры особенно удобны для адаптивного управления, решения проблемы разрешимости, обучения правил, обработки неточных сведений, в архивном и издательском деле, для поиска информации и массового контроля.

В заключение подчеркнем еще раз, что сетевые автоматы находятся сейчас еще на самом раннем — «младенческом» — этапе своего развития, поэтому сравнение их возможностей с возможностями современного компьютера было бы несправедливо. С нашей точки зрения существуют веские доводы, позволяющие считать новый тип обработки информации весьма перспективным. Кроме того, развитие автоматов нового типа по затратам сопоставимо с усовершенствованием компьютеров традиционного типа.

«Все это приводит к теории (вычислимости), гораздо менее жесткой и не похожей на формальную логику структур типа “все или ничего”. Новая теория носит в меньшей степени комбинаторный и в большей — аналитический характер. В действительности существуют многочисленные указания, позволяющие считать, что новая система формальной логики близка к дисциплине, имевшей в прошлом мало общего с логикой. Речь идет о термодинамике, прежде всего в той форме, которую ей придал Больцман, термодинамике, составлявшей часть теоретической физики, наиболее близкую к некоторым аспектам обработки и измерения информации» (Джон фон Нейман).

В тот симбиоз статистической термодинамики с информатикой, который предвидел фон Нейман, ответы на вопросы, затронутые в этом разделе, вносят решающий вклад.

Послесловие к русскому изданию

Когда в 1990 году наша книга «Физика процессов эволюции» вышла в свет на немецком языке, Германия переживала переломный период. Создавалось впечатление, что никто не считал возможным объединение двух немецких государств с совершенно различными общественными системами и возникновение существующей ныне единой Германии. В смысле синергетического метода, положенного в основу этой книги, происходившее тогда было кинетическим фазовым переходом, подобно тому, как при фазовых переходах в физике происходили большие флуктуации. Эволюционирующие системы, спектр которых простирается от космоса через биологическое разнообразие до социоэкономических систем, подчиняются определенным общим закономерностям, рассмотренным в нашей книге с точки зрения физики и синергетики. Эта новая область научных исследований в настоящее время весьма быстро развивается во всем мире, и ей посвящены многочисленные новые публикации. Часть наиболее важных работ указана в списке литературы, но, разумеется, учесть в тексте все результаты, полученные с 1990 года, не представлялось возможным. В целом применение к процессам эволюции методов нелинейной динамики и синергетики оказалось весьма плодотворным, и мы глубоко убеждены, что эта область и далее заслуживает большого внимания. Именно потому, что русская наука внесла столь фундаментальный вклад в развитие синергетического подхода к изучению процессов эволюции (упомянем лишь о фундаментальных работах незабвенного «первопроходца» М. В. Волькенштейна), авторам доставляет особое удовольствие, что отныне их книга становится доступной русским читателям. Мы выражаем признательность Ю. А. Данилову за большой труд по переводу и переработке текста книги и благодарны Ю. Л. Климонтовичу и Ю. М. Романовскому за многолетние дружеские связи и многочисленные полезные советы.

Авторы

Литература

- Альфен Г, Аррениус Дж Эволюция Солнечной системы М Мир, 1987
- Эйген М Самоорганизация материи и эволюция биологических макромолекул М Мир, 1973
- Амбарцумян В А, Кардашев А С, Троицкий В С (ред) Проблемы поиска жизни во Вселенной М Наука, 1986
- Аптер М А Кибернетика и эволюция М Мир, 1970
- Арнольд В И Обыкновенные дифференциальные уравнения 3-е изд М, 1984
- Арнольд В И Теория катастроф М Наука, 1990
- Арнольд В И, Варченко А Н, Гуссеин-Заде С М Особенности дифференцируемых отображений М Наука, 1981
- Балкарей Ю И, Евтихов М Г, Элисон М И Локализованные стационарные состояния в распределенной волновой среде // Микроэлектроника 1979 Т 8, № 6 С 493
- Балкарей Ю И, Евтихов М Г, Элисон М И Локализованные стационарные состояния в двумерных диффузионных волнах // Микроэлектроника 1980 Т 9, № 2 С 144
- Балкарей Ю И, Евтихов М Г, Элисон М И Исследование возникновения самоструктуризации в активных диффузионных средах // Микроэлектроника 1981 Т 10, № 1 С 78
- Балкарей Ю И, Захарова А А, Элисон М И О возможности создания метастабильной активно распределенной системы // Микроэлектроника 1981 Т 10, № 3 С 260
- Балкарей Ю И, Элисон М И Метастабильные автоволновые системы // Микроэлектроника 1979 Т 8, № 5 С 428
- Балкарей Ю И, Элисон М И Перспективы автоволновой электроники // Электронная промышленность 1983 Т 8 С 11
- Балкарей Ю И, Элисон М И, Никулин М С Автоволновая среда с памятью // Микроэлектроника 1977 Т 6, № 2 С 152
- Барьяхтар В Г и др (ред) Проблемы теоретической физики Киев Наукова думка, 1986
- Бауэр Э С Теоретическая биология М —Л ВИЭМ, 1935
- Будыко М И Энергетический баланс Земли Л Гидрометеиздат, 1978
- Букатова И Л, Элинсон М И, Шаров А М Оптоэлектронный подход для наблюдения двумерных изображений // Микроэлектроника 1978 Т 7, № 5 С 397
- Вайнберг С Первые три минуты М Энергоатомиздат, 1981
- Васильев В А, Романовский Ю М, Чернавский Д С, Яхно В Г Автоволновые процессы М Наука, 1987
- Веденов А А, Левченко Е Б Об одном классе нелинейных систем с памятью // Письма ЖЭТФ 1985 Т 11 С 328
- Велихов Е П, Ковалёв А С, Рахимов А Т Физические явления в лампах газового разряда М Наука, 1987
- Гаркунов А И (ред) Избирательный перенос в сильнонагруженных элементах с трением М, 1982
- Хессин Р Б Непостоянство генома М Наука, 1985
- Грехов М Т (ред) Автоволновые процессы в системах с диффузией Горький Академия, 1981
- Гуревич А В, Минтц В Г Локализованные волны в неоднородных средах // УФН 1984 Т 142 С 61
- Гуревич Л Е, Чернин А Д Возникновение звезд и галактик М Наука, 1987
- Дарвин Ч Собрание сочинений М —Л Изд-во АН СССР, 1938
- Деборин Г, Павловская Т, Доуз К, Фокс С Происхождение жизни и эволюционная биохимия М Наука, 1975
- Долгопальский А Какие языки родственны с европейскими? // Наука 1971/1972 С 106—119 М Знание, 1972
- Евтихов М Г, Балкарей Ю И, Никулин М Г, Элисон М И Математическое моделирование динамической фазовой памяти автоволновой системы // Микроэлектроника 1978 Т 7, № 5 С 421
- Жаботинский А М Концентрации—автоколебания М Наука, 1974
- Жарков В Н Внутренняя структура Земли и планет М Наука, 1983
- Звонкин А К, Левин Л А О комплексной массе // УФН 1970 Т 26 С 85
- Зельдович Я Б Об импульсах, приводящих к разделенным областям // Доклады АН СССР 1983 Т 270 С 1369
- Зельдович Я Б, Новиков И Д Релятивистская астрофизика М Наука, 1967
- Зельдович Я Б, Новиков И Д Структура и эволюция Вселенной М Наука, 1975
- Зотин А И Термодинамический подход к проблемам развития, роста и старения М Наука, 1974
- Иваницкии Г Р, Кринский В И, Селков Е Е Математическая биофизика клетки М Наука, 1978
- Кадомяцев Б Б Коллективные феномены в плазме М Наука, 1976
- Калнин А А Динамические структуры и нелинейные спектральные трансформации в реакциях структурного взаимодействия // Изв Акад наук СССР Сер Биофизика 1982 Т 27 С 509, 1984 Т 29 С 78
- Калнин А А, Лухинин В В Структурное программирование в микроэлектронике Л РИО ЛЕТИ, 1981
- Калнин А А, Лухинин В В Физические основы функциональной электроники в самоорганизующихся средах // Электронная промышленность 1983 Т 8 С 6

- Камшилов М М Жизнь на Земле М Наука, 1974
- Кастлер Г Возникновение биологической организации М Мир, 1967
- Кернер В С, Осипов В В Прерывание тока в рабочей области вольт амперных характеристик с отрицательным и положительным дифференциальным напряжением // Микроэлектроника 1974 Т 3, № 1
- Кернер В С, Осипов В В Нелинейная теория неизотермических прерываний тока в транзисторных структурах // Микроэлектроника 1977 Т 6, № 4 С 337
- Кернер В С, Осипов В В Стационарные и бегущие диссипативные структуры в активных кинетических средах // Микроэлектроника 1981 Т 10, № 5 С 407
- Климонтович Ю Л Кинетическая теория неидеального газа и неидеальной плазмы М Наука, 1975
- Климонтович Ю Л Статистическая теория открытых систем Т 1, 2 М Янус, 1995, 1999
- Колмогоров А Н Три подхода к определению понятия «количество информации» // Проблемы передачи информации 1965 Т 1
- Корнберг А Синтез ДНК М Мир, 1977
- Кринский В А, Михайлов А С Автоволны М Знание, 1984
- Лампрехт И, Зотин А И (ред) Термодинамика биологических процессов М Наука, 1976
- Ландау Л Д, Лифшиц Е М Курс теоретической физики Т III Квантовая механика М Наука, 1989
- Ландау Л Д, Лифшиц Е М Курс теоретической физики Т IV Гидродинамика М Наука, 1988
- Ландау Л Д, Лифшиц Е М Курс теоретической физики Т V Статистическая физика М Наука, 1995
- Лифшиц И М, Градескул С А, Пастур Л А Введение в теорию неорганизованных систем М Наука, 1982
- Лопушанская А И Термодинамика необратимых процессов М Наука, 1987
- Миркин Б Г, Родин С Н Графы и гены М Наука, 1977
- Михайлов А С Неравновесные фазовые переходы в биологических популяциях // Доклады АН СССР 1978 Т 243 С 786
- Монин А С, Шишков Ю С История климата Л Гидрометеиздат, 1979
- Монин А С Ранняя геологическая история Земли М Недра, 1987
- Николис Г, Пригожин И Самоорганизация в неравновесных системах М Мир, 1979
- Николис Г, Пригожин И Познание сложного М Мир, 1990
- Николис Дж Динамика иерархических систем М Мир, 1987
- Новиков И Д Эволюция Вселенной М Наука, 1979 (2-е изд 1983)
- Опарин А И Жизнь ее происхождение, сущность и развитие М Изд во АН СССР, 1960
- Паитген Х О, Рихтер П Х Красота фракталов М Мир, 1993
- Паташинский А С, Шумило Б И ЖЭТФ 1979 Т 77 С 1417
- Пиблс П Крупномасштабная структура Вселенной М Мир, 1983
- Полуэктов Р А (ред) Динамическая теория биологических популяций М Наука, 1974
- Полуэктов Р А, Рык Ю А, Швытов И А Динамические модели экологических систем Л Гидрометеиздат, 1980
- Постон Т, Стюарт И Теория катастроф и ее приложения М Мир, 1980
- Пригожин И От существующего к возникающему М Наука, 1985
- Раизер Ю А Основы современной физики плазмы газовых разрядов М Наука, 1980
- Ратнер В А Молекулярная генетика принципы и механизмы Наука Новосибирск, 1983
- Ратнер В А Методологические аспекты самовоспроизводства в молекулярно генетических системах управления // Методологические вопросы теоретической биологии Пушино АН СССР, 1986
- Ратнер В А, Шанин В В Самовоспроизводство молекулярно генетических систем управления // Журнал общей биологии 1983 Т 44 С 51
- Редько В Г Оценка скорости эволюции в моделях Эйгена и Куна // Биофизика 1986 Т 31 С 511
- Розенблатт Ф Принципы нейродинамики М Мир, 1965
- Розенталь И Л Геометрия динамика и Вселенная М Наука, 1987
- Романовский Ю М, Степанова Н В, Чернавский Д Математическая биофизика М Наука, 1984
- Садовский М В Электрон в случайном поле Теория фазовых переходов и нелинейные решения с конечным действием // Физика твердого тела 1979 Т 21 С 743
- Сватков Н М Основы планетарного географического прогноза М Мысль, 1974
- Саврежев Ю М, Логофет Д О Стабильность биологических систем М Наука, 1978
- Сергин В Я, Сергин С Я Системный анализ проблемы больших колебаний климата и оледенения земли Л Гидрометеиздат, 1978
- Стратонович Р Л Теория информации М Сов радио, 1975
- Стратонович Р Л Нелинейная термодинамика необратимых процессов М Наука, 1985
- Тимофеев Рессовский Н В, Воронцов Н Н, Яблоков А Н Краткий очерк эволюционной теории М Наука, 1969
- Феигельман М В О распространении плоского фронта в неоднородной среде // ЖЭТФ 1983 Т 85 С 1851
- Ферстер Г фон, Зонф Г В Принципы самоорганизации М Мир, 1964
- Волькенштейн М В О количестве и значении информации в биологии // Журнал общей биологии 1976 Т 37 С 483

- Волькенштейн М В Молекулярная биофизика М Наука, 1978
- Волькенштейн М В Мутации и значение информации // Журнал теоретической биологии 1979 Т 80 С 155
- Волькенштейн М В Биофизика М Наука, 1981
- Волькенштейн М В Сущность биологической эволюции // УФН 1984 Т 143 С 429
- Волькенштейн М В Пунктуальность, неадаптируемость, нейтральность и эволюция // Изв АН СССР 1986 Т 3 С 325
- Волькенштейн М В, Чернавский Д С Физические аспекты применения теории информации в биологии // Изв АН СССР Сер Биология 1979 Т 4 С 531
- Фомин П И Гравитационная неустойчивость вакуума и квантового происхождения Вселенной // Проблемы теоретической физики Киев Наукова думка, 1986
- Хаазе Р Термодинамика необратимых процессов М Мир, 1967
- Хакен Г Синергетика М Мир, 1980
- Хинчин А Я Математические основы статистической механики М, 1943
- Хомский Н Синтаксические структуры // Новое в лингвистике Вып II М Издательство иностранной литературы, 1962
- Шмальгаузен И И Факторы эволюции М Наука, 1968
- Шкловский И С Звезды Их возникновение, жизнь и смерть М Наука, 1984
- Шналь С Е Физикохимические факторы биологической эволюции М Наука, 1979
- Шредингер Э Что такое жизнь с точки зрения физика М Мир, 1962
- Эбелинг В Образование структур при необратимых процессах М Мир, 1979
- Эбелинг В, Зандер С Роли саморепродукции в эволюционном процессе // Вопросы современной биологии и биофизики Пушкино АН СССР, 1986
- Эбелинг В, Крафт В Д, Крем Д Теория связанных состояний и ионизационное равновесие в плазме и твердом теле М Мир, 1979
- Эйген М Самоорганизация материи и эволюция биологических макромолекул М Мир, 1973
- Эйген М, Винклер Р Игра жизни М Наука, 1979
- Эйген М, Шустер П Гиперцикл Принципы самоорганизации макромолекул М Мир, 1982
- Ефетов А, Ларкин А Волны плотности заряда в стохастических потенциалах // ЖЭТФ 1977 Т 72 С 2350
- Яглом А М, Яглом И М Вероятность и информация М Наука, 1973
- Ablay P Optimieren mit Evolutionsstrategien *Spektrum der Wissenschaft*, Juli 1987, 104
- Abramson H Information Theory and Coding McGraw-Hill, New York 1963
- Ackley D H, Hinton G E, Sejnowski T J A Learning Algorithm for Boltzmann Machines *Cognitive Science* 9 (1985) 147–169
- Ahlers G, Behringer R P The Rayleigh—Benard Instability and the Evolution of Turbulence *Progr Theor Phys* (Japan), Suppl 64 (1978) 186
- Aleksander I Fused Electronic Circuit which Learn by Example *Electronic Letters* 1 (1965) 6
- Aleksander I The Logic of Connectionist Systems *IEEE Computer* (1988)
- Allen P M Darwinian Evolution and a Predator-Prey Ecology *Bull Math Biol* 37 (1975) 389
- Allen P M Evolution, Population Dynamics, and Stability *Proc Nat Acad Sci USA* 73 (1976) 665
- Allen P M, Cahn J W A Microscopic Theory of Antiphase Boundary Motion and its Application to Antiphase Domain Coarsening *Acta Metall* 27 (1979) 1085
- Allen P M, Ebeling W On the Stochastic Description of a Predator-Prey Ecology *Biosystems* 16 (1983) 113.
- Alpher R A, Herman R G, Gamov G The Early Universe *Phys Rev* 74 (1948) 1198
- Amari S, Arbib M A Competition and Cooperation in Neural Nets Springer-Verlag, Berlin 1982
- Ambarzumyan V A Probleme der modernen Kosmogonie Akademie-Verlag, Berlin 1976
- Amot D, Gutfreund H, Sompolinsky H Storing Infinite Numbers of Patterns in a Spin-Glass Model of Neural Networks *Phys Rev Lett* 55 (1985) 1530
- Amot D, Gutfreund H, Sompolinsky H Statistical Mechanics of Neural Networks Near Saturation *Ann Physik* (N Y) 173 (1987) 30
- Anderson P W Absence of Diffusion in Certain Random Lattices *Phys Rev* 109 (1958) 1492
- Anderson P W More is Different — Broken Symmetry and the Nature of the Hierarchical Structure of Science *Science* 177 (1972) 393
- Anderson P W Local Moments and Localized States *Rev Mod Phys* 50 (1978) 191
- Anderson P W Suggest Model for Prebiotic Evolution The Use of Chaos *Proc Nat Acad Sci USA* 80 (1983) 3386
- Anderson P W Basic Notions in Condensed Matter Physics Cummings, Menlo-Park, 1984
- Andresen B, Berry R, Nitzan A, Salamon P Thermodynamics in Finite Time I Step-Carnot-Cycle *Phys Rev A* 15 (1977) 2086
- Atlan H L'Organisation Biologique et la Theorie de l'Information Herman et Cie, Paris 1972
- Babloyantz A Selforganization in Multiple Unit Systems In *Dynamics of Synergetic Systems* (Ed Haken H) Springer-Verlag, Berlin/Heidelberg/New York 1980

- Babloyantz A, Nicolis G Chemical Instabilities and Multiple Steady States in Monod-Jacob-Type Models *J Theor Biol* **34** (1972) 185
- Baras F, et al Stochastic Theory of Adiabatic Explosion Preprint ULB, Brussel 1983
- Barrow J, Silk J The Structure of the Early Universe *Sci American*, April 1980, 98
- Bartholomay A F *Bull Math Biophys* **20** (1958) 97 175, **21** (1959) 363
- Bausch R, Dohm V, Janssen Z K, Zia R K P Critical Dynamics of an Interface in $1 + \epsilon$ Dimensions *Phys Rev Lett* **47** (1981) 1837
- Benson S W The Foundation of Chemical Kinetics McGraw-Hill Book Company, New York/Toronto/London 1960
- Bernal I D Die Wissenschaft in der Geschichte VEB Deutscher Verlag der Wissenschaften, Berlin 1967
- Bertalanffy L von Theoretische Biologie Verlag von Gebrüder Bornträger, Berlin-Zehlendorf 1942
- Bertalanffy L von Biophysik des Fließgleichgewichts Vieweg, Braunschweig 1956
- Bertalanffy L von Principles and Theory of Growth In *Fundamental Aspects of Normal and Malignant Growth* (Ed W W Nowinski) Elsevier, Amsterdam 1960
- Bertalanffy L von, Beier W, Laue R Biophysik des Fließgleichgewichtes Akademie-Verlag, Berlin 1977
- Beyer W A, Stein M L, Smith T F, Ulam S M A Molecular Sequence Metric and Evolutionary Tree *Math Biosciences* **19** (1974) 9
- Binder K, Young A P Spin Glasses Experimental Facts, Theoretical Concepts and Open Problems *Rev Mod Phys* **58** (1986) 801
- Bittrich H J Leitfaden der chemischen Thermodynamik VEB Deutscher Verlag der Wissenschaften, Berlin 1971
- Bittrich H J, Haberland D, Just G Leitfaden der chemischen Kinetik VEB Deutscher Verlag der Wissenschaften, Berlin 1973
- Blome H J, Priester W Urknall und Evolution des Kosmos *Naturwiss* **71** (1984) 515
- Blum H F *American Scientist* **49** (1962) 474
- Blum H F Times Arrow and Evolution Princeton University Press, Princeton 1968
- Blumenfeld L A Probleme der molekularen Biophysik Akademie-Verlag, Berlin 1977
- Boseniuk T, Ebeling W, Engel A Boltzmann and Darwin Strategies in Complex Optimization *Phys Lett* **125** (1987) 307
- Bourbaki N Espaces Vectoriels Topologiques Paris 1966
- Bray A J, Moore M A Metastable States in Spin Glasses *J Phys C* **33** (1980) L 469
- Bremermann H J A Method of Unconstrained Global Optimization *Math Biosciences* **9** (1970) 1
- Bremermann H J Complexity of Automata, Brain and Behaviour In *Lecture Notes in Bio mathematics*, Vol 4 (Ed M Conrad) Springer Verlag, Berlin/Heidelberg/New York 1974
- Bresch C Zwischenstufe Leben Evolution ohne Ziel Piper Verlag, München/Zürich 1977
- Bruckner E Die intensiv erweiterte Reproduktion gesellschaftlicher Arbeitsmöglichkeiten — ein einfaches Modell *Rostocker Physikalische Manuskripte* **3/II** (1978) 55
- Bruckner E, Ebeling W, Scharnhorst A The Application of Evolution Models in Scientometrics *Scientometrics* **18** (1990) 21
- Bruckner E, Ebeling W, Scharnhorst A The Dynamics of Instabilities *System Dynamics Review* **5** (1989) 176
- Callen E, Shapiro D A Theory of Social Imitation *Physics Today* **27**, July (1974)
- Calvin M Chemical Evolution Clarendon Press, Oxford 1969
- Cardy J L Electron Localization in Disordered Systems and Classical Solutions in Ginzburg—Landau Field Theory *J Phys C* **11** (1978) L 321
- Cartanu D Synchronization Phenomena in Reaction Diffusion System BSB B G Teubner Verlagsgesellschaft, Leipzig 1986
- Chaitin G Randomness and Mathematical Proof *Sci American* June 1975, 47
- Chandrasekhar S Hydrodynamic and Hydromagnetic Stability Clarendon Press, Oxford 1961
- Chomsky N Topics in the Theory of Generative Grammar Mouton, Den Haag 1966
- Chomsky N Studies on Semantics in Generative Grammar Mouton, Den Haag 1972
- Chomsky N Die formale Natur der Sprache In *Biologische Grundlagen der Sprache* (Hrsg Lenneberg, E H) Suhrkamp-Verlag, Frankfurt (Main) 1972
- Coleman S Fate of the False Vacuum Semiclassical Theory *Phys Rev D* **15** (1977) 2929
- Conrad M Thermodynamic Extremal Principles in Evolution *Biophysik* **9** (1973) 191
- Conrad M The Thermodynamic Meaning of Ecological Efficiency *Am Nat* **111** (1977) 99
- Conrad M Evolutionary Adaptability of Biological Macromolecules *J Mol Evol* **10** (1977) 87
- Conrad M Evolution of the Adaptive Landscape *Lect Notes in Biomathematics* **21** (1978) 147
- Conrad M Adaptability The Significance of Variability from Molecule of Ecosystem Plenum Press, New York 1983
- Conrad M Macroscopic macroscopic Interface in Biological Information Processing *Biosystems* **16** (1984) 345
- Conrad M On Design Principles for a Molecular Computer *Communications of the ACM* **28** (1985) 464
- Conrad M The Lure of Molecular Computing *IEEE Spectrum* **11** (1986) 55
- Crick F C, Mitchison G The Function of Dream Sleep *Nature* **304** (1983) 111
- Crow J F, Kimura M An Introduction to Population Genetics Theory Harper and Row, New York 1970

- Czajkowski G, Ebeling W Non-Equilibrium Phase Transitions in Enzyme Reaction Systems *J Non-Equilibrium Thermodynamics* 2 (1977) 1
- Damaschun G, Buder E Präbiologische Evolution von Biopolymerkonfigurationen *Biol Rundschau* 16 (1978) 1
- Dautcourt G Relativistische Astrophysik (WTB) Akademie-Verlag, Berlin 1976
- Davidson R C Methods in Nonlinear Plasma Theory Academic Press, New York/London 1972
- Dawkins R Das egoistische Gen Springer-Verlag, Berlin/Heidelberg/New York 1978
- Dayhoff M O Computer Analysis of Protein Evolution *Sci American*, July 1969, 265
- Dayhoff M O Atlas of Protein Sequences and Structure The National Biomedical Research Foundation Silver Spring 1969
- De Dominicis C, Grebay M, Garel T, Orland H Number of Metastable States in Spin Glasses *H Phys (Paris)* 41 (1980) 923
- De Groot S R, Mazur P Non-Equilibrium Thermodynamics North Holland, Amsterdam 1962
- Decker P The Origin of Molecular Asymmetry through the Amplification of "Stochastic in Formation" (Noise) in Bioids *J Mol Evol* 4 (1974) 49
- Decker P Evolution in Bioids *Origins of Life* 6 (1975) 211
- Dehn H Effect of Bromine Derivative of Malonic Acid on the Oscillating Reaction of Malonic Acid Cerium Ions and Bromate *Nature* 213 (1967) 589
- Demetrius L Statistical Mechanics and Population Biology *J of Stat Phys* 30 (1983) 709
- Demetrius L An Extremal Principle of Macromolecular Evolution *Physica Scripta* 36 (1987) 693
- Demetrius L, Schuster P, Stigund K Polynucleotide Evolution and Branching Processes *Bull Math Biol* 47 (1985) 239
- Dickerson R E The Structure and History of an Ancient Protein *Sci American* April 1972, 58
- Diehl H W, Kroll D M, Wagner H Interface Solution of the Ginzburg—Landau—Wilson Model at Low Temperature *Z Phys B* 36 (1980) 329
- Dolgopolski A Boreisch — Ursprache Eurasiens? *Ideen des exakten Wissens* 1 (1973) 19 *Bild der Wissenschaft* (1973) 1140
- Dolgov A D, Zeldovich Ya B Cosmology and Elementary Particles *Rev Mod Phys* 53 (1981) 1
- Dorschner J Die Entstehung des Planetensystems *Wissenschaft und Fortschritt* 26 (1976) 275
- Dose K, Rauchfuss H Chemische Evolution und der Ursprung lebender Systeme Wissenschaftliche Verlagsgesellschaft, Stuttgart 1975
- Dosi G, et al Technical Change and Economic Theory Pinter, London 1988
- Dress A, Hendricks H, Küppers G (Hrsg.) Selbstorganisation Die Entstehung von Ordnung in Natur und Gesellschaft Piper Verlag, Munchen/Zürich 1986
- Dress A, Jaeger N I, Plath P J Zur Dynamik idealer Speicher Ein einfaches mathematisches Modell *Theor Chim Acta* 61 (1982) 437
- Dyson F J Time without End Physics and Biology in an Open Universe *Rev Mod Phys* 51 (1979) 447
- Ebeling W Ein stochastisches Modell für das Wirken des Marxschen Wertgesetzes im Kapitalismus der freien Konkurrenz *Rostocker Physikalische Manuskripte* 3/1 (1978) 47
- Ebeling W Stochastic Model of Free Competition in Economic Systems Preprint Universidad Veracruzana, Xalapa Ver, Mexico 1979, Handbook of Systems Analysis
- Ebeling W Pattern Dynamics and Optimization by Reaction Diffusion Systems *J of Stat Phys* 45 (1986) 891
- Ebeling W, Molgedey L, Reimann A Stochastic Models of Search Dynamics *Physica A* 287 (2000) 599–612
- Ebeling W, Czajkowski G Phase Separation in Bistable Enzyme Reaction Systems *Studia biophysica* 60 (1976) 201
- Ebeling W, Engel A Models of Evolutionary Systems and Their Application to Optimization Problems *Syst Anal Model Simul* 3 (1986) 377
- Ebeling W, Feistel R Zur Kinetik molekularer Replikationsprozesse mit Selektionscharakter *Studia biophysica* 46 (1974) 183
- Ebeling W, Feistel R Zur nichtlinearen Kinetik von homogenen Konkurrenzreaktionen *Z Phys Chemie (Leipzig)* 257 (1976) 705
- Ebeling W, Feistel R Physikalische Aspekte der qualitativen Sprünge bei Entwicklungsprozessen II Modellsysteme *Wiss Z Humboldt Universität Berlin*, GSR 25 (1976) 25
- Ebeling W, Feistel R Stochastic Theory of Molecular Replication Processes with Selection Character *Ann Phys (Leipzig)* 34 (1977) 91
- Ebeling W, Feistel R Physik der Selbstorganisation und Evolution Akademie-Verlag, Berlin 1982 2 Auflage 1986
- Ebeling W, Feistel R Continuous Lotka—Volterra Models for Evolutionary Models In *Lotka—Volterra Approach to Cooperation and Competition in Dynamic Systems* (Eds W Ebeling, M Peschel) Akademie-Verlag, Berlin 1985
- Ebeling W, Freund J, Schweitzer F Entropie und Information Teubner, Stuttgart, Leipzig 1998
- Ebeling W, Jiménez-Montaño M A On Grammars Complexity and Information Measures of Biological Macromolecules *Math Biosci* 52 (1980) 53
- Ebeling W, Klimontovich Yu L Selforganization and Turbulence in Liquids BSB B G Teubner Verlagsgesellschaft, Leipzig 1984
- Ebeling W, Mahnke R On the Complexity of Cytochrome c and the Influence of the Genetic Code *Studia biophysica* 71 (1978) 173
- Ebeling W, Malchow H Bifurkations in a Bistable Reaction System *Ann Phys (Leipzig)* 36 (1979) 121 40 (1983) 68, 151
- Ebeling W, Schimansky Geier L Stochastic Dynamics of a Bistable Reaction System *Physica* 98 A (1979) 587, *Ann Phys (Leipzig)* 40 (1983) 10, 227

- Gatlin L L Information Theory and the Living System Columbia Univ Press, New York 1972
- Geissler E, Lippert E, Nitschmann J, Thomas-Petersen G (Hrsg.) Kleine Enzyklopädie Leben VEB Bibliographisches Institut, Leipzig 1976
- Gibbs H, et al (Eds) Optical Bistability III Springer-Verlag, Berlin/Heidelberg/New York 1986
- Glansdorff P, Prigogine I Thermodynamics of Structure, Stability and Fluctuations Wiley-Interscience Publishers, New York 1971
- Glaser R Biophysik VEB Gustav Fischer-Verlag, Jena 1986
- Goel N S, Maitra S C, Montroll E W On the Volterra and Other Nonlinear Models of Interacting Populations *Rev Mod Phys* 43 (1971) 231
- Gould S J Ontogeny and Phylogeny Harvard University Press, Cambridge/Mass 1977
- Greenwood P J, Harvey P S, Slatkin M (Eds) Evolution Essays in Honour of John Maynard Smith London u a Cambridge Univ Press 1985
- Gunther E Grundriss der Genetik VEB Gustav Fischer Verlag, Jena 1978
- Guth A 10^{-32} Seconds After the Big Bang Vortrag auf dem Second Monond Astrophysics Meeting Les Arcs/Savoie 1983
- Guth A H, Steinhart P J Das inflationäre Universum *Spektrum der Wissenschaft* Juli 1984
- Gyarmati I Non-Equilibrium Thermodynamics Field Theory and Variational Principles Springer-Verlag, Berlin 1970
- Haddon R, Lomola A The Organic Computer Can Microchips be Built by Bacteria? *Sciences N Y* 23 (1983) No 3, 40
- Haken H (Ed) Synergetics Teubner, Stuttgart 1973, Dynamics of Synergetic Systems Springer-Verlag, Berlin/Heidelberg/New York 1980
- Haken H Cooperative Effects in Systems Far from Equilibrium and in Nonphysical Systems *Rev Mod Phys* 47 (1975) 67
- Haken H Erfolgsgeheimnisse der Natur Deutsche Verlagsanstalt, Stuttgart 1981
- Haken H (Ed) Chaos and Order in Nature Springer-Verlag, Berlin/Heidelberg/New York 1981
- Haken H Advanced Synergetics Springer-Verlag, Berlin/Heidelberg/New York 1983
- Haken H (Ed) Complex Systems-Operational Approaches in Neurobiology Physics and Computers Springer-Verlag, Berlin 1986
- Hassenstein B Biologische Kybernetik Gustav Fischer Verlag, Heidelberg 1970
- Heidmann W, Decker P, Pohlmann R Cooperative Kinetics in the Formaldehyd Bioid *Proc 2nd Issol Meeting*, Kyoto 1977
- Hemmen J L van, Morgenstern I (Eds) Proceedings of the Heidelberg Colloquium on Glassy Dynamics 1986 *Lecture Notes in Physics* 275 (1987) Springer-Verlag, Berlin 1987
- Hiernaux J, Babloyantz A Dissipation in Embryogenesis *J Nonequil Thermodyn* 1 (1976) 33
- Hinton G E, Anderson J A Parallel Models in Associative Memories Erlbaum, Hillsdale/N J 1981
- Hörz H Marxistische Philosophie und Naturwissenschaften Akademie-Verlag Berlin 1974
- Hörz H, Wessel K F Philosophische Entwicklungstheorie VEB Deutscher Verlag der Wissenschaften, Berlin 1983
- Hörz H, Wessel K F (Hrsg.) Struktur—Bewegung—Entwicklung VEB Deutscher Verlag der Wissenschaften, Berlin 1985
- Hofbauer J, Sigmund K Evolutionstheorie und dynamische Systeme Parey-Verlag, Hamburg 1984
- Hofbauer J, Schuster P, Sigmund K Competition and Cooperation in Catalytic Selfreplication *J Math Biol* 11 (1981) 155
- Hoffmann G W, et al A Teachable Network Based on an Unorthodox Neuron *Physica* 220 (1986) 233
- Hofmann E Dynamische Biochemie Teile I bis III, 3 bearbeitete und erweiterte Auflage Akademie-Verlag, Berlin 1979 (Teile I und II), 1980 (Teil III)
- Hogg T, Hubermann A Parallel Computing Structures *J Stat Phys* 41 (1985) 115
- Holland J H Adaptation in Natural and Artificial Systems Ann Arbor University of Michigan Press 1975
- Hopfield J J Neural Networks and Physical Systems with Emergent Collective Computational Properties *Proc Nat Acad Sci USA* 79 (1982) 2554
- Hopfield J J, Tank D W Computing RC-Networks *Biol Cybernetics* 52 (1985) 141
- Hopfield J J, Tank D W Computing with Neural Circuits A Model *Science* 233 (1986) 625
- Hopfield J J, Feinstein D I, Palmer R G Unlearning Has a Stabilizing Effect in Collective Memories *Nature* 304 (1983) 158
- Hoppensteadt F C An Introduction to the Mathematics of Neurons Cambr Univ Press, Cambridge 1986
- Ivachnenko A, Müller A Selbstorganisation von Vorhersagemodellen VEB Verlag Technik, Berlin 1984
- Iwainsky A Dynamische Optimierung VEB Verlag Technik, Berlin 1984
- Iwainsky A (Ed) Optimization of Connection Structures in Graphs Publ des ZKI der AdW, Berlin 1985
- Jacob F, Monod I Cold Spring Harbour Sympos *Quantitat Biol* 26 (1961) 389
- Jantsch E Die Selbstorganisation des Universums Vom Urknall zum menschlichen Geist Hanser-Verlag, München 1979
- Jiménez-Montaña M A, Ebeling W A Stochastic Evolutionary Model of Technological Change *Collective Phen* 3 (1980) 107
- Jiménez-Montaña M A, Zamora-Cortina L Evolutionary Model for the Generation of Amino Acid Sequences and Its Application to the Study of Fragment of Mammalhemoglobin Chains 7th Intern Biophys Cong and 3th Pan-American Biochem Cong Mexico City 1981
- Jones B L Analysis of Eigen's Equations for Selection of Biological Molecules with Fluctuating Mutation Rates *Bull Math Biol* 39 (1977) 311
- Jones B L A Solvable Selfreproductive Hypercycle Model for the Selection of Biological Molecules *J Math Biol* 4 (1977) 187

- Ebeling W, Schimansky Geier L Stochastic Processes in Highdimensional Systems and Models of Evolution In *Noise in Nonlinear Dynamical Systems* (Eds F Moss, P E V McClintock), Vol I Cambridge University Press, Cambridge 1988
- Ebeling W, Schmelzer J Koexistenz von Sorten in nichtlinearen autokatalytischen Parallelreaktionen *Z Phys Chemie* (Leipzig) **261** (1980) 667, Contributions to the Theory of Competing Predators *Studia biophysica* **98** (1983) 31, **80** (1980) 53
- Ebeling W, Ulbricht H (Eds) *Selforganization by Nonlinear Irreversible Processes* Springer-Verlag, Berlin 1986
- Ebeling W, Engel A, Mazenko V G Modelling of Selection Processes with Age-Dependent Birth and Death Rates *Biosystems* **19** (1986) 213
- Ebeling W, Engel A, Mazenko V G Selection between Aging Species In *Thermodynamics of Biological Processes* (Eds I Lamprecht, A I Zotin) Walter de Gruyter, Berlin 1988
- Ebeling W, Feistel R, Jiménez-Montaño M A On the Theory of Stochastic Replication and Evolution of Molecular Sequences *Rostocker Phys Manuskripte* **2** (1977) 105
- Ebeling W, Feistel R, Herzel H *Physica Scripta* **36** (1987) 761
- Ebeling W, Engel A, Esser B, Feistel R Diffusion and Reaction in Random Media and Models of Evolution Processes *J of Stat Phys* **37** (1984) 369
- Ebert D Dialog zwischen Teilchenphysik und Kosmologie *Spektrum* **6** (1984) 5
- Eigen M Wie entsteht Information? Prinzipien der Selbst-Organisation in der Biologie *Berichte der Bunsenges* **80** (1976) 1059
- Eigen M Darwin und die Molekularbiologie *Angew Chemie* **93** (1981) 221
- Eigen M, Winkler R Das Spiel R Piper & Co Verlag München/Zürich 1975
- Engel A Selektion und Diffusion in Stochastischen Feldern Diplomarbeit, Humboldt-Universität Berlin 1983
- Engel A Reaktions-Diffusionsprozesse in Stochastischen Medien Diss A, Humboldt Universität Berlin 1985
- Engel A Frustration in komplexen Systemen *Wiss Z Humboldt-Universität Berlin*, MNB **35** (1986) 418
- Engel A, Ebeling W Interaction of Moving Interfaces with Obstacles *Phys Lett A* **122** (1987) 20
- Engel A, Ebeling W, Feistel R, und Schimansky-Geier L Dynamics of Interfaces in Random Media In *Selforganization by Nonlinear Irreversible Processes* (Eds W Ebeling, H Ulbricht) Springer-Verlag, Berlin 1986
- Feistel R Ein dynamisches Modell der Funktion des Wertgesetzes *Rostocker Phys Manuskripte* **1** (1977) 48
- Feistel R On the Evolution of Biological Macromolecules *Studia biophysica* **93** (1983) 113, 121, **95** (1983) 107, **95** (1983) 133
- Feistel R Physik und Evolution *Wiss Z Humboldt-Universität Berlin* **35** (1986) 412
- Feistel R, Ebeling W Dynamische Modelle zum Selektionsverhalten offener Systeme *Wiss Z Wilhelm Pieck Universität Rostock*, MNR **25** (1976) 507
- Feistel R, Ebeling W On the Eigen—Schuster Concept of Quasispecies in the Theory of Natural Self-Organization *Studia biophysica* **71** (1978) 139
- Feistel R, Ebeling W On the Thermodynamics of Irreversible Processes in Ecosystems *Studia biophysica* **86** (1981) 23k7
- Feistel R, Ebeling W Models of Darwin Processes and Evolution Principles *Biosystems* **15** (1982) 291
- Feistel R, Ebeling W Stochastic Models of Evolutionary Processes In *Thermodynamics and Regulation of Biological Processes* (Eds I Lamprecht, A I, Zotin) Walter de Gruyter, Berlin 1984
- Feistel R, Ebeling W Evolution of Complex Systems VEB Deutscher Verlag der Wissenschaften, Berlin 1989, Kluwer Academic Publ, Dordrecht 1989
- Feistel R, Romanovsky Yu M, Vasiliev V A Über die eigen'schen Hyperzyklen in Koazervaten *Biophysica* **86** (1981) 237
- Feistel R Abstammungsgeschichte des Menschen VEB Gustav Fischer Verlag, Jena 1978
- Field R J, Noyes R M *J Chem Phys* **60** (1974) 1877
- Fisher J C, Pry H R A Simple Substitution Model of Technological Change General Electric Report 70 C 215 (1970), *Practical Applications of Technological Forecasting in Industries* (Ed Cetron, M J) John Wiley & Sons, Inc New York 1971
- Fisher R A The Genetical Theory of Natural Selection Clarendon Press, Oxford 1930
- Fitch W M, Margoliash E Construction of Phylogenetic Trees *Science* **155** (1967) 279
- Fleischer L Der Beitrag der Thermodynamik zur physikalischen Interpretation der Strukturbildung und Evolution URANIA-Prasidium, Berlin 1979
- Foerster H von Some Remarks on Changing Populations In *The kinetics of Cellular Proliferations* New York 1959
- Foerster H von In *Self Organizing Systems* (Eds M C Yovits, et al) Pergamon Press, Oxford 1960
- Foksterling H D, Kuhn H, Tews K H Computermodell zur Bildung selbstorganisierender Systeme *Angew Chemie* **84** (1972) 862
- Fontana W, Schuster P A Computer Model of Evolutionary Adaption Preprint 1988
- Fortak H G Entropy and Climate *Development in Atmospheric Science* **10** (1979) 1
- Fox S W, Dose K Molecular Evolution and the Origin of Life W H Freeman & Co, San Francisco 1972
- Franck U F Chemical Oscillations *Angew Chem Int Ed Engl* **17** (1978) 1
- Frese W Zeugen der Genesis *Naturwiss Rundschau* **32** (1979) 406
- Fritsch H Vom Urknall zum Zerfall Piper-Verlag, München/Zürich 1983
- Fritsch H Quarks — Urstoff unserer Welt Piper Verlag, München/Zürich 1981
- Fuchs Kittowski K Probleme des Determinismus und der Kybernetik in der molekularen Biologie 2. Auflage VEB Gustav Fischer Verlag, Jena 1976
- Gardiner C W, Chaturvedi S The Poisson Representation *J Stat Phys* **17**, (1977) 429

- Jones B L Some Selection Models in Catalytically Coupled Selfreproducing Macromolecular Systems *Bull Math Biol* 41 (1979) 119
- Jones B L, Enns R H, Rangnekar S S On the Theory of Selection of Coupled Macromolecular Systems *Bull Math Biol* 38 (1976) 15
- Judson H F The Eighth Day of Creation London 1979
- Kahrig E, Besserlich H Dissipative Strukturen, VEB Georg Thieme, Leipzig 1977
- Kanter I, Sompolinsky H Associative Recall of Memory without Errors *Phys Rev A* 35 (1987) 380
- Kaplan R W Der Ursprung des Lebens Georg Thieme Verlag, Stuttgart 1978
- Karpman V I Nichtlineare Wellen Akademie-Verlag, Berlin 1977
- Kaufmann S A The Origins of Order Oxford Univ Press, New York, Oxford 1993
- Kawasaki K, Ohta T Kinetic Drummhead Model of Interface *Prog Theor Phys* 67 (1982) 147 and 68 (1982) 129
- Kepper P de, Pacalt A, Rossi A C R Acad Sci Ser C 282 (1976) 199
- Kerner E H A Statistical Mechanics of Interacting Biological Species *Bull Math Biophys* 19 (1957) 121.
- Kimball J C Localization and Spectra in Solid State Systems *J Phys C* 11 (1978) 4347
- Kimura M Evolutionary Rate at the Molecular Level *Nature* 217 (1968) 624
- King J L, Jukes T H Non-Darwinian Evolution *Science* 164 (1969) 788
- Kinzel W Learning and Pattern Recognition in Spin Glasses Models *Z Phys B* 60 (1985) 205
- Kirby K G, Conrad M The Enzymatic Neuron as a Reactiondiffusion Network of Cyclic Nucleotides *Bull Math Biol* 46 (1984) 765
- Kirby K G, Conrad M Intraneuronal Dynamics for Evolutionary Learning *Physica* 16D (1986)
- Kirkpatrick S Models of Disordered Systems Lecture Notes in Physics 149 (1981) 280 Springer-Verlag, Berlin 1984
- Kirkpatrick S, Toulouse G Configuration Space Analysis of the Travelling Salesman Problems *J Physique* 46 (1985) 1277
- Kirkpatrick S, Gelatt C D, Vecchi M P Optimization by Simulated Annealing *Science* 220 (1983) 671
- Khinchin A Ya Mathematical Foundations in Information Theory Dover Publ, New York, 1957
- May R M Stability and Complexity in Model Ecosystems University Press, Princeton, N J (1973)
- Klimontovich Yu L, Kremp D, Kraeft W D Kinetic Theory for Chemically Reacting Gases and Partially Ionized Plasmas *Adv Chem Phys* 50 (1987) 175
- Klix F Erwachsendes Denken VEB Deutscher Verlag der Wissenschaften, Berlin 1980, 1985
- Klix F (Hrsg.) Organismische Informationsverarbeitung Akademie-Verlag, Berlin 1974
- Korner K Probleme der Biogenese VEB Gustav Fischer Verlag, Jena 1974 (2, erweiterte Aufl 1979)
- Korner K Wie das Leben entstand Urania-Verlag, Leipzig/Jena/Berlin 1977
- Kohonen T Selforganization and Associative Memory Springer-Verlag, Berlin 1984
- Kraeft W D, Kremp D, Ebeling W, Ropke G Quantum Statistics of Charged Particie Systems Akademie-Verlag, Berlin 1986, Plenum Press, London/New York 1986
- Krumbiegel G Tiere und Pflanzen der Vorzeit Urania-Verlag Leipzig/Jena/Berlin 1977
- Küppers B The General Principle of Selection and Evolution at the Molecular Level *Progr Biophys Molec Biol* 30 (1975) 1
- Küppers B Towards an Experimental Analysis of Molecular Self-Organization and Precellular Darwinian Evolution *Naturwiss* 66 (1979) 228
- Küppers B O Molecular Theory of Evolution Springer-Verlag, Berlin 1983
- Küppers B O Der Ursprung biologischer Information Piper-Verlag, München/Zürich 1986
- Küppers B O (Hrsg.) Ordnung aus dem Chaos Prinzipien der Selbstorganisation und Evolution des Lebens Piper-Verlag, München/Zürich 1987
- Kuhnert L Photochemische Manipulation von chemischen Wellen *Naturwiss* 73 (1986) 96
- Kuhnert L A New Optical Photochemical Memory Device in a Light Sensitive Chemical Active Medium *Nature* 319 (1986) 393
- Landauer R Fundamental Limitations in the Computational Process *Berichte der Bunsenges* 80 (1976) 1048
- Landauer R $dQ = TdS$ Far from Equilibrium *Phys Rev* 18A (1978) 255
- Lanius K Physik der Elementarteilchen Akademie-Verlag, Berlin 1980
- Lee P A, Ramakrishnan T V Disordered Electronic Systems *Rev Mod Phys* 67 (1985) 287
- Leung-Yang-Cheong S K, Cover T M Some Equivalence between Shannon Entropy and Kolmogorov Complexity IEEE Transaction, *Inform Theory* 24 (1978) 331
- Leuthäusser I An Correspondence of Eigen's Evolutions Models and Ising Systems *J Chem Phys* 84 (1986) 1884
- Leuthäusser I Statistical Mechanics of Eigen's Evolution Model *J Stat Phys* 48 (1987) 343
- Ley H Zum Stand der Entwicklungstheorie in den Naturwissenschaften *Di Z Philos* 23 (1975) 964
- Leven R Zur nichtlinearen Theorie der Strahl-Plasma-Wechselwirkung I, II, III, IV *Beitr Plasmaphys* 9 (1969) 29, 165, 293, 10 (1970) 27
- Linde A D Elementary Particles and Cosmology *Rep Progr Phys* 47 (1984) No 8, *Proc XXII Int Conf High Energy Physics* Vol II, p 125 Leipzig 1984

- Linde H Marangoni Instabilities In *Convective Transport and Instability Phenomena* (Eds J Zierep, H Oertel) Braun, Karlsruhe 1983
- Linde H Konzeption einer Komplexaufgabe auf dem Wissensgebiet Dissipative Strukturen ZIPC, Berlin 1984
- Linde H, Schwarz E Freie Oberflächenkonvektion beim Stoffübergang an fluiden Grenzen *Z Phys Chem* (Leipzig) **224** (1963) 331
- Linde H, Winkler K Hydrodynamische Stabilität einer fluiden Phasengrenze bei Stoffübergang *Z Phys Chem* (Leipzig) **225** (1964) 223
- Lothar R Die Beherrschung der Mannigfaltigkeit VEB Gustav Fischer Verlag, Jena 1972
- Lorenz K Das sogenannte Böse Zur Naturgeschichte der Aggression Dr G Borotha Schöler Verlag, Wien 1963
- Lumsden C J, Wilson E O Genes, Mind and Culture Harvard University Press, Cambridge/Mass 1981
- Lurie D, Wagensbebg J Entropy Balance in Biological Development and Heat Dissipation in Embryogenesis *J Nonequilibrium Thermodyn* **4** (1979) 127
- Malchow H, Schimansky-Geier L Noise and Diffusion in Bistable Nonequilibrium Systems BSB B G Teubner Verlagsgesellschaft, Leipzig 1986
- Mansfield E Technological Change Norton & Co, New York 1971
- Mauersberger P Entropie und freie Enthalpie in aquatischen Ökosystemen *Acta Hydrodynamica* **26** (1981) 67
- May R M Stability and Complexity in Model Ecosystems University Press, Princeton/New Jersey 1973
- May R M (Ed) Theoretical Ecology Blackwell Scientific Publications, Oxford/London/ Edinburgh/Melbourne 1974
- Mayr E, et al Evolution — A Scientific American Book Freeman & Co, San Francisco 1978
- McCaskill J S A Stochastic Theory of Macromolecular Evolution *Biol Cybern* **50** (1984) 63
- McCulloch W S, Pitts W A Logical Calculus of the Ideas Imminent in Nervous Activity *Bull Math Biosci* **5** (1943) 115
- McKendrick A G Applications of Mathematics to Medical Problems *Proc Edinburgh Math Soc* **51** (1926) 98
- McLaughlin J B, Martin P C *Phys Rev A* **12** (1975) 186
- Mezard M, Nadal J P, Toulouse G Solvable Models of Working Memories *J Physique* **47** (1986) 1457
- Mezard M, Parisi G, Virasoro M A Random Free Energies in Spin Glasses *J Phys Lett* (Paris) **46** (1986) L 217
- Mezard M, Parisi G, Sourlas N, Toulouse G, Virasoro M A Replica Symmetry Breaking and the Nature of the Spin Glass Phase *J Phys* (Paris) **45** (1984) 843
- Mikhailov A S, Schimansky Geier L, Ebeling W Effect of Fluctuations on Plane Front Propagation in Bistable Nonequilibrium Systems *Phys Lett A* **96** (1983) 453
- Miller S L, Orgel L E The Origins of Life on Earth Prentice Hall, Englewood Cliffs, N J 1973
- Minsky M, Papert S Perceptrons an Introduction to Computational Geometry MIT Press, Boston 1969
- Monod J Zufall und Notwendigkeit R Piper & Co Verlag, München 1971 (Original Edition du Seuil, Paris 1970)
- Morowitz H J Foundations of Bioenergetics Academic Press, New York/San Francisco/London 1978
- Mott N F Electrons in Disordered Systems *Adv Physics* **16** (1967) 49
- Nadal J P, Toulouse G, Changeu J P, Dehaene S Networks of Formal Neurons and Memory Palimpsests *Europhys Lett* **1** (1986) 535
- Nattermann T Domain Walls Fluctuations and the Incommensurate-Commensurate Transition in Two and Three Dimensions *J Phys* **43** (1982) 631
- Nattermann T Interface Roughening in Random Field Models In *Proceedings of the XXth Winterschool on Theoretical Physics "Karpacz 1984"* (Ed A Pekalski) Lect Notes in Phys, Springer Verlag, Berlin 1984
- Neumann J von The General and Logical Theory of Automata In *Cerebral Mechanism of Behaviour* (Ed L A Jeffres) New York/London 1951
- Neumann J von Theory of Self Reproducing Automata University of Illinois Press, Urbana 1966
- Nicols G, Malek-Mansour M Nonequilibrium Phase Transitions and Chemical Reactions *Suppl Theor Phys* **64** (1978) 249
- Odum E P The Strategy of Ecosystem Development *Science* **164** (1969) 262
- Odum, E P Fundamentals of Ecology Saunders, Philadelphia 1971
- Oster S F, Perelson A S, Katchalsky A Network Thermodynamics *Quart Rev Biophys* **6** (1973) 1
- Ostwald W Gedanken zur Biosphäre 1931 Wiederabdruck BSB B G Teubner Verlagsgesellschaft, Leipzig 1978
- Pacault A, Vidal C (Eds) Synergetics Springer Verlag, Berlin/Heidelberg/New York 1979
- Papentin F A Darwinian Evolutionary System I Definition and Basic Properties *J Theor Biol* **39** (1973) 397
- Parisi G The Order Parameter for Spin Glasses A Function on the Interval 0–1 *J Phys A* **13** (1980) 1101
- Parisi G Order Parameter for Spin Glasses *Physic Rev Lett* **50** (1983) 1946
- Parisi G A Memory which forgets *J Phys A* **19** (1986) L 617
- Peebles P J E Physical Cosmology Princeton University Press, Princeton 1971
- Penzlin H Grundprinzipien lebendiger Systeme In *Kompendium der allgemeinen Biologie* (Hrsg E Libbert) VEB Gustav Fischer Verlag, Jena 1976
- Personnaz L, Guyon I, Dreyfus G Information Storage and Retrieval in Spin Glasses Like Neural Networks *J Physique Lett* **46** (1985) L 359

- Personnaz L, Guyon I, Dreyfus G, Toulouse G Logically Constrained Learning Mechanism in Networks of Formal Neurons *J Stat Phys* **43** (1986) 411
- Peschel M, Mende W Leben wir in einer Volterra-Welt? Akademie-Verlag, Berlin 1983
- Peschel M, Mende W The Predator Prey Model Akademie-Verlag, Berlin 1986
- Polanyi M Life's Irreducible Structure *Science* **160** (1968) 1308
- Ponnamperuma C The Origin of Life New York 1972
- Prigogine I Introduction to Thermodynamics of Irreversible Processes Thournes, Chicago 1955, Wiley Interscience, New York 1967
- Prigogine I, Defay R Chemische Thermodynamik VEB Deutscher Verlag für Grundstoffindustrie, Leipzig 1962
- Prigogine I, Stengers I Dialog mit der Natur Piper-Verlag, München/Zürich 1981
- Prigogine L, Wiame J M Biologie et Thermodynamique des Phenomenes Irreversibles *Experientia* **2** (1946) 451
- Prigogine I, Nicolis G, Babloyantz A Thermodynamics and Evolution *Physics Today* Nov 1972, p 23, Dec 1972, p 38
- Rammal R, Toulouse G, Virasoro M A Ultrametricity for Physicists *Rev Mod Phys* **58** (1986) 765
- Ratner V A The Genetic Language In *Progress in Theoretical Biology* (Eds R Rosen, F M Snell), Vol 3 Academic Press, New York 1974
- Rechenberg I Evolutionsstrategien — Optimierung technischer Systeme nach Prinzipien der biologischen Information Friedrich Frommann Verlag (Gunter Holzboog K G), Stuttgart-Bad Cannstatt 1973
- Reich J G, Selkov E E Energy Metabolism of the Cell Academic Press, London 1981
- Richter P H, Ross J The Efficiency of Engines Operating Around a Steady State at Finite Frequency *J Chem Phys* **69** (1978) 5521
- Richter P H, Rehms P, Ross J Control and Dissipation in Oscillatory Chemical Engines *Prog of Theor Phys* **66** (1981) 385
- Riedl R Die Ordnung des Lebendigen Systembedingung der Evolution Verlag Parey, Hamburg und Berlin 1975
- Rigos A A, Deutsch J M A Simple Model for the Influence of Fluctuations on Explosive Reactions *J Chem Phys* **76** (1982) 5180
- Rokhsar D S, Anderson P W, Stein D L Selforganization in Prebiological Systems A Model for the Origin of Genetic Information Princeton University, Preprint 1983
- Rompe R, Steenbeck M Der Plasmazustand der Gase *Ergebnisse der exakten Naturwissenschaften* **18** (1939) 275
- Rompe R, Treder H -J Grundfragen der Physik Akademie-Verlag, Berlin 1979
- Ruelle D, Takens F *Comm Math Phys* **20** (1971) 167
- Rumelhart D E, McClelland J L (Eds) Parallel Distributed Processing MIT Press, Cambridge/Mass 1986
- Rumschitzky D Spectral Properties of Eigen Evolution Matrices Preprint 1986
- Runcorn S K (Eds) Continental drift Academic Press, New York 1962
- Sagan C Die Drachen von Eden Das Wunder der menschlichen Intelligenz Drömersche Verlagsgesellschaft Th Knaur Nacht, München/Zürich 1978
- Sander S Matrizen der Evolution Preprint, Berlin 1988
- Sankoff D Minimal Mutation Trees of Sequences *SIAM J Appl Math* **28** (1975) 35
- Sankoff D, Morel A, Cedergren B On the Evolution of Ribosomal RNA-Sequences *Nature* **241** (1973) 111
- Saunders P T, Ho M W On the Increase in Complexity in Evolution *J Theor Biol* **63** (1976) 375.
- Scharf H (Hrsg) Evolution Nova Acta Leopoldina Bd **42** Halle 1975
- Scheidegger U Zur Beschreibung strukturierter Objekte mit kontextfreien Grammatiken In *Organismische Informationsverarbeitung* (Hrsg F Klix) Akademie-Verlag, Berlin 1974
- Scheler W, Mohr P Lebensprozesse — Ziele und Modelle der Chemie *Wiss u Fortschritt* **29** (1979) 99
- Schimansky Geier L, Ebeling W Nucleation in bistable Systems *Ann Phys (Leipzig)* **40** (1983) 10
- Schleicher A Die Darwinsche Theorie und die Sprachwissenschaft Verlag H Böhlaus, Weimar 1863 (Переиздание в *Nova Acta Leopoldina* NF **42** (1975) 377)
- Schleicher A Über die Bedeutung der Sprache für die Naturgeschichte des Menschen Verlag H Böhlaus, Weimar 1865
- Schmutzer E Relativistische Physik BSB B G Teubner Verlagsgesellschaft, Leipzig 1968
- Schnakenberg J Network Theory of Microscopic and Macroscopic Behaviour of Master Equation Systems *Rev Mod Phys* **47** (1976) 571
- Schopf J W Vor 600 Millionen Jahren — Explosion hochentwickelter Lebensformen *URANIA* **7** (1976) 36
- Schuster P Vom Makromolekül zur primitiven Zelle *Chemie in unserer Zeit* **6** (1972) 1
- Schuster P Structure and Dynamics of Replication-Mutation Systems *Physica Scripta* **35** (1987) 402
- Schuster P, Sigmund K, Wolff R Dynamical Systems under Constant Organization *Bull Math Biophys* **40** (1978) 743, *J Diff Equ* **32** (1979) 357, *Biol Cybern* **40** (1981) 1
- Schwefel H P Numerische Optimierung von Computermodellen mittels der Evolutionsstrategie Birkhäuser Verlag, Basel/Stuttgart 1977
- Schweitzer F, et al A Stochastic Approach to Nucleation in Finite Systems Part I in *Physica* **150A** (1988) 261, Part II in **153A** (1988) 573

- Segal J, Körner U, Leiterer K P Die Entstehung des Lebens — ein biophysikalisches Problem VEB Gustav Fischer Verlag, Jena 1982
- Sejnowski T J, Kienker P K, Hinton G E Learning Symmetry Groups with Hidden Units Beyond the Perceptron *Physica* **220** (1986) 260
- Sejnowski T, Rosenberg C NETalk a Program which Converts Text into Spoken English Technical Report JHU/EECS 86/01 Baltimore/Maryland 1987
- Simon H The Architecture of Complexity *Proc Am Phil Soc* **106** (1962) 467
- Slobodkin L B Energy Relationship at the Population Level *Amer Natur* **94** (1960) 213
- Sonntag I, Feistel R, Ebeling W Random Networks of Catalytic Biochemical Reactions *Biom J* **23** (1981) 501
- Stein D S, Anderson P W A model for the Origin of Biological Catalysis *Proc Natl Acad Sci USA* **81** (1984) 1751
- Strobeck C N-Species Competition *Ecology* **54** (1973) 650
- Stugren B *Grundlagen der Allgemeinen Ökologie* VEB Gustav Fischer Verlag, Jena 1978
- Swetina J, Schuster P Self Replication without Errors *Biophys Chem* **16** (1982) 329
- Tanaka F, Edwards S F Analytic Theory of the Ground State Properties of a Spin Glass I Ising Spin Glasses *J Phys F* **10** (1980) 2471
- Tembrock G Biokommunikation (WTB) Akademie-Verlag, Berlin 1971
- Tembrock G Grundlagen des Tierverhaltens Akademie-Verlag, Berlin 1977
- Tembrock G Informationswechsel und Evolution In *Struktur und Prozeß* (Hrsg K F Wessel) VEB Deutscher Verlag der Wissenschaften Berlin 1977
- Thiele H Zur Definition von Kompliziertheitsmassen für endliche Objekte In *Organismische Informationsverarbeitung* (Hrsg F Klux) Akademie-Verlag, Berlin 1974
- Thielemann F K, Metzinger J, Klapdor H V New Actinide Chronometer Production Ratios and the Age of the Galaxy *Astrophys J* **123** (1983) 162
- Thom B Stabilité Structurelle et Morphogenese W A Benjamin Inc, New York 1972
- Thompson C J, McBride J L On Eigen's Theory of Self Organization of Matter and the Evolution of Biological Macromolecules *Math Biosci* **21** (1974) 127
- Thouless D J Electrons in Disordered Systems and the Theory of Localization *Phys Rep* **13** (1974) 93
- Thouless D J Electrons in a Four Dimensional Disordered System *J Phys C* **9** (1976) L 603
- Toulouse G Theory of the Frustration Effect in Spin Glasses *Commun Phys* **2** (1977) 115
- Treder H -J *Relativität und Kosmos* (WTB) Akademie-Verlag, Berlin 1968
- Treder H J Zur Relaxationszeit des Kosmos *Ann Phys (Leipzig)* **43** (1986) 621
- Unsöld A Kosmische Evolution *Naturwiss Rundschau* **28** (1975) 3
- Uyeda S The New View of the Earth W H Freeman & Co, San Francisco 1978
- Velarde M G Hydrodynamic Instabilities In *Isotopic Fluids* (Eds R Balian, J L Peube) Gordon and Breach, New York 1976
- Volz H Information I, II Akademie-Verlag, Berlin 1982, 1983
- Voigt H M Structure Formation Based on Self Organization Principles with Application to Electronics Design Akademie der Wissenschaften, Berlin 1985
- Voigt H M Structuring and Evolution Processes *Proc 5th Meeting of the UNESCO Working Group in System Theory* Akademie-Verlag, Berlin 1985
- Voigt H M Structure Formation Processes Vectoroptimization and Evolution An Ecological Approach ZKI Preprint, Berlin 1985
- Volkenstein, M V *Biochim Biophys Acta* **180** (1969) 562
- Wheatley J, Hoffler T, Swift S W, Miglion A Experiments with an Intrinsically Irreversible Acoustic Heat Engine *Phys Rev Letters* **50** (1983) 499
- Whitehead J A Cellular Convection *American Scientist* **59** (1971) 444
- Wiener N, Rosenblüth A The Mathematical Formulation of the Problem of Conduction of Impulses in a Network of Connected Excitable Elements, Specifically in Cardiac Muscle *Arch Inst Cardio (Mexico)* **16** (1946) 205
- Wilczek F Foundations and Working Pictures in Microphysical Cosmology XVIII, Solvay Conf, Austin/Texas 1983
- Wolfram S Cellular Automata as Models of Complexity *Nature* **311** (1984) 419
- Wright S The Roles of Mutation and Selection in Evolution *Proc 6th Int Congr Genetics* **1** (1932) 356–366
- Wunsch G *Moderne Systemtheorie* Akademische Verlagsgesellschaft, Geest & Portig, K G, Leipzig 1975
- Yockey H P Calculation of the Probability of Spontaneous Biogenesis by Information Theory *J Theor Biol* **67** (1977) 377, 74 (1978) 148
- Zaikin A N, Kawczynski A L Spatial Effects in Active Chemical Systems *J Nonequil Thermodynamics* **1** (1976) 100
- Zuckerkandl E The Appearance of New Structures and Functions of Proteins During Evolution *J Mol Evol* **7** (1975) 1–97
- Zuckerkandl E Evolutionary Processes and Evolutionary Noise at the Molecular Level *J Mol Evol* **7** (1976) 269

Предметный указатель

- автоматы, 114–116, 296, 305
аминокислоты, 38–40, 73, 199, 203, 263, 268, 274
биоид; 121
биомасса, 74, 176
биосфера, 48, 75
Большой Взрыв, 5, 10, 19, 23, 31, 32, 41, 43, 83
вандерваальсовские газы, 34
вероятность выживания, 114, 155, 156
возрастные структуры, 186, 189, 191, 193
гёделевское число, 264, 280
гиперотбор, 145, 167, 174, 175
гравитационные неустойчивости, 32, 37, 89
грамматика, 39, 283, 284
диссипативная структура, 7, 12, 54, 81, 100, 103, 122, 293, 294
диффузия, 60, 62, 63, 88, 97, 104, 232, 253, 301
ДНК, 40, 107, 128, 137, 184, 202, 229, 273, 275
игра Конуэя, 115, 120
имитация, 138, 216, 223
индивидуальное развитие, 72, 183, 240
инновация, 198, 212, 225, 226, 228, 293
информация, 10, 13, 37, 105, 121, 184, 257
календарь эволюции, 41, 66
квазивид, 232, 235, 238
квант эволюции, 239
кинетика Михаэлиса—Ментен, 126, 127, 171
коацерват, 39, 83, 199, 215
конкуренция, 38, 125, 143
критерии эволюции, 59, 63, 64, 233
линейные соотношения Онсагера, 57, 62
матрица популяции (сообщества), 175
машина Тьюринга, 116
модель Лотки—Вольтерры, 175, 177
— старения Маккендрика—фон Фёрстера, 185
морфогенез, 13
мутация, 39, 40, 175, 196, 197
неравенство Клаузиуса—Карно, 51
неустойчивость Марангони, 93
— Тейлора, 90
нуклеиновые кислоты, 38, 199, 200, 207, 213, 274, 283, 300
оптимизация, 13, 143, 242
отбор, 13, 40, 125, 143
отношения хищник—жертва, 76, 175, 241
планковский формфактор, 68, 71
Праземля, 37, 183, 198, 199, 203, 207, 254
принцип гетеротрофии, 40, 41
— Дарвина, 11
— сегрегации, 146, 147
— усиления, 197, 211
принципы эволюции, 65, 121, 145–147, 149–151, 161, 170, 178, 202, 206, 296, 297
производство энтропии, 53–55, 57, 59, 62–65, 70, 78, 93, 147, 295
протеноиды, 39, 199, 207
процессы эволюции, 5–8, 10, 11, 36, 41, 46, 64–66, 89, 97, 153, 195, 196, 198, 210, 223, 226, 233, 237, 242
распределение Пуассона, 113
— функций, 41
расстояния Хэмминга, 248, 264, 305
расширение космоса, 23
реакция Белоусова—Жаботинского, 103, 310
— Брея, 97, 98
— Бриггса—Раушера, 98
РНК, 40, 106, 202, 268, 275
самовоспроизведение, 12, 37–39, 105, 116, 121, 141, 199, 204, 230, 240, 242
самоорганизация, 5, 7, 8, 35, 48, 53, 257, 258, 293, 294, 297
синергетика, 8
сложность, 269, 280
соотношения симметрии Онсагера, 62
текущее равновесие, 55, 70
теорема Гленсдорфа—Пригожина, 65, 147
— Пригожина, 59, 65
теория Эйгена—Шустера, 5, 142, 153, 202, 207
тепловая смерть, 11
технология, 8, 137
течение Куэтта, 89, 90
— Пуазейля, 89, 90
трофическая структура, 176
турбулентность, 90–92, 104
уравнение Навье—Стокса, 63, 90, 91
— Фишера—Эйгена, 151
фазовый переход, 13, 14, 83, 101, 109, 294
фенотип, 143, 197, 229, 230, 232
фоновое излучение, 19, 20, 23, 31, 32, 66
фотонная мельница, 10, 30, 34–36
фронт, 87, 88
фрустрация, 222, 223, 242, 243, 254
фундаментальное уравнение Гиббса, 49, 50, 52, 61, 77
циклы жизни, 184
эволюционная спираль, 198
эволюция, 5, 7, 8, 10, 12, 16–18, 24, 34, 37, 41, 46, 116, 143, 183, 196, 229, 242
экологическая картина, 163
экосистема, 10, 36, 75, 78, 194
экспорт энтропии, 12, 35, 51, 53, 54, 68, 69, 78, 83, 122, 199
экстремальный принцип, 94, 145, 146
эффект Бенара, 69, 92–94, 101, 108
— Ганна, 96, 97
язык, 9, 16, 39

Содержание

Предисловие	5
Глава 1. Введение	6
1.1. Возникновение сложного	6
1.2. Какую цель ставит перед собой физика процессов эволюции?	7
1.3. Об общей структуре процессов эволюции	11
Глава 2. Сценарий эволюции Метагалактики, Земли и жизни	19
2.1. Большой Взрыв и расширение	19
2.2. Многократное нарушение симметрии и возникновение элементов	24
2.3. Структурирование протоплазмы и образование фотонной мельницы	30
2.4. Возникновение жизни	36
Глава 3. Термодинамические условия эволюции	46
3.1. Энергия и энтропия	46
3.2. Начала термодинамики и самоорганизация	48
3.3. Глобальные условия самоорганизации	53
3.4. Локальные равновесия и критерии эволюции	59
3.5. Термодинамические действующие силы эволюции Земли	66
3.6. Термодинамика самоорганизации живых организмов	69
3.7. Термодинамика экологических систем	74
Глава 4. Неустойчивости и пространственно-временные структуры	83
4.1. Термодинамические неустойчивости	83
4.2. Кинетические неустойчивости в жидких средах	89
4.3. Кинетические неустойчивости в химических системах	97
Глава 5. Самовоспроизведение	105
5.1. Динамика простого самовоспроизведения	105
5.2. Стохастика простого самовоспроизведения	110
5.3. Самовоспроизводящиеся автоматы	114
5.4. Динамика нелинейного самовоспроизведения	121
5.5. Стохастика нелинейного самовоспроизводства	128
5.6. Самовоспроизводство в сложных системах	136
Глава 6. Процессы конкуренции и отбора	143
6.1. Эволюция как проблема оптимизации	143
6.2. Отбор в случае простой конкуренции	147
6.3. Конкуренция и отбор в сетях	158
6.4. Сосуществование и гиперотбор	167
6.5. Отбор в сложных системах	175
Глава 7. Индивидуальное развитие как новая стратегия эволюции	183
7.1. Роль онтогенеза в эволюции	183
7.2. Дискретные модели циклов жизни	184
7.3. Модель старения Маккендрика—фон Фёрстера	185
7.4. Процессы отбора в моделях с непрерывным старением	187
7.5. Сложные возрастные структуры	193

Глава 8. Мутации и спирали эволюции	196
8.1. Частота ошибок при последовательной репликации	196
8.2. Основы теории Эйгена—Шустер	202
8.3. Стохастика процессов эволюции	210
8.4. Игра в имитацию	216
8.5. Инновации в сложных системах	225
Глава 9. Эволюция в пространстве фенотипов	229
9.1. Адаптивный ландшафт	229
9.2. Скачкообразный характер эволюции в простой модели	234
9.3. Индивидуальное развитие фенотипа	239
Глава 10. Фрустрация и иерархическое упорядочение	242
10.1. Дарвиновская эволюция как задача фрустрированной оптимизации	242
10.2. Некоторые результаты теории спиновых стекол	243
10.3. Модель Андерсона пребиологической эволюции	252
Глава 11. Эволюция информации	257
11.1. Переработка информации как процесс самоорганизации	257
11.2. Информация и жизнь	260
11.3. Биопоследовательности и дерево эволюции	264
11.4. Энтропия и сложность	269
11.5. Генетический язык и грамматика биопоследовательностей	283
Глава 12. Перспективы использования	293
12.1. Проблемы энергии, материалов и физических элементов	293
12.2. Проблемы автоматизации и оптимизации	296
12.3. Проблемы биотехнологии	300
12.4. Ассоциативная память	302
12.5. Сетевые автоматы и нейροкомпьютеры	305
Послесловие к русскому изданию	312
Литература	313
Предметный указатель	324