BIOPOLYMERS VOL. 15, 1345–1375 (1976)

Theoretical Calculations of the Helix–Coil Transition of DNA in the Presence of Large, Cooperatively Binding Ligands

JAMES D. McGHEE,* *Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403*

Synopsis

Theoretical calculations are conducted on the helix–coil transition of DNA, in the presence of large, cooperatively binding ligands modeled after the DNA-binding proteins of current biological interest. The ligands are allowed to bind both to helix and to coil, to cover up any number of bases or base pairs in the complex, and to interact cooperatively with their nearest neighbors. The DNA is treated in the infinite homogeneous Ising model approximation, and all calculations are done by Lifson's method of sequence-generating functions. DNA melting curves are calculated by computer in order to explore the effects on the transition of ligand size, binding constant, free activity, and ligand–ligand cooperativity. The calculations indicate that (1) at the same intrinsic free energy change per base pair of the complexes, small ligands, for purely entropic reasons, are more effective than are large ligands in shifting the DNA melting temperature; (2) the response of the DNA melting temperature to increased ligand binding constant K and/or free ligand activity L is adequately represented at high values of KL (but not at low KL) by a simple independent site model; (3) if curves are calculated with the total amount of added ligand remaining constant and the free ligand activity allowed to vary throughout the transition, biphasic melting curves can be obtained in the complete absence of ligand–ligand cooperativity. In an Appendix, the denaturation of poly[d(A-T)] in the presence of the drug, netropsin, is used to verify some features of the theory and to illustrate how the theory can be used to obtain numerical estimates of the ligand binding parameters from the experimental melting curves.

INTRODUCTION

Proteins that regulate biological processes by interacting with DNA can be divided into two classes. One class contains what can be called "specific" binding proteins exemplified by the *lac* and λ repressors; these proteins bind with a great affinity to one (or a few) genetically defined "operator" sites on the genome. The second class, treated in the present paper, contains what can be called "nonspecific" binding proteins; these proteins do not have a genetically defined binding site on the genome. Rather, any and all of the DNA bases or base pairs can potentially be involved in a binding site. This class of nonspecific binding proteins can be further divided into those proteins that interact preferentially with the native con-

* Present address: Laboratory of Molecular Biology, NIAMDD, NIH, Bethesda, Md. 20014.

formation of DNA and thus stabilize the DNA helix, and those proteins that interact preferentially with the denatured or coil form of DNA and thus destabilize the helix.

Examples of nonspecific helix stabilizers are the histones, protamines, basic polypeptides, perhaps such exotic molecules as hormone receptors,[1] and even the specific repressor molecules when bound "nonspecifically" to nonoperator DNA. (Indeed it has recently been pointed out[2] that the specific repressor–operator interaction can be quantitatively understood only when the large background of weaker nonspecific binding is taken into account.) Also to be included in a more general class of helix stabilizers are drugs such as actinomycin and netropsin, dyes such as ethidium and proflavine, and, for completeness, the small cations.

Proteins that belong to the second class of nonspecific DNA-binding proteins, the helix destabilizers, have only been discovered fairly recently. The first such "DNA-unwinding" or "melting protein" to be found was the gene 32-protein produced in T4-infected *E. coli;*[3] since then, proteins that bind strongly to single-stranded DNA have been isolated from a wide variety of organisms.[4–9] These proteins all appear to be present in large amounts *in vivo,* to have little DNA base specificity, and are usually presumed, although it is actually known in only a few cases,[3,7] to function in DNA synthesis and/or recombination. Other members of the class of helix destabilizers are proteins such as the gene 5-protein of filamentous phages,[10,11] the enzyme bovine ribonuclease,[12,13] and, for completeness, chemical denaturants such as OH^-, H^+, and formaldehyde.

Even the simplest model of the interaction of these nonspecific ligands with DNA must contain the following two features. First, most of these proteins are large, and may cover up to 20 or more bases or base pairs in the bound complex. Second, in many cases the binding process is cooperative, i.e., neighboring ligands interact favorably and tend to cluster on the DNA; this is particularly important with the unwinding proteins such as T4 gene 32-protein.[14] These two properties of the ligand make the isothermal binding even to one form of DNA (i.e., to either all helix or all coil) rather complex, although a number of theoretical treatments are available, both for noncooperative[15,16] and cooperative ligands.[17–19] The present paper considers the further levels of complexity introduced when this binding is imposed upon the helix–coil transition of DNA, where the proteins can, in general, bind both to the helix and to the coil forms, and where all processes vary with temperature in a complicated and interlocked fashion.

Over the last 15 years or so, the theory of the helix–coil transition of DNA has been studied extensively (see, for example, Refs. 20, 21). Furthermore a number of these studies have coupled the melting transition to reversible ligand binding, both for DNA[22–24] and for the analogous case of polypeptides.[25–27] However, the emphasis in all these binding theories has been on using small simple ligands such as acid or alkali as a means to investigate DNA by itself. Attempts to extend such theories to large noncooperative ligands only allow ligand binding in what has come to be known as the

"Scatchard" approximation;[24,28] the extent of this approximation will be assessed below. Although these treatments have been found to be inadequate for the present purposes, nevertheless they indicate, at least qualitatively, the type of behavior to be expected in more complex ligand–DNA systems. Using a more comprehensive theoretical model, the present paper attempts to explore quantitatively this more complex behavior and to isolate the effect that each feature of the model has on the calculated DNA transition. In all the calculations, the DNA is represented by an infinitely long, homogeneous polynucleotide and treated by the nearest neighbor Ising model, both for reasons of conceptual simplicity and because the number of parameters required to describe even this simple model is close to exceeding the number that can be uniquely determined from an experimental melting curve.[13] At least for the present, a more complex model would seem unwarranted. An attempt is made to use, in the calculations, numerical values that might be representative of real DNA-binding proteins, in order that some of the calculated curves might serve as rough "standard" curves with which to compare experimental transitions. As will be illustrated in the Appendix, comparison of experimental and calculated melting curves can easily be used to define ligand size within a factor of 2 and binding constants to within an order of magnitude.

THEORY

Model and Statistical Weights

Figure 1 shows the model to be used to represent protein ligands binding to a partially melted DNA molecule.

The DNA is taken to be an infinitely long homogeneous polynucleotide and treated by the simple nearest neighbor Ising model. As is well known,[20,21] only two parameters are required to describe this model. The first parameter is the stability constant, s, which is the equilibrium constant for forming a helix base pair from a coil base pair, at the end of a long helical stack. If, as is conventional, the statistical weight of a coil base pair is defined as unity, the statistical weight of a helix base pair is s; at the T_m of free DNA, $s = 1$. The temperature dependence of s is given by the van't Hoff equation, using an experimental value for the enthalpy change for melting a base pair. The second parameter required in the DNA model is the cooperativity or nucleation parameter σ; this is the statistical weight assigned to a boundary between a helix region and a coil region. The ostensible purpose of σ is to correct for the loss of an extra stacking interaction on forming a coil region; however σ will also contain some sort of average of the loop-weighting function[20,21] and in general will absorb anything that makes the DNA-only transition cooperative. For simplicity, σ is taken to be temperature independent.

As seen in Figure 1, the ligand is considered to be a featureless object, which can exist either bound to the DNA or free in solution. In the most

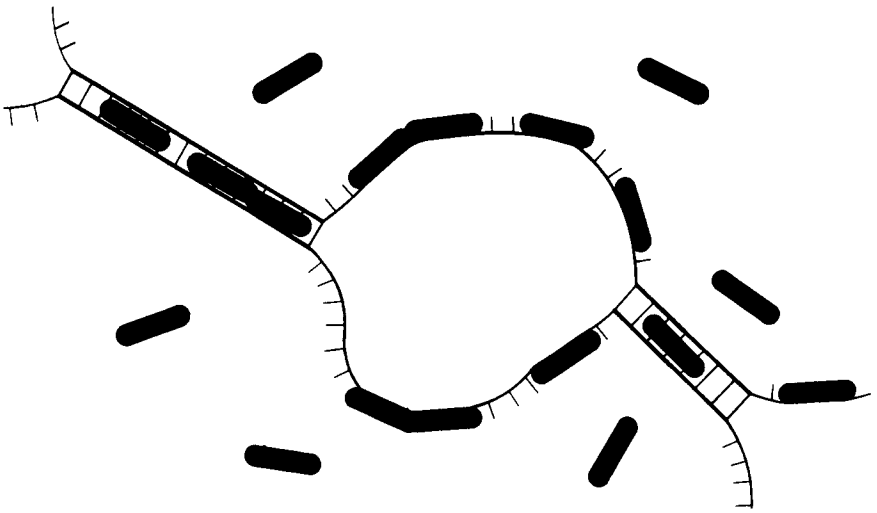

Fig. 1. Diagram of the model of the DNA–ligand systemto be used for the calculations. The homogeneous DNA consists of alternating regions of helix and coil. The ligand exists either free in solution, bound to helix, or bound to coil; bound ligands can interact with their nearest neighbors.

general case, the ligand is allowed to bind both to helix and to coil forms of the DNA. However in a majority of the later calculations it will be simpler to let the binding constant to either coil or helix be zero (to describe pure stabilizers or destabilizers, respectively). It is emphasized that the model assumes that the ligand is able to transfer freely around the DNA molecules, i.e., the binding process must be at equilibrium at all points in the helix–coil transition.

The free ligand concentration is taken to be equivalent to ligand activity and denoted by L. The ligand is considered to bind to any particular stretch of helical base pairs with an (intrinsic) association constant of K_h (M^{-1}). This binding could either be strictly polar with respect to the DNA (as might occur with coil-binding proteins) or twofold rotation could be allowed (as might occur with helix-binding proteins); in the latter case, the symmetry factor of 2 is contained in K_h. In the complex, the bound ligand sterically covers up, or makes inaccessible to a second ligand, a stretch of n_h contiguous base pairs. The statistical weight of a helix–ligand complex, relative to the statistical weight of the n_h free helical base pairs, is K_hL.[18,27] Since, as noted above, all the statistical weights are assigned relative to the statistical weight of the coil configuration (defined as unity), a helix–ligand complex is assigned the overall weight $s^{n_h}K_hL$. Bound ligands are allowed to interact only with nearest neighbor bound ligands. The ligand–ligand cooperativity parameter ω_h is defined as the equilibrium constant for the process of moving an isolated ligand, bound with no nearest neighbors, into a position where it now has one nearest neighbor.[19] Thus the statistical weight assigned to a pair of bound nearest neighbor ligands is a factor of ω_h greater than for a pair of bound but isolated ligands.

When a ligand binds to the coil form of DNA, it is assumed that both strands in the loop region can bind ligand independently. The statistical weights assigned to such coil–ligand complexes are similar to those assigned to complexes with the helix, except that the subscript c rather than h is used. Thus a ligand is allowed to bind to any particular stretch of n_c *bases* on one strand, with an intrinsic association constant, K_c (M^{-1}); the statistical weight assigned to the complex is then K_cL. Nearest neighbor ligands are allowed to interact with an equilibrium constant of ω_c, i.e., a contiguously bound ligand is assigned an additional statistical weight ω_c. The approximation by which opposite sides of the loop are allowed to bind ligands independently is described below.

It will sometimes be convenient to use symbols without subscripts to refer to either helix or to coil parameters. However it must be remembered that, in the current model, the ligand binds to *base pairs* in the helix regions but to *bases* in the coil regions. An alternative model, in which the ligand binds to *bases* in the helix, will be mentioned below.

System Partition Function

The statistical weights (or local partition functions) defined in the last section must now be combined into the overall, semigrand canonical partition function Z of the DNA–ligand system. The procedure to be used is Lifson's method of sequence-generating functions;[29] this has been fully described, both in the original paper, and by several subsequent investigators,[15,18,20,30] and thus will be given here only in the briefest outline.

As seen in Figure 1, there are four different types of sequences of bases or base pairs in which all the members are in the same configurational state: a stretch or sequence of (one or more) helix base pairs; a sequence of helix base pairs complexed with ligands; a sequence of free coil bases; and finally a sequence of coil bases complexed with ligands. For each of these types of sequences and for any particular sequence length, the statistical weights can be assigned according to the last section. For example, the statistical weight of a sequence of one helix base pair is s, of a sequence of two helix base pairs is s^2, and of a sequence of j helix base pairs is s^j. The first step in applying the method of sequence-generating functions is to write down, for each type of sequence, an infinite series in increasing powers of a variable $1/x$, where the coefficient of each term in the series is the statistical weight assigned to the corresponding length of that particular type of sequence. If $T(x)$ is the power series, or sequence-generating function, associated with sequences of uncomplexed helical base pairs, then:

$$T(x) = \frac{s}{x} + \left(\frac{s}{x}\right)^2 + \left(\frac{s}{x}\right)^3 + \ldots$$

$$= \frac{s}{x - s}$$

under conditions where the series converges.

If $U(x)$ is the sequence-generating function for sequences of helix base pairs complexed with ligand, then:

$$U(x) = \left(\frac{s}{x}\right)^{n_h} K_h L + \left\{\left(\frac{s}{x}\right)^{n_h} K_h L\right\}^2 \omega_h + \left\{\left(\frac{s}{x}\right)^{n_h} K_h L\right\}^3 \omega_h{}^2 + \ldots$$

$$= \frac{s^{n_h} K_h L}{x^{n_h} - \omega_h s^{n_h} K_h L}$$

Two things should be noted about $U(x)$: (1) since the bound ligand covers up n_h base pairs in the complex, the length of any sequence of bound ligands must be an integral multiple of n_h base pairs and hence the variable in the denominator of each term must always be a power of x^{n_h}; (2) since a sequence of j bound ligands has $(j-1)$ nearest neighbor ligand–ligand interactions, the statistical weight of such a sequence contains the factor $\omega_h{}^{(j-1)}$.

There is an important difference between the sequence-generating functions for a helix region and those for a coil region, since coil regions have two strands, each of which can bind ligand independently. This feature can be treated by a simple approximation in which the sequence-generating functions for coil regions are constructed in increasing powers of $1/\sqrt{x}$, rather than $1/x$ (i.e., if the average partition function of a base pair is x, then that of a base is $\sqrt{x}$). Physically, the approximation replaces the two separate strands of a loop with one single strand, twice as long. (This has only a minor effect on the calculated results, underestimating the average helix length by about 30% and shifting the melting temperature by a few tenths of a degree.) If $V(x)$ is the sequence-generating function for free coil regions, then:

$$V(x) = \frac{1}{\sqrt{x}} + \left(\frac{1}{\sqrt{x}}\right)^2 + \left(\frac{1}{\sqrt{x}}\right)^3 + \ldots$$

$$= \frac{1}{\sqrt{x} - 1}$$

Finally, if $W(x)$ is the sequence-generating function for sequences of ligands bound to coil, then:

$$W(x) = \frac{K_c L}{\sqrt{x}^{n_c}} + \omega_c \left\{\frac{K_c L}{\sqrt{x}^{n_c}}\right\}^2 + \omega_c{}^2 \left\{\frac{K_c L}{\sqrt{x}^{n_c}}\right\}^3 + \ldots$$

$$= \frac{K_c L}{\sqrt{x}^{n_c} - \omega_c K_c L}$$

Lifson has shown[29] that, in the limit of an infinitely long DNA molecule, the partition function of the DNA–ligand system can be written as:

$$Z = x_1{}^N \tag{1}$$

where N is the number of base pairs in the DNA molecule ($N \to \infty$) and x_1 is the largest root of the following determinantal equation:

$$f(x) = \begin{vmatrix} -1 & U & V & W \\ T & -1 & V & W \\ \sigma \cdot T & \sigma \cdot U & -1 & W \\ \sigma \cdot T & \sigma \cdot U & V & -1 \end{vmatrix} = 0 \tag{2}$$

where the various entries are the sequence-generating functions defined above. The statistical weight σ is inserted every time there is a coil-to-helix junction, with no distinction being made whether the coil or helix has bound ligand.

Using Eqs. (1) and (2) to define the partition function, expressions are now derived for average quantities, such as fraction of the base pairs that are helix, coil, or complexed with ligand, etc., as well as the average lengths of each such sequence. In general, if any base pair can have a configuration denoted by Y, with an assigned statistical weight y, then the average number of Y configurations per DNA molecule is[20,21] $\bar{Y} = (\partial \ln Z/\partial \ln y)$, or using Eq. (1) above, the average *fraction* of base pairs in configuration Y is $\bar{Y}/N = (\partial \ln x_1/\partial \ln y)$. For example, if Y is the helical configuration and hence $y = s$, then fraction helix = $(\partial \ln x_1/\partial \ln s)$. All average quantities do not have to be obtained by direct differentiation of the partition function, but rather can be obtained by implicit differentiation of Eq. (2), $f(x) = 0$, used to define x_1. Thus:

$$\frac{\bar{Y}}{N} = -\frac{(\partial f/\partial \ln y)}{(\partial f/\partial \ln x)} \tag{3}$$

where the derivatives are evaluated at $x = x_1$.[30] The formal derivatives of Eq. (3) are easily translated into more detailed expressions involving the assigned statistical weights (s, K_h, etc.).

Calculation Methods

To calculate a melting curve of a homogeneous polynucleotide in the presence of a ligand, the following procedure is used: (1) to describe the DNA at any one temperature, the parameters s and σ must be specified; alternatively, T_m, ΔH, and σ are specified, where T_m is the melting temperature of pure DNA and ΔH is the enthalpy change for melting a mole of base pairs; (2) to describe the ligand, the parameters that must be specified are: K_h, K_c, n_h, n_c, ω_h, ω_c, and L; (3) at any one temperature, the sequence-generating functions are constructed and Eq. (2) is solved numerically for the largest root x_1; (4) all average quantities, such as fraction helix, are calculated using Eq. (3); (5) the temperature is changed and the above processes are repeated until the entire melting curve is constructed.

To limit the number of calculations, a "standard DNA" is used throughout, and is defined by the following parameters: $T_m = 50°C$, $\Delta H = 8$ kcal/mol of base pair,[31] $\sigma = 10^{-4}$, and where required, a total DNA concentration of 5×10^{-5} M base pairs. These parameters give a DNA-only melting curve (fraction coil versus temperature) that is superficially similar

to that of poly[d(A-T)] in about 0.03 M Na^+, with a transition slope and average helix length measured at the T_m of about 0.7 ($°C^{-1}$) and 70 base pairs, respectively (see below). A shift in T_m of 40°C corresponds to a free energy change of about 1 kcal/base pair. Although the theoretical treatment given above is for an infinitely long DNA molecule, the length at which an actual DNA becomes experimentally "infinite" will depend upon the particular property being studied. For example, the melting temperature of a series of alternating d(A-T)oligomers approaches that of the polymer at about 50 base pairs,[32] and the T_m of natural DNA is unchanged above a molecular size of several thousand base pairs.[33] With cooperative ligands, however, much longer DNA molecules might be required before end effects can be neglected.[18]

Ligand parameters are chosen so as to cover the range of values that have been experimentally observed in several DNA–protein systems. Examples of the site size n_h found for helix-binding ligands are 1–2 base pairs for ions, dyes, etc.; 3–6 base pairs for drugs such as netropsin[34,35] and actinomycin;[36,37] 13 base pairs for the nonspecific binding of *lac* repressor (P. H. von Hippel, manuscript submitted for publication); probably 20–30 base pairs for histones and protamines;[38] and perhaps even larger for the highly polymerized basic polypeptides (although it is doubtful in this last case whether the ligands conform to the current model). Examples of the site sizes n_c found for coil-binding ligands are one base for chemical denaturants such as acid, alkali, and formaldehyde; 4–5 bases for fd gene 5-protein;[10,11,39,40] 10–12 bases for ribonuclease,[13,41] and 5–10 bases for the various unwinding proteins.[4,6,13,14,42] Thus for a protein, 10 bases or base pairs might be taken as roughly representative of the coil and helix site sizes, respectively. The binding constants (K_h and K_c) might be expected to range from very low (say $<10^3$ M^{-1} for gene 32-protein binding to helix DNA or *lac* repressor binding to coil DNA) up to at least 10^8–10^{10} M^{-1} (e.g., for gene 32-protein binding to coil[13,14] or *lac* repressor binding to helix DNA[43,44]). The binding constants of histones to DNA may be even several orders of magnitude higher. There is greater uncertainty in appropriate values to be used for ligand–ligand cooperativities, but ω_h and ω_c probably range from unity (for independent ligands) upwards to about 1000.[13,14,45] In the following calculations, all these parameters are taken to be temperature independent. A further important parameter in the calculations is the total ligand concentration; this may range from several molar, for ligands such as formaldehyde and small ions, down to 10^{-7} M or even below, for the large tightly binding proteins.

RESULTS AND DISCUSSION

Using the above theory, it is now possible to determine the effect that each feature of the ligand has on the shape and position of the melting transition of a homogeneous polynucleotide. To isolate the effect of one variable, such as ligand size, the calculations must be arranged in such a

way that the effects of other variables such as ligand binding constant, cooperativity, and total concentration, are maintained constant. This may lead to somewhat artificial and experimentally inaccessible choices of conditions and it will soon become apparent that other normalization procedures are possible. It is simplest first to consider either a pure helix stabilizer (i.e., one which has absolutely no affinity for the coil form of DNA and thus $K_c = 0$) or alternatively, a pure helix destabilizer (one which has no affinity for the helix form of DNA and thus $K_h = 0$). The more general case, where ligand is allowed to bind both to helix and to coil ($K_h \neq K_c \neq 0$), will be considered later.

Ligand Size

At the melting temperature of a homogeneous polynucleotide, the free energy of a helical base pair is equal to the free energy of a coil base pair. If a ligand that binds to the helix is now added to the system, the average free energy of a helical base pair is lowered. Thus, to reach a new T_m the temperature must be raised, to an extent depending on the free energy change per base pair caused by ligand binding. This quantity, the amount by which a helix stabilizer lowers the average free energy of a helical base pair, can be considered to consist of two parts: 1) an intrinsic free energy change, describing the molecular details of the actual binding process; and 2) a mixing or shuffling free energy change, arising from the number of different ways in which the bound ligands can be arranged on the helix. Consider, for example, a region of a DNA helix which is N_h base pairs long and to which are bound B_h helix stabilizers, each covering n_h base pairs. If, for the moment, all ligand binding is considered to be noncooperative (i.e., $\omega_h = 1$), the overall free energy change of these (independent) ligands binding to this stretch of helix can be written as:

$$\Delta G_{\text{bind}} = B_h \, \Delta G_{\text{int}} + \Delta G_{\text{mix}} \tag{4}$$

where ΔG_{int} is the intrinsic free energy change of binding one ligand, and is given by[24,27] $\Delta G_{\text{int}} = -RT \ln (K_h L)$; ΔG_{mix} is the shuffling or mixing free energy and is given by $\Delta G_{\text{mix}} = -RT \ln (\Omega)$, where Ω, the number of different ligand arrangements,[15] is:

$$\Omega = \frac{(N_h - B_h n_h + B_h)!}{(N_h - B_h n_h)!\, B_h!} \tag{5}$$

One way to compare ligands of different sizes is to keep the intrinsic free energy change per base pair of the complex the same for all sized ligands. Thus a ligand that complexes with 10 base pairs provides 10 times more intrinsic free energy of binding than does a ligand that complexes with one base pair. By designing the binding constants in this manner, the intrinsic free energy contribution will be the same for all sized ligands at the same fractional saturation of the helix. Thus the difference between ligands of different size will only appear through differences in ΔG_{mix}. From Eq. (5),

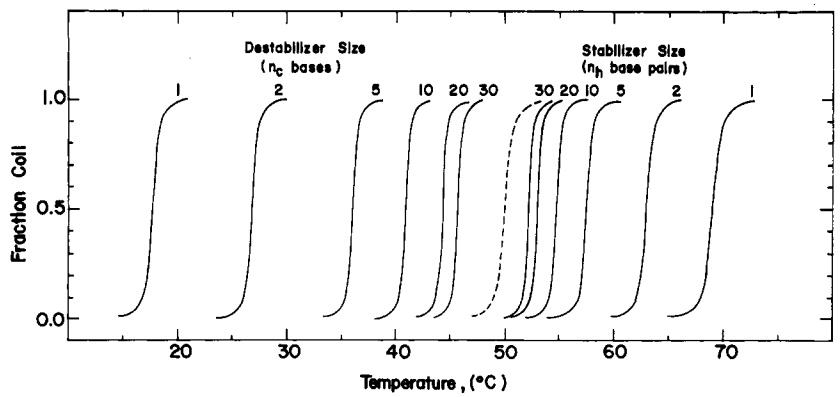

Fig. 2. Melting curves, plotted as fraction total coil vs. temperature, calculated for DNA in the presence of various sized helix stabilizers and helix destabilizers. The product of binding constant and free ligand activity (KL) is maintained constant at unity for all curves. The dashed line is the DNA-only transition, with T_m = 50°C.

ΔG_{mix} is seen to be a fairly strong function of ligand size, with smaller ligands able to be arranged in many more ways than equivalent numbers of larger ligands. Thus, even under conditions where the intrinsic free energy change per base pair of the complex is the same for all sized ligands, the overall DNA stabilization is expected to be greater, the smaller the ligand. The same type of effect is also expected with helix destabilizers, with smaller ligands being more effective than larger ligands in lowering the T_m.

The magnitude of this effect can now be investigated quantitatively. The intrinsic free energy change per base or base pair in the complex is given by $\Delta G_{int}/n = -(1/n)\, RT \ln (KL)$ (where, as noted earlier, lack of subscripts refers to either helix stabilizers or helix destabilizers). If this quantity is to be kept constant for all sized ligands, then $(KL)_{n=1} = \sqrt[n]{(KL)_{n\geq 1}}$. This condition is most simply met if the product KL is maintained at unity for all ligands. Figure 2 shows examples of such melting curves calculated, for both helix stabilizers and helix destabilizers, with site sizes ranging from one to 30 bases or base pairs, and all with $(KL) = 1$. (All these calculations are done with free ligand concentration or activity remaining constant throughout the transition.) As seen in Figure 2, the transition of the standard DNA melting in the absence of ligand takes place at 50°C. In the presence of a helix stabilizer, of size n_h = 1 base pair (and with KL = 1), the helix–coil transition temperature is increased to about 69°C. However, as expected from the above argument, with increasing ligand size, the helix is stabilized less and less until at n_h = 30 base pairs, the DNA T_m is increased only about 2°C. Clearly ligand size can exert a large effect on the DNA melting temperature, simply through changing the number of ways that the bound ligands can be arranged. The same type of effect is also seen with helix destabilizers, although in the current model and as will be discussed below, the magnitude of T_m shift is roughly twice that of a helix stabilizer at the same values of n and KL.

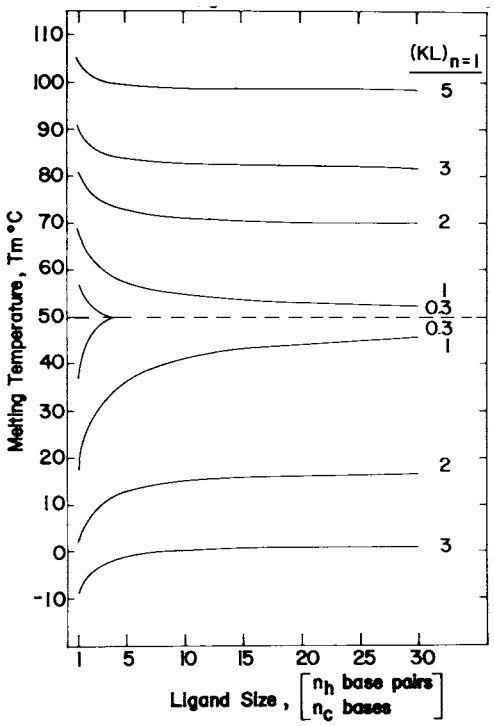

Fig. 3. The relation between melting temperature T_m(°C) and ligand size (either helix stabilizer or destabilizer), when the intrinsic free energy change per base pair of the complex is maintained constant. The DNA-only transition is at 50°C.

The effect of ligand size is also important at choices of KL other than unity. Figure 3 plots melting temperatures for various sized ligands, both stabilizers and destabilizers, calculated with $(KL)_{n=1}$ ranging from 0.3 to 5; for a ligand of size $n = 10$, this corresponds to $(KL)_{n=10}$ ranging from less than 10^{-5} up to about 10^7. It is seen that large ligands are always less effective than small ligands in perturbing the DNA melting temperature. However, Figure 3 also shows that the dependence of T_m on ligand size decreases with increasing KL. At high values of KL, the mixing free energy (which by design is the only distinction between different sized ligands) becomes a smaller proportion of the overall free energy change of ligand binding, both because the intrinsic free energy change increases (i.e., helix saturation at T_m increases with increasing KL) and because, at these high levels of lattice saturation, the number of ways in which the bound ligands can be arranged, and hence the mixing free energy, decrease.

Several secondary effects of ligand size on the helix–coil transition can also be detected in Figure 2 and in the melting curves used to construct Figure 3. In particular, with increasing ligand size, the slope of the melting curves, the average helix length, and the degree of helix (or coil) saturation at the T_m all increase. These effects can perhaps be most clearly illustrated by comparing different sized ligands, not at the same intrinsic free energy

change as was done above, but at the same melting temperature. For example, consider a DNA molecule in the presence of a helix stabilizer of size either $n_h = 1$ or $n_h = 30$ base pairs, but with the resultant T_m of 55°C in both cases. For a helix stabilizer of size $n_h = 30$, a value of $(KL)_{n=30}$ of 52 is required to raise the DNA T_m from 50° to 55°C; for a ligand of $n_h = 1$ on the other hand, a value for $(KL)_{n=1}$ of only 0.2 is required for the same degree of stabilization. At the T_m the DNA helix is over 80% saturated with the larger ligand whereas, with the smaller ligand, the helix is less than 20% saturated. In addition, the average length of a helix region is over twice as long for the larger ligands; from Eq. (5) above, it is seen that large ligands tend to "spread." (For the same reason, the degree of helix stabilization caused by a large ligand depends to some extent on the average helix length associated with the unperturbed DNA-only transition, as determined by the cooperativity parameter σ. However, for $\sigma < 10^{-3}$, the T_m becomes essentially independent of ligand size.) All the above effects reflect the difference between large and small ligands embodied in Eq. (4). If two (independently binding) ligands have different site sizes but yet cause the same T_m shift in DNA, i.e., they both contribute the same overall free energy change per DNA base pair, the free energy change contributed by the smaller ligand consists of a smaller intrinsic free energy change but a larger change in the free energy mixing than does the free energy change contributed by the larger ligand.

Intrinsic Binding Constant and Free Ligand Activity

A common experiment is to vary ligand activity and to observe any resulting changes in the DNA melting temperature. In this section, calculations are performed in just such a manner except that, since the association constant and ligand activity always appear together in the statistical weights, the product KL is varied rather than K or L individually. (The association constant K refers, as originally defined, to the overall intrinsic ligand binding constant with no compensation being made for different sized ligands, as was made in the previous section; L refers to the free ligand activity at the T_m; again the ligands are considered to bind noncooperatively.)

The relation between DNA T_m and the product KL is plotted in Figure 4 for both stabilizers and destabilizers, with site sizes ranging from 1 to 30 bases or base pairs. As expected, the absolute value of the shift in T_m increases with increasing ligand binding constant and/or ligand activity. At the same value of KL, small ligands are considerably more effective in shifting the T_m than are large ligands. (This is much more evident in Figure 4 than in Figure 3 since, as was just noted, the overall binding constants are now the same for the different sized ligands and are not normalized to the same intrinsic free energy change per complexed base or base pair.) The shapes of the curves are also different for different sized ligands, with the curves for large ligands being flatter at high KL. As noted earlier, the mixing free energy for large ligands rapidly decreases as the lattice

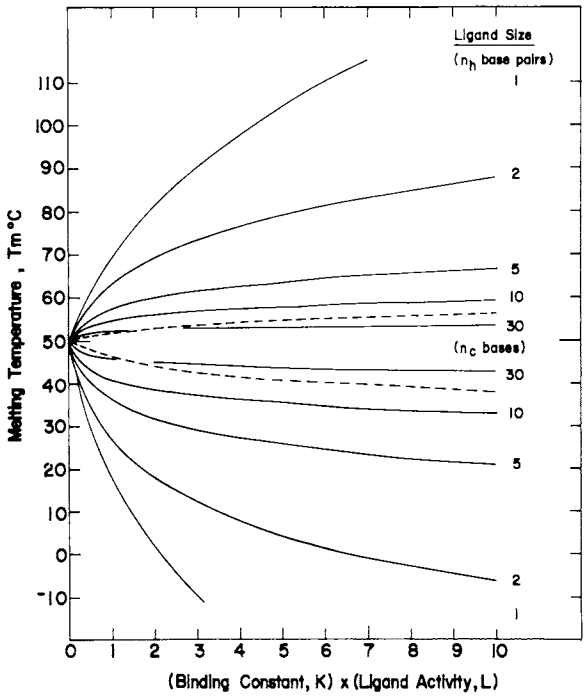

Fig. 4. DNA melting temperatures T_m(°C) plotted vs. the product of the ligand binding constant and free ligand activity, for various sized helix stabilizers and helix destabilizers. The dashed curves are plots of Eq. (6), calculated for ligands of size $n = 10$.

approaches saturation binding, reflecting the accumulation of gaps between bound ligands that are smaller than the ligand size.

The curves in Figure 4, for $n_h = 1$ base pair and $n_c = 1$ base, follow plots of an equation previously derived for a number of situations in which reversible ligand binding is coupled to a helix–coil transition.[23–28] In the present nomenclature:

$$\frac{1}{T_m^\circ} - \frac{1}{T_m} = \frac{R}{\Delta H} \ln \left\{ \frac{(1 + K_h L)^{1/n_h}}{(1 + K_c L)^{2/n_c}} \right\} \tag{6}$$

where for the moment $n_h = n_c = 1$ and either K_h or $K_c = 0$. T_m° is the unperturbed DNA melting temperature measured in the absence of ligand and both T_m° and T_m are expressed in degrees kelvin; ΔH is the enthalpy change for melting a base pair. The term containing $K_c L$ in Eq. (6) is squared since, for a ligand of size $n = 1$ in the current model, there are twice as many potential binding sites per base pair in the coil as in the helix. Thus, as can be seen in Figures 2–4, at equivalent values of KL, the T_m shift caused by a helix destabilizer is always substantially greater than that caused by a helix stabilizer.

With $n_h > 1$ or $n_c > 1$, Eq. (6) has also been used[24,28] to describe the response of DNA T_m to the binding of *large* ligands. However, this treats

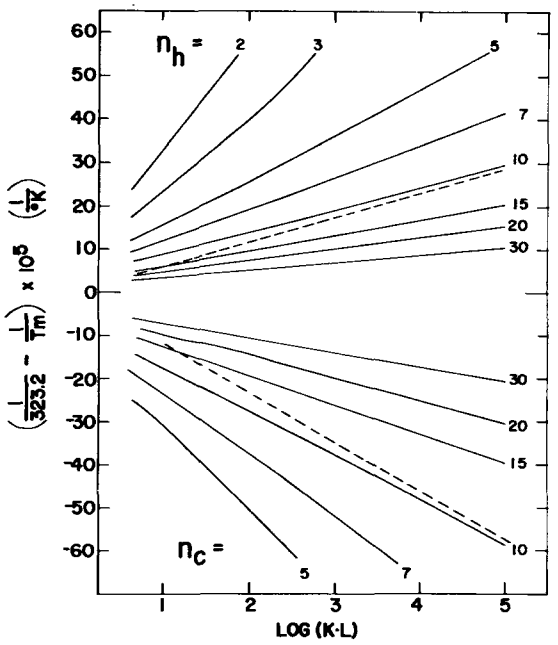

Fig. 5. The difference between the reciprocals of DNA melting temperatures (in degrees kelvin) with and without ligand, plotted vs. $\log_{10}(KL)$. Dashed lines are plots of Eq. (6) for a ligand size of 10.

large ligand binding only in the Scatchard approximation, which in essence stipulates that ligands can bind only at regularly spaced intervals of n_c bases or n_h base pairs. In the current model, any base or base pair can start a binding site and the bound ligands can be arranged in many more ways than are allowed by the Scatchard approximation, especially at low levels of lattice saturation. Thus eq. (6) should underestimate the effect that large ligands have on the T_m. The dashed lines in Figure 4 are calculated with Eq. (6) for either a helix stabilizer or destabilizer of size $n = 10$, and, as expected, are seen to deviate substantially from the curves produced by the more extensive calculations. In this range of KL, more error is introduced through using Eq. (6) than by a twofold error in the ligand size.

The curves of Figure 4 extend only to a value of $KL = 10$. However, with proteins that bind very tightly, values of KL as high as 10^5 or so might conceivably occur. As noted earlier, under such conditions the mixing free energy becomes a smaller fraction of the total free energy of interaction, and hence Eq. (6), which underestimates this mixing free energy, might be expected to become a better approximation as KL is increased. Figure 5 plots the calculated relations between DNA melting temperature and $\log_{10}(KL)$ for various sized stabilizers and destabilizers and for $10 \leq KL \leq 10^5$. The ordinate in Figure 5 is expressed as the difference between the reciprocals of the melting temperatures (in degrees kelvin) measured in the absence and presence of ligand. This function of T_m is directly pro-

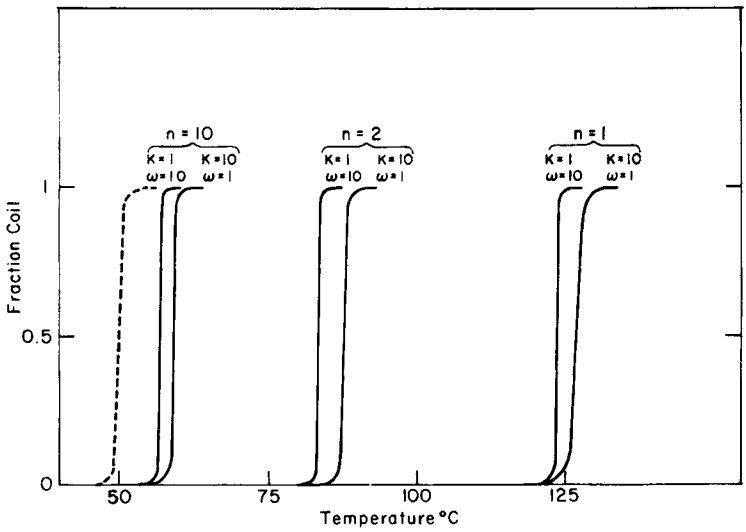

Fig. 6. The effect of ligand–ligand cooperativity on the shape and position of the DNA transition. Curves are calculated for helix stabilizers of size $n = 1, 2$, or 10 base pairs, each present at a constant free activity of 1 M.

portional to the stabilization or destabilization free energy, is approximately equal to $\Delta T_m \times 10^{-5}$, and, for $KL \gg 1$, is predicted by Eq. (6) to be linear in $\log_{10}(KL)$. The calculated plots are seen to be fairly adequately represented by straight lines, with slopes inversely proportional to the size of the ligand. The dashed lines are calculated from Eq. (6) for ligands of size $n = 10$, and are seen, as expected, to approach the more detailed calculations as KL is increased. Indeed, at $KL = 10^3$ and for $n = 10$, the T_m predicted by Eq. (6) would be less in error than would the T_m predicted by the detailed calculations but with a 10% error in ligand size.

If the ligand size can be determined (either from an independent titration experiment or, as discussed below, from the behavior of the melting curves at limiting ligand concentration), then Eq. (6) can readily be used to obtain a preliminary estimate of KL from the observed T_m. If KL is large (say $KL > 100$), this will be a fairly good estimate. If KL is low, curves such as Figure 4 should be used for a closer estimate of KL. If the free ligand activity at the T_m can be determined (either by some propitious optical property of the ligand, by working at ligand excess, or by estimating where the T_m would be if all added ligand remained unbound), the ligand binding constant can at least be estimated.

Ligand–Ligand Cooperativity

In previous sections, the calculations have been done with independently binding noncooperative ligands ($\omega_h = \omega_c = 1$). In this section, bound ligands are allowed to interact with their nearest neighbors.

For any size of helix stabilizer and for any particular value of K_hL, increasing ligand–ligand cooperativity ($\omega_h > 1$) leads to a further increase in T_m since ligands bound contiguously now have an added free energy of binding not previously available, i.e., their binding constant is effectively larger. For any value of ω_h, this effect is greater the smaller the ligand; i.e., the ratio of potential ligand–ligand contacts to ligand–DNA contacts is higher the smaller the ligand. Similar effects apply to helix destabilizers; increasing ω_c decreases the T_m.

Perhaps a better way to assess the effect of ligand–ligand cooperativity on the melting transition is to vary both KL and ω but to keep the product $KL\omega$ fixed; in this way, all ligands can potentially contribute the same amount of stabilization free energy to the DNA. Such melting curves are shown in Figure 6, and are calculated for a helix stabilizer of size $n_h = 1$, 2, or 10 base pairs, with either ($K_h = 10\ M^{-1}$, $\omega_h = 1$) or ($K_h = 1$, $\omega_h = 10$). Ligand activity is maintained constant at unity through each curve. In all cases, the curves calculated with independent ligands show a greater T_m shift than do the curves calculated with the cooperative ligands; in the former case, all ligands bind tightly whereas in the latter case, only those ligands bind tightly that also bind with nearest neighbors. With increasing ω_h, the melting curves also become steeper and the average length of a helix region at T_m increases since now the ligands tend to bind in long stretches. For example, at the T_m of the $n_h = 1$ curves of Figure 6, the average helix length is about threefold larger and the average number of contiguously bound ligands is about sixfold larger for the cooperative as compared to the noncooperative ligand.

If nearest neighbor ligands interact unfavorably, ω is less than unity and the binding is termed "anticooperative." In the current model, the effects of $\omega < 1$ are not very revealing and in the limit of $\omega = 0$, i.e., infinite nearest neighbor anticooperativity, the ligands simply appear one residue longer. Conversely, if the current model is used to fit data from a ligand that actually exhibits anticooperative binding (e.g., adjacent ligands repel each other electrostatically), this anticooperativity will be absorbed into an "effective site size" and the ligand will appear longer than its actual physical occlusion size.

As described briefly in the previous section, with independent ligands the magnitude of the T_m shift can be used to estimate the product KL, if the ligand size is known. The same procedure applied to cooperatively binding ligands will underestimate the product $KL\omega$ as can be seen in Figure 6; the amount of error introduced will depend upon the exact conditions.

Melting Transitions Calculated at a Constant Total Amount of Ligand

The melting curves in previous sections have all been calculated with the free ligand concentration (activity) constant throughout the transition; this

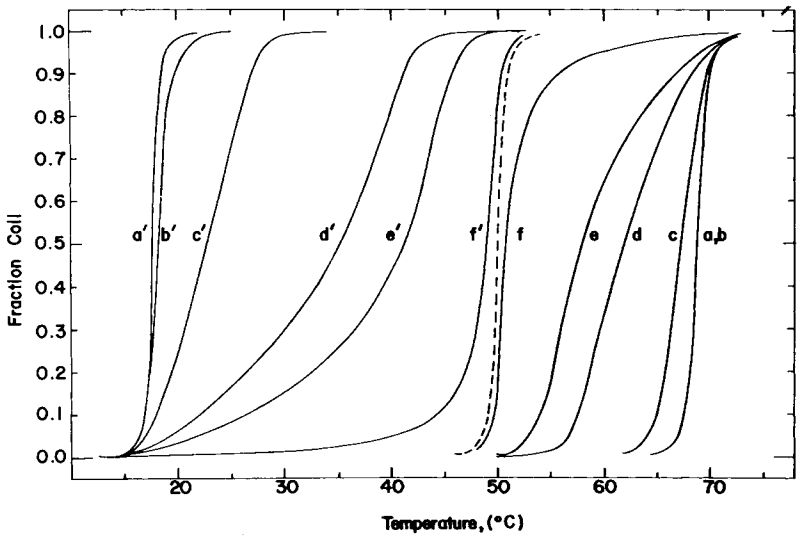

Fig. 7. The effect of total added ligand concentration on the shape and position of the DNA melting transition. Curves labeled with primed letters refer to a helix destabilizer of size n_c = 1 base; curves labeled with unprimed letters refer to a helix stabilizer of size n_h = 1 base pair. For all curves the product of the binding constant and total added ligand concentration is unity, as detailed in Table I. The DNA-only transition is the dashed curve at T_m = 50°C.

condition would apply if the melting took place inside a DNA-impermeable ligand-permeable membrane immersed in a large reservoir of free ligand, or, less contrived, if the ligand were present in great excess over the DNA such that the bound ligand concentration were a negligible fraction of the total ligand concentration. In DNA–protein systems, however, the total protein concentration and the total DNA concentration will usually be roughly comparable, and in the range of say one protein per 1–10^3 DNA base pairs. Since the protein binding will usually be quite tight, the amount of protein bound will certainly *not* be a negligible fraction of the total amount of protein added; in many circumstances most of the added protein will be bound, leaving very little ligand free in solution. Furthermore when the DNA helix and coil interconvert during a melt, the concentration of

TABLE I
Parameters Used in Figure 7

Curve	K (M^{-1})	L_{total} (M)
a,a′	1	1
b,b′	10^3	10^{-3}
c,c′	10^4	10^{-4}
d,d′	5×10^4	2×10^{-5}
e,e′	10^5	10^{-5}
f,f′	10^6	10^{-6}

bound and free ligand (and hence the statistical weight KL) will also be expected to change. To investigate how the melting curves reflect this phenomenon, an iteration cycle has to be added to the calculation method in order to adjust $L_{free} + L_{bound} = L_{total}$ at each temperature. The DNA concentration now becomes an important parameter and the following calculations have been done with the standard polynucleotide at a concentration of 5×10^{-5} molar base pairs (equivalent to an A_{260} of about 0.6).

Figure 7 illustrates the quite striking effects to be expected when the total ligand concentration is comparable to the total DNA concentration and the *free* ligand activity is allowed to change continuously during the melt. Curves are calculated for (independent) stabilizers and destabilizers, both of site size $n = 1$. In order to be able to compare, even roughly, calculations made over a large range of ligand concentrations, the curves of Figure 7 are calculated with the product of binding constant, and *total* ligand concentration fixed at unity; thus if the total added ligand is decreased, the binding constant is increased (see Table I). If the ligand concentration and binding constant are such that the amount of ligand bound is a negligible fraction of the total amount of ligand added, then all curves would be the same as the "constant-free-ligand" curves calculated above. This is seen to be the case in Figure 7 where, for total added ligand concentrations ranging from 1 to 10^{-3} M (with K thus ranging from 1 to 10^3 M^{-1}) the melting curves are almost super-imposable. In the other extreme, of very high binding constant but very dilute ligand, the curves approach the DNA-only curve, since there is very little ligand, even if tightly bound, to influence DNA behavior. Between these two extremes, however, the melting curves broaden and exhibit a variety of asymmetric shapes depending upon the exact conditions. For helix stabilizers this transition broadening is due, as has been realized for some time,[46,47] to ligand transferring from the melted helix to the remaining, as yet unmelted, helix, thereby causing it to be further stabilized; the free ligand concentration increases on going through the transition. For helix destabilizers, the explanation is analogous, except now the free ligand concentration decreases on going through the transition, and the coil regions become progressively less stabilized as temperature is increased.

For a helix stabilizer, all the added ligand is free in solution at the completion of the melting transition; for a helix destabilizer the same is true at the beginning of the transition. Thus each "constant-total-ligand" melting curve of Figure 7 has as an upper limit (for helix stabilizers) or as a lower limit (for helix destabilizers) a "constant-free-ligand" curve calculated as if all the added ligand remained free in solution. As remarked earlier, it is useful to bear these limits in mind when interpreting experimental melts.

The calculations of Figure 7 were arranged so as to cover a large range of both total ligand concentration and ligand binding constant. However in any experimental study of the effect of a particular ligand on the DNA helix–coil transition, the binding constant is (more or less) fixed and a series

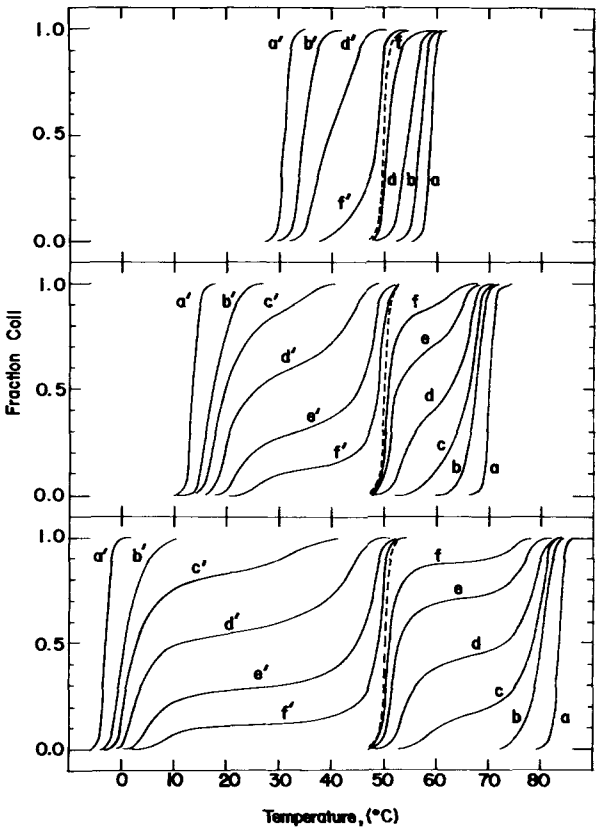

Fig. 8. Melting curves of DNA at a series of total added ligand concentrations, ranging from 10 to 200% potential saturation of either the helix or the coil. Site size of helix stabilizer is $n_h = 10$ base pairs, of helix destabilizer is $n_c = 10$ bases. The ligand binding constants are top: $K = 10^6\ M^{-1}$; middle: $K = 10^8\ M^{-1}$; bottom: $K = 10^{10}\ M^{-1}$. Total DNA concentration is $5 \times 10^{-5}\ M$ base pairs = $10^{-4}\ M$ bases. DNA transition is shown as dashed line. The total ligand concentration corresponding to each curve is given in Table II.

TABLE II
Parameters Used in Figure 8

	Stabilizer		Destabilizer	
Potential Saturation(%)	Curve	Concentration (M)	Curve	Concentration (M)
10	f	5×10^{-7}	f′	10^{-6}
25	e	1.25×10^{-6}	e′	2.5×10^{-6}
50	d	2.5×10^{-6}	d′	5×10^{-6}
75	c	3.75×10^{-6}	c′	7.5×10^{-6}
100	b	5×10^{-6}	b′	10^{-5}
200	a	10^{-5}	a′	2×10^{-5}

of melting curves are measured over a usually rather limited range of ligand-to-DNA input ratios. Curves calculated so as to mimic this situation are shown in Figure 8. A site size of $n = 10$ is chosen as representative of a protein, either a helix stabilizer or destabilizer, and three different binding constants are used: 10^6, 10^8, and 10^{10} M^{-1}; in all cases the ligand binding is noncooperative. For each binding constant, melting curves are calculated for input ratios of *total* ligand to DNA ranging from 10 to 200% *potential* saturation (i.e., if all the added stabilizer or destabilizer were bound, 10 to 200% of the DNA base pairs or bases, respectively, would be complexed). The total ligand concentration corresponding to each curve is given in Table II. As expected, with the same binding constant, increasing the total added ligand concentration increases the calculated shift in T_m; at the same total ligand concentration, increasing the binding constant also increases the calculated shift in T_m. As the total added ligand is increased from low levels, the melting curves can become very broad; however, as more ligand is added they eventually sharpen until at about 200% potential saturation (if not before) the curves have about the same slope as the DNA-only transition. (If such a series of curves are experimentally available, this observation can be used to estimate a ligand site size, at least to within a factor of 2.) However the most striking feature of Figure 8 is that, above a certain value of the binding constant, the calculated transition curves become very definitely biphasic, with the relative sizes of the two phases approximately reflecting the ligand-to-DNA input ratios. Furthermore, this biphasicity is observed in the complete absence of any cooperative interaction between the bound ligands.

To determine the origins of this noncooperative biphasicity, one particular curve from Figure 8 can be taken as an example and investigated in greater detail. Consider the 50% potential saturation curve for a helix stabilizer, curve d in the bottom panel of Figure 8, where the ligand binds to the helix with an intrinsic binding constant of 10^{10} M^{-1}, covers up ten base pairs in the complex, and is present at a total concentration of 2.5×10^{-6} M. (These particular parameter values actually describe rather closely the nonspecific low-salt binding of *lac* repressor to poly[d(A-T)].[44] Figure 9 plots four different aspects of this melting transition.

Consider first what the DNA molecule looks like at a temperature below the DNA-only transition, for example 45°C. The detailed calculations indicate that the fraction coil is essentially zero, essentially all of the added ligand is bound, and the helix is thus 50% saturated with bound ligand. The ligand is so tightly bound that the free ligand concentration, before the start of the transition, is calculated to be 2.4×10^{-11} M. When the temperature is raised, the DNA molecule can melt by two nonexclusive processes. The first type of process does not change the amount of bound ligand, i.e., the uncomplexed base pairs can melt around the bound ligands (by overcoming the DNA-only cooperativity) and/or the bound ligands can be rearranged to allow the DNA to melt in larger loops. The process that actually occurs will be some mixture of melting in small loops and rearranging the bound

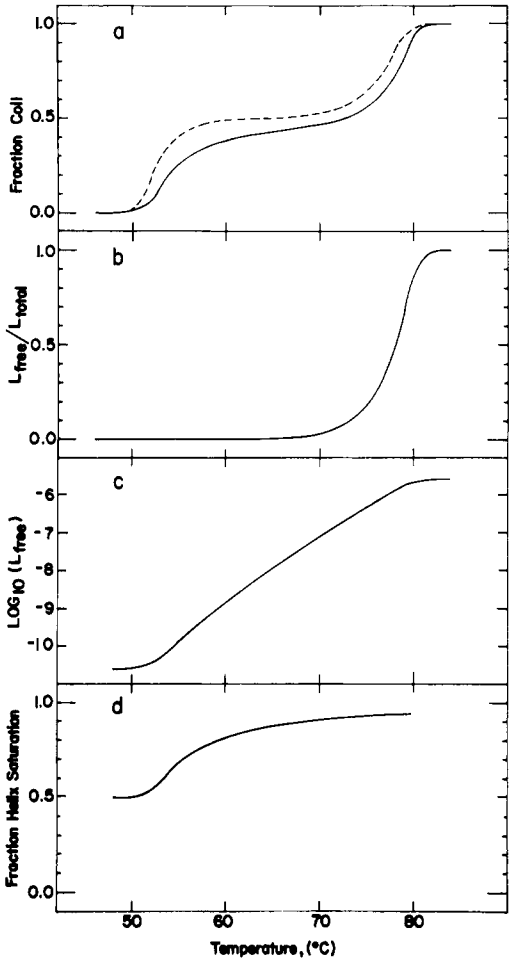

Fig. 9. Melting transition of DNA in the presence of a helix stabilizer, of size $n_h = 10$ base pairs, $K_h = 10^{10} M^{-1}$, and total ligand concentration of $2.5 \times 10^{-6} M$. (a) Fraction coil. (b) Fraction of total added ligand that is free in solution. (c) Logarithm of the free ligand concentration. (d) Fraction of saturation of the remaining helix with bound ligand. Total DNA concentration is $5 \times 10^{-5} M$ base pairs; DNA-only transition has $T_m = 50°C$. The dashed line in (a) is calculated using the approximation described in the text.

ligands but both ways require free energy and thus even the start of the transition occurs at higher temperatures than the DNA-only transition. From Figure 9a, it is seen that uncomplexed base pairs melt reasonably abruptly and the saturation of the remaining helix (Figure 9d) climbs concomitantly.

However, to melt more than 50% of the DNA molecule, a second process must obviously occur, in which the bound ligands are expelled from their complexes. Indeed, as seen on the logarithmic scale of Figure 9c, the free ligand concentration is increasing throughout the transition, even in the

earliest portion. This increase in free ligand concentration must lead, however, to a further stabilization of the remaining helix. From consideration of Figure 5 for a ligand of $n = 10$, a change in the product KL by one order of magnitude causes an increase in DNA stabilization of about 5°C. In the present case, since K is fixed, if free ligand increases by a factor of 10, the T_m of the remaining helix must similarly be increased by roughly 5°C. However, since the ligand is initially bound so tightly, and the free ligand concentration is initially so low, the free ligand concentration can increase by several orders of magnitude (and thus increase the helix stability by several tens of degrees) before the actual free ligand concentration is of the same order of magnitude as the DNA concentration. Only when this happens does the further release of bound ligand show up as an observable increase in melted DNA. The DNA stability depends roughly logarithmically on free ligand concentration, whereas the fraction of the DNA that is uncomplexed depends arithmetically; thus the second phase, when it occurs, is quite abrupt.

This same type of biphasicity is also expected with small ligands but only under experimentally unrealizable conditions. Thus for a stabilizer of size $n = 1$, added at a total concentration of $2.5 \times 10^{-5}\ M$ (50% potential saturation), a binding constant of $4 \times 10^{9}\ M^{-1}$ is required to have the initial free ligand concentration be $2.5 \times 10^{-10}\ M$ and thus comparable to the conditions of Figure 9. However from Eq. (6) it can be calculated that the end of the transition will now occur at about 4000°C.

The possibility of having biphasic transitions with independent ligands was suggested earlier by Crothers[24] from a consideration of the form of Eq. (6). Equation (6) however, besides being based on the Scatchard approximation, applies only to the T_m where the fractions of helix and coil are equal. With the Scatchard model, it is very straightforward to construct the DNA–ligand partition function in terms of binding constants, free ligand concentrations, etc., essentially as was done above for the more complete model; [actually the Scatchard model requires only that the statistical weight s of the nearest neighbor Ising model be replaced by the term $s^n(1 + KL)$].[23,24,28] If complete melting curves are then calculated with the same parameters as in Figures 8 and 9, adjusting to constant total ligand concentration at each temperature, biphasic curves are indeed obtained as predicted,[24] and as can be seen from the dashed lines in Figure 9a, are quite a reasonable approximation. Moreover the two phases of the melt are substantially more distinct than those calculated with the more extended model. In the Scatchard approximation, where discrete binding sites occur every n_h base pairs on the helix, each such binding site can saturate independently of its neighbor. Thus in the initial phases of the melt, the ligand can be readily rearranged and the remaining helix rapidly approaches complete saturation. In the more extended model however, as the initially uncomplexed DNA attempts to melt by rearranging the bound ligands, these transferred ligands interfere with each other, small

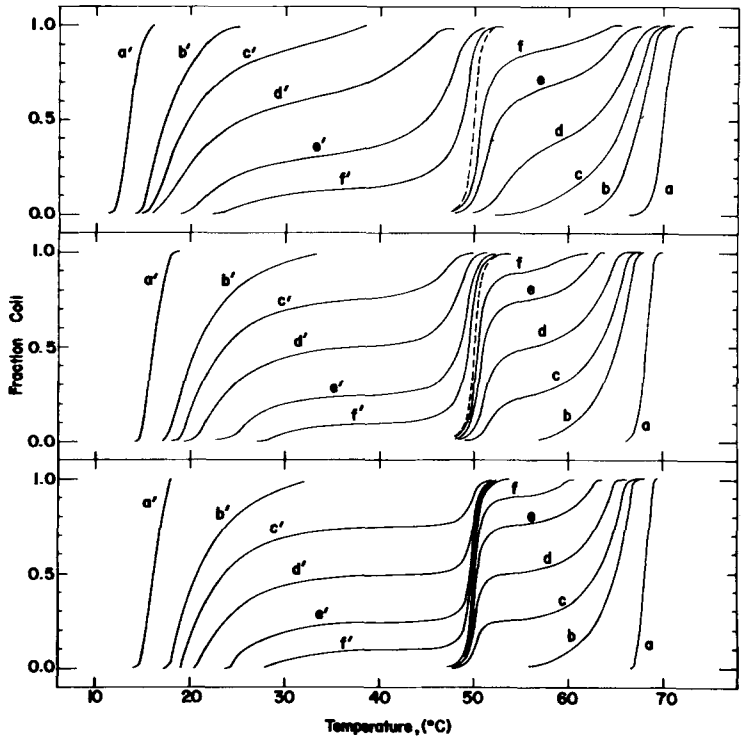

Fig. 10. Melting curves of DNA in the presence of various total added concentrations of either a helix stabilizer (n_h = 10 base pairs) or a helix destabilizer (n_c = 10 bases). Ligand concentrations range from 10 to 200% potential saturation of either helix or coil. Top: $K = 10^8\ M^{-1}$, $\omega = 1$; middle: $K = 10^6\ M^{-1}$, $\omega = 10^2$; bottom: $K = 10^4$, $\omega = 10^4$. Total DNA concentration is $5 \times 10^{-5}\ M$ base pairs. DNA-only transition is shown as dashed line. The curve labels and corresponding ligand concentrations are the same as in Fig. 8 and Table II.

gaps accumulate that are smaller than the ligand size, and the remaining helix saturates only gradually. As can be seen in Figure 9d, the saturation of the remaining helix reaches 90% only with difficulty.

Although these calculations demonstrate that biphasic melts cannot be taken as diagnostic of ligand–ligand cooperativity, there is ample evidence from other experimental techniques that some ligands do indeed interact in a cooperative fashion.[14,48–50] The effect of ligand cooperativity at constant total ligand concentration is examined in Figure 10. Helix stabilizers and destabilizers, both of site size n = 10, are used as examples and the total ligand input again varies from 10 to 200% potential saturation. In the three series of Figure 10, the product $K\omega$ is maintained at $10^8\ M^{-1}$ but ω is increased from 1 to 10^4 from top to bottom. It is seen that with increasing ligand–ligand cooperativity, although the overall T_m shift is somewhat less for reasons discussed earlier, the two phases of the melting curve become more distinct, with the middle region becoming quite flat and the end re-

gions (especially the limb closer to the DNA-only transition) becoming quite steep. The relation between curve shape and ligand cooperativity can be used to estimate ω from experimental melts.[13]

For these cooperative ligands, the free ligand concentration can be shown to follow a temperature profile similar to that of noncooperative ligands illustrated in Figure 9b; however helix saturation (for a stabilizer) now rapidly reaches 100% with increasing temperature; favorable ligand interactions offset mixing entropy effects and cause the bound ligands to cluster, allowing any uncomplexed DNA to melt unhindered. At the T_m of the 50% potential saturation curves of Figure 10, the average number of helix stabilizers bound in an uninterrupted string is about 2, 20, and 250 as ω increases from 1 to 10^2 to 10^4.

For a homopolynucleotide melting in the presence of less than saturating amounts of ligand, there are thus at least three causes of biphasic melting curves: 1) as discussed in the present section, large independent tightly binding ligands, which are nevertheless free to transfer and are at binding equilibrium throughout the transition; 2) "irreversible ligand binding" in which ligands cannot transfer, such as may occur with histones and polyamino acids bound to DNA at low salt;[48,51,52] 3) true cooperative interaction between bound ligands, whether mediated through DNA distortion or direct ligand–ligand contact. There are also a number of artifactual origins of melting biphasicity; one which must always be considered with proteins is the denaturation of the ligand with increasing temperature, leading to an abrupt "cooperative" change in the DNA transition. Thus it is seen that a biphasic melting curve is neither a sufficient nor a necessary requirement of ligand–ligand cooperativity.

An Alternative Model for Helix Stabilizers

In the model used throughout the present paper, ligands are counted as if they bind to base pairs in the helix. A concrete example would be a ligand bound in the major groove of DNA, such that any transverse section through the helix would intersect at the most one bound ligand. An alternative model would be needed if, for example, helix stabilizers bound to the DNA-phosphate backbone; in the simplest case the two backbones could bind independently of each other and thus any transverse section through the helix could intersect two bound ligands. The stabilizer site size should now be counted in bases rather than base pairs. Such a model could apply to many ligands, such as small ions.

The problem as to which model to use logically arises if melting curves are calculated using numerical parameters determined independently in isothermal titration experiments. If normalized as fractional lattice saturation, plots such as the Scatchard plot customarily used to estimate binding parameters would probably be experimentally indistinguishable for all sized ligands above a certain size, around five lattice residues long.[19] Thus, unless there is independent evidence, it is arbitrary whether the site

size of a helix-binding ligand is expressed as n base pairs or $2n$ bases. (The estimate of the intrinsic binding constant would however be the same for both normalizations.) By rederiving all the above theory, it can be shown that all the qualitative effects discussed in previous sections must also apply to this alternate model. However, for purely entropic reasons, a ligand that binds to $2n$ bases in the helix raises T_m by several degrees more than does a ligand that binds to n base pairs (but has the same intrinsic binding-constant).

The Melting of a Homogeneous Polynucleotide in the Presence of Ligands that Bind Both to Helix and to Coil

Previous sections considered ligands that bound exclusively either to the helix or to the coil form of DNA. Although this simplification is very convenient, in general a ligand can be expected to have at least some affinity for both forms of DNA; in addition, site sizes and cooperative interactions can in principle differ between helix and coil binding. In this section, this more general model is very briefly considered; only enough calculations are done to illustrate the added complexity and to indicate under what conditions the above helix-binding-only or coil-binding-only approximation can be expected to fail.

In the current model, a ligand binds to base pairs on the helix and to bases in the coil form. Since, at the melting temperature, there are twice as many coil bases as there are helix base pairs, the binding constant of a ligand to the coil form can be much smaller than the binding constant to the helix and still keep the overall T_m unperturbed. For example, for a ligand of size $n_h = n_c = 1$, Eq. (6) indicates that (for K_cL and $K_hL \gg 1$) the condition to keep T_m unshifted is $K_cL \sim \sqrt{K_hL}$. As can be seen from Eq. (5) above, there are many more possible ways of arranging the same number of bound ligands on the coil regions than on the helix regions, and this consideration becomes more important for larger ligands. For example, for a ligand of $n_h = n_c = 10$, the value of K_cL required to maintain T_m unperturbed ranges from only $\frac{1}{3}$ to $\frac{1}{2}\sqrt{K_hL}$. If ligand size also changes ($n_h \neq n_c$), at least the direction of the effect can be judged from Eq. (5). These shuffling entropy effects are less important for cooperative binding ligands, and as cooperativity increases, $K_c\omega_cL$ approaches $\sqrt{K_h\omega_hL}$.

A more concrete illustration of the antagonism between the helix- and coil-binding properties of a ligand is shown in Figure 11, for a ligand of site size $n_h = n_c = 10$ bases or base pairs, and present at a total concentration of $5 \times 10^{-6}\,M$ (100% potential helix saturation). In the top panel of Figure 11, the helix-binding constant K_h is kept at $10^6\,M^{-1}$ and melting curves are calculated with the coil-binding constant K_c being increased from 0 to $10^6\,M^{-1}$. It is seen that for $K_c \leq 10^2\,M^{-1}$ there is no detectable effect on the transition; however, for $K_c = 10^3\,M^{-1}$, the transition is perceptibly lowered; at $K_c = 10^4\,M^{-1}$, the lowering is substantial; and somewhere between $K_c = 10^5$ and $10^6\,M^{-1}$ the two binding processes balance and the T_m

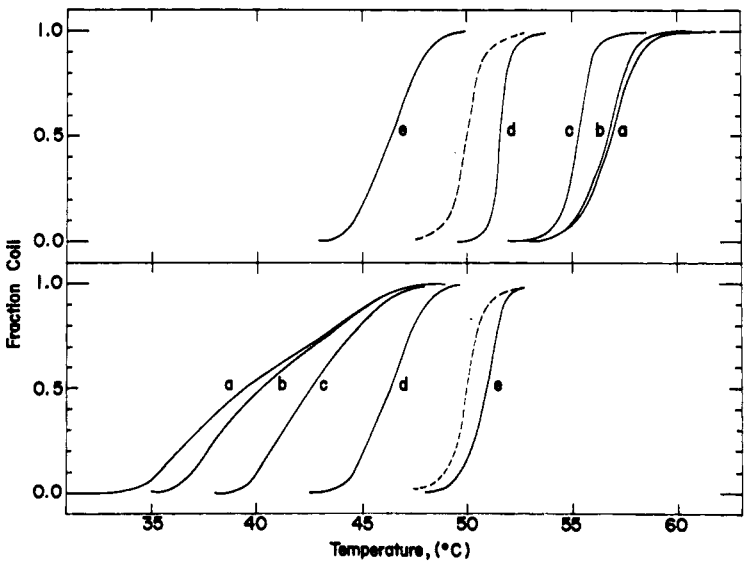

Fig. 11. The melting of DNA in the presence of a ligand that can bind both to helix and to coil. The ligand covers $n_h = 10$ base pairs on the helix and $n_c = 10$ bases on the coil, and is present at a total added concentration of 5×10^{-6} M. Top: $K_h = 10^6$ M^{-1}; K_c = curve a: 0, b: 10^3, c: 10^4, d: 10^5, e: 10^6 M^{-1}; bottom: $K_c = 10^6$ M^{-1}; K_h = curve a: 0, b: 10^4, c: 10^5 d: 10^6, e: 10^7 M^{-1}.

is unperturbed. Thus in this particular case, coil binding would have an experimentally observable effect when K_c is only 0.1% of K_h. The bottom panel of Figure 11 considers the opposite calculations, with K_c kept at 10^6 M^{-1} and K_h increased. In this case, a helix-binding constant of 10^4 M^{-1} is necessary before the transition is noticeably shifted, and a K_h of between 10^6 and 10^7 M^{-1} is necessary to balance the coil binding.

As discussed above, for a ligand of size $n = 10$ the rough condition for the T_m to be unperturbed is that $K_cL \simeq \frac{1}{2}$ to $\frac{1}{3}$ $\sqrt{K_hL}$ and thus the free ligand concentration must influence the interplay between helix and coil binding. For example, in Figure 11, the difference between the T_m at $K_c = 0$ and the T_m at $K_c = 10^4$ M^{-1} is roughly halved if the total ligand concentration is halved and roughly doubled if the total ligand concentration is doubled. In view of these added complexities it is reassuring that the paradigms of biological stabilizers and destabilizers, *lac* repressor and T4 gene 32-protein, respectively, show little affinity for the opposite form of DNA.

APPENDIX

The Melting of Poly[d(A-T)] in the Presence of Netropsin

The interaction of the antiobiotic netropsin with DNA has been extensively studied.[34,53,54] The results of these investigations indicated that the melting, at low ionic strength, of the synthetic polynucleotide poly[d(A-T)] in the presence of varying amounts of netropsin might be an appropriate experimental system with which to investigate certain aspects of the theory described in the main part of this paper.

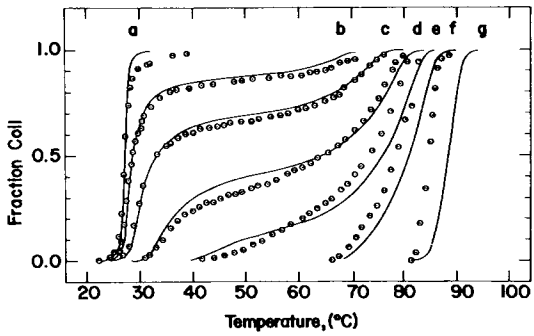

Fig. 12. Comparison between experimental and theoretical melting curves for poly[d(A-T)] melting in the presence of various amounts of netropsin. The points are experimental and the different curves correspond to the following input ratios of drug/base pair: curve a—0 (DNA-only control); curve b—1/36.6; curve c—1/14.9; curve d—1/7.6; curve e—1/5.1; curve f—1/3.7; curve g—1/1.9. The DNA concentration is 4.0×10^{-5} M base pairs for all curves. The solid lines are theoretically generated, using a site size of four base pairs and a helix-binding constant of 5×10^{8} M^{-1}.

Poly[d(A-T)] (sized at 9S), obtained from PL Biochemicals, was extracted twice with phenol and dialyzed into the buffer used for the melting experiments: 4×10^{-3} M cacodylic acid, 2×10^{-3} M NaOH, 1×10^{-4} M Na_2EDTA, pH 6.25. Netropsin (kindly provided by Drs. Roger Wartell and A. M. Craig) was stored at −20°C and dissolved in the above buffer just before use; drug concentrations were determined from the absorbance at 296 nm, using a molar extinction coefficient[34] of 2.02×10^{4} M^{-1} cm^{-1}. Solutions of poly[d(A-T)] (final concentration 4.0×10^{-5} M base pairs) and netropsin (final concentrations ranging from 1×10^{-6} to 21×10^{-6} M) were prepared by direct mixing, bubbled with helium, and transferred to Teflon-stoppered cuvettes. Melting transitions were observed at 260 nm in a Gilford 2000 automatic spectrophotometer, at a rate of temperature increase of about 0.3°C/min. Weight losses during the runs were about 0.5%.

To allow comparison with calculated melting curves, the data were normalized as fraction coil versus temperature, by means of the following manipulations. Absorbance readings were first corrected for thermal expansion of the solvent and then expressed as a fraction of the total absorbance change occurring between 20° and 95°C. (The average overall change in absorbance on melting was 0.30 ± 0.01 without drug, compared to 0.32 ± 0.02 with drug. The absorbance change brought about by the release of bound drug was small, contributing at the most about 10% of the total absorbance change of melting and usually much less. An absorbance change due to drug release could also be observed at 325 nm[34,53,54] and, in principle, could be used to correct the absorbance change at 260 nm to that representing DNA only; however this correction was usually negligible.) To remove the dependence of the coil absorbance on temperature, a least squares straight line was fitted to the poly[d(A-T)]-only transitions between 40° and 95°C and used as the fully melted limit at each temperature to correct all the other melts.[55] A series of such normalized curves was obtained with netropsin inputs ranging from 1/37 to 1/2 drug molecules per base pair, and is shown as the points in Figure 12. The solid lines in the figure are theoretically generated using the experimental DNA and netropsin concentrations, assuming that the drug does not bind to the coil form of DNA, and that the helix binding is noncooperative. Thus the only two adjustable parameters are the site size on the helix (n_h base pairs) and the helix binding constant K_h. The values used to generate the curves in Figure 12 are $n_h = 4$ base pairs and $K_h = 5 \times 10^{8}$ M^{-1}; the process of choosing these parameters will be discussed below.

Before using the model of independent helix stabilizers to interpret the curves of Figure 12, the following qualifications and alternative explanations must be considered.

1) It is unlikely that the biphasicity in the transitions of Figure 12 is due to cooperative ligand binding. In isothermal studies of netropsin binding to various DNA's, Wartell et al.[34]

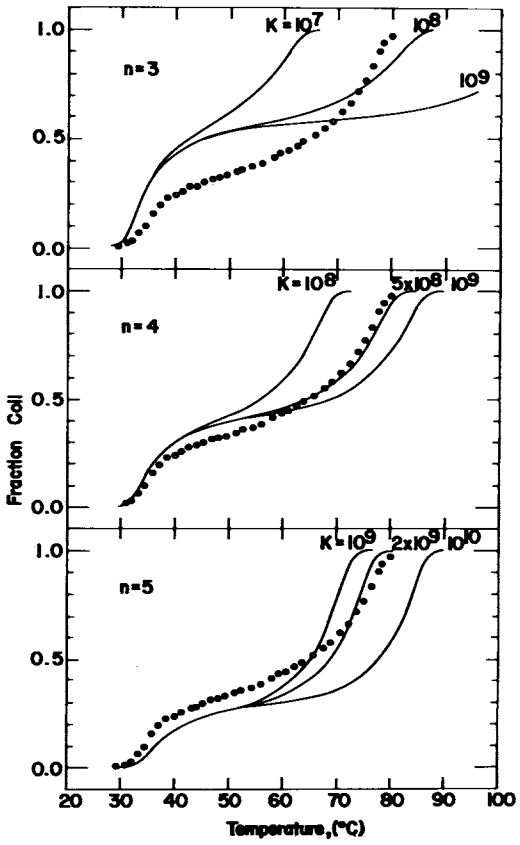

Fig. 13. Refinement of the numerical estimates of the model by computer simulation. The points represent the experimental melting curve obtained at 1 netropsin molecule/7.6 DNA base pairs. The solid lines are computer generated using a ligand site size of either n_h = 3, 4, or 5 base pairs and a range of association constants, as marked on each curve.

found no evidence for humped Scatchard plots, indicative of ligand–ligand cooperativity. Furthermore, when the solution of poly[d(A-T)] and about half-saturating netropsin used in the above melting experiments was sedimented in the analytical ultracentrifuge, there was no evidence for the two phases in the sedimentation boundary that might be expected if ligand binding were cooperative.

2) The observed biphasicity is not due to irreversible ligand binding but rather ligand is free to transfer throughout the melting transition. For example, if solutions are prepared with equal concentrations of poly[d(A-T)] and calf thymus DNA and roughly half-saturating levels of netropsin, the resulting melting curves are independent of the order of addition of DNA, polynucleotide, or drug. Furthermore, if a solution of calf thymus DNA and netropsin (about one drug per seven base pairs with virtually all being bound) is held at 35°C (i.e., above the poly[d(A-T)]-only transition, but below the poly[d(A-T)]–netropsin or calf thymus DNA–netropsin transitions) and poly[d(A-T)] is then added (also at 35°C and hence melted; at a final concentration equal to the calf thymus DNA), the poly[d(A-T)] regains hypochromicity within the experimental mixing time, about 30 sec. This shows that netropsin must be able to transfer and partition between the individual DNA molecules at a rate that is rapid relative to the heating rate and thus the experimental melts can be regarded as being at equilibrium with respect to ligand binding.

3) Wartell et al.[34] found no evidence, under their conditions, for any interaction of netropsin with single-stranded DNA. Although this is less certain under the present conditions of low ionic strength, this supports the assumption that $K_c = 0$.

4) Several experimental factors caution against over-interpretation of the melting curves. For example, netropsin is known to be unstable.[54] If a solution of netropsin is boiled for 10 min the absorbance at 296 nm is reduced by about 15%; if this same solution is then added to poly[d(A-T)], the melting curves are shifted to lower temperatures relative to the unboiled netropsin controls, as if about 15% of the drug had been destroyed. This effect is difficult to correct for, since under the melting conditions, the drug is usually completely bound to the DNA, where it is probably protected against thermal destruction. In any event, a sudden drug "denaturation" is certainly not the cause of the second melting phase since biphasic curves are also seen on cooling. A further source of error is that, at these low ionic strengths and slightly acid pH's, poly[d(A-T)] by itself undergoes some hydrolysis during the melting; (the S value drops by about 40% with 5–10% of the absorbance becoming acid soluble). For these two reasons, the transition curves are not exactly reproduced on cooling and the fraction coil values shown in the figures are probably accurate only to within about 5%.

5) Poly[d(A-T)] has the property of forming hairpins[56] and thus it may be an inadequate representative of the idealized infinitely long homopolynucleotide used in the theoretical model. At the moment it is difficult to assess the associated errors.

Within the above limitations, the melting curves of Figure 12 can now be interpreted quantitatively in terms of the simple noncooperative helix-binding-only model. The DNA-only transition is fit reasonably well using $\sigma = 10^{-4}$ and $\Delta H = 8$ kcal/bp; the same DNA parameters were used throughout the calculations of the text. Since the DNA concentration and drug concentrations are given, there are thus only two adjustable parameters, n_h and K_h, with which to fit the observed melting curves. These two parameters can be first estimated roughly and then their numerical values refined by computer simulation.

The curves of Figure 12 are at first very broad but sharpen with increasing ligand concentration. The slope of curve g is fairly close to the slope of the DNA-only transition and, according to observations made in the main text, this should occur at about 200% potential saturation of the DNA with ligand. The ligand input for curve g is 1 drug per 1.9 base pairs of DNA, and thus n_h can be roughly estimated as four base pairs (say three to five base pairs).

In curve g of Figure 12, all the added netropsin (2.1×10^{-5} M) is free in solution just at the completion of the transition, say 88°C. This temperature can be used to approximate the melting temperature that would have occurred if the free ligand concentration at the T_m were also 2.1×10^{-5} M. Using Eq. (6), with $n_h = 4$, $L = 2.1 \times 10^{-5}$ M, $T_m° = 27$°C, $T_m = 88$°C, allows K_h to be estimated as about 4×10^8 M^{-1}. [The product KL is about 10^4, and thus the use of Eq. (6) is justified.]

In Figure 13, the range of acceptable parameter values is explored more closely, by attempting to reproduce by computer the experimental curve with 1 drug per 7.6 base pairs, i.e., about half-saturation. Three site sizes are used, $n_h = 3$, 4, and 5 base pairs. For each site size, several values of the binding constants are used to generate the curves; outside this chosen range of binding constants, the fit becomes much worse. It is seen that a site size of $n_h = 4$ and a binding constant of around 5×10^8 M^{-1} best fit the data. The curves generated for $n_h = 5$ fit the data less well for all choices of K_h; the curves with $n_h = 3$ are considerably worse. From this example, it is seen that there is remarkably little latitude in choosing the adjustable parameters. For a curve such as was used in Figure 13, where the ligand is about half-saturating, the size or vertical position of the first phase of the melt is determined by the size of the ligand, almost independently of the choice of binding constant. Once n_h is determined K_h is quite accurately determined by the temperature (or horizontal position) of the second phase of the melt. The fit could certainly be improved by allowing nonintegral numbers of base pairs in the binding site, and a temperature-dependent binding constant; however both these refinements would seem to be presently unjustified.

The value of $n_h = 4$ base pairs for the site size can be compared to $n_h = 3$ base pairs determined by Wartell et al.[34] and $n_h = 5$ base pairs determined by Zasadatelev et al.[35] The value of the binding constant of 5×10^8 M^{-1} is about 100-fold higher than that determined by

Wartell et al.[34] (at about a 100-fold higher ionic strength), and coincides closely to the low ionic strength value estimated by Zimmer.[54] At the present stage, the agreement between the above theory and the melting behavior of at least this one selected model system, can be regarded as encouraging.

The author thanks Dr. Peter von Hippel for his support, advice and encouragement, and Dr. J. A. Schellman for his advice and criticism and for introducing him to the method of sequence-generating functions. He also thanks Ms. R. Hicks for the figures and Ms. Laura Lewis and Tim Kingan for their help in conducting the computer calculations. This study was supported in part by USPHS Research Grant GM-15792 (P. H. von Hippel, principal investigator).

References

1. Yamamoto, K. R. & Alberts, B. (1974) *J. Biol. Chem.* **249,** 7076–7086.
2. von Hippel, P. H., Revzin, A., Gross, G. A. & Wang, A. (1974) *Proc. Nat. Acad. Sci. U.S.* **71,** 4808–4812.
3. Alberts, B. M., Amodio, F. J., Jenkins, M., Gutmann, E. D. & Ferris, F. L. (1968) *Cold Spring Harbor Symp. Quant. Biol.* **XXIII,** 289–305.
4. Sigal, N., Delius, H., Kornberg, T., Gefter, M. L. & Alberts, B. (1972) *Proc. Nat. Acad. Sci. U.S.* **69,** 3537–3541.
5. Scherzinger, E., Litfin, F. & Jost, E. (1973) *Mol. Gen. Genet.* **123,** 247–262.
6. Reuben, R. C. & Gefter, M. L. (1973) *Proc. Nat. Acad. Sci. U.S.* **70,** 1846–1850.
7. Van der Vliet, P. C., Levine, A. J., Ensinger, M. J. & Ginsberg, H. A. (1975) *J. Virol.* **15,** 348–354.
8. Herrick, G. & Alberts, B. (1973) *Fed. Proc.* **32,** Abstract 497.
9. Hotta, Y. & Stern, H. (1971) *Develop. Biol.* **26,** 87–99.
10. Alberts, B., Frey, L. & Delius, H. (1972) *J. Mol. Biol.* **68,** 139–152.
11. Oey, J. L. & Knippers, R. (1972) *J. Mol. Biol.* **68,** 125–138.
12. Felsenfeld, G., Sandeen, G. & von Hippel, P. H. (1963) *Proc. Nat. Acad. Sci. U.S.* **50,** 644–651.
13. Jensen, D. E. (1974) Ph.D. dissertation, University of Oregon. Jensen, D. E., Kelly, R. C. & von Hippel, P. H., manuscript submitted.
14. Alberts, B. M. & Frey, L. (1970) *Nature* **227,** 1313–1318.
15. Latt, S. A. & Sober, H. A. (1967) *Biochemistry* **6,** 3293–3306.
16. Crothers, D. M. (1968) *Biopolymers* **6,** 575–584.
17. Zasedatelev, A. S., Gurskii, G. V. & Volkenshtein, M. V. (1971) *Mol. Biol. (Russ.)* **5,** 245–251.
18. Schellman, J. A. (1974) *Israel J. Chem.* **12,** 219–238.
19. McGhee, J. D. & von Hippel, P. H. (1974) *J. Mol. Biol.* **86,** 469–489.
20. Poland, D. & Scheraga, H. A. (1970) *Theory of Helix-Coil Transitions in Biopolymers,* Academic, New York.
21. Birshtein, T. M. & Ptitsyn, O. B. (1966) *Conformations of Macromolecules,* Interscience, New York.
22. Litan, A. (1966) *J. Phys. Chem.* **70,** 3107–3115.
23. Lazurkin, Yu. A., Frank-Kamenetskii, M. D. & Trifonov, E. N. (1970) *Biopolymers* **9,** 1253–1306.
24. Crothers, D. M. (1971) *Biopolymers* **10,** 2147–2160.
25. Schellman, J. A. (1958) *J. Phys. Chem.* **62,** 1485–1494.
26. Peller, L. (1959) *J. Phys. Chem.* **63,** 1199–1206.
27. Bixon, M. & Lifson, S. (1966) *Biopolymers* **4,** 815–821.
28. Frank-Kamenetskii, M. D. & Karapetyan, A. T. (1972) *Mol. Biol. (Russ.)* **6,** 621–627.
29. Lifson, S. (1964) *J. Chem. Phys.* **40,** 3706–3710.
30. Bradley, D. & Lifson, S. (1968) in *Molecular Associations in Biology,* Pullman, B., Ed., Academic, New York, pp. 261–270.

31. Scheffler, I. E. & Sturtevant, J. M. (1969) *J. Mol. Biol.* **42,** 577–580.
32. Scheffler, I. E., Elson, E. L. & Baldwin, R. L. (1970) *J. Mol. Biol.* **48,** 145–171.
33. Crothers, D. M., Kallenbach, N. R. & Zimm, B. H. (1965) *J. Mol. Biol.* **11,** 802–820.
34. Wartell, R. M., Larson, J. E. & Wells, R. D. (1974) *J. Biol. Chem.* **101,** 6719–6731.
35. Zasedatelev, A. S., Gurskii, G. V., Zimmer, Ch. & Thrum, H. (1974) *Mol. Biol. Rep.* **1,** 337–342.
36. Sobell, H. M. & Jain, S. C. (1972) *J. Mol. Biol.* **68,** 21–34.
37. Müller, W. & Crothers, D. M. (1968) *J. Mol. Biol.* **35,** 251–290.
38. Phillips, D. M. (1971) *Histones and Nucleohistones,* Plenum, London.
39. Day, L. A. (1973) *Biochemistry* **12,** 5329–5339.
40. Anderson, E., Nakashima, Y. & Konigsberg, W. (1975) *Nucl. Acid Res.* **2,** 361–371.
41. Raju, E. V. & Davidson, N. (1969) *Biopolymers* **8,** 743–755.
42. Weiner, J. H., Bertsch, L. L. & Kornberg, A. (1975) *J. Biol. Chem.* **250,** 1972–1980.
43. Lin, S.-Y. & Riggs, A. D. (1972) *J. Mol. Biol.* **72,** 671–690.
44. Revzin, A., Wang, A. & von Hippel (1975), in preparation.
45. Kelly, R. C. (1975) Ph.D. dissertation, University of Oregon.
46. Dove, W. F. & Davidson, N. (1962) *J. Mol. Biol.* **5,** 467–478.
47. Lerman, L. S. (1964) *J. Cell. Comp. Physiol.* **64,** Supl. 1, 1–18.
48. Olins, D. E., Olins, A. L. & von Hippel, P. H. (1967) *J. Mol. Biol.* **24,** 151–176.
49. Shapiro, J. T., Leng, M. & Felsenfeld, G. (1969) *Biochemistry* **8,** 3219–3232.
50. Shapiro, J. T., Stannard, B. S. & Felsenfeld, G. (1969) *Biochemistry* **8,** 3233–3241.
51. Ansevin, A. T. & Brown, B. W. (1971) *Biochemistry* **10,** 1133–1142.
52. Li, H. J. & Bonner, J. (1971) *Biochemistry* **10,** 1461–1470.
53. Zimmer, Ch., Reinert, K. E., Luck, G., Wähnert, U., Löber, G. & Thrüm, H. (1971) *J. Mol. Biol.* **58,** 329–348.
54. Zimmer, Ch. (1975) *Progr. Nucl. Acid Res. Mol. Biol.* **15,** 285–318.
55. Bloomfield, V. A., Crothers, D. M. & Tinoco, I. (1974) *Physical Chemistry of Nucleic Acids,* Harper & Row, New York, p. 301.
56. Spatz, H. Ch. & Baldwin, R. L. (1965) *J. Mol. Biol.* **11,** 213–222.

Received October 10, 1975
Returned for revision December 17, 1975
Accepted January 28, 1976