Investigations on the alpha-helix to coil transition in HP heterogeneous polypeptides using network rigidity

2004 J. Phys.: Condens. Matter 16 S5035

(http://iopscience.iop.org/0953-8984/16/44/001)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **16** (2004) S5035–S5046 PII: S0953-8984(04)85848-X

Investigations on the alpha-helix to coil transition in HP heterogeneous polypeptides using network rigidity

M S Lee, G G Wood and D J Jacobs[1]

Department of Physics and Astronomy, California State University Northridge, Northridge, CA 91330, USA

E-mail: donald.jacobs@csun.edu

Received 1 September 2004
Published 22 October 2004
Online at stacks.iop.org/JPhysCM/16/S5035
doi:10.1088/0953-8984/16/44/001

Abstract

The alpha-helix to coil transition in homogeneous polypeptides has recently been described using a distance constraint model (DCM) that employs network rigidity as an underlying mechanical interaction. The DCM accounts for intramolecular hydrogen bonding, hydrogen bonding to solvent and hydration. These interactions are parametrized to reflect the dependence on the conformational state of the polypeptide backbone. The DCM is capable of describing both heat and cold denaturation. As a function of temperature from low to high, a generic re-entrant response of a homogeneous polypeptide chain in aqueous solution is predicted to make a transition from coil (hydrated state) to helix and then to coil (disordered state). Here we study the thermodynamic stability of heterogeneous polypeptides that include hydrophobic (H) and polar (P) residue types. We explore the nature of the transition by adjusting the overall HP composition using transfer matrix methods that take into account long-range effects due to network rigidity.

1. Introduction

The alpha-helix to coil transition in polypeptide chains is a well studied problem. Beginning with the seminal work of Doty and Yang [1] in the 1950s, experimental studies have found many surprises [2], and more continue to appear [3] because intrinsic complexity is being uncovered by studying different polypeptides (both natural and synthetic) under different kinds of solvent conditions with ever improving methodologies. Likewise, computational techniques using molecular dynamics simulations [4–6] are providing detailed information at the molecular level. Alternatively, coarse grain Ising-like models, such as the Zimm–Bragg [7] and Lifson–Roig [8] models, have proved useful for understanding the essential features

[1] Author to whom any correspondence should be addressed.

of the helix–coil transition found in polypeptides. Many embellishments have since been considered [2, 9, 10] that take into account a variety of additional interactions. Nevertheless, the underlying formalism is based on two kinds of parameters that reflect a nucleation process of formation, then propagation, of helical structure.

Throughout the biochemistry literature and in standard physical-chemistry textbooks, language adopted from early works [11] classifies parameters as either describing the nucleation or propagation of helix. In spite of known inconsistencies [12, 13] between predictions and experiments, and misconceptions of the meaning of the nucleation parameter [14], the Lifson–Roig model (and its derivatives) are in widespread use—due to the combined qualitative correctness and quantitative simplicity. Perhaps fewer surprises would appear if Ising-like models were improved fundamentally. Therefore, we pose the question: what improvements can be incorporated in Ising-like models to more correctly capture the essential physics?

We suggest that the answer lies in taking proper account of network rigidity, which we formulate in terms of a distance constraint model (DCM). Upon the celebration of the 60th birthday of Michael Thorpe, an advocate of rigidity theory and its widespread applications [15], it is fitting to say that given the above question, it would have been difficult to find the answer [13] in the absence of rigidity theory. Moreover, we found additional open problems in the biochemistry literature involving the notion of free energy decomposition. In physical chemistry it is found that additivity of free energy decomposition works well in applications to small molecules, but in biochemistry, additivity generally fails for component entropies and free energies [16, 17]. The common practice of assuming additivity in free energy decompositions in coarse grain models has been identified as a problem that can lead to breaking the second law of thermodynamics when modelling cold denaturation [18]. By invoking network rigidity as an underlying mechanical interaction that is explicitly accounted for in the DCM, the open problems plaguing free energy decompositions in our opinion are also resolved.

We claim that proper application of network rigidity is the answer to the above-stated problems. The DCM employs a computationally tractable approximation scheme [19] that restores the utility of a free energy decomposition. It has been found that the DCM describes a large collection of experimental thermodynamic data [20–22] on polypeptides and proteins markedly well. The DCM, however, is evolving to capture more experimental complexity, such as differences in the basic 20 amino acids. Here, we extend prior work to heterogeneous polypeptides. In this paper a brief review of the DCM is given with emphasis placed on features that result from polypeptide heterogeneity. A theoretical description is presented that involves only two kinds of residues labelled as H (hydrophobic) or P (polar). Predictions for thermodynamic response functions such as helix content and heat capacity for different chain lengths and compositions are given with discussions and then conclusions.

2. The distance constraint model (DCM)

Applying a free energy decomposition directly in terms of specific interaction type, denoted by t, allows a microscopic partition function, Q_t, to be defined. The corresponding microscopic free energy is given by $G_t = -RT \ln Q_t$, where R is the universal gas constant and T is absolute temperature. A common modelling assumption (even in the current literature) is that the total free energy of a conformation (denoted by $\mathcal{F}$) is given as $G(\mathcal{F}) = \sum_t N_t G_t$, where N_t is the number of interactions of type t present in the system. Both N_t and G_t depend on conformation $\mathcal{F}$. The linearity assumption corresponds to independent component parts, such that $Q(\mathcal{F}) = \prod_t Q_t^{N_t}$. For the solid state, such a free energy decomposition is routinely made using a normal mode analysis, where each interaction type can be assigned to a particular mode. This is possible because phase space volume is decomposed into

orthogonal coordinates. More generally, linearity of free energy components is valid whenever a decomposition reduces coupling between subsystems to a negligible level. Interaction types identified by local conformational properties will generally strongly couple to one another rendering a product function for $Q(\mathcal{F})$ to be incorrect [23]. Consequently, total conformational entropy is overestimated because phase space is not partitioned into disjoint parts, which results in 'double counting' of regions of configuration space.

Network rigidity is explicitly calculated to obtain an approximation for $Q(\mathcal{F})$. To accomplish this task, the interactions are not only assigned a free energy, G_t, but are also represented as a constraint on the system. Each constraint is assigned one or more elementary distance constraint(s). The conformation of a biopolymer is then specified by constraint topology representing a generic mechanical framework, hence the use of $\mathcal{F}$ to label conformations. The coarse grain nature of the DCM expresses constraint type t in terms of discrete states using Ising-like variables η_{α_t} and σ_{α_t}. The α_t index is used to denote a particular constraint of interaction type t. The variable η can be 1 or 0 if the constraint is present or not respectively. The variable σ can take integer values between 0 and m_t, where m_t gives the maximum number of independent distance constraints possible. The σ-variables are non-trivial to determine for specified $\mathcal{F}$, requiring application of generic rigidity calculations.

Decomposing G_t into enthalpy and entropy components and treating all elementary distance constraints within constraint type t equivalently, yields

$$G_{\alpha_t}(\mathcal{F}) = H_t\eta_{\alpha_t} - RT\frac{S_t}{m_t}\sigma_{\alpha_t}. \tag{1}$$

Here H_t, S_t respectively represent the enthalpy and maximal entropy contribution from constraint type t. Since no pressure dependence is considered here, the enthalpy term will be regarded as energy. The entropy contribution is divided into discrete levels of contributions, ranging from S_t when the constraint is completely independent down to 0 if the constraint is completely (all its elementary distance constraints) redundant. Choosing ϵ_t in place of H_t and γ_t in place of S_t/m_t yields a generic equation for the free energy contribution of framework $\mathcal{F}$ given as

$$G(\mathcal{F}) = \sum_t \epsilon_t N_t(\mathcal{F}) - RT\sum_t \gamma_t I_t(\mathcal{F}) \tag{2}$$

where $N_t(\mathcal{F}) = \sum_{\alpha_t}\eta_{\alpha_t}$ gives the total number of constraints of type t present in the framework, and $I_t(\mathcal{F}) = \sum_{\alpha_t}\sigma_{\alpha_t}$ gives the corresponding total number of independent distance constraints. $G(\mathcal{F})$ is the total free energy of a conformation that accounts for degeneracy in atomic coordinates consistent with the specified constraint topology.

From equation (2) the total energy, $E(\mathcal{F})$, is given by the first sum. Additivity of energy is valid and does not pose any problems. The second summation is over a set of independent constraints to account for non-additivity of entropy components. Unlike a normal mode decomposition, an independent set of constraints will not generally provide an orthogonal basis. As a result, not all 'double counting' of configuration space is eliminated. Therefore, only an upper bound estimate for total conformational entropy can be obtained, which depends on the selected independent set of constraints, $\{I_t\}$. However, all independent constraint sets provide upper bound estimates. Therefore, a preferential set of constraints, $\{I_t^{(\mathrm{p})}\}$ is determined by requiring distance constraints with the lowest pure entropy components to be placed prior to constraints having greater entropy components. This procedure is mathematically well defined [24], yielding a lowest upper bound estimate.

Separating equation (2) into parts, it follows that

$$E(\mathcal{F}) = \sum_t \epsilon_t N_t \qquad \text{and} \qquad S(\mathcal{F}) = R\sum_t \gamma_t I_t^{(\mathrm{p})}. \tag{3}$$

The partition function is constructed over an ensemble of accessible constraint topologies given by

$$Q = \sum_{\mathcal{F}} \mathrm{e}^{-\beta G(\mathcal{F})} = \sum_{\mathcal{F}} \mathrm{e}^{\tau(\mathcal{F})} \mathrm{e}^{-\beta E(\mathcal{F})} \tag{4}$$

where $\beta = 1/RT$ and $\tau(\mathcal{F}) = S(\mathcal{F})/R$. The free volume [25] accessible to the network of constraints is quantified by the degeneracy factor e^{τ}. The ensemble of constraint topologies reflect constraints breaking and forming. Physically, equation (3) expresses an enthalpy–entropy compensation mechanism tied directly to network rigidity. Redundant constraints stabilize rigid regions at low temperature but are destabilizing at high temperatures. Determination of whether a constraint is redundant or independent depends on all constraints within the framework and the long-range character of network rigidity [26, 27]. The net effect of competing constraint topologies governs molecular cooperativity, including dramatic topological rearrangements associated with structural changes in folding and unfolding. The sub-ensemble of most probable constraint topologies characterize the state of a biopolymer, and fluctuations about the equilibrium state are related to flexibility.

2.1. Free energy decomposition for polypeptide chains

Each amino acid (residue) has a hydrogen bond (H bond) acceptor atom, a donor atom, and a side chain functional group that distinguishes it as a specific amino acid. The molecular structure is constrained by inter-residue and intra-residue covalent bonds such that two *degrees of freedom* (dof) remain along the backbone. The two dof correspond to the phi and psi bond angles[2]. The angles phi and psi mostly fall into well defined Ramachandran [28] regions corresponding to secondary structure in proteins.

We employ a free energy decomposition scheme used in prior work [20] that successfully describes experimental data on heat and cold denaturation in polypeptides in mixed solvent conditions [29]. The conformation of the backbone is discretized into three states: alpha-helical, coil, or hydrated, denoted by letters a, c, and h respectively. The first two states are invoked in the Lifson–Roig model. The hydrated state accounts for a structural cage or clathrate of solvent molecules surrounding the residue. Clathrate formation is energetically favourable, but decreases polypeptide entropy.

Constraint types and their parameters are defined in table 1. Covalent bonds are quenched and have the lowest entropic weights. Therefore, covalent bond constraints are preferentially assigned as independent before all other constraint types. Generally, fluctuating constraint types are assigned energy and entropy parameters. However, if the constraint type is always redundant for all network topologies, then its entropy parameter has no effect on the calculations for conformational entropy. H bonding to unstructured solvent is modelled as such a constraint because a decrease in polypeptide flexibility from mobile solvent is not expected. Other interactions, such as electrostatics, side chain interactions, and self-avoidance are neglected as is done in the Lifson–Roig model. However, unlike for all prior models, qualitative discussions of rigidity and flexibility of polypeptides found in earlier works [7, 8, 11, 30] are replaced with explicit calculations to obtain better estimates for the conformational entropy.

The backbone H bond parameters (U_{xyz}, γ_{xyz}) and U_0 are considered to be independent of residue type. All intramolecular H bond parameters used in prior works [19, 20] are employed here. The three consecutive conformations given as {aac, aca, caa} are treated equivalently with respect to intramolecular H bonds. Likewise {cca, cac, acc} are treated equivalently.

[2] Proline is an exception that we do not consider here. Proline has one degree of freedom because a fivefold covalent bonded ring connects the side chain back to the main chain.

Table 1. Constraint list: covalent bonds are quenched requiring no parametrization. The third column gives m_t, the maximum number of distance constraints. For intramolecular H bonding, x, y, and z can be a or c conformations of three consecutive residues that the H bond spans. No loss of generality occurs by setting $V_a = 0$ and $U_{ccc} = 0$ as arbitrary reference energies. Similarly, $\gamma_{aaa} \equiv 2$ is an arbitrary entropy reference.

Type of interaction	Energy	Entropy	Constraints
Covalent bonds	—	—	$\surd$
Coil conformation	V_c	δ_c	2
Helical conformation	$V_a \equiv 0$	δ_a	2
Hydration shell	V_h	δ_h	2
Intramolecular H bond	U_{xyz}	γ_{xyz}	3
Solvent H bond	U_0	—	—

The intramolecular H bond entropy parameters are: $\gamma_{ccc} = 2.92$, $\gamma_{cac} = 2.76$, $\gamma_{aca} = 2.15$, and $\gamma_{aaa} \equiv 2$. The energy parameters are: $U_{ccc} \equiv 0$, $U_{cac} = -2.34$, $U_{aca} = -2.83$, and $U_{aaa} = -4.64$. These parameters were originally obtained by fitting DCM predictions to simulation data [4, 5] on polyalanine chains in different environments (vacuum and water). The H bond to solvent energy parameter, U_0, was shown to be strongly dependent on the solvent conditions. Here it is selected to be $U_0 = -2$ kcal mol^{-1}. The five parameters $\{V_c, V_h, \delta_c, \delta_a, \delta_h\}$ need to be specified for each residue type. For H residues the parameters are: $V_c = 1.1$ kcal mol^{-1}, $V_h = -2.8$ kcal mol^{-1}, $\delta_c = 3.6$, $\delta_a = 2.6$, and $\delta_h = 0.7$. For P residues the parameters are: $V_c = 0.3$ kcal mol^{-1}, $V_h = -3.4$ kcal mol^{-1}, $\delta_c = 4.0$, $\delta_a = 1.8$, and $\delta_h = 0.7$. These numerical values were selected to produce a working example having qualitative features commonly observed in experiments on polypeptides in aqueous solution. Our focus is on presenting general properties that follow from the DCM formalism.

2.2. The transfer matrix approach for heterogeneous polypeptide chains

For the free energy decomposition scheme presented in table 1 the partition function expressed in equation (4) is calculated exactly using a transfer matrix approach. In prior works, complete mathematical and computational details [19] and several detailed hand examples [20] can be found. It is worth noting here that the size of the transfer matrix depends on the entropy parameters because of the rule that assigns independent constraints preferentially with lowest entropy weights. For the entropy parameters selected in this work, a *homogeneous* polypeptide consisting of the (H, P) residue type requires a transfer matrix size of $(306 \times 306, 150 \times 150)$. The large matrix sizes compared to the 3×3 matrix for the Lifson–Roig model are mainly a reflection of the long-range nature of network rigidity, and the fact that the DCM accounts for the mechanisms responsible for both normal and inverted helix–coil transitions (not just one).

In an infinitely long random HP *heterogeneous* polypeptide chain with 50% H composition, the transfer matrix size is 617×617. The transfer matrix is constructed as a direct product of a rigidity state space and a conformational space [19]. The only modification required to handle different residue types is the annotation of the conformational states {h, a, c} with a residue index as $\{h_{aa}, a_{aa}, c_{aa}\}$, where in this case the possible amino acids, aa, can be either H or P. As a result of the annotation, the conformational space is enlarged, thus giving rise to a larger transfer matrix size.

For finite chains the transfer matrix size depends on the particular composition. For example, a HP polypeptide chain with sequence HPHHPHHHPHHHHPHHHHHP requires a transfer matrix size of 509×509. In this case, the size of the transfer matrix remains the same upon sequence inversion. However, the HP polypeptide HHHHHPPPPP requires a transfer

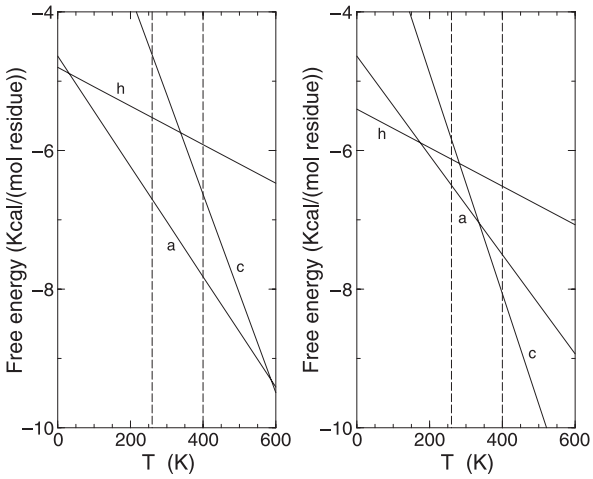

Figure 1. The free energy $G(\mathcal{F}, T)$ per residue as a function of temperature for an infinitely long chain having a conformation that is 100% hydrated (h), alpha-helical (a), or a coil (c). The (left, right) panels show results for (H, P) homogeneous chains. The region between vertical dashed lines marks an experimentally accessible region. General equations for obtaining these lines are given in [20].

matrix size of 395×395. Upon inverting this sequence to PPPPPHHHHH, the transfer matrix size becomes 437×437. In spite of the different matrix sizes and elements, the partition function and all thermodynamic properties are calculated to be identical. Different matrix sizes (and elements) appear because the transfer matrix is constructed by direct propagation of rigidity down the chain [19]. In some cases, subsets of rigidity states are not accessible traversing from left to right as they were in traversing from right to left. Results of DCM calculations are invariant under change of the direction of propagation as physics demands[3]. This is not the case for the Zimm–Bragg and Lifson–Roig models generalized to heterogeneous polypeptides [13] because they are based on ad hoc local nucleation parameters that do not properly take into account long-range effects of rigidity propagation. Nucleation of helix formation in the DCM is an outcome of the rigidity calculation—not input as a model parameter.

3. Results and discussion

3.1. H and P homogeneous polypeptide chains

DCM predictions based on our selected parameters are summarized in figure 1, showing the free energy per residue for a homogeneous (H, P) chain in three different extreme macrostates. The H polypeptide has features intentionally representative of alanine, which is a hydrophobic residue with an unusually high propensity for forming an alpha-helix [31]. Over the experimentally accessible temperature range, long H polypeptides will be alpha-helical, exhibiting no transition. Simulations on polyalanine show that the transition from helix to coil is well above boiling (even though liquid water is non-existent at atmospheric pressure!) In this paper, we too extend temperature ranges to understand general features. At extremely (low, high) temperatures the hypothetical (hydrated, coil) macrostate is the most stable.

[3] Symmetry is broken by dipole interactions along the backbone, but this effect is not in the model.

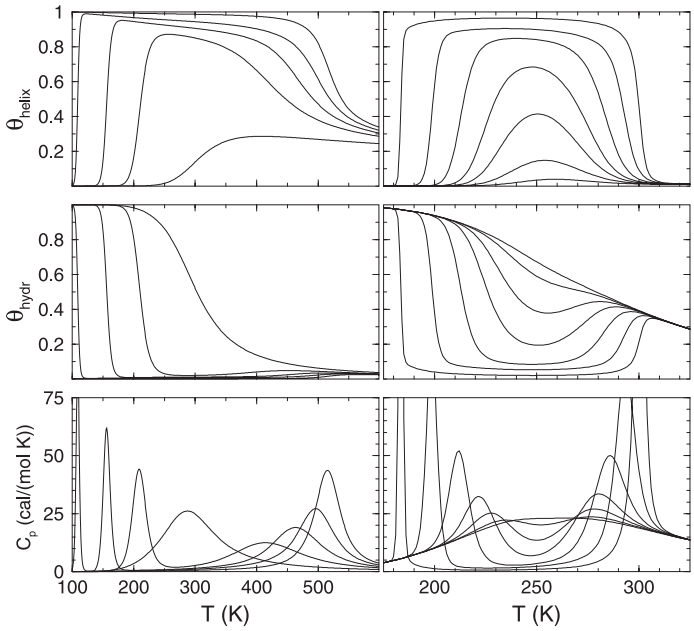

Figure 2. Results for (H, P) homogeneous chains are given in (left, right) panels. Helix content, Θ_{helix}, hydration content, Θ_{hydr}, and heat capacity, C_p, are shown in the top, middle, and bottom rows. These thermodynamic response functions are plotted for chain lengths (no of residues) 5, 10, 15, 25, 100 for the H polypeptide and 25, 30, 35, 40, 50, 75, 200 for the P polypeptide. The orderings of the lines are such that for larger chain lengths: (i) the greater the helix content, (ii) the greater the depletion in hydration content, and (iii) the greater the separation of the two heat capacity peaks.

Free energy lines for long P polypeptide chains, also shown in figure 1, have similar properties. However, the low temperature line-crossing is close to being experimentally accessible, and the high temperature crossing is accessible. For long P polypeptides, it can be expected that a sharp transition from helix to coil as temperature is increased will occur at $\approx$335 K in accordance with Schellman's two-state model [30]. Plotting free energy lines for limiting cases of an infinite chain provides insight into thermodynamic stability. Solvent and pressure conditions modify y-intercepts and slopes[4] that shift free energy lines and change locations of line-crossings. If the free energy for the completely hydrated chain is greater than the free energy for both helix and coil macrostates, then cold denaturation is not possible. This is the case considered here for both H and P chains within the experimentally accessible range.

Key signatures in the thermodynamic response for (H, P) homogeneous polypeptides of finite length are shown in figure 2. Over an extended temperature range both hydrophobic and polar residues exhibit heat and cold denaturation. Helix stability has re-entrant behaviour, where, upon cooling or heating, helix content is lost. Helix content serves as an order parameter, but the mechanism for cold denaturation is ascribed to hydration. Therefore, hydration content is monitored, and serves as a second order parameter. It is seen that depleting hydration content corresponds to increasing helix content, indicating that intramolecular H bonding is more favourable at intermediate temperatures than H bonding to solvent. In the DCM, polypeptide H bonding to solvent is generically modelled as fluctuations in H bonding between mobile solvent molecules and rigid clathrate structures that act like local pockets of ice.

[4] The DCM does not require these lines to be straight. Temperature dependent parameters would be expected in an accurate description of real amino acids in mixed solvents.

Starting below the cold denaturation temperature, heating the polypeptide results in breaking apart clathrate structures; this is akin to the process of melting ice. Hydration is modelled here as a short-range interaction, but it breaks intramolecular backbone H bonds. As chain length increases the helix becomes more stable as the degree of cooperativity between backbone H bonds increases due to the long-range nature of the network rigidity. Thus, the propensity toward hydration dramatically decreases to prevent disruption of backbone H bonds. However, the stabilizing effect of backbone H bonds diminishes as temperature increases. Upon further heating, gain in conformational entropy drives the rigid helical structure to break apart into a flexible coil.

Cooperativity is best observed by tracking the transition (heat and cold) as a function of chain length. Heat capacity curves in figure 2 show peaks separating further apart as chain length increases, until saturation occurs. Saturation is a result of exceeding the rigidity correlation length. The high temperature transition saturates at 512 K for the H polypeptide, which is lower than the 587 K predicted from figure 1. Similarly, a saturation at 302 K is found for the P polypeptide instead of 335 K. Low temperature saturation values are in agreement with free energy line-crossing in both cases. Therefore, the high temperature transition does not reflect a two-state process. All the heat capacity features found here, including magnitudes and the non-two-state behaviour in the normal helix–coil transition, are consistent with experiments.

3.2. HP heterogeneous polypeptide chains

Experimentally, polypeptides with high composition of alanine are not soluble in aqueous solutions. Therefore, as a matter of necessity, alanine rich polypeptides having high alpha-helix propensity must contain polar residues. Therefore it is important to investigate the helix–coil transition as a function of H composition using the HP heterogeneous DCM. We begin by considering a block of five residues, and all possible H–P permutations. For each permutation, the block can be repeated an arbitrary number of times to produce chains of length 5, 10, 15, etc. A phase diagram giving the transition temperatures (for cold and heat denaturation) is given in figure 3 for chains having an H content of 20%, 40%, 60%, and 80%. The general feature of the phase diagram reflects how the heat capacity peaks separate further apart as chain length increases. That is, the (low, high) temperature transition point (decreases, increases) to some saturation temperature as chain length increases. For fixed H composition, the curves look like a sideways parabola. Between the lines defining the parabola is the alpha-helix phase. Above the high temperature transition line is the coil phase. Below the low temperature transition line is the hydrated phase. No phase can be defined for extremely short polypeptides. Actually, the phase boundaries are fuzzy, since the polypeptides represent a small finite system (not in the thermodynamic limit).

The phase diagram shown in figure 3 shows that the experimentally accessible transition (normal helix–coil transition) interpolates between the two limiting values of the homogeneous H or P polypeptides. The locations of the transitions and the shape of the helix content curves are in qualitative agreement with experiments (by construction when choosing parameter values). At fixed H composition, the symbols used in defining the phase boundary in figure 3 show remarkable consistency in spite of there being as many copies as there are distinct arrangements of H and P. A natural question to ask is: how much dependence is there on the transition temperature through sequence rearrangement at fixed H composition? This question pertaining to the DCM predictions was originally motivated by experimental work by Baldwin and co-workers [32] on how much sequence rearrangement affects the alpha-helix transition. They concluded that there is a strong dependence on the order of residues in the polypeptide.

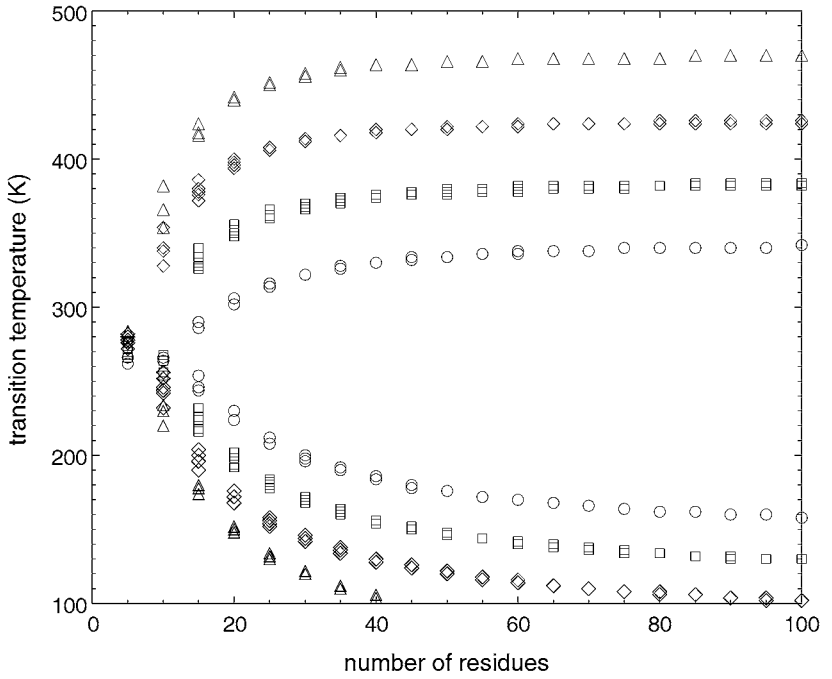

Figure 3. A phase diagram expressing the transition temperature dependence on chain length for HP block polypeptides with 20, 40, 60, 80% H content—shown as circles, squares, diamonds, upwards-pointing triangles respectively. The same symbols are used independently of the arrangement of H- and P-type residues within the chain. These data were generated by considering all possible H–P arrangements in a five-residue block, which is then repeated to obtain all chain lengths shown (as multiples of 5).

The distribution of transition temperatures for HP heterogeneous block polypeptides of various lengths is shown in figure 4 for a block of five residues, and in figure 5 for a block of 13 residues. In both cases, compositions at (or near) 60% H and 40% H (more polar) are considered. The general trend is that the dependence on HP arrangements is greater in shorter chains for fixed block size, and greater for larger block sizes. These results are consistent with the idea that HP arrangements play an important role in governing the transition temperature whenever permutations are made within the rigidity correlation length. If the correlation length is small, HP arrangements will not be very influential. The full effects of network rigidity are not felt in small blocks, suggesting that HP arrangements are not as effective in scrambling up how rigidity propagates. The repeating block units do not increase the variance as much because a regular periodicity sets in. However, as the block length increases, there are more ways to influence how rigidity propagates. Our results show that shorter chains are more dependent on the HP arrangement. Perhaps this is because there are fewer ways for a variety of affects to self-average out. Although it was expected that HP arrangements would affect the transition temperature, the results presented here show the degree to which this is true.

In figure 5, the 40% H composition for chain length of 13 shows two well separated distributions of transition temperatures. It is surprising to see such a dramatic separation. We therefore investigated the nature of the sequences to determine any special characteristics responsible for such a jump. Although further investigation using statistical tools is required, preliminary inspection shows that the sequences giving rise to low transition temperatures

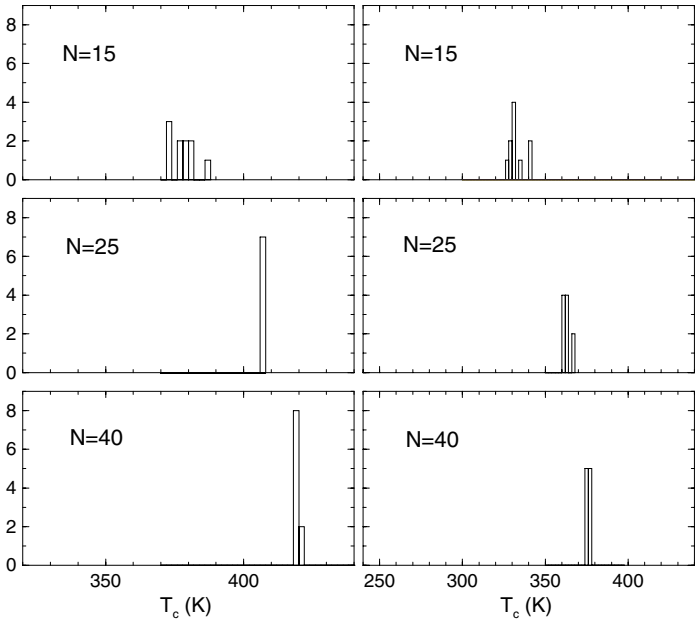

Figure 4. A histogram of transition temperatures, T_c, from ten HP block polypeptides for chain lengths 15, 25, 40 with block size 5. The left panel shows all possible HP arrangements with 60% H composition corresponding to three H and two P residues within the block. The right panel shows all HP arrangements at 40% H composition.

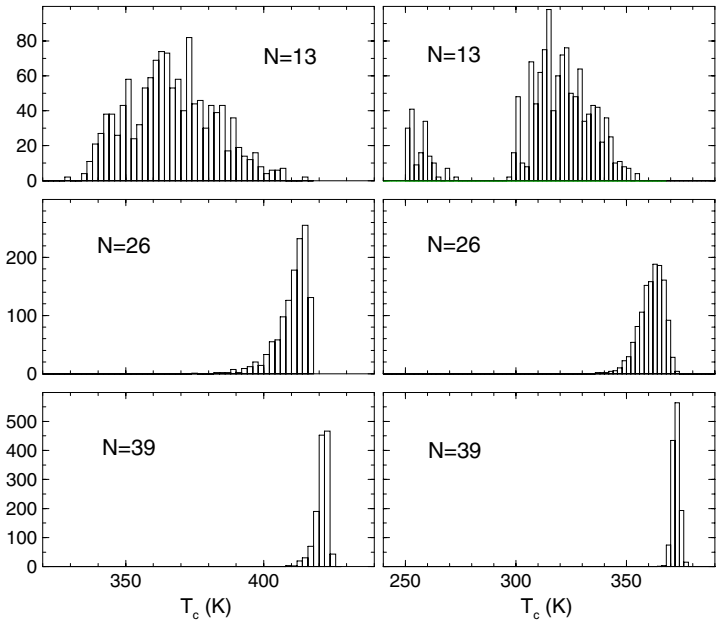

Figure 5. A histogram of transition temperatures, T_c, from 1287 HP block polypeptides for chain lengths 13, 26, 39 with block size 13. The left panel shows all possible HP arrangements with ≈60% H composition corresponding to eight H and five P residues within the block. The right panel shows histograms for ≈40% H composition.

have long consecutive strings of P residues near one or both ends of the polypeptide. Since there is a clustering of Ps there is also a clustering of Hs. In all sequences, there are always at least three consecutive Hs. The prediction that such a dramatic change can occur is quite interesting. Future investigations along these lines appear to be warranted.

4. Conclusions

The DCM provides a framework for constructing Ising-like models based on free energy decomposition. Although an approximation scheme, the DCM is computationally tractable and provides a general approach not limited to the alpha-helix–coil transition, or to biopolymers. Indeed, the approach is applicable to physical systems in general, including structural glass networks where the ideas of rigidity applied to microscopic systems were born [33, 34]. The impetus for constructing the DCM was a desire to understand both protein stability and flexibility self-consistently. In this context, the underlying assumption of the DCM is that the calculated lowest upper bound estimate for conformational entropy is sufficient for accurately describing thermodynamic stability in biopolymers. Considerable support for this approach is mounting; it has been demonstrated that the DCM accurately models essential features of the thermodynamic response in applications to polypeptides undergoing helix–coil transitions in mixed solvents [20] and in proteins [21, 22].

The DCM does not suffer from drawbacks that the traditional Lifson–Roig models have, namely modelling helix nucleation directly. Instead, the DCM models various interaction types that interact via network rigidity, which provides a general mechanism for enthalpy–entropy compensation. As a result, nucleation is an outcome of the network rigidity calculation. Furthermore, the only special attribute of the alpha-helix to coil transition is that the problem can be simplified to a point where it is solved exactly using transfer matrix methods. The DCM transfer matrix formalism originally applied to the alpha-helix–coil transition in mixed solvent conditions has been extended to heterogeneous polypeptides. A simple HP model was introduced for capturing the essential nature of hydrophobic (H) and polar (P) residues. Exact matrix calculations were performed to determine the degree to which the DCM predicts transition temperature dependence on rearrangements of residues in a sequence. It was found that in some cases dramatic changes result. Future work will explore better model parametrization and, also important, investigate sequence signatures that might appear in nature for either stabilizing or destabilizing regions causing them to become alpha-helical.

Acknowledgments

The authors thank Sargis Dallakyan for helping with the manuscript preparation and providing a critical review. DJJ is grateful to have been introduced to rigidity theory by Mike Thorpe and to have had the privilege of working with him for over a decade. It is a pleasure for DJJ to present this paper in celebration of Mike's 60th birthday.

References

[1] Doty P and Yang J T 1956 *J. Am. Chem. Soc.* **78** 498
[2] Doig A J 2002 *Biophys. Chem.* **101/102** 281
[3] Shi Z *et al* 2002 *Proc. Natl Acad. Sci.* **99** 9190
[4] Mitsutake A and Okamoto Y 1999 *Chem. Phys. Lett.* **309** 95
[5] Peng Y *et al* 2003 *J. Phys. Chem.* **118** 2374
[6] Aremin R *et al* 2003 *Protein Sci.* **12** 1145

Garcia A E and Sanbonmatsu K Y 2002 *Proc. Natl Acad. Sci. USA* **99** 2782
Kise K J and Bowler B E 2002 *Biochemistry* **41** 15826
[7] Zimm B H and Bragg J K 1958 *J. Chem. Phys.* **28** 1247
[8] Lifson S and Roig A 1961 *J. Chem. Phys.* **34** 1963
[9] Muñoz V and Serrano L 1994 *Nat. Struct. Biol.* **1** 299
Muñoz V and Serrano L 1995 *J. Mol. Biol.* **245** 275
[10] Gans P J *et al* 1991 *Biopolymers* **31** 1605
[11] Poland D and Scheraga H A 1970 *Theory of Helix–Coil Transitions in Biopolymers* (New York: Academic)
[12] Bierzynski A 1987 *Comment. Mol. Cell. Biophys.* **4** 189
[13] Bierzynski A and Pawlowski K 1997 *Acta Biochim. Pol.* **44** 423
[14] Wetlaufer D B 1990 *Trends Biochem. Sci.* **15** 414
[15] Thorpe M F and Duxbury P M (ed) 1999 *Rigidity Theory and Applications* (New York: Plenum)
[16] Mark A E and van Gunsteren W F 1994 *J. Mol. Biol.* **240** 167
[17] Dill K A 1997 *J. Biol. Chem.* **272** 701
[18] Hallerbach B and Hinz H J 1999 *Biophys. Chem.* **76** 219
[19] Jacobs D J *et al* 2003 *Phys. Rev.* E **68** 061109
[20] Jacobs D J and Wood G G 2004 *Biopolymers* **75** 1–31
[21] Jacobs D J and Dallakyan S 2004 *Biophys. J.* submitted
[22] Livesay D R *et al* 2004 *FEBS Lett.* at press
[23] Brady G P and Sharp K A 1995 *J. Mol. Biol.* **254** 77
[24] Cormen T H *et al* 1990 *Introduction to Algorithms* (Cambridge, MA: MIT Press) See the greedy algorithm for determination of independent basis set.
[25] Grest G S and Cohen M H 1981 *Adv. Chem. Phys.* **48** 455
[26] Guyon E *et al* 1990 *Rep. Prog. Phys.* **53** 373
[27] Jacobs D J and Thorpe M F 1995 *Phys. Rev. Lett.* **75** 4051
[28] Ramachandran G N *et al* 1963 *J. Mol. Biol.* **7** 95
Lovell S C *et al* 2003 *Proteins* **50** 437
[29] Andersen N H *et al* 1996 *J. Am. Chem. Soc.* **118** 10309
[30] Schellman J A 1955 *C. R. Lab. Carlsberg, Sér. Chim.* **29** 230
Schellman J A 1958 *J. Phys. Chem.* **62** 1485
[31] Baldwin R L 2003 *J. Biol. Chem.* **278** 17581
[32] Padmanabhan S *et al* 1990 *Nature* **344** 268
[33] Phillips J C 1979 *J. Non-Cryst. Solids* **34** 153
[34] Thorpe M F 1983 *J. Non-Cryst. Solids* **57** 355