Published in final edited form as:

Biopolymers. 2011 April ; 95(4): 240–253. doi:10.1002/bip.21572.

Nonadditivity in the α -helix to coil transition

Gregory G. Wood^a, Drew A. Clinkenbeard^a, and Donald J. Jacobs^{*,b}

^a Department of Mathematics and Applied Physics, CSU Channel Islands, Camarillo, CA 93012 USA

^b Department of Physics and Optical Science, UNC Charlotte, Charlotte, NC 28223 USA

Abstract

The Lifson-Roig Model (LRM) and all its variants describe the α -helix to coil transition in terms of additive component free energies within a free energy decomposition scheme, and these contributions are interpreted through sequence-context dependent nucleation and propagation parameters. Although this phenomenological approach is able to adequately fit experimental data on helix content and heat capacity, the number of required parameters increases dramatically with additional sequence variation. Moreover, due to nonadditive competing microscopic effects that are difficult to disentangle within a LRM, large uncertainties within the parameters emerge. We offer an alternative view that removes the need for sequence-context parameterization by focusing on individual microscopic interactions within a free energy decomposition and explicitly account for nonadditivity in conformational entropy through network rigidity using a Distance Constraint Model (DCM). We apply a LRM and a DCM to previously published experimental heat capacity and helix content data for a series of heterogeneous polypeptides. Both models describe the experimental data well, and the parameters from both models are consistent with prior work. However, the number of DCM parameters is independent of sequence-variability, the parameter values exhibit better transferability, and the helix nucleation is predicted by the DCM explicitly through the nonadditive nature of conformational entropy. The importance of these results is that the DCM offers a system-independent approach for modeling stability within polypeptides and proteins, where the demonstrated accuracy for the α -helix to coil transition over a series of heterogeneous polypeptides described here is one case in point.

Introduction

As a precursor to protein folding [1], the mechanisms that control the α -helix to coil transition and stability of the α -helix structure [2] have been intensely investigated [3]. Despite much experimental effort, establishing a transferable amino acid helix propensity scale with respect to a standard state of aqueous solution has been elusive. A comprehensive review [4] based on experimental work that measure the change in Gibbs free energy upon helix formation, ΔG , show significant variation per amino acid. To construct a universal scale, the ΔG_x values for amino acid, x , are measured relative to ΔG_{ALA} for alanine to define the change in free energy by $\Delta\Delta G \equiv \Delta G_x - \Delta G_{ALA}$. An average estimate for $\Delta\Delta G_x$ varies from zero (for alanine) up to 1.11 kcal/mol for glycine. Disregarding proline, deviations ranged from 0.21 kcal/mol for threonine to 0.56 kcal/mol for leucine. Measurements on individual polypeptides typically have random error of about 0.05 kcal/mol [5,6], indicating that the large deviations in $\Delta\Delta G_x$ are due to sequence dependence.

*To whom correspondence should be addressed: djacobs1@uncc.edu.

To model this observed sequence-context dependence is a difficult challenge based on additive free energy decomposition schemes. Component free energies are strictly additive over independent subsystems [7,8], and approximately additive over weakly coupled subsystems. Systematic error between model predictions and experiment derive from the assumption of additivity between short segments along the chain. Numerous examples demonstrate a single amino acid substitution with negligible disturbance in local side-chain interactions can sometimes cause a dramatic change in stability [9,10,11,12,13]. These observations indicate amino acids are coupled through long-range interactions.

Surveying the literature, one finds the ubiquitous strategy of employing a Lifson-Roig model [14] (LRM) or some *modified* LRM (mLRM) to describe experimental data or simulation results on the α -helix to coil transition. The advantage of this approach is that the mathematical form of the transfer matrix is easy to construct. However, a modified form is required to account for different amino acids and numerous other factors, such as end-cap perturbations, interactions between side-chains, partial 3-10 helix content, hydration and salt-bridges. Unfortunately, for any mLRM that is relatively simple using a few parameters per amino acid type, the parameters are found not to be transferable [15]. However, Herculean versions [16,17,18] that account for myriad effects using many hundreds of parameters to describe arbitrary heterogeneous chains have predictability power. Although many more parameters are used, they are characterizing local effects. This type of approach assumes that sequence-context dependence can be explained by localized interactions within small windows along the sequence.

Other early work [19] and recent work [20] have considered modifying the Zimm-Bragg model [21] to extend to larger window sizes involving k-neighbors with respect to a given reference amino acid. It was found that employing greater window sizes adjust the nucleation and propagation parameters, but no change in qualitative predictions was established. On the other hand, for heterogeneous polypeptides the nominal number of model parameters exponentially increases as a function of window size to reflect sequence-context combinations. Greater window sizes quickly leads to a model that is over-parameterized to explain available experimental data. As such, in order to keep the number of required parameters manageable using larger windows, simplifications were made to the matrix elements of the transfer matrix [19], or the model was only applied to a homogeneous polypeptide [20], where sequence-context dependence is lost altogether. In this type of approach, simplifications to reduce the number of parameters defeat the purpose of larger windows in the first place. Of critical concern, is that beyond an arbitrarily selected window size, interactions between amino acid residues are uncoupled (additive), and potential spatial long-range context dependence is lost.

Alternative View

Now consider the second option of calculating long-range interactions explicitly. Previously, we introduced a Distance Constraint Model (DCM) [22] as an alternative approach to account for nonadditivity in free energy decomposition schemes using constraint theory. The DCM directly relates the independent degrees of freedom governing the dynamics of molecular structure to conformational entropy. The DCM provides a good estimate for conformational entropy when correlated motions are present because the long-range nature of how flexibility and rigidity propagate through molecular structure [23] is explicitly calculated. In REF [22] we show how to solve the DCM exactly using a transfer matrix for the α -helix to coil transition. The disadvantage of this approach is that the mathematical form of the transfer matrix is complicated, and is typically one to two orders of magnitude larger than the 3×3 matrix of a LRM. The large size of the matrices is a reflection of the long-range nature of interactions that span far beyond small window sizes along the polypeptide chain. Albeit the mLRM is simpler mathematically, it misses the essential

physics that conformational entropy critically depends on the long-range nature of network rigidity. On the other hand, recently, a simplified DCM was presented that clearly demonstrates the connection between nonadditivity in conformational entropy to network rigidity as an underlying mechanical interaction responsible for helix nucleation [24]. Here, we strive to employ the simplest model possible, but not simpler than it should be. Specifically, the inherent long-range mechanical interaction of network rigidity determines the nucleation of helix, nonadditivity in conformational entropy and molecular cooperativity.

In this paper, we focus on a collection of experimental data [25] with conditions, sequences, and mutations that approach the ideal limit for a mLRM to work well. The experimental data consist of three excess heat capacity curves and helix content for six polypeptides. We show that a mLRM reproduces both helix content and heat capacity using 8 fitting parameters that fall within the large parameter range reported in the literature. The DCM describes the same data equally well using transferable parameters from earlier studies on two very different systems [22,26], plus 7 fitting parameters that account for new amino acid types. The fitting parameters for both the DCM and mLRM are all within physically reasonable ranges and produce helix-propensities that are consistent with consensus experimental rankings. From these results, both models represent an adequate description of the data, but the generality and underlying assumptions between the models are very different.

Unlike the Zimm-Bragg [21] and Lifson-Roig [14] type models, the DCM ascribes nonadditivity in conformational entropy to network rigidity properties of molecular structure that is not limited to helix and coil interconversion. The DCM applies to the beta-hairpin to coil transition [27] and to proteins [28,29,30] where heat capacity is reproduced well [31]. In other work on proteins [32] we found that nonadditivity in free energy cycles for distant double mutants correlate with propagation of rigidity. The significance of this work further suggests the DCM is well founded as a general theory for folding cooperativity.

Theory: Distance Constraint Model

Free Energy Decomposition

Free energy components depend on the conformation of the polypeptide. The DCM applied to the α -helix to coil transition represents backbone conformations, as shown in Fig. 1. Similar to the LRM, the backbone conformation is described in terms of local α -helical or coil states, labeled respectively as a and c . These states define allowed ranges of the ϕ and ψ dihedral angle *degrees of freedom* (DOF) similar to the LRM. However, unlike the LRM [14], only interactions within an amino acid are considered. Thus, the free energy of an amino acid residue in the a or c state is *independent* of neighboring amino acid conformational states. The free energy of the α -helical state is modeled as $G_a = v_a - TR2\delta_a$ where v_a is its enthalpy, T is temperature in Kelvin, R is the universal gas constant, and $R2\delta_a$ is its entropy. In the DCM, each DOF is assigned an entropy value. For simplicity, the ϕ and ψ DOF are treated equivalently, hence the factor of 2. The physical interpretation is that when the amino acid is in the α -helical state it has an effective constant enthalpy, v_a , and degeneracy $e^{2\delta_a}$. The free energy of the coil state is modeled as $G_c = V_c - TR2\delta_c$, with analogous explanations.

Free energy is also attributed to backbone H-bonds spanning a triplet between the $(i-2)$ -th and $(i+2)$ -th amino acids. Following previous work [33] and depicted in Fig. 1, the H-bond interaction is modeled using three distance constraints related to the distance between donor and acceptor atoms (denoted as d), the distance between hydrogen and acceptor atoms (denoted as b), and the distance between the hydrogen atom with the carbon atom that covalently bonds to the backbone carboxyl oxygen atom (denoted as c). H-bond geometries

are coarse grained into discrete states based on the conformations of the 3 amino acids located at $i - 1$, i and $i + 1$. Out of 8 possible states (i.e. 2^3 triplet conformations), the free energy of a H-bond that spans a full α -helical stretch is $G_{aaa} = u_{aaa} - TR3\gamma_{aaa}$ where u_{aaa} is its enthalpy, and $R3\gamma_{aaa}$ is its entropy. For simplicity, we treat the 3 DOF equivalently, hence the factor of 3. We assign an enthalpy u to model the *disruption* of backbone H-bonds due to H-bonding to solvent. The free energy decomposition scheme used in this work is summarized in Table 1.

Constraint Theory

Following common practice, covalent bonding within a molecular structure “freezes out” atomic DOF. As such, covalent bonds are modeled as quenched distance constraints, leaving only dihedral angles (i.e. ϕ , ψ , and side chain χ angles) as free. The specification of all ϕ and ψ angles defines the conformation of the backbone. Fluctuating constraints are used to set internal coordinates to *fixed values* that are contained within an allowed binned range given by either the *a*- or *c*-conformational states. After transforming the dihedral angle constraints into equivalent distance constraints [33], one can ask: How many distance constraints are needed to make the backbone of a polypeptide chain rigid? The minimum number is $3N - 6$, where N is the number of amino acids. However, most of these constraints are quenched from covalent bonding, and play a passive role throughout the helix-coil transition.

Fluctuating constraints associated with side-chains are in principle important, but side-chain to side-chain interactions are neglected in this work. Within this approximation, only fluctuating constraints applied to the backbone DOF are responsible for the helix-coil transition. Therefore, it follows that $2n$ is the minimum number of distance constraints that need to be applied to *fix* the backbone conformation consisting of n amino acids (assuming proline is not present). In addition to the intrinsic amino acid interactions involving the ϕ and ψ dihedral angles, backbone H-bonds are modeled as fluctuating constraints. In all cases except when no H-bond forms along the backbone, the molecular structure has more constraints along the backbone than the available $2n$ DOF.

A distance constraint may be independent or redundant. Physically, if a distance constraint is added into a flexible region, it will be *independent* because the number of DOF required to specify the conformation is reduced by one. If a distance constraint is added to a rigid region, then there is no reduction of DOF within the system, and the constraint is *redundant*. Mathematically, the position vector of the i -th atom is given as $\vec{r}_i$ and the distance between a pair of atoms is specified as d_{ij} . The system of distance constraint equations of the form $d_{ij} = |\vec{r}_i - \vec{r}_j|$ can be solved algebraically (or using graph theory [33]) to identify linearly independent constraints. In the application here, we start with $2n$ DOF that allows the backbone to take on any accessible conformation. Based on the coarse grained scheme, the backbone conformations of a given amino acid residue will fall in a helix or coil state, and for each triplet of residues, a spanning backbone H-bond may or may not form.

To begin with, consider a case where no backbone H-bonds form. The $2n$ -DOF are each fixed by applying a distance constraint to lock each specified internal coordinate (i.e. ϕ and ψ). After $2n$ constraints are placed to pin down the ϕ and ψ angles, the backbone conformation is fully specified as a particular *constraint topology*. The constraint topology of the polypeptide chain is represented as a string of local states, such as *ccaccaacc* for $n = 9$. Irrespective of whether a particular residue is in an α -helix or coil state, two distance constraints are necessary to specify its local geometry. As backbone H-bonds form, additional constraints must be added to specify their local geometrical requirements. In general, the various distance constraints are not compatible in regions throughout the structure. In these cases, over-constrained regions lead to redundancy, where the geometrical

forms preferred by weaker interactions will more readily accommodate to the geometrical forms preferred by the stronger interactions.

The employed free energy decomposition scheme is based on a set of internal coordinates (2 dihedral angles per amino acid, and 3 distances per H-bond) that define an over specified coordinate system. Specifying the values of the coordinates represents a constraint. For n residues, only $2n$ coordinates need to be specified, leading to $2n$ independent constraints, and the remaining constraints will be redundant. For example, when N_{hb} intramolecular backbone H-bonds form, there will be $3N_{hb}$ additional distance constraints. Out of the $2n + 3N_{hb}$ distance constraints, $2n$ will be independent, and $3N_{hb}$ redundant. Although this count is unique, identification of which constraints are independent is not unique. Both the distribution and strength of constraints are important to determine the degree that a region is rigid or flexible. The strength of a constraint is tied to its assigned local entropy value. Smaller entropy values correspond to stronger, less accommodating, interactions. By explicitly accounting for these effects, the DCM provides a good estimate for conformational entropy.

Non-additivity of Entropy

Interactions are modeled as distance constraints. Based on the conformation of structure, different combinations of interactions lead to different constraint topologies. A constraint topology is defined after all DOF are fixed using distance constraints. The total entropy of a constraint topology is estimated by a sum over entropies from only *independent distance constraints*. This would give the exact answer if each of the independent constraints were 100% decoupled (i.e. orthogonal). However, since linearly independent constraints are generally not 100% decoupled, we only obtain an upper bound estimate. Moreover, many upper bound estimates are possible depending on which constraints are identified as independent. However, a rigorous lowest upper bound estimate is obtained by preferentially selecting constraints with lowest assigned entropies to be independent [22,26]. Suppose when comparing two distance constraints, only one of the two is found to be linearly independent with respect to the system of constraints, while the other is redundant. Then the preferential rule is to select the constraint with lowest assigned entropy as independent, or select arbitrarily when both distance constraints have equal entropy assignments.

We show how to calculate the enthalpy and entropy for the following two example sequences:

$$\begin{array}{cccccccccccccccc} E_1^c & A_2^c & A_3^a & A_4^c & R_5^c & A_6^a & A_7^a & E_8^a & A_9^a & R_{10}^a & A_{11}^a & A_{12}^c & E_{13}^c \\ E_1^c & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ E_1^c & A_2^c & A_3^a & A_4^c & R_5^c & A_6^a & A_7^a & E_8^a & A_9^a & R_{10}^a & A_{11}^a & A_{12}^c & E_{13}^c \end{array}$$

These two examples describe a 13 amino-acid chain with the same backbone conformation but differing in constraint topology because of intramolecular H-bonding. One letter codes define the amino acid type, and their subscripts label the location within the sequence. Superscripts define the local conformational state as either α -helix (^A) or coil (^C). The number of H-bonds that span an amino acid residue is represented by dots over the one-letter codes. In the top sequence (example 1) there are no intramolecular H-bonds, while in the bottom sequence (example 2) five H-bonds form between amino acid pairs {1-5, 5-9, 6-10, 7-11, 8-12}. Contrasting these two sequences show how for a given backbone conformation, H-bond fluctuations are possible due to the competition between solvent H-bonds and intramolecular H-bonds. Therefore, for a 13-mer polypeptide there are 2⁹ accessible H-bond patterns for the same backbone conformation.

Referring to these two example sequences, the total enthalpy is given in Eq. 1 as

$$\begin{aligned}
 E_1 &= 9u + 2v_c^E + v_a^E + v_c^R + v_a^R + 3v_c^A + 5v_a^A \\
 E_2 &= E_1 + u_{cac} + 4u_{aaa} - 5u
 \end{aligned}
 \tag{1}$$

The important point to take notice is that the total enthalpy is a *linear function* of all component interactions present, which depends on the composition of the sequence and its conformation. The parameter, u , is the enthalpy of a H-bond that forms between the backbone of the polypeptide and solvent molecules. Example 1 describes a particular constraint topology where all 9 possible intramolecular H-bonds are broken in favor of 9 H-bonds forming between the polypeptide and solvent. The constraint topology of example 1 can be used as a reference state. Then, as shown in example 2, the enthalpy of the second constraint topology is expressed as the change in enthalpy due to H-bond differences with respect to the first constraint topology. In example 2, five intramolecular H-bonds form at the expense of breaking five H-bonds that interact with solvent. The intramolecular H-bond energies depend on the conformational states of the three residues it spans across, which is why one u_{cac} and four u_{aaa} terms arise.

Using dimensionless “pure entropies”, we calculate $S_k = R\sigma_k$ where S_k is the entropy of constraint topology k , and σ_k is its pure entropy. It is seen from Eq. 2 that the total entropy is generally nonadditive, although for example 1, σ_1 is additive because all constraints present within the system are independent. More interestingly, in example 2, there are 15 more distance constraints than DOF, indicating that 15 of these constraints are redundant and that a *nonlinear function* over component entropies will come into play. Assuming $\delta_a^A < \delta_c^A$, the preferential rule for selecting redundant constraints must be applied. By inspection the distance constraints that are imposed by dihedral angles versus that of a H-bond within the small stretch of α -helix between and including residues 6 to 11 directly compete with one another. The constraints with lowest entropy are preferentially assigned to be independent, which can be expressed in terms of $\min()$ and $\max()$ functions for this simple case.

$$\begin{aligned}
 \sigma_1 &= 4\delta_c^E + 2\delta_a^E + 2\delta_c^R + 2\delta_a^R + 6\delta_c^A + 10\delta_a^A \\
 \sigma_2 &= [4\delta_c^E + 2\delta_c^R + 2\delta_c^A] + [2\delta_a^A + 4\delta_c^A + 3\gamma_{cac} - 3\max(\delta_c^A, \gamma_{cac})] + [8\min(\delta_a^A, \gamma_{aaa}) + 2\min(\delta_c^E, \gamma_{aaa}) + 2\min(\delta_a^R, \gamma_{aaa})]
 \end{aligned}
 \tag{2}$$

As it can be seen in Eq. 2, the total number of contributing entropy components is always 26 because this is exactly the number of independent constraints that is required to specify the conformation. In the expression for σ_2 , the top bracket corresponds to additive contributions because no H-bond crosslinking is present, the middle bracket accounts for the potential effect of a single H-bond that spans residues 2 to 4, and the bottom bracket accounts for the 6 residue-long α -helical stretch. Despite the complexity of Eq. 2 for the second example, the equation for σ_2 can be obtained by inspection just by considering local competitions between constraints. Note that while the long-range nature of rigidity is not revealed in the second example, nonadditive effects already appear due to local competitions. Additional examples involving long-range effects can be found in appendix A in REF [26].

Helix Nucleation

Starting from the disordered coil state, nucleation of helix structure must overcome a large entropy loss. The compensation for this is a lowering of enthalpy, as helical conformations and intramolecular H-bonds form. The origin of cooperatively naturally occurs in the DCM [24]. Once a few constraints pay an entropy cost to rigidify a region, additional constraints lower enthalpy without an entropy penalty. Therefore, a consecutive group of intramolecular H-bonds along the chain will initiate helix structure better than having the same number of

H-bonds dispersed along the chain. However, it is generally not obvious where the optimal nucleation site or competing sites will be located because many interactions work cooperatively as far as network rigidity is concerned. That is, the nucleation process is governed by the rigidification of structure involving the long-range property of how rigidity propagates [23]. Rigidity propagation within a polyaniline chain was characterized by a rigidity correlation length of 40 amino acids [22]. Meaning, the rigidity or flexibility of the conformation at a given amino acid within polyaniline can be changed by removing or adding a distance constraint up to 40 amino acids away. This result suggests that for polypeptides less than 40 amino acids, helix nucleation will depend on the sequence-context of the entire chain.

Partition Function

The free energy of an accessible constraint topology, k , is given by $G_k = E_k - TR\sigma_k$. For n amino acids, there are $2^n \times 2^{n-4}$ constraint topologies. The partition function is given by

$Z = \sum_k e^{-\frac{G_k}{RT}}$. A transfer matrix is employed to calculate the partition function exactly. The size of the transfer matrix depends on the specific sequence of the polypeptide, and ranges from 16×16 in noncooperative cases, to 400×400 in highly cooperative cases for a homogenous polypeptide [22]. For the heterogeneous polypeptides considered here (Table 2) the largest transfer matrix is found to be 771×771 , indicating greater degree of nuance in cooperativity. The transfer matrix enforces the entropy-based preferential rule to identify independent constraints as rigidity or flexibility is propagated down the chain. Mathematical details are given in previous publications for homogenous polypeptides [22,26], and heterogeneous polypeptides [34].

Results

Sequences

Our results are based on experimental DSC and CD data on six polypeptides reported by Richardson and Makhatadze [25]. Using similar notation, the sequences {A4, A6, V5, G6, V6, V345} listed in Table 2 were designed so that at pH 2 the effects of sidechain interactions are minimal. We optimized DCM parameters by fitting to their published heat capacity curves for A4, A6 and V5. For a given set of DCM parameters, the DCM predictions are calculated using a transfer matrix method. The size of the transfer matrix not only depends on the specific parameters, but also on the arrangement of amino acids within the sequence and the direction of propagation (i.e. from N- to C-terminus or vice versa) as Table 2 summarizes. Although the transfer matrix depends on propagation direction, the partition function does not.

Model Parameters

In the present work, the DCM has 7 fitting parameters. All intramolecular hydrogen bond (IHB), solvent, alanine and endcap parameters are taken from prior work [22] based on fitting to simulation data on homogeneous alanine chains of different lengths and solvents.

One alanine parameter dealing with the enthalpy of its coil state, v_C^A , was found to depend on solvent conditions in this prior work, and it is considered to be a fitting parameter here.

Interestingly, the best fit value for v_C^A was found to be only 0.2 kcal/mol less favorable than before, which is well within the expected range of variance. In another work [26], it was found the IHB parameters exhibited transferability. We continue to employ these IHB parameters, listed in Table 3 for convenient reference. Because we are considering new amino acid types in this work, we must specify for amino acid type x the enthalpy difference

between coil and α -helix state, v_c^x , and the entropy of the α -helix state, δ_{α}^x and entropy δ_c^x for the coil state.

To keep the number of adjustable parameters to a minimum, three fitting parameters are assigned for valine, and only three more are assigned for both glutamic acid and arginine. There is not enough data to justify more parameters, and treating the parameters of glutamic acid and arginine as effective parameters for the sequences given in Table 2 is justified because their positions are conserved. Furthermore, the alanine endcap parameters from previous work have been employed for both endcaps. It is worth noting that one endcap is alanine and the other is tyrosine, so using different parameters for each is justified, but not required for adequate fitting using the DCM (or the mLRM). All DCM parameters are listed in Table 4, where they are also translated in terms of ΔG .

Within the mLRM, there are two types of statistical weights, given by w^x for helix propagation and v^x for helix nucleation that must be specified for each amino acid type x . The corresponding free energies are given by $\Delta G_w = -RT \ln(w)$ and $\Delta G_v = -RT \ln(v)$ respectively. Each free energy is expressed in terms of two variables describing enthalpy and entropy, such as $\Delta G_w = E_w - TS_w$. To keep the number of fitting parameters to a minimum, identical parameters are used for both endcaps. It was found that only the nucleation parameters at the endcaps were necessary to adjust in order to fit the data.

Meaning, only E_v^{end} and S_v^{end} where allowed to vary. Next, entropy and enthalpy changes from coil to helix and coil to nucleation state are required for alanine, valine, and arginine/glutamic acid, which would tally in total to twelve additional independent parameters given by $\{E_v^x, S_v^x, E_w^x, S_w^x\}$. Treating glutamic acid and arginine with the same set of effective parameters is justified based on sequence conservation, which is a model independent approximation to minimize free parameters.

In the case of the mLRM, a great reduction of free parameters can be made by noting that the nucleation parameters are commonly taken as uniform across the amino acid types [35]. Therefore, all the nucleation enthalpies and entropies are considered to be the same across all the amino acids (and endcaps) to yield a total of 8 fitting parameters. Six of these parameters model the propagation characteristics of alanine, valine, and arginine/glutamic acid, and two parameters characterize the nucleation process that is assumed to be amino acid independent. It is worth noting that many other combinations of free parameters that included using more and less numbers of fitting parameters were explored. The final set of 8 fitting parameters described above reflects the best results obtained, and these parameters are listed in Table 4. In addition, Table 4 includes the conversion of all mLRM parameters to effective Zimm-Bragg values using well-known transformation equations [36] and translated into ΔG values.

Heat Capacity

Using the fitting parameters listed in Table 4, the mLRM and DCM both adequately reproduce the heat capacity respectively shown in Fig. 2 and Fig. 3 for three polypeptides. Baselines were included in the fitting procedure for both models even though the reported experimental results were adjusted (baselines were already subtracted out) to reflect excess heat capacity [25] to characterize the structural transition between a disordered conformation and alpha-helical structure. The mLRM required no C_p baseline, while the DCM contributes about 2.75 cal/mol/K that needed to be subtracted off. This non-zero contribution from the DCM reflects the fluctuations that occur between solvent H-bond to intra-molecular H-bonds, and therefore can still be regarded as being part of the solvation process. Unlike the mLRM, the DCM does not completely separate the two effects. The discrepancies between the predicted curves and experimental data are within expected tolerances based on the

simplicity of the models and the error bars within the experimental data itself. In general, the DCM does better fitting in the high temperature region of the heat capacity, while mLRM fits better near the peaks. The three heat capacity curves were fit simultaneously with the reported helix content at 0 °C. Without the helix content data, both models can fit just to the heat capacity better than shown, but the predicted helix content would not be as good. Consequently, both heat capacity and helix content data is used in the fitting process to justify the number of fitting parameters, which for the mLRM becomes 2 fitting parameters per curve. The fitting parameters for both models do not vary much based on the constraints imposed during the fitting process (see methods).

Helix Content

The helix content provides an estimate for the fraction of amino acids in the α -helix state within the chain. This estimate is based on a linear relationship between the raw experimental CD data and a model prediction. Richardson and Makhatadze [25] used a standard linear-formula (Eq. 2 and 3 in REF [25]) to convert their CD data into an estimated helix fraction, which we denote as $f_{exp,H}$. For the DCM, the linear transformation:

$$f_{calc,H} = 2.49(f_H - 0.422) \quad (3)$$

is employed to obtain a calculated fraction of helix, $f_{calc,H}$ to directly match with the experimental fraction of helix, $f_{exp,H}$. As shown in Fig. 4, the DCM predicted helix fraction, f_H at 0°C, for the six polypeptide chains is found to be linearly related to the corresponding experimental estimates, $f_{exp,H}$. This result indicates excellent agreement with experiment regarding helix content. Employing the same process for the mLRM, the linear transformation equation is given as:

$$f_{calc,H} = 0.821(f_H + 0.091) \quad (4)$$

Based on the linearity of the predicted values versus experimental estimates, it is clear that the DCM helix content predictions describe the experimental helix content better than does the mLRM.

Despite the improved linear relationship, it appears that the DCM predicted helix content is on a very different scale compared with the mLRM, and it predicts a negative helix-content for the G6-polypeptide, which by definition is impossible. However, both of these concerns is a non-serious artifact caused by comparing the DCM predictions (and our mLRM predictions) not to the raw CD data, but rather to the standard-linear formula for helix-content used by Richardson and Makhatadze [25]. Since this is a linear transformation, mathematically there is absolutely no problem in comparing one linear transformation to another. However, these differences (shift in slope and y-intercept) do point to differences in the model predictions. The most interesting difference between the mLRM and DCM predictions is that at high temperatures where the polypeptides are mainly coil, the helix-content predicted by the DCM is about 30%, while the mLRM is about 10%. It is worth noting that a random coil is expected to have more helix-content than 10% of its amino-acids in an alpha-helical conformation [37,38,39], although 30% saturation levels are higher than expected. For low temperatures, the helix-contents for all three polypeptides reach nearly to the same maximum value, although the DCM is slightly lower, by less than 5%. Direct model-to-model comparisons will be discussed in the next section.

Discussion

The DCM parameters for the amino acids show that $\nu_c^x < \nu_a^x$ and $\delta_c^x > \delta_a^x$. Without formation of any IHB, this implies the α -helix state has lower conformational entropy and *greater* enthalpy than the coil state. Since the DCM lumps solvation interactions into these parameters, this condition reflects the interactions with solvent. Consequently, without intramolecular H-bonding the coil state would be stable at all temperatures with no transition. Therefore, for this set of sequences in aqueous solution, the DCM predicts that *intramolecular H-bonding is necessary for the helix to coil transition*. This result is physically reasonable, but can there be another DCM parameter set that would indicate a different conclusion? Although we cannot rule out such a situation, we have explored fitting the data with other parameter sets that would lead to a different conclusion. No such parameter set was found that simultaneously fits all the data.

Presumably, the DCM parameters for the various amino acids are dependent on the choice of the IHB model. To obtain maximum consistency, we have treated the IHB model and its parameters (see Table 3) as transferable in all previous works [22,26,34]. The results found here further support the notion that IHB interactions maintain transferability across diverse systems. Since this aspect has been quite useful, it is worth mentioning that in earlier work [26], a simpler alternative H-bond parameterization similar to the LRM that assumes an H-bond can only form in the aaa conformation state was constructed. Although this simpler H-bond model was able to fit one set of data, it was not found to exhibit transferability [26]. Moreover, the assumption that H-bonds only form when there are consecutive α -helix conformations is not correct.

The transferable IHB model was originally motivated by a structural bioinformatics check [22] showing that an IHB can form for any of the eight accessible backbone triplet states. This at first surprising result was rationalized by realizing triplets represent coarse grained bins, and the range of allowed H-bond geometries is broad. There will always be some ccc conformations that are not very different than aaa conformations, for example. On the other hand, the enthalpy parameter, u , for H-bonding between solute and solvent is not a transferable parameter, which provides the competition between intramolecular H-bonding and H-bonding to solvent.

The employed DCM accounts for solvent effects through the ν_c^x parameters, which cannot be transferable since they depend on the nature of the solvent. In principle, for a fixed type of solvent, the parameters should be transferable and independent of sequence. However, as mentioned above, many details about how amino acids interact with solvent have been neglected in the DCM employed here. Therefore, these parameters are required to be adjustable to absorb systematic errors caused by working with an oversimplified model. Nevertheless, essential mechanisms are readily captured using the DCM. For example, the excess heat capacity is principally concerned with the H-bond fluctuations coupled to the conformational changes from α -helix to coil, which strongly influence the rigidity and flexibility of the structure.

mLRM and DCM Comparison

Using the parameters given in Tables 3 and 4, the heat capacity and helix content based on the mLRM and the DCM are compared in Fig. 5 for three different sequences of the form (Y-(XEARA)_n) where X=A,V and G, and each sequence has its length varied with $n=4,5,6,7,8$. As mentioned above, the DCM has a non-zero residual baseline in its heat capacity prediction. When this is subtracted out, the heat capacity predictions not only agree well with the experimental curves that they were fitted to, but there is also qualitative agreement with the mLRM. The main differences are: For X=valine, the DCM heat capacity

curves peak at slightly higher temperatures than the corresponding mLRM heat capacity curves, and the peak heights are higher. For X=glycine, just the reverse situation is seen, where the DCM heat capacity curves are more flat, and the structural transition to alpha-helix begins at lower temperature, which is experimentally inaccessible.

In comparing the helix content between the two models for the same sequence comparisons, the DCM predicts up to 30% residual alpha-helical structure at high temperatures, while the mLRM has a lower residual of about 10% helix content. Moreover, for the three sequences considered, the helix-coil transition has a broader helix content curve within the DCM because the overall change in helix content is not as dramatic. Interestingly, for the case X=valine, the DCM has sharper heat capacities for the various size peptide, but it still maintains a broader helix content curve. In general, Fig. 5 shows that there are similar qualitative behaviors between the two models, and linear relationships for helix content between model predictions and experimental measurements are just about equally good. Nevertheless, the minimum helix content of the DCM appears to be larger than what is normally inferred experimentally. For example, the application of Eq. (4) scales the experimental results (saturation baselines) in a way that assigns low helix content at temperatures well above the transition temperature (or melting temperature) of the polypeptide. This suggests that the DCM is not correct in its helix content predictions. Therefore, we next consider uncertainties in model parameters and experimental interpretations.

Uncertainties in Model Parameters

The Zimm-Bragg and Lifson-Roig models are based on an additive free energy decomposition scheme. These two models have been routinely employed to describe key features of the α -helix to coil transition. For example, in this work we showed that the mLRM can be parameterized to produce a good simultaneous fit to the excess heat capacity for the A4, A6 and V5 polypeptides. An overview of published LRM parameters with scrutiny clearly indicates that relatively simple modified forms of the LRM fail to describe the helix-coil transition [15,22] using a consistent set of parameters. Inconsistency and lack of transferability in parameters has endured over 50 years. Indeed, parameter dependence on chain length and sequence-context has been widely accepted as an inevitable consequence of a phenomenological theory. However, for the present sequences, it is unnecessary to consider length or sequence dependence to attain adequate fits to joint heat capacity and helix content data. The parameters we obtained from fitting are consistent within the broad variation found within the literature.

The alternative view is that the assumption of additivity in free energy decomposition is an over-simplification. The DCM confronts nonadditivity in free energy decomposition schemes by explicitly accounting for how flexibility and rigidity propagate through molecular structure. At the expense of some mathematical complexity, the free energy decomposition becomes markedly simpler. In this work, we forced several DCM parameters to maintain complete transferability, where there is no dependence on local interaction types, sequence-context or chain length. Many of these parameters were based on fitting to molecular dynamics simulation results [22]. As such, there is too much uncertainty in the parameters to make a strong statement that the helix content predictions of the DCM will always be greater than that of the mLRM. If we were to allow all DCM parameters to be free, we would over fit the data. Therefore, as a check, we added an auxiliary term in our least squares function to drive the DCM to produce a helix content at high temperature to be the same as the mLRM, which is 10%. In doing this, we considered 8 non-transferable parameters to be free. The result is that the DCM can produce quantitatively similar results as the mLRM, but the overall least squares error increases because the heat capacity predictions worsen by about 20%. We conclude from this check, that there is nothing

inherent within the DCM that it cannot predict low helix content (if this were actually the case).

Uncertainties in Experimental Estimates

Much experimental work has shown that the statistical coil state in small polypeptides does not mimic well the random coil. In particular, it was shown by NMR that a small 7-mer polyalanine peptide has little helix content [40]. This result may not be too surprising considering the intrinsic uncertainties related with estimating the helix content in short polypeptides has always been problematic [41,42], especially in terms of adjusting for the finite size of the polypeptides. Basically, there is a lack of a universal “zero” used in circular dichroism (CD) measurements, which is related to the lack of a universal “k” value which describes the length dependence [43] where it typically ranges from 2.5 to 4.3, and as high as 6.3 has been used. In particular, a k-value of 2.5 was used in Eq. (3) found in REF [25]. Furthermore, UV resonance Raman (UVR) spectroscopy can measure more accurately low amounts of helix content compared to CD spectroscopy because the magnitude of the molar ellipticity per peptide bond is nearly linear with the number of peptide bonds [44]. As a result, it is known that CD measurements usually underestimate the amount of helix content in short polypeptides [44]. An added complication is that the solvent conditions between the NMR experiment [40] and the heat capacity measurements and CD measurements [25] were all different. In particular, the NMR study was done in 10 mM phosphate buffer, pH 7. The specific heat and CD measurements were both done at pH 2.0 but they used different buffers: The DSC used 20mM glycine buffer while the CD experiment used 10mM NaCl and 1 mM sodium phosphate, sodium borate and sodium citrate (which they term the CD buffer). After noting these experimental difficulties in finding absolute numbers, we performed a 100 ns molecular dynamics simulation (after equilibration) in explicit water at 300K for a 7-mer polyalanine peptide using Gromacs [45,46] with the OPLS-AA force field [47]. The statistics we counted for the center alanine was that it was found in a helix conformation 42% of the time.

In view of our molecular dynamics results, the other computational results [37–39] stated earlier, and the intrinsic experimental uncertainties, we simply leave the DCM as predicting higher alpha-helix content than might otherwise be expected. It is interesting to point out that the DCM predictions show that the lowest value of helix content is not chain length dependent. This means that there is no cooperativity involved, and any discrepancies that may reside in the model parameters reflect the intrinsic property of the amino acids. Nevertheless, the main results of this work can be stated without pinning down what the exact parameterization should be. Explicit calculation of network rigidity allows the nucleation process to be predicted, irrespective of chain length, sequence, solvent properties and even different types of topology in the molecular structure. The DCM is a general theory that has been successfully applied to proteins [28,31] and the beta-hairpin to coil transition [27]. Based on the current implementation of the DCM here and in prior works, only a few non-transferable parameters remain to account for solvent effects to fit to experimental data, such as heat capacity, as expected for a phenomenological theory. Encouraged by these results, and other related works, a large scale effort to determine a complete DCM parameterization for all 20 amino acids that is statistically significant is underway. Only in this type of large-scale process can the number of required parameters within the DCM win over that needed in Lifson-Roig type models.

Conclusions

Both a modified Lifson-Roig model (mLRM) and a Distance Constraint Model (DCM) adequately reproduce the heat capacity and helix content data taken on several polypeptides by Richardson and Makhatadze[25]. The mLRM provides slightly better fits to heat capacity

data, but the DCM fits better to the helix content in terms of linearly correlating with the data. Both models are consistent with the experimental helix propensity scale developed by Pace and Scholtz. Based on only the results presented here, and at the level of quantifying which model describes the experimental data better, neither model can be objectively discarded over the other. However, between these two competing models, there is a difference in parameterization and interpretation of the parameters.

The results presented here together with previous results, strongly suggests the DCM exhibits considerable transferability in its parameterization. Moreover, the DCM is based on a system-independent principle that invokes a long-range mechanical interaction that links the nonadditivity of conformational entropy to how rigidity and flexibility propagate through molecular structure. This alternative viewpoint offers a new modeling scheme that is much less complicated regarding parameterization of individual types of microscopic interactions within a free energy decomposition because it directly accounts for sequence-context dependence due to nonadditive effects.

Methods

Considering the six sequences listed in Table 2, there are seven different types of amino acid residues that need to be parameterized (ALA, GLU, ARG, VAL, GLY, and the TYR and ALA endcaps). For the DCM, H-bonds must be parameterized as well. Relying on parameter transferability, previously determined IHB, ALA and ALA-endcap parameters are invoked, except for the coil state energy, v_c^A , which is taken as a fitting parameter.

This procedure reduces the number of unknown DCM parameters to 17. However, this nominal count is drastically reduced to only 7 parameters because the sequence with GLY is not used in any fitting, and the same endcap parameters are used regardless of amino acid type. An eighth free parameter, the enthalpy of bond formation to solvent (which replaces a intramolecular hydrogen bond), was initially allowed to vary with little effect, thus it has been fixed to a previously used value (see the u parameter in Table 3). In general, the solvent H-bond energy is a nontransferable solvent-dependent parameter.

The ranking of helix propensities obtained by Pace and Scholtz [4] is invoked, which is given as:

$$\Delta\Delta G^A < \Delta\Delta G^E \approx \Delta\Delta G^R < \Delta\Delta G^V < \Delta\Delta G^G. \quad (5)$$

Based on this consensus ranking [4] that glutamic acid (E) and arginine (R) have similar helix propensity, and because these two amino acid residues are identically spaced for each sequence (see Table 2), glutamic acid and arginine are assigned identical parameters. Since the ΔG^A cancels out in all comparative terms in Eq. 5, we also enforce two inequalities given as $\Delta G^A < \Delta G^E$ and $\Delta G^E < \Delta G^V$ where:

$$\Delta G^x = v_c^x - v_a^x + u - u_{aaa} - 2T(\delta_c^x - \min(\delta_a^x, \gamma_{aaa})) \quad (6)$$

at $T = 300$ K. Two extreme states of either all α -helix or all coil with no IHB are compared using Eq. 6 for a hypothetical homogenous chain. Under this circumstance, the rank ordering of helix propensities will be significant in spite of context dependencies. These inequalities are sufficiently restrictive to ensure consistency over multiple best fits, especially when the heat capacity and helix content data are simultaneously fitted.

When simultaneously fitting two disparate types of experimental data, there is no universal way to form a single measure of goodness of fit. First, the three heat capacity curves were fit without regard to helix content data. Second, a reasonable, but not perfect fit to the helix content data was identified. The experimental uncertainties on the helix content data ranged from 2–4% and a value slightly larger was used. The reasoning is that since the fits to heat capacity were imperfect, a similar degree of imperfection in helix content was acknowledged. Third, the joint chi-squared (goodness of fit) measure is taken as the sum of the two terms: the chi-squared of the heat capacity curves plus a chi-squared for the helix content data scaled such that: the reasonable helix content fit would have an equal chi-squared contribution to the best heat capacity fits found. With this measure, the two models yielded roughly identical total chi-squared values. The relative weights of the two chi-squared terms were adjusted until the best fits of the two models yielded near-identical total chi-squared values. The use of both types of measurements together (instead of purely helix content) constrains the model predictions enough to eliminate problems related to multiple good solutions for different sets of model parameters.

The sequence, G6, containing glycine is also analyzed, but it does not have an α -helix state. As such, we simply adjusted the parameters of GLY (given in Table 4) to be consistent with the helix propensity scale without additional fitting. The DCM parameters yield ΔG^A to be -1.50 kcal/mol. Pace and Scholtz [4] indicate that the ΔG^G is as much as 2 kcal/mol greater. Selecting this greatest estimate, we use ΔG^G of 0.50 kcal/mol at $T = 300$ K. Following Pace and Scholtz, we linearly scale our ΔG^x to arrive at a helix propensity scale (hp) given in Table 4, where by definition, alanine is set at 0 and glycine is set at 1. By holding ΔG^G to a fixed value, allows us to adjust v_c^G and δ_c^G independently. Multiple good solutions exist, but we report the parameters where δ_c^G has the greatest conformational entropy in the coil state compared to all other amino acids. The enthalpy was adjusted to keep the helix-content prediction to be as low as possible. Alternatively, we could have optimized the GLY parameters to fit well along the helix content line. However, this was not done in order to use the helix content predictions as an independent check of the DCM consistency across amino acid types.

In the case of the mLRM, after a similar process of reducing the number of free parameters to 8, the largest reasonable gap in helix propensity between alanine to glycine in ΔG , which is 2 kcal/mol, is selected to drive down the helix content of the G6 curve. With this large gap, the ΔG of glycine is 1.7 kcal/mol, making the helix state very unfavorable. Nevertheless, the G6 peptide has a greater degree of helix content remaining than found in by experiment. Therefore, there is an overall inconsistency in attempting to reduce the helix content on the G6 sequence (the only one with GLY) on the one hand, and to match the helix propensity scale as listed in Table 4. The values selected split the difference in error.

Acknowledgments

This work was supported in part by California State University, Channel Islands to GGW and by NIH Grant R01 GM073082 to DJJ. We thank Dennis Livesay for fruitful discussions and for helpful comments on this manuscript. We also thank Hui Wang for running the 100 ns long MD simulation for us to check helix content within a 7-mer polyalanine chain in pure water.

References

1. Scheraga HA. Biopolymers. 2007; 89:479–485. [PubMed: 18008324]
2. Pauling L, Cory RB, Branson HR. Proc Natl Acad Sci USA. 1951; 37:205–210. [PubMed: 14816373]
3. Scholtz JM, Baldwin RL. Annu Rev Biophys Biomol Struct. 1992; 21:95–118. [PubMed: 1525475]

4. Pace CN, Scholtz JM. *Biophys J*. 1998; 75:422–427. [PubMed: 9649402]
5. Blaber M, Zhang XJ, Lindstrom JD, Pepiot SD, Baase WA, Matthews BW. *J Mol Biol*. 1994; 235:600–624. [PubMed: 8289284]
6. O'Neil KT, Degradó WF. *Science*. 1990; 250:646–651. [PubMed: 2237415]
7. Mark AE, van Gunsteren WF. *J Mol Biol*. 1994; 240:167–176. [PubMed: 8028000]
8. Dill KA. *J Biol Chem*. 1997; 272:701–704. [PubMed: 8995351]
9. Park S-H, Shalongo W, Stellwagen E. *Biochemistry*. 1993; 32:7048–7053. [PubMed: 8334134]
10. Chakrabarty A, Kortemme T, Baldwin RL. *Prot Sci*. 1994; 3:843–852.
11. Myers JK, Pace NC, Scholtz JM. *Proc Natl Acad Sci USA*. 1997; 94:2833–2837. [PubMed: 9096306]
12. Chang C-F, Zehfus MH. *Peptide Sci*. 1996; 40:609–616.
13. Yang J, Spek EJ, Gong Y, Zhou H, Kallenbach NR. *Protein Sci*. 1997; 6:1264–1272. [PubMed: 9194186]
14. Lifson S, Roig A. *J Chem Phys*. 1961; 34:1963–1974.
15. Bierzynski A, Pawlowski K. *Acta Biochim Pol*. 1997; 44:423–432. [PubMed: 9511955]
16. Mu noz V, Serrano L. *J Mol Biol*. 1995; 245:275–296. [PubMed: 7844817]
17. Ho BK, Thomas A, Brasseur R. *Prot Sci*. 2003; 12:2508–2522.
18. Schmidler SC, Lucas JE, Oas TG. *J Comp Biol*. 2007; 14:1287–1310.
19. Wako H, Saito N, Scheraga HA. *J Prot Chem*. 1983; 2:221–249.
20. Murza A, Kubelka J. *Biopolymers*. 2008; 91:120–131. [PubMed: 18814306]
21. Zimm BH, Bragg JK. *J Chem Phys*. 1959; 31:526–535.
22. Jacobs DJ, Dallakyan S, Heckathorne A, Wood GG. *Phys Rev E*. 2003; 68(061109):1–22.
23. Guyon E, Roux S, Hansen A, Bideau D, Troadec J-P, Crapo H. *Rep Prog Phys*. 1990; 53:373–419.
24. Vorov OK, Livesay DR, Jacobs DJ. *Biophysical J*. 2009; 97:3000–3009.
25. Richardson JM, Makhataдзе GI. *J Mol Biol*. 2004; 335:1029–1037. [PubMed: 14698297]
26. Jacobs DJ, Wood GG. *Biopolymers*. 2004; 75:1–31. [PubMed: 15307195]
27. Jacobs, DJ.; Fairchild, M. *Progress in Biopolymer Research*. Sanchez, Pablo C., editor. Nova Publishers; NY:
28. Jacobs DJ, Dallakyan S. *Biophys J*. 2005; 88:903–915. [PubMed: 15542549]
29. Livesay DR, Jacobs DJ. *Proteins*. 2006; 62:130–143. [PubMed: 16287093]
30. Jacobs DJ, Livesay DR, Hules J, Tasayco ML. *J Mol Biol*. 2006; 358:882–904. [PubMed: 16542678]
31. Livesay DR, Dallakyan S, Wood GG, Jacobs DJ. *FEBS Letters*. 2004; 576:468–476. [PubMed: 15498582]
32. Istomin AY, Gromiha MM, Vorov OK, Jacobs DJ, Livesay DR. *Proteins*. 2008; 70:915–924. [PubMed: 17803237]
33. Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF. *Proteins*. 2001; 44:150–165.
34. Lee MS, Wood GG, Jacobs DJ. *J Phys: Cond Mat*. 2004; 16:S5035–S5046.
35. Rohl CA, Chakrabarty A, Baldwin RL. *Prot Sci*. 1996; 5:2623–2637.
36. Qian H, Schellman JA. *J Phys Chem*. 1992; 96:3987–3994.
37. Daggett V, Levitt M. *J Mol Biol*. 1992; 223:1121–1138. [PubMed: 1538392]
38. Fitzkee NC, Rose GD. *Proc Natl Acad Sci USA*. 2004; 101:12497–12502. [PubMed: 15314216]
39. Beck DAC, Alonso DOV, Inoyama D, Daggett V. *Proc Natl Acad Sci USA*. 2008; 105:12259–12264. [PubMed: 18713857]
40. Shi Z, Olson CA, Rose GD, Baldwin RL, Kallenbach NR. *Proc Natl Acad Sci USA*. 2002; 99:9190–9195. [PubMed: 12091708]
41. Gans PJ, Lyu PC, Manning MC, Woody RW, Kallenbach NR. *Biopolymers*. 1991; 31:1605–1614. [PubMed: 1814507]
42. Manning MC, Woody RW. *Biopolymer*. 1991; 31:569–586.

43. Firestine AM, Chellgren VM, Rucker SJ, Lester TE, Creamer TP. *Biochemistry*. 2008; 47:3216–3224. [PubMed: 18266321]
44. Ozdemir A, Lednev IK, Asher SA. *Biochemistry*. 2002; 41:1893–1896. [PubMed: 11827535]
45. Lindahl E, Hess B, van der Spoel D. *J Mol Modeling*. 2001; 7:306–317.
46. Hess B, Kutzner C, van der Spoel D, Lindahl E. *J Chem Theory Comp*. 2008; 4:435–447.
47. Jorgensen WL, Maxwell DS, Rives TJ. *J Am Chem Soc*. 1996; 118:11225–11236.

Figure 1.

Top: Schematic diagram of a section of alpha-helix. Along the backbone, there are ϕ_i and ψ_i dihedral angles for the i -th residue. The staggered thick solid lines represent $i-2$ to $i+2$ H-bonding along the chain, showing that a H-bond spans three complete residues. **Middle Center:** The local geometry of the i -th residue is characterized by five degrees of freedom, consisting of two dihedral angles $\{\phi_i, \psi_i\}$ and three distances $\{b_i, c_i, d_i\}$. **Bottom Left:** The two dihedral angles characterize the local backbone conformation of the i -th residue. **Bottom Right:** The three distances characterize the local geometry of the H-bond that spans across the $i-1$, i and $i+1$ residues. The geometry of a H-bond depends on the ϕ_{i-1} , ψ_{i-1} , ϕ_i , ψ_i , ϕ_{i+1} and ψ_{i+1} , dihedral angles of each of the three residues that a H-bond spans.

Figure 2.
Best fit results for the heat capacity data from Richardson and Makhatadze [25] with the mLRM using 8 free parameters.

Figure 3. Best fit results for the heat capacity data from Richardson and Makhatadze [25] with the DCM using 7 free parameters.

Figure 4.

The DCM and mLRM predicted helix fraction at 0 C is plotted versus corresponding experimental values. Note that both models require arbitrary linear transformation from the actual calculated helix content to this scale. Equations are given by Eq. 3 and Eq. 4 for the DCM and mLRM, respectively.

Figure 5.

Direct predictions from the mLRM (black lines) and the DCM (red lines) for three types of polypeptide sequences of the form $(Y-(XEARA)_n)$ where $X=ALA, VAL$ and GLY are compared from the top to bottom rows respectively. On the left and right columns, C_p and helix content as a function of temperature are respectively plotted. The size effect of the polypeptides is monitored, where for each sequence its length is varied with $n=4, 5, 6, 7, 8$. The progression of chain length is obvious because the smaller the “ n ” the less cooperative the transition.

Table 1

The free energy decomposition scheme is defined. There are 2 amino acid states (α -helix and coil), and 8 H-bond states. An H-bond spanning a triplet of amino acids is characterized based on the number of α -helix and coil states encountered. Only symmetric triplets are listed. For example, triplets with 2 α -helix states and 1 coil state (i.e. $\frac{1}{3}$ coil) implies *aac*, *aca* and *caa* have the same parameters.

Interaction	Amino acid x	Backbone hydrogen bond						
Allowed state	α -helix	Coil	No coil			1/3 coil	2/3 coil	3/3 coil
Enthalpy	$0^{\dagger}$	1^x_c	u_{aaa}			u_{aca}	u_{cac}	$0^{\dagger}$
Entropy	δ^x_a	δ^x_c	$2^{\dagger}$			γ_{aca}	γ_{cac}	N/A
Relevant DOF	2	2	3			3	3	3
								Break
								u
								N/A
								None

[†] Arbitrarily defined to set a reference without loss of generality.

Table 2

From Richardson and Makhatadze [25] their sequences have similar form given by $Y(XEARA)_n$, with X being ALA, GLY or VAL, and with $n = 6$ for all chains except A4, for which $n = 4$. The dimension of the transfer matrix to propagate rigidity information from the C-to N-terminus ($N \leftarrow C$) is smaller than that for $N \rightarrow C$. This is because TYR at the N-terminus requires additional calculation for its affect on other amino acids down stream along the chain.

Chain	Sequence	$\dim_{N \rightarrow C}$	$\dim_{N \leftarrow C}$
A4	YAEARAAEARAAEARAAEARA	570	547
A6	YAEARAAEARAAEARAAEARAAEARA	570	547
V5	YAEARAAEARAAEARAAEARAVEARAAEARA	771	741
G6	YGEARAGEARAGEARAGEARAGEARAGEARA	589	527
V6	YVEARAVEARAVEARAVEARAVEARAVEARA	637	594
V345	YAEARAAEARAVEARAVEARAVEARAAEARA	771	741

Table 3

DCM parameters for an intramolecular H-bond (IHB). Enthalpies are in units of kcal/mol when not specified. Entropies are dimensionless. Quantities indicated with a dagger superscript are arbitrarily defined to set a reference without loss of generality.

IHB	<i>aaa</i>	<i>aca</i>	<i>cac</i>	<i>ccc</i>
u_{xyz}	-4.637	-2.827	-2.399	0 [†]
γ_{xyz}	2 [†]	2.149	2.760	2.917

Table 4

The top eleven rows of the table are mLRM enthalpy and entropy parameters: E_v , σ_v , E_w and σ_w translated to Lifson-Roid (v, w) and Zimm-Bragg equivalents (σ, s) at $T = 300$ K. All enthalpies have units of kcal/mol and all entropies are dimensionless “pure” entropies. Free energy differences are calculated by Eq. 6 at $T = 300$ K. The middle six lines are DCM parameters: δ_a and δ_c are the dimensionless entropy parameters of the alpha helix and coil states (respectively) and v_c is the enthalpy of the coil state. The bottom two lines of the table are average helix propensity from Pace and Scholtz [4] is given by $\langle hp \rangle$ and their standard deviation given by σ_{hp} . For GLU and ARG, the Pace and Scholtz values were averaged. Quantities indicated with a dagger superscript are arbitrarily defined to set a reference without loss of generality.

Parameter	Endcap	ALA	Glu/Arg	VAL	GLY
E_v	-0.676	1.06	1.06	1.06	1.06
σ_v	-3.30	-1.40	-1.40	-1.40	-1.40
E_w	-	-1.09	-0.711	-0.312	-0.09
σ_w	-	-1.38	-1.38	-1.38	-1.38
v	0.114	0.042	0.042	0.042	0.042
w	-	1.553	0.826	0.424	0.293
σ	0.013	1.77×10^{-3}	1.77×10^{-3}	1.77×10^{-3}	1.77×10^{-3}
s	-	1.395	0.792	0.407	0.294
ΔG_{LRM}	-	-0.264	0.515	0.115	1.736
$\Delta \Delta G_{LRM}$	-	0	0.78	0.38	2.0
hp_{LRM}	-	0	0.39	0.19	1.0
δ_a^x	2 [†]	2.60	0.39	1.60	0.39
v_c^x	-3.10	-1.27	-0.53	-2.54	-1.43
δ_c^x	3.50	3.60	3.53	2.40	3.53
ΔG_{DCM}	0.35	-1.045	0.055	-0.725	0.955
$\Delta \Delta G_{DCM}$	1.85	0	1.10	0.32	2.0
hp_{DCM}	0.93	0	0.32	0.16	1.0
$\langle hp \rangle$	-	0	0.6	0.18	1.0
σ_{hp}	-	0	0.20	0.22	0