THE JOURNAL OF PHYSICAL CHEMISTRY B

ARTICLE

pubs.acs.org/JPCB

A New Two-State Polymer Folding Model and Its Application to α-Helical Polyalanine

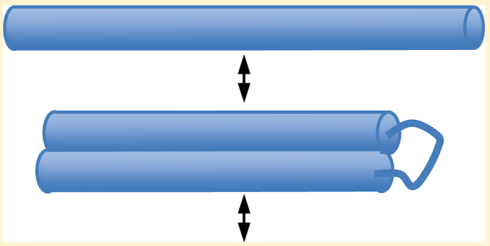
Per Linse,*[†] Peter Palenčár,[‡] and Tomáš Bleha[‡]

[†]Physical Chemistry, Department of Chemistry, Lund University, Box 124, SE-221 00 Lund, Sweden

[‡]Polymer Institute, Slovak Academy of Science, 845 41 Bratislava, Slovakia

Supporting Information

ABSTRACT: A new two-state polymer folding model is proposed, in which the folding of a stiff helical polymer is enabled by allowing for short sequences of coils connecting shorter and separated helices. The folding is driven by short-range attraction energy among stacked helices and is opposed by the free-energy cost of forming coils from helical monomers. Principal outcomes of the model are equilibrium distribution of the number of helices and their length in helical polymers. The proposed model is applied to α-helical polyalanine. The distribution of the number of α-helices as a function of number of alanine residues is fitted to the corresponding result from molecular dynamics simulation employing an all-atom potential model with very good agreement. The influence and significance of the fitting parameters and possible use of the two-state folding model are discussed.



INTRODUCTION

The classic helix−coil transition theories by Zimm and Bragg[1] and by Lifson and Roig[2] are based on the notion that the stability of a (straight) helix is independent of the chain length. This notion is increasingly challenged by recent investigations. For example, a whole family of integral membrane proteins that includes rhodopsin and bacteriorhodopsin forms, in membranes, compact bundles of seven approximately antiparallel α-helices. Bowie[3] has performed a statistical analysis of packing preferences of transmembrane helices for three protein structures including bacteriorhodopsin by using the Protein Data Bank files. The survey revealed that helix length varied from 14 to 36 residues with a preference for lengths greater than 20 residues. A more recent statistical analysis[4] based on sampling a large set (over 2000) of proteins reports a sharp reduction of the helix abundance with the length of protein helices, reinforcing the notion that the helix stability depends on its length.

In this contribution, we propose a new two-state polymer folding model for describing certain folding of stiff polymers. Our basic assumption is that the polymer folding is enabled by conformational changes in a stiff structural element (such as an helix) into several and shorter stiff elements linked by flexible elements (such as a coils). The folding is driven by short-range attraction among stacked stiff elements and is opposed by a free-energy cost of forming the flexible elements from parts of the stiff element. Our model differs from the classical ones in two important aspects. First, our model departs from a stiff polymer and examines the effects of introducing flexibility, whereas the classical ones starts from a random coil that is turned into a helix. Second, we include short-range attraction among spatially neighboring stiff elements of the polymer.

The usefulness of the proposed model is examined by applying it on properties of an all-atom potential model of α-helical polyalanine. The formation of secondary structure elements such as α-helices and β-sheets belongs to the elementary protein folding processes. The tendency of helix formation is largely a local property of the peptide sequence. Alanine is a small and neutral residue commonly found in transmembrane proteins. Experiments by Chakrabartty et al.[5] have shown that alanine is the best helix stabilizer of all naturally occurring amino acids. Hydrophobic peptides with high alanine content are known for their inclination to form helical structures in aqueous solution. It has been therefore assumed that the lowest free-energy structure of polyalanine is α-helical. The experimental determination of the stable conformation of polyalanine is hindered by its low solubility in water at ambient temperature.

Considerable efforts have been devoted in coarse-grained and atomistic simulation studies[6−12] to understand the conformational properties of $(Ala)_l$ polypeptides. For short ones, typically of $l < 22$, the computational studies revealed that the abundance of individual secondary structures in Ala peptides at room temperature is significantly affected by the medium.[10−12] The α-helix was verified to be a stable structure of $(Ala)_l$ in vacuum and hydrophobic medium, whereas in aqueous solution extended conformations of β-sheets dominated.[11,12] Levy et al.[11] have argued that the helical propensity concept has to be extended by an incorporation of the effect of medium to properly account for the thermodynamic preferences of alanine peptide conformations.

Received: February 28, 2011
Revised: August 17, 2011
Published: August 22, 2011

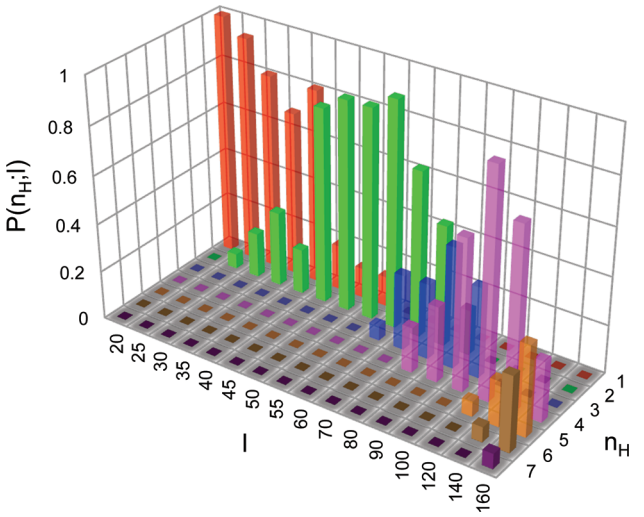


Figure 1. Normalized probability distribution $P(n_H;l)$ as a function of the number of α-helices n_H and the number of residues l from all-atom molecular dynamics simulation in the absence of solvent (data for $l = 40$, 45, and 60 are taken from ref 14 and the remaining data are unpublished, Palenčár and Bleha).

Recently, studies on the folding of polyalanine in vacuum (absence of solvent) have been performed using a molecular dynamics (MD) simulation technique.[13,14] It was found that aside from the environment, the structure of polyalanine peptides is greatly affected by the number of amino residues (refs 13 and 14 and references therein). In particular, the equilibrium between (i) a long α-helix and (ii) more compact conformations comprising two or more stacked α-helices joined by small coils and stabilized by short-range attraction between neighboring α-helices were investigated in detail.[14] The degree of folding was found to depend on the medium, temperature, and the number of alanine residues. The probability of n_H α-helices at various numbers of alanine residues l from all-atom simulations in vacuum is presented in Figure 1. In the case of a short polyalanine ($l \leq 20$), only one α-helix appeared with one nonhelical residue at each helix end; hence there was no fold of the helical polyalanine. At increasing length of polyalanine, a progressively larger number of α-helices was formed, which were separated by coils forming U-turns. Throughout, for a given number of residues a distribution among polyalanine peptides having a different number of α-helices was established, signifying a not too large free-energy difference between them. The observation of bundles of stacked α-helices in simulation of long-chain polyalanine in vacuum[13,14] can be directly relevant to an assembly of peptides in hydrophobic membrane-like environment and to the organization of integral membrane proteins.[3] Such proteins traverse the cell membrane in a zigzag fashion and often fold in a form of antiparallel-oriented α-helical strands separated by a short U-turn composed of a flexible coil.

TWO-STATE POLYMER FOLDING MODEL

Description. We start with a polymer composed of l monomers forming a single and stiff helix composed of $l - 2l_E$ monomers with a short and nonhelical tail composed of l_E monomers at each end. Consider now that sections of the polymer undergo helix-to-coil transitions and form flexible coils. These coils will divide the original helix into several and shorter helices. All coils are assumed to be of the same length, and for a given l all helices are also of same length. Furthermore, the flexibility of the coils facilitates the helices to stack (see Figure 2). The nature of the short-range attraction driving the stacking is in the present context of no concern. We assume that the free-energy change of this process is composed by (i) a free-energy cost associated with the $n_H - 1$ helix-to-coil transitions of regions involving l_C monomers and (ii) a free-energy gain from the short-range attraction among the n_H stacked helices. In the following, we propose expressions of these two free-energy contributions and a concomitant expression of the probability of finding n_H helices separated by $n_H - 1$ coils as a function of the number of monomers l. For conveniences, Table 1 contains a summary of some important variables.

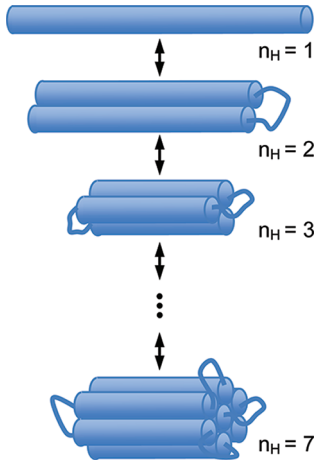


Figure 2. Schematic illustration of (top to bottom) a single α-helix, two, three, and seven parallel and adjacent α-helices packed in a two-dimensional hexagonal structure. Coils joining the helices are also illustrated.

The conservation of the l monomers residing in helices, coils, and at the polymer ends gives the relation

$$l = n_H l_H + (n_H - 1)l_C + 2l_E \quad (1)$$

where l_H is the number of residues in one helix, l_C is the number of residues in one coil (assumed to be constant), and l_E is the number of nonhelical residues (assumed to be constant) at *each* polymer end. By utilizing the two free-energy contributions as discussed above, the free-energy change of forming n_H helices separated by $n_H - 1$ coils from a single helical polymer comprising l residues can be expressed according to

$$\Delta F(n_H; l) = (n_H - 1)\Delta f + E(n_H, l_H) \quad (2)$$

where Δf denotes the free-energy change of forming *one* coil from a section of the helix and $E(n_H,l_H)$ the short-range attraction among n_H helices of equal length. Here, and in the following, all energies are given in kT units. Furthermore, assume

$$E(n_H, l_H) = g(n_H)l_H\varepsilon \quad (3)$$

where $g(n_H)$ is (mainly) a geometrical factor denoting the number of helix−helix contacts and ε the short-range attraction between two neighboring helices per monomer. As the attraction is short-range, (i) the values of $g(n_H)$ are insensitive to the number of monomers of an helix, and (ii) the relative values of $g(n_H)$ are insensitive to the precise surface-to-surface distance between neighboring helices. For example, a packing of the helices in a 2D-hexagonal lattice would give $g_{hex}(n_H) = 0$, 1, 3, 5, 7, 9, 12, etc. for $n_H = 1$, 2, 3, 4, 5, 6, 7, etc.

dx.doi.org/10.1021/jp2019395 |J. Phys. Chem. B 2011, 115, 11448–11454

Table 1. Important Variables and Their Description of the Two-State Polymer Folding Model[a]

variable	description[b]
n_H	number of helices
l	number of monomers
l_H	number of monomers in one helix
l_C	number of monomers in one coil
l_E	number of nonhelical monomers at each polymer end
$\Delta F(n_H; l)$	free energy of a polymer consisting of l monomers making n_H helices relative to a single helical polymer
$E(n_H, l_H)$	short-range attraction among n_H helices each consisting of l_H monomers
Δf	free energy change of forming a coil from a section of a helix
ε	strength of short-range interaction between two neighboring helices per monomer
$g(n_H)$	number of helix−helix contacts for a bundle of n_H helices
$P(n_H; l)$	probability of a polymer comprising of l monomers to form n_H helices of equal length and separated by coils containing l_C monomers

[a] All energies are given in kT units, where k is the Boltzmann constant and T the temperature. [b] In an application on α-helical polyalanine the following conversions should be made: (i) "polymer" → "polyalanine", (ii) "monomer" → "alanine residue", and (iii) "helix" → "α-helix".

The insertion of eq 3 into eq 2 and the subsequent use of eq 1 give the total free-energy change of folding a helical polymer from one into n_H shorter and stacked helices according to

$$\Delta F(n_H; l) = (n_H - 1)\Delta f + g(n_H)\varepsilon \frac{l - [(n_H - 1)l_C + 2l_E]}{n_H}, \quad \left(0 < n_H < n_{H,max} \approx \frac{l}{l_C} + 1\right) \tag{4}$$

Finally, assuming a Boltzmann distribution, the probability $P(n_H; l)$ of having n_H stacked helices for a given l becomes

$$P(n_H; l) = \frac{\exp\left[-\Delta F(n_H; l)\right]}{\sum_{n'_H} \exp\left[-\Delta F(n'_H; l)\right]} \tag{5}$$

with

$$\sum_{n_H} P(n_H; l) = 1, \ \forall\ l \tag{6}$$

as being the (conventional) normalization condition of probabilities.

By the use of the probability distribution $P(n_H; l)$, the average number of helices $\langle n_H(l)\rangle$, the average length of helices, $\langle l_H(l)\rangle$, and the average fraction of the residues residing in helices, $\langle x_H(l)\rangle$, can be expressed according to

$$\langle n_H(l)\rangle = \sum_{n_H} n_H P(n_H; l) \tag{7}$$

$$\langle l_H(l)\rangle = \sum_{n_H} l_H P(n_H; l) \tag{8}$$

$$\langle x_H(l)\rangle = \left\langle \left(\frac{n_H l_H}{l}\right)(l)\right\rangle = \frac{1}{l}\sum_{n_H} n_H l_H P(n_H; l) \tag{9}$$

where l_H from eq 1 has to be used in eqs 8 and 9 to eliminate l_H. Whereas $\langle n_H(l)\rangle$ depends on l only through the probability distribution, $\langle l_H(l)\rangle$ depends on l though both $l_H(l)$ and $P(n_H; l)$, and $\langle x_H(l)\rangle$ depends on l though both $(n_H l_H/l)(l)$ and $P(n_H; l)$.

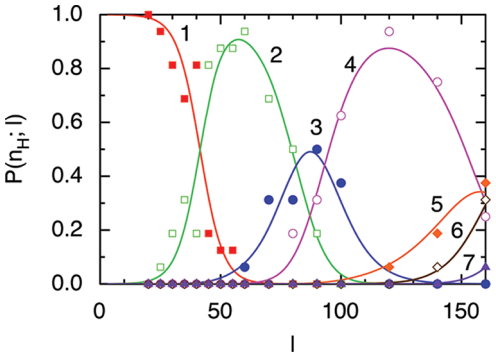


Figure 3. Normalized probability distribution $P(n_H; l)$ of indicated number of α-helices n_H as a function of the number of residues l from all-atom molecular dynamics simulation (symbols, data from Figure 1) and the two-state folding model fitted to the simulation data (curves).

Application to Polyalanine. The two-state polymer folding model will be applied to α-helical polyalanine, and the following changes of notation: "polymer" → "polyalanine", (ii) "monomer" → "alanine residue", and (iii) "helix" → "α-helix" become appropriate. In the absence of detailed experimental data, we use recent data from MD simulation of an all-atom model of polyalanine (see also Supporting Information for further description of the MD simulations). It is important to note that we *do not* asses the quality of the simulation data as such, rather we use the results of that study as an *input* to examine to which degree a complicated many-body system can be captured by such a simple folding model as ours.

RESULTS

Fit to All-Atom MD Simulation Data. Our two-state folding model was first fitted to results from MD simulation of an all-atom model. For that reason, the simulation data behind Figure 1 have been replotted in Figure 3 (symbols) as probability distribution functions $P(n_H; l)$, $1 \le n_H \le 7$. Thus, for a given number of alanine residues l, the distributions provide the probability of the polyalanine to possess n_H α-helices of equal length. For example, for l = 80 the probability of having two α-helices is ≈50%, three α-helices ≈40%, and four α-helices ≈10%.

Our two-state folding model as described in the previous section by eqs 4 to 6 contains $4 + n_{H,max}$ parameters, namely, l_C, l_E, Δf, ε, and $g(n_H)$, $1 \le n_H \le n_{H,max}$. With $n_{H,max} = 7$ (cf. Figure 3), we obtain 11 parameters. To reduce the number of degrees of freedom, we have used the following constraints: (i) $l_C = 4$ and $l_E = 1$ (both values taken from further analysis of all-atom simulations reported in ref 14), (ii) $g(1) = 0$ (no short-range attraction with one helix present), and (iii) $g(2) = 1$ (one helix−helix contact with two helices present). The values of the remaining $11 - 4 = 7$ parameters [Δf, ε, and $g(n_H)$, $3 \le n_H \le 7$] of a least-squared fit to the all-atom MD simulation data are given in Table 2, whereas Figure 3 also provides the fitted probability distributions (curves) together with the MD data. Within the uncertainty of the simulation data, we see that the characteristic features such as (i) successive increase of the number α-helices at increasing l, (ii) the locations of the maxima, and (iii) the shape and intensity of the maxima are all very well-described by the two-state folding model.

Properties of Folded Polyalanine. We will now examine some properties of folded polyalanine as deduced from the two-state folding model. Figure 4 shows (i) the average number of helices, (ii) the average length (number of residues) of helices, and (iii) the

Table 2. Values of Variables from a Least-Squared Fit to All-Atom MD Simulation Data of Polyalanine

variable	value
l_C	4[a]
l_E	1[a]
$g(n_H)$, n_H = 1, 2	0, 1[b]
Δf	7.36[c]
ε	−0.415[c]
$g(n_H)$, n_H = 3, 4, 5, 6, 7	2.33, 4.16, 5.98, 8.14, 10.51[c]

[a] Fixed and taken from further analysis of the all-atom simulations reported in ref 14. [b] Fixed. [c] Fitted.

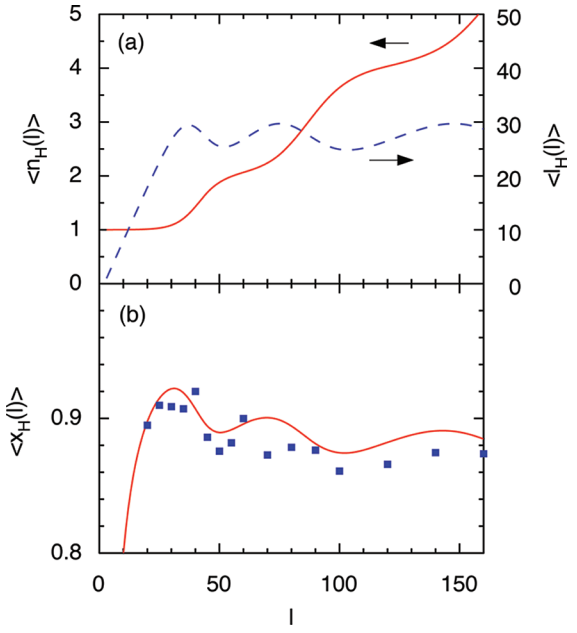


Figure 4. (a) Average number $\langle n_H(l)\rangle$ (solid curve) and average length $\langle l_H(l)\rangle$ (dashed curve) of helices and (b) average fraction of residues residing in α-helices $\langle x_H(l)\rangle$ (solid curve) as a function of the number of residues l from the two-state folding model with values of variables given in Table 2. In part b, corresponding data from all-atom molecular dynamics simulations (symbols).

average fraction of residues in α-helices as a function of the number of residues in the polyalanine as obtained from the two-state folding model with parameters fitted to the all-atom MD simulation data.

First, the average number of helices $\langle n_H(l)\rangle$ is a monotonously increasing function of the number of residues l (Figure 4a, solid curve). The slope is relatively small at $l \approx 60$ and 120 where $\langle n_H(60)\rangle \approx 2$ and $\langle n_H(120)\rangle \approx 4$, which is in perfect agreement with the location of the maxima of the $P(n_H;l)$ distributions occurring at $l \approx 60$ and ≈120 shown in Figure 3. Moreover, we notice that the probability of having three helices is too small to give rise to an inflection point at $l \approx 90$. After the formation of one extended helix, the average length of the α-helices $\langle l_H(l)\rangle$ oscillates between 25 and 30 residues for the l-interval investigated (Figure 4a, dashed curve). Maximal helical lengths appear when $P(n_H; l)$ versus l display a negative slope roughly between the maximum in $P(n_H; l)$ and the following significant transition to a conformation with a larger number of α-helices.

Finally, the fraction of residues being in helices $\langle x_H(l)\rangle$ has the most complicated behavior (Figure 4b). It was computed from probability data $P(n_H; l)$ of the two-state folding model obtained from the fit to corresponding probability data of the all-atom potential model (see Figure 3). The main feature comprises (i) an initial steep rise followed by a maximum and a gentle decay. On a more detailed level, (ii) a stepwise (or oscillatory) behavior is overlaid on the decay with the first step appearing at $l \approx 40$ and a second one at $l \approx 88$. (iii) Finally, between the steps, $\langle x_H(l)\rangle$ displays upturns. Observation (i) arises from the fact that the number of nonhelical end-residues is fixed, and hence their influence on $\langle x_H(l)\rangle$ at increasing l is declining and becomes only important at small l. (ii) After referring to Figure 4a, we conclude that the step at $l \approx 40$ is related to the transition between conformations with one helix and two helices, and the second step at $l \approx 88$ corresponds to a transition between conformations with two and four helices. As before, the probability distribution of the conformation with three helices is too small for this conformation to have a visible impact on $\langle x_H(l)\rangle$. Thus, the detailed pattern in Figure 4b is related to the transition from a dominance of n_H to n_H' helices with $n_H' = n_H + 1$ or $n_H + 2$. (iii) Just at a transition, new coil(s) is (are) formed over a small increase of l, leading to a sizable reduction of $\langle x_H(l)\rangle$. However, between transitions, the change of the number of coils is small; most of the additional residues just lengthen the helices and bring about an increase in $\langle x_H(l)\rangle$ over a short interval of l.

Figure 4b also contains corresponding data obtained from the underlying all-atom MD simulations (symbols). A comparison of the fraction of residues being in helices from the all-atom potential model and the two-state folding model shows reasonable similarities, but the numerical uncertainty of the former data makes it difficult to analyze $\langle x_H(l)\rangle$ in detail. Apparently, the use of a fit of the simulated probability data $P(n_H; l)$ of the all-atom potential model to an appropriate folding model (see Figure 3) leads to a numerically more *robust* evaluation of $\langle x_H(l)\rangle$ (curve) than a direct evaluation from the simulation (symbols). This is a significant achievement of the two-state folding model.

Parameter Variation. We will now examine the effect of varying the fitting parameters on the probability distribution $P(n_H; l)$ with the least-squared fit as a reference. Figure 5 (dotted curves) displays the probability distributions after an increase of the free energy change of forming a coil from a section of an α-helix from $\Delta f = 7.36$ to 8.50, a decrease of the strength of short-range interaction between two neighboring α-helices per alanine from $\varepsilon = -0.415$ to -0.501, and a simultaneous increase of Δf and decrease of ε. An increase in Δf causes a positive shift of the interval of l at which peptides with a specific number of α-helices dominates (Figure 5a). Thus, and as expected, a larger free-energy cost of forming a four residues long coil from an α-helix (larger Δf) stabilizes polyalanine with fewer α-helices. An increase of the short-range attraction between α-helices (more negative ε) favors polyalanine conformations with a larger number of α-helices (Figure 5b). Finally, a simultaneous change in Δf and ε such that their shifts in $P(n_H = 1;l)$ cancel each other leaves (i) the location of *all* maxima practically unshifted and (ii) sharpens the transition between polypeptide conformations having n_H and n_H' α-helices (Figure 5c).

We have also examined the significance of $g(n_H)$, which in the two-state folding model describes the number of (effective) pairs of α-helices being close and parallel to each other. First, Figure 6 provides probability distributions for $g(n_H) = g_{hex}(n_H)$, which represents an ideal 2D-hexagonal packing. A comparison of these probability distributions with those obtained from the least-squared fit shows in particular that an ideal hexagonal stacking

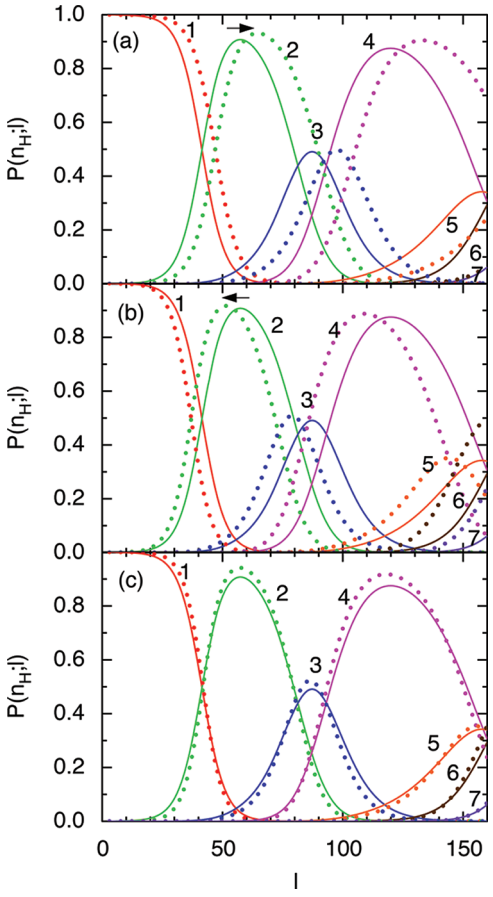


Figure 5. Normalized probability distribution $P(n_H; l)$ at indicated number of α-helices n_H as a function of the number of residues l from the two-state folding model with values of variables given according to Table 2 (solid curves) and (a) an increase of Δf from 7.36 to 8.50 (dotted curves), (b) a decrease of ε from −0.415 to −0.501 (dotted curves), or (c) a simultaneously change of Δf from 7.36 to 8.50 and ε from −0.415 to −0.501 (dotted curves).

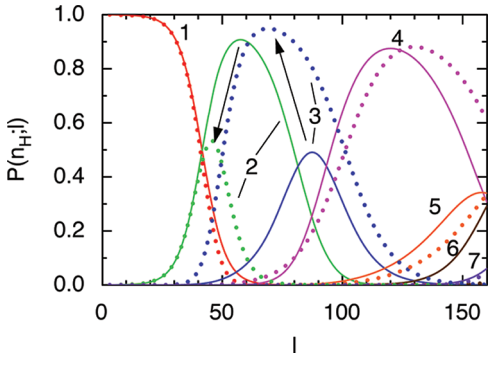


Figure 6. Normalized probability distribution $P(n_H; l)$ at an indicated number of α-helices n_H as a function of the number of residues l from the two-state folding model with values of variables according to Table 2 (solid curves) and values of variables given in Table 2 but $g(n_H) = g_{hex}(n_H) = 0$, 1, 3, 5, 7, 9, and 12 (dotted curves).

increases the probability of bundles with three helices on the expense of bundles with two helices.

In Figure 7, we compare the fitted number of effective helix−helix contacts with those of an ideal hexagonal packing.

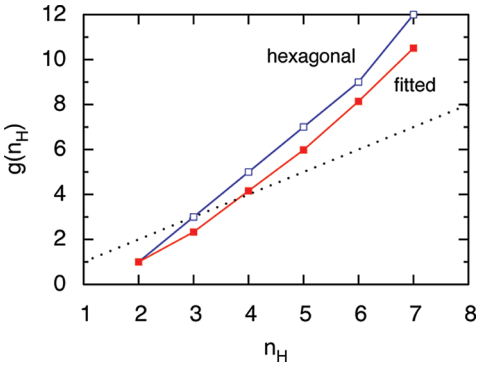


Figure 7. Number of effective helix−helix contacts $g(n_H)$ as a function of the number of helices in a bundle of α-helices from the two-state folding model fitted to all-atom molecular dynamics simulation data (filled symbols) and an ideal two-dimensional hexagonal packing (open symbols). See the text for further details. The linear relation $g(n_H) = n_H$ is also shown (dotted line).

We recall that the value of $g(n_H)$ at $n_H = 2$ is set to unity; thus, the data give the number of effective helix−helix contacts measured in terms of the effective helix−helix contact formed with two helices. Thus, the interaction energy of the helix−helix contact with $n_H = 2$ is not necessarily needed to be equal in the all-atom potential model and in the ideal case. First, we have $g(n_H) > (n_H)^x$, $x = 1$, for $n_H \geq 4$, due to the formation of a 2D-bundle of parallel α-helices. Second, the difference between the fitted and ideal values is fairly constant. The marked difference between the fitted and the ideal behavior in Figure 7 lies formally in the larger number of effectively interacting polyalanine α-helices in the 2D-hexagonal lattice as compared to the simulated data for polyalanine with three α-helices, namely, 3 versus 2.33. This issue will be discussed in more detail in a subsequent report.

To conclude, the parameters Δf and ε, representing the free energy change of forming a coil from a section of an α-helix and the strength of short-range interaction between two neighboring and parallel α-helices per alanine, respectively, individually shift the location of the $P(n_H; l)$, whereas the ratio of Δf and ε affects the width of the probability peaks. The values of $g(n_H)$ regulate the relative importance of conformations with a different number of α-helices.

DISCUSSION

Since solvent is absent in our consideration, the formation of several and folded α-helices should be an intrinsic property of the polypeptide. Our model complements the conventional helix−coil treatments by considering an equilibrium between a single and several folded α-helices. Then, for example, ΔH_m, the melting enthalpy of $(Ala)_l$, involves apart from the conventional helix-to-coil melting term also the stabilization energy of several α-helices forming a bundle relative to a single one. Our data suggest that an $(Ala)_l$ helix is divided into two or several α-helices of length of about 28 residues (Figure 4a) and a single α-helix is practically absent for $l > 55$ (Figure 3). This is in very good agreement with the average experimental value of 26 residues for some transmembrane helices found by Bowie.[3]

Model Parameters. Our thermodynamic two-state model invokes the parameters Δf denoting the free energy change of forming a coil from a section of an α-helix, ε denoting the strength of short-range interaction between two neighboring

α-helices per alanine, and $g(n_H)$, $3 \leq n_H \leq 7$, representing the number of effective helix−helix contacts for a bundle of n_H α-helices. On the basis of (i) the relatively large scatter of the simulation data (Figure 3) and (ii) the fairly large covariance of Δf and ε (Figure 5c), we estimate the relative precision of Δf and ε to be ≈10%. Moreover, the least-squares fit is very sensitive to a change of a single value of $g(n_H)$, $3 \leq n_H \leq 7$; a change of a percent deteriorates the fit considerably making the determination of $g(n_H)$ more accurate. However, note that the fit is invariant of individual scaling of ε and of the function g while keeping their product constant (cf. eq 3).

The fitted values of $g(n_H)$, $3 \leq n_H \leq 7$ are consistently smaller than $g_{hex}(n_H)$, but not too far off. Thus, our analysis is basically in agreement with the all-atom simulation study,[13,14] in which the helices appeared parallel and stacked in a 2D-hexagonal manner (cf. Figure 4 of ref 14). The smaller fitted values of $g(n_H)$, $3 \leq n_H \leq 7$, implies a less efficient alignment for three or more helices as compared to two aligned α-helices in the underlying all-atom simulation study.

Finally, the fitted $\Delta f = 7.36$ and $\varepsilon = -0.415$ (in kT units), implying a cost of nearly 5 kJ/mol per alanine residue to form a disordered coil of α-helical residues and an attraction of 1 kJ/mol per alanine residue, are both generally reasonable. Unfortunately, experimental data on the value of parameters Δf and ε for folding of long $(Ala)_l$ peptides are not yet available. However, the conformational equilibrium of the charged (protonated) Ala-rich peptide $AcA_{14}KG_3A_{14}K + 2H^+$ was investigated by ion mobility measurements in the gas phase.[15] This peptide involving lysine (K) was designed to form two antiparallel Ala helices connected by a short glycine (G) loop. From the equilibrium constant between straight and folded conformations, the enthalpy change $\Delta H = -45$ kJ mol^{-1} for docking the two helices together was determined. This gives an attraction between helices per residue of about 3 kJ mol^{-1}, in reasonable consistency with our estimation of ε for $(Ala)_l$ peptides.

As in all coarse-graining procedures, the final assessment of a coarse-grained model lies in its usefulness to predict certain (and relevant) properties of the underlying and more detailed model *and* at the same time being less cumbersome and/or less time-consuming to use. Here, we have found that the equilibrium distribution of the number and size of α-helices predicted by an all-atom potential model of polyalanine studied by simulations can accurately be represented by a two-state folding model.

Relevance to the Protein and Synthetic Polymer Modeling. The proposed two-state folding model is topical and closely related to recent advances in the structural characterization of both complex proteins and simple synthetic polymers. Regarding proteins, the compact bundles of helices treated in the model resemble a structural motif "helix-turn-helix" that is considered a super-secondary structural element in membrane proteins. The proteins form the bundles of α-helices when transverse the membrane in a zigzag way, in close analogy to our two-state folding model. The concept of self-organization is omnipresent also in polymers, for example, as polymer crystallization. The regular chain folding is a principal mode of polymer crystallization, resulting into lamellar structure of polymer materials. The folded structure of long $(Ala)_l$ peptides is analogous to that found by MD simulations at incipient lamellar crystallization of synthetic helical polymers poly(oxymethylene) and isotactic poly(propylene).[16]

Implications and Possibilities. The ability of our two-state folding model to represent properties of an all-atom potential model of polyalanine rises several implications and possibilities:

(A) An important conclusion is that the coarse structure of a complicated many-body system such as a polyalanine peptide can be understood in terms of two effective interactions: (i) helix stability with respect to coil and (ii) short-range helix−helix attraction. The change of representation from an all-atom potential model to the two-state folding model involves a drastic reduction of the number of variables needed to describe the equilibrium between conformations involving different number of stiff structure elements of the polymer. This is a significant finding with conceptual implications.

(B) By fitting the data of an all-atom potential model obtained from simulation on the two-state folding model, a more robust analysis of other quantities such as the fraction of residues being in helices could be performed. Similar approaches describing experimental and computational data of other hydrophobic polypeptides are straightforward to make.

(C) By augmenting how the size of the stiff element depends on the number of constituting monomers, the model could be used to predict the suppression or promotion of stiff elements in confined geometries as well as in a heterogeneous environment, such as membranes.

(D) Furthermore, by having temperature-dependent data, free-energy variables can be divided into enthalpic and entropic contributions, enabling a modeling of temperature dependence. Hence, the model can assist in elucidating the key factors and interactions driving the self-assembly of polymers and in this way provide insight into the mechanism of self-organization of complex polymers.

CONCLUSIONS

A simple two-state polymer folding model of α-helical polyalanine, of which key elements are a free-energy cost of forming coils from a helix and a reduction of the potential energy arising from parallel and stacked helices, is proposed. The model is able to quantitatively describe the probability of having various numbers of α-helices as a function of the number of alanine residues for polyalanine of different lengths as predicted by an all-atom potential model solved by MD simulations. In particular, it was found that the probability of having two α-helices stacked was larger and the probability of having three α-helices stacked was lower as compared to an ideal 2d-hexagonal stacking.

Here, we have used properties of an all-atom potential model as input to our two-state folding model as similar experimental data are not yet available. We emphasize that the assessment of the usefulness of the two-state model does not depend on the accuracy of the all-atom potential model as long it reasonably well describes properties underlying a more detailed all-atom potential model. Furthermore, we envision that the presented two-state folding model and obvious extensions of it can be of significant value to describe certain conformational equilibria of secondary structures of regular hydrophobic polypeptides in nonpolar media, in confined space, and in heterogeneous systems as such membranes.

ASSOCIATED CONTENT

Supporting Information. Details of underlying molecular dynamics simulation of an all-atom potential model of L-polyalanine in vacuum. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*E-mail: per.linse@fkem1.lu.se.

ACKNOWLEDGMENT

Financial support by the Swedish Research Council (VR) through the Linnaeus Grant Organizing Molecular Matter (OMM) Center of Excellence (239-2009-6794) and through an individual grant to P.L. (621-2007-5251) is gratefully acknowledged. The research of P.P. and T.B. was supported by Science and Technology Assistance Agency grants (APVV-0079-07 and APVV-0607-07).

REFERENCES

(1) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1958**, *28*, 1246. Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526.
(2) Lifon, S.; Roig, A. *J. Chem. Phys.* **1961**, *34*, 1963.
(3) Bowie, J. U. *J. Mol. Biol.* **1997**, *272*, 780.
(4) Zhang, W.; Lei, H.; Chowdhury, S.; Duan, Y. *J. Phys. Chem. B* **2004**, *108*, 7479.
(5) Chakrabartty, A.; Kortemme, T.; Baldwin, R. L. *Protein Sci.* **1994**, *3*, 843.
(6) Shental-Bechor, D.; Kirca, S.; Ben-Tal, N.; Haliloglu, T. *Biophys. J.* **2005**, *88*, 2391.
(7) Mortenson, P. N.; Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2002**, *117*, 1363.
(8) Nymeyer, H.; Garcia, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13934.
(9) Garcia, A. E. *Polymer* **2004**, *45*, 669.
(10) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472.
(11) Levy, Y.; Jortner, J.; Becker, O. M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2188.
(12) Soto, P.; Baumketner, A.; Shea, J.-E. *J. Chem. Phys.* **2006**, *124*, 134904.
(13) Palencar, P.; Bleha, T. *Macromol. Theor. Simul.* **2010**, *19*, 488.
(14) Palencar, P.; Bleha, T. *J. Mol. Model.* **2011**No. 10.1007/s00894-011-0997-4.
(15) Kaleta, D. T.; Jarrold, M. F. *J. Am. Chem. Soc.* **2003**, *125*, 7187.
(16) Yamamoto, T. *Polymer* **2009**, *50*, 1973.