JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

On the Role of Entropy in the Stabilization of α-Helices

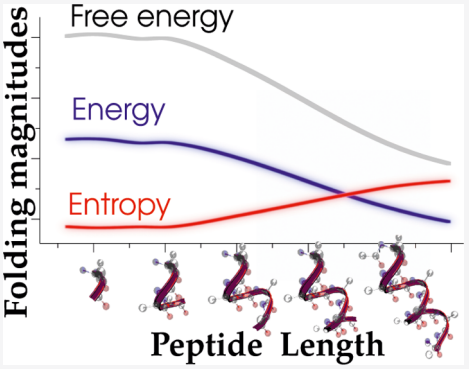
Adolfo Bastida,* José Zúñiga, Alberto Requena, Beatriz Miguel, and Javier Cerezo*

Cite This: *J. Chem. Inf. Model.* 2020, 60, 6523−6531

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Protein folding evolves by exploring the conformational space with a subtle balance between enthalpy and entropy changes which eventually leads to a decrease of free energy upon reaching the folded structure. A complete understanding of this process requires, therefore, a deep insight into both contributions to free energy. In this work, we clarify the role of entropy in favoring the stabilization of folded structures in polyalanine peptides with up to 12 residues. We use a novel method referred to as K2V that allows us to obtain the potential-energy landscapes in terms of residue conformations extracted from molecular dynamics simulations at conformational equilibrium and yields folding thermodynamic magnitudes, which are in agreement with the experimental data available. Our results demonstrate that the folded structures of the larger polyalanine chains are stabilized with respect to the folded structures of the shorter chains by both an energetic contribution coming from the formation of the intramolecular hydrogen bonds and an entropic contribution coming from an increase of the entropy of the solvent with approximate weights of 60 and 40%, respectively, thus unveiling a key piece in the puzzle of protein folding. In addition, the ability of the K2V method to provide the enthalpic and entropic contributions for individual residues along the peptide chain makes it clear that the energetic and entropic stabilizations are basically governed by the nearest neighbor residue conformations, with the folding propensity being rationalized in terms of triads of residues.



1. INTRODUCTION

Like any other physical process, protein folding is governed by minimization of free energy. What is, however, quite remarkable is the fact that the change in energy is nearly compensated by a change in entropy, which results in a small change of free energy.[1] This happens, for example, in the formation of α-helices, which is energetically favored and entropically hindered,[2−8] with both contributions equilibrated at temperatures close to 25 °C. At the molecular level, this delicate balance between energy and entropy comes from changes of the intra- and intermolecular interactions, which determine the energy contribution and from the sizes of conformational spaces explored by the protein and the solvent, which account for the entropy variation of the folding process. The high flexibility of proteins, in which each residue may adopt different conformations defined by the backbone dihedral angles ϕ and ψ, and the importance of the intermolecular interactions with the solvent water molecules make it impossible to characterize these intrinsically complex systems accurately with simple models. An additional difficulty arises from the fact that the conformation of each residue is mediated by the conformation of its nearest neighbor residues[9−17] (nnrs), so the secondary structure of the protein has to be accounted for as the result of conformational preferences of amino acid triads, instead of the isolated dynamic of each residue.

Molecular dynamics (MD) simulations emerge then as the ideal tool to study the general principles governing the protein-folding processes since they can predict the protein's native structure from its amino acid sequence.[18,19] The evaluation of the folding thermodynamic magnitudes from atomistic simulations is, however, far from straightforward.[20] In principle, one just needs to estimate accurately the energetic and entropic contributions to free energy in terms of the dihedral coordinates. In practice, the evaluation of the mean potential energy is problematic due to the large number of interactions involved,[16] and the entropic contribution of the solvent is, to say the least, uncertain.[18,21] Furthermore, the delicate energy−entropy balance in the folding process implies that both contributions have to be evaluated with a small margin of error in order to reach reliable conclusions on their roles in the formation of α-helices.

In a recent work,[16] we have presented an effective method named K2V to calculate the mean (ϕ, ψ) energy landscapes of peptides from MD simulations. The application of this method

Received: October 9, 2020
Published: December 7, 2020

https://dx.doi.org/10.1021/acs.jcim.0c01177
J. Chem. Inf. Model. 2020, 60, 6523−6531

to the initial folding process, in which the peptide has not yet reached its conformational equilibrium, allowed us to unveil the existence of an energetic self-folding mechanism that guides the molecule toward its helical conformation. In the present work, we extend our study to the evaluation of the folding thermodynamic magnitudes of peptides after equilibrium is reached, revealing the fundamental role that energy and entropy play in the stabilization of the α-helices. We have chosen polyalanine peptides as test models since they have a remarkable α-helix propensity.[22] In particular, we have used the capped polymers of alanine ACE-$(Ala)_m$-NME with m = 1−4, 6, 8, 10, and 12 which allows us to analyze the evolution of the folding magnitudes with the peptide chain length. It is shown that the folding energy per residue increases with the total number of alanine residues, while the entropic term, which includes the favorable contribution of the solvent, decreases as the molecule lengthens, thus favoring the stabilization of the α-helix. The ability of the K2V method to provide folding thermodynamic magnitudes for each residue of the peptide allows us, in addition, to analyze quantitatively the energetic and entropic influence of the nnrs.

2. METHODS

2.1. MD Simulations. MD simulations of the ACE-$(Ala)_m$-NME molecules with m = 1−4, 6, 8, 10, and 12, dissolved in water, were carried out using the GROMACS package v2016.4.[23,24] Each solute molecule was surrounded by a number of water molecules ranging from 600 to 2000 (depending on the length of the polyalanine chain) and placed in a cubic box of a size chosen to reproduce the experimental density of the liquid at room temperature. All polyalanine molecules were described using the CHARMM36[25] force field, and the flexible TIP3P model was used for the solvent water molecules. This force field has been recently shown[26] to produce good results in reproducing experimental J-coupling constants and amide I′ profiles for alanine in GAG and AAA tripeptides. Periodic boundary conditions were imposed in the simulations using the particle-mesh Ewald method to treat the long-range electrostatic interactions. The equations of motion were integrated using a time step of 0.5 fs. All simulations were carried out in a NVT ensemble at 298 K by coupling to a thermal bath.

Every system was equilibrated following a two-step process. In the first step, the system was propagated during 0.4 ns keeping the solute molecules frozen in a conformation in which every residue was in an α_R conformation. In the second step, the velocities of the polyalanine atoms were reset randomly using a Boltzmann distribution, after which an additional equilibration of 10 ns followed. The last step was repeated 112 times with different sets of velocities. Each of these 112 initial configurations was propagated during 10 ns generating the same number of trajectories. During these production runs, the kinetic energy of the solute atoms and the values of the dihedral angles were written every 5 fs. This computational strategy allowed us to obtain thermally and conformationally equilibrated systems, as tested by monitoring the time evolution of the average conformational populations of the residues during the 10 ns production runs, which remained basically constant.

2.2. Calculation of the Folding Thermodynamic Magnitudes. The usual procedure to calculate the folding thermodynamic magnitudes from MD simulations consists of fitting the data to a model for the helix−coil transition.[27] The Lifson−Roig model[28] characterizes the helix formation in terms of the equilibrium constant (w) which accounts for enlarging an existing helix into one residue. This equilibrium constant is directly related to the folding Helmholtz free energy through $\Delta F_f = -k_B T \ln w$ in a NVT ensemble. By performing MD simulations at different temperatures and fitting the results to the Helmholtz free energy expression

$$\Delta F_f = \Delta E_f - T\Delta S_f \tag{1}$$

it is possible to evaluate both the folding energy (ΔE_f) and entropic contribution ($T\Delta S_f$). From a practical point of view, this procedure is, however, computationally very demanding because it requires a large number of MD simulations at different temperatures. Also from a conceptual point of view, the method is compromised both by the use of a nucleation-elongation model whose validity is not fully assured and by the assumption that the folding energies and entropies remain fully constant in the temperature range used in the simulations.

An alternative way to calculate the Helmholtz free energy difference between two conformational regions, i and j, from MD simulations is by a simple counting method,[20] using the expression

$$\Delta F_{ij} = F_j - F_i = -k_B T \ln \frac{p_j}{p_i} \tag{2}$$

where p_i (p_j) is the number of times that the system visits the i (j) region during the simulation or, equivalently, the normalized population of the region. For this method to work properly, the MD simulations have to be long enough to sample the conformational space adequately, so that the magnitudes p_i and p_j provide true equilibrium conformational populations. If the energy barriers are substantially higher than the thermal energy, it is also necessary to use enhanced sampling techniques in the MD simulations. In the folding process of polypeptides, the folding free energy is given by

$$\Delta F_f = -k_B T \ln \frac{p_{\Omega_f}}{1 - p_{\Omega_f}} \tag{3}$$

where p_{Ω_f} is the fraction of the total population found in the folded region (Ω_f) spanned by the 2D conformational space of the (ϕ, ψ) dihedral angles. More generally, the (ϕ, ψ) conformational space can be divided into a $N \times N$ grid, and eq 2 can be used to evaluate the free energy differences between any couple of the resulting N^2 grid points, thus providing a free-energy map like that shown in Figure 1.

The selection of the size of the folded region deserves some attention. Typically,[8,27,29] the folded or helical region is identified with the α_R region, defined as that enclosing the maximum of the α_R population with $\delta \simeq \pm 30°$ intervals. The choice of this δ interval requires a detailed justification, since the definition of the helical region may alter substantially the balance between the folded and unfolded states and modify the values of the thermodynamic folding magnitudes. We take advantage in this respect of the well-known fact[30] that the formation of an intramolecular hydrogen bond (HB) interaction between the peptide CO of residue i and the peptide NH of residue $i + 4$ requires that the three consecutive residues $i + 1$, $i + 2$, and $i + 3$ are in the α_R conformation. The size of the helical region, located around the maximum of the α_R conformation, is then determined by conciliating such requirement with a geometrical description of HB interactions,

https://dx.doi.org/10.1021/acs.jcim.0c01177
J. Chem. Inf. Model. 2020, 60, 6523−6531

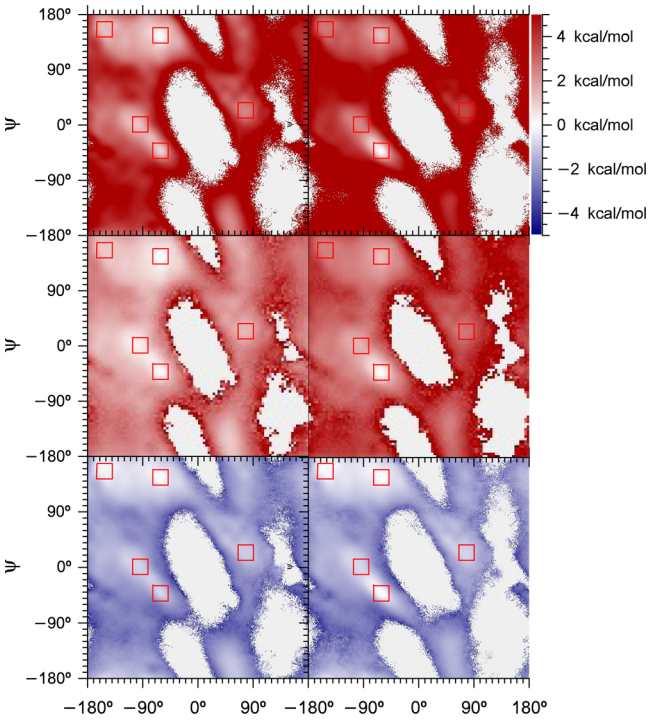


Figure 1. Helmholtz free energy (top), energy (medium), and entropy ($T\Delta S$) (bottom) maps for the m = 1 (left) and m = 12 (right) polyalanine molecules. Squares around the maxima of the α_R, α', α_L, β, and pPII conformations are plotted as guides for the eye. All maps were calculated using a 90 × 90 grid. Zero point energies were arbitrarily set at the maximum of the α_R conformation.

namely the definition of Zewail et al.[22] We thus set the helical region by ensuring that, for each geometrically identified HB, most of the conformations of the i + 1, i + 2, and i + 3 residues lie on this region. In order to carry out such analysis, in Figure 2 we show the fraction of HB interactions based on geometric

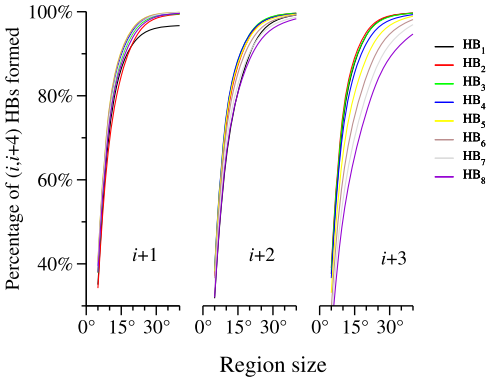


Figure 2. Fraction of geometrically defined (i, i + 4) HBs corresponding to α_R conformation of the i + 1, i + 2, and i + 3 residues, defined within the square regions of different size centered at (−61, −42°) for the decalanine (m = 10) molecule. The eight possible (i, i + 4) HBs are depicted separately.

considerations that correspond to α_R conformations of the i + 1, i + 2, and i + 3 residues, using a square region centered at (−61, −42°) for the decalanine (m = 10) molecule. As observed, it is necessary to consider a size region around ±30° in order to account for >90% of the HBs formed. This result is in agreement with the work by Best and Hummer[29] where an energetic criterium for the estimation of the HBs was applied. To facilitate the comparison with previous studies[27,29] we define the conformational regions using the $\phi \pm 35°$ and $\psi \pm 30°$ intervals.

Once the folded region has been set, the calculation of the folding energies should be simple in principle and just requires the evaluation of the mean energy of the system in the folded and unfolded regions. At thermal equilibrium, the mean kinetic energies are identical through all conformational space, since they depend solely on the number of degrees of freedom and the temperature of the system. It is only necessary to calculate then the mean potential energies, which in practice, though, is not viable due to the huge statistics required and the subsequent numerical errors associated.

We have recently shown,[16] nevertheless, that it is possible to determine accurately folding energies through potential-energy maps derived from the kinetic energy of the atoms. The calculations are carried out using the so-called K2V algorithm, which is based on the fact that the averaged changes in the kinetic energy of any internal coordinate of the system in a short time interval are equal to minus the potential-energy difference, that is

$$\overline{\Delta T}_{i\to j} = -\overline{\Delta V}_{i\to j} \tag{4}$$

because the kinetic energy changes arising from internal energy fluxes vanish at thermal equilibrium, in which the average kinetic energy per degree of freedom must be equal to $\frac{1}{2}k_BT$. In eq 4, the averaged kinetic (ΔT) and potential (ΔV) energy differences run over all $i \to j$ transitions occurring along the trajectories. We have shown also in our earlier work that only a small number of atoms have a real kinetic energy contribution to the displacement of the (ϕ_i, ψ_i) dihedral angles of the ith residue. In fact, the five atoms which define the dihedral angles (C_{i-1}–N_i–$C_{\alpha,i}$–C_i–N_{i+1}) plus the $H_{\alpha,i}$ and $H_{N,i}$ atoms, account all together for 99.5% of the total kinetic energy involved in the displacement of the (ϕ_i, ψ_i) angles of polyalanine chains. This important finding greatly reduces the statistical noise and allows us to evaluate accurately the potential-energy differences in eq 4 and more importantly, makes it possible to calculate the potential-energy difference for each residue independently. The division of the (ϕ, ψ) conformational space in a $N \times N$ grid, with N being high enough to ensure that the potential energy in every grid point remains constant, allows us to evaluate the (ϕ, ψ) potential-energy maps (see Figure 1) by simple regression analysis. More details about the K2V algorithm and some illustrative examples are given elsewhere.[16] The folding energies can then be calculated from these maps as the difference between the mean values of the potential energy inside and outside the folded region, as follows

$$\Delta E_f = \frac{\sum_{i\in\Omega_f} p_i V_i}{\sum_{i\in\Omega_f} p_i} - \frac{\sum_{i\notin\Omega_f} p_i V_i}{\sum_{i\notin\Omega_f} p_i} \tag{5}$$

Once the energy and Helmholtz free-energy maps are evaluated, the entropy map is directly obtained by subtraction of the former (see Figure 1) as can be done for calculating the folding entropy using eq 1. We should emphasize that this procedure allows us to calculate the folding thermodynamic magnitudes from MD simulations solely at the temperature of interest. It provides also independent values for every residue

https://dx.doi.org/10.1021/acs.jcim.0c01177
J. Chem. Inf. Model. 2020, 60, 6523−6531

in the polypeptide and requires no model to evaluate equilibrium folding constants.

3. RESULTS AND DISCUSSION

3.1. Folding Magnitude Maps. The main results obtained in this work are the average folding magnitudes, either global or per residue, which we present and discuss in the following sections. The central tools to obtain such magnitudes are the free-energy and the potential-energy maps, so we start by paying some attention to them before describing the average magnitudes.

We have already shown in Figure 1 the maps obtained for the $m = 1$ and $m = 12$ polyalanine molecules, with the magnitudes being averaged for the latter over the 12 residues. As observed, the energy barriers between the different conformers are typically a few times higher than the thermal energy for the two molecules, so the barriers can be overpassed when thermal fluctuations concentrate enough energy in the dihedral angles.[31] The depth of the α_R conformer energy well is, however, substantially greater than the depths of the other conformers for the $m = 12$ molecule. As explained in our previous work,[16] the α_R conformation is favored during the folding process by an energetic self-folding mechanism in which the successive formation of intramolecular $(i, i + 4)$ HBs plays a central role.

The availability of free-energy and potential-energy maps with a similar, and high, accuracy makes it possible to calculate the entropic contribution maps, shown in the bottom panels of Figure 1 by simple subtraction. The most striking observation in the entropy maps is that for the shorter $(m = 1)$ molecule, the entropy around the α_R conformation is lower than that for the extended β and pPII conformations, while for the $m = 12$ molecule, the entropy is similar in both cases. Therefore, the formation of the α_R helix increases the entropy of the system in long peptide chains. This fact can be interpreted in terms of the mobility of solvent degrees of freedom. In such a folded conformation, in which strong intramolecular contacts are formed within the peptide, the interaction with the surrounding solvent molecules becomes weaker and this, in turn, increases the mobility of the solvent and eventually raises its entropic contribution. Additionally, the entropy of the extended conformations is not basically altered by the length of the molecular chain. We will return to this conceptual image when describing the average magnitudes.

We should emphasize that the entropic maps presented provide the relative entropy between residue conformations coming from the effect of the rest of the degrees of freedom. Therefore, it is not straightforward to guess the folding entropy just from visual inspection of the maps. In fact, this thermodynamic magnitude is not derived from the entropic map in the present work but from the combination of the average potential-energy and free-energy terms (see eqs 1, 3, and 5). The evaluation of these magnitudes requires the definition of the folded and the unfolded states, which are associated to different regions of the 2D conformational space of the residues. The folded state spreads over a smaller region around the α_R conformer, while the unfolded state corresponds to the remaining conformational space. The much larger unfolded state region explains why the entropy of folding becomes eventually negative, even though the entropy at the small region associated to α_R conformer is higher.

3.2. Global Folding Magnitudes. The folding thermodynamic magnitudes of alanine-enriched polypeptides have been measured using different experimental techniques such as circular dichroism,[2−5] NH exchange,[2] isothermal titration calorimetry,[6,32] and $^{13}C{=}O$ NMR chemical shifts.[7] In spite of the fact that the experimental conditions were different in these works, using for instance distinct buffered water solutions and different peptide sizes and sequences, all of them practically agree in the experimental values provided, which are $\Delta H_f = -1.0 \pm 0.1$ kcal/mol (the difference between ΔE_f and ΔH_f is negligible in solutions at room temperature), $T\Delta S_f = -0.80 \pm 0.35$ kcal/mol, and $\Delta G_f = -0.20 \pm 0.05$ kcal/mol per alanine residue. This coincidence in the experimental results demonstrates that the measured folding magnitudes for the alanine residue are governed by the formation of the α_R-helix, a process which is basically unaltered by changes in the saline composition or the presence of a small fraction of other residues within the peptide molecule. The folding process is therefore energetically favored and entropically hindered. These results are usually rationalized by assuming that the α-helices of polyalanines are energetically stabilized by intramolecular $(i, i + 4)$ hydrogen bonds. Additional experimental data,[33] which show that alanine has a much smaller propensity to adopt α_R conformations in short peptides with no $(i, i + 4)$ HBs, seem to support such an interpretation. As for the decrease in entropy, it can be interpreted as a consequence of the more ordered structure of the α_R-helices, compared with the extended conformations of the molecule, which reduces its ability to explore the conformational space.[34] These two arguments undervalue, however, the role of the solvent molecules. In the first place, although water molecules form HBs with the C═O and N−H groups of the polypeptide when the peptide is in an extended conformation, these intermolecular HBs are absent in the helical conformation, and this has an energetic cost opposite to the energetic gain coming from the formation of the intramolecular HBs.[35−38] Second, apart from the conformational entropy due to the motion of the backbone and side-chains of the polypeptide,[34] there is a non-negligible entropic contribution from the water molecules which is not easy to evaluate.[39] The lack of direct experimental information about the folding magnitudes for short peptides, due to its low propensity to achieve fully helical states,[39] makes the MD simulations an appropriate tool to check the validity of the preceding arguments.

In Figure 3 we plot the folding magnitudes of the polyalanine molecules averaged over all residues, as extracted from our simulations and compare them with the experimental values. As observed, the folding free energies calculated reproduce the experimental tendency of being positive for short-chain polyalanines, with low propensity to achieve helical conformations, and decreasing toward negative values as the molecular chain lengthens and the helical structure becomes dominant. For instance, the average population of the α_R region for the $m = 1$ and $m = 12$ molecules are 8.1 and 71.1%, respectively. We see also that the estimated folding magnitudes for the longer $(m = 10, 12)$ polyalanine molecules agree well with the experimental results and with previous theoretical calculations by Best et al.[27] ($\Delta H = -1.2$ kcal/mol, $T\Delta S = -1.0$ kcal mol) obtained using the Lifson−Roig model and the same force field.

As for the evolution of the folding energetic and entropic contributions with the chain length, we observe that the folding magnitudes remain basically constant for the shorter $(m = 1{-}4)$ molecules with a small percentage of α_R populations, in line with previous studies[40,41] showing that

https://dx.doi.org/10.1021/acs.jcim.0c01177
J. Chem. Inf. Model. 2020, 60, 6523−6531

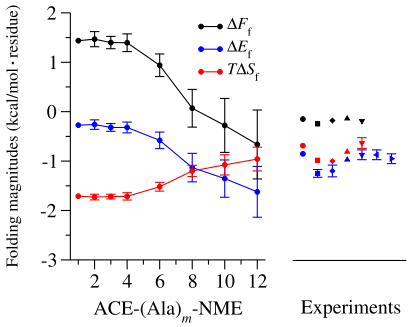


Figure 3. Folding Helmholtz free energy (ΔF_f), energy (ΔE_f), and entropic contribution at 25 °C ($T\Delta S_f$) for the polyalanine molecules obtained from MD simulations. Error bars correspond to the standard deviations of the magnitudes calculated for every residue in the molecule. Experimental results from refs 7 (circles),[2] (square, circular dichroism results),[2] (diamond, NH exchange results),[3] (triangle up),[4] (triangle down),[6] (triangle left), and[5] (triangle right).

peptides with up to five alanines are far away from folding. For larger polyalanine molecules, a simultaneous decrease of the folding energies and an increase of the entropic contribution toward less negative values are observed, with both tendencies being favorable for the formation of the α_R-helix. These results indicate that the formation of the intramolecular $(i, i + 4)$ HBs in the larger polyalanine molecules alters significantly the energetic and entropic balance between the folded and unfolded conformations. By comparing the energetic and entropic contributions for the shorter and larger molecules, we estimate that the formation of the folded alanines is favored in approximately 1.3 and 0.8 kcal/mol per residue, respectively. Accordingly, the formation of the α_R-helices has an energetic contribution which accounts for 60% of the change of the Helmholtz free energy, while the remaining 40% comes from the entropic term. The importance of the entropic contributions deserves a more detailed analysis.

The entropic contribution includes both the conformational entropy of the polyalanine molecules and the entropy of the solvent. The former contribution, which accounts for the transition from any unfolded conformation to the α_R region, is expected to be negative, as confirmed by the recent work by Sosnick et al.,[34] who obtained an average value of −1.5 kcal/mol for the conformational entropic contribution per alanine residue in helices based on MD simulations using the same version of the CHARMM force field. The entropic term of the solvent is expected to be positive if the mobility of water molecules increases upon folding, that is, if the number of HBs between the peptide and the solvent decreases with respect to that of the unfolded state. Our results can therefore be interpreted in terms of the different contributions of the solvent to entropy depending on the size of the peptide which affects its ability to form peptide−solvent HBs in the folded conformation. Short peptides in particular are unable to form intramolecular $(i, i + 4)$ HBs, thus having free sites to form HBs with the solvent in any conformation. This explains why the folding entropies obtained in our simulations for the shorter peptides, for which solvent plays only a minor role, are similar to the conformational entropy calculated by Sosnick et al.[42] which can be assumed to depend only slightly on the peptide length. The increase of the entropic term observed for the larger chains (∼0.8 kcal/mol) reveals then an entropic contribution from the water solvent, that is, a positive entropic contribution from the solvent during the folding process as expected from the loss of intermolecular HBs, which increases the mobility of the water molecules. We note that a recent study has attributed the origin of the entropic stabilization of the α-helices to the presence of delocalized soft low-frequency modes.[43] This explanation is also consistent with the independence shown by the entropy of extended conformation from the chain length, as already observed in the corresponding maps in Figure 1. Contrarily to α-helices, such extended structures are essentially equivalent regardless of the length of the chain, displaying a similar HB network, and therefore the contribution of the solvent to the entropic term does not significantly vary among peptides of different lengths.

3.3. Folding Magnitudes for Each Residue. One notable advantage of the K2V method is its ability to provide individual mean potential-energy maps for every residue, thus allowing us to study changes of folding magnitudes along the peptide chain, as shown in Figure 4. The overall profiles of the

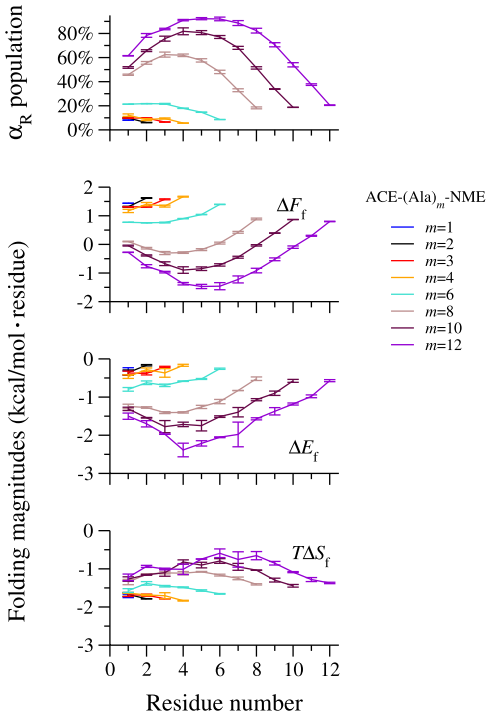


Figure 4. α_R populations, folding Helmholtz free energy (ΔF_f), energy (ΔE_f), and entropic contribution at 25° ($T\Delta S_f$) for every residue in the polyalanine molecules obtained from MD simulations. Error bars correspond to the standard deviations of the results obtained by block analysis by performing the statistics in two independent groups of trajectories.

Helmholtz free-energy changes between different polyalanine molecules, and also within each chain, are dictated by both the energetic and the entropic contributions, the former being a factor of ∼1.5 larger as discussed in the previous section. We observe in Figure 4 that the results for the shorter (m = 1−4) peptides are quite similar, and they evolve for the larger ones as the molecules tend to adopt the helicoidal structure.

We also observe in Figure 4 that the folding of the inner residues is favored over the outer ones, as shown in previous simulations[22] and experiments[44−46] and that the folding in the N-terminus residues is higher than the folding in the C-terminus residues, again in agreement with previous experimental observations[47] and theoretical simulations,[29,48] prob-

https://dx.doi.org/10.1021/acs.jcim.0c01177
J. Chem. Inf. Model. 2020, 60, 6523−6531

ably due to the intrinsic molecular asymmetry of both ends of the peptide. It is also interesting that the folding magnitudes of the extreme residues are much less affected by the length of the molecule than for the inner residues, being practically constant for the larger molecules. As expected, the folding process modifies more the thermodynamics of the inner residues, which participate more actively in the formation of the intramolecular $(i, i + 4)$ HBs, than the extreme ones. This also partially explains why the thermodynamic magnitudes in Figure 3 are not fully converged with the increasing length of the polyalanine molecules. As more residues are included, the percentage of inner residues increases and the global magnitudes get closer to the values obtained for the inner residues. Additionally, we observe that the percentage of folded inner residues is still increasing, although modestly for the larger peptides studied.

A closer inspection of the entropic term (bottom panel in Figure 4) reveals two clearly different trends, depending on the length of the polypeptide. First, the shorter molecules have more negative entropic terms, in agreement with the mean values shown in Figure 3. Interestingly, while the entropic term remains nearly constant for all residues along the chain for short peptides, it fluctuates notably along the chain for the largest peptides. The entropy change is therefore smaller for the inner residues than for the residues on the edges of the chain. The reason for this comes from the balance between conformational entropies of the peptide and the solvent. As far as the solvent contribution is concerned, the entropy profiles can be interpreted in terms of the different ability of inner/edge residues to unlock solvent molecules upon the formation of the helix, as discussed for the average values. While folded inner residues leave no HB sites to bind water molecules, the more solvent-exposed edge residues still face HB sites to water molecules. As a result, solvent molecules have more freedom to move upon folding for inner residues, and the entropic term of the solvent eventually becomes more positive for them, as derived from our simulations.

3.4. Interresidual Cooperative Effects on Folding Propensity. There is increasing evidence[9−17] supporting the idea that the nnrs' conformation affects the dynamics and energetics of a given residue and that the secondary structure of polypeptides must be interpreted in terms of the preferential conformations of amino acids triads. In order to test this possibility, we show in Figure 5 the average α_R-folding percentage of the inner residues of the polyalanine molecules as a function of the folding state of their nnrs. As seen, the general trend is that a given residue has a higher percentage of α_R-folding as more nnrs are already folded. Therefore, the α_R-folding of the nnrs of a given residue increases its own folding propensity. This behavior is observed for all polyalanine peptides analyzed, regardless of their length, that is, also including those that are not able to form strong intramolecular H bonds, although the magnitude of the effect is strongly favored for the larger molecules. Therefore, the effect is mainly originated from the formation of the intramolecular HBs, but there is also an intrinsic tendency of the alanine residues to facilitate the α_R-folding of their nearest neighbors. Moreover, this effect is slightly more pronounced when the nnr is placed on the left side with respect to the right side. Overall, these results prove that the folding energetics of a residue is greatly influenced by its nnrs.

Therefore, we applied the K2V method to calculate mean potential-energy maps for every residue conditioned to the α_R-

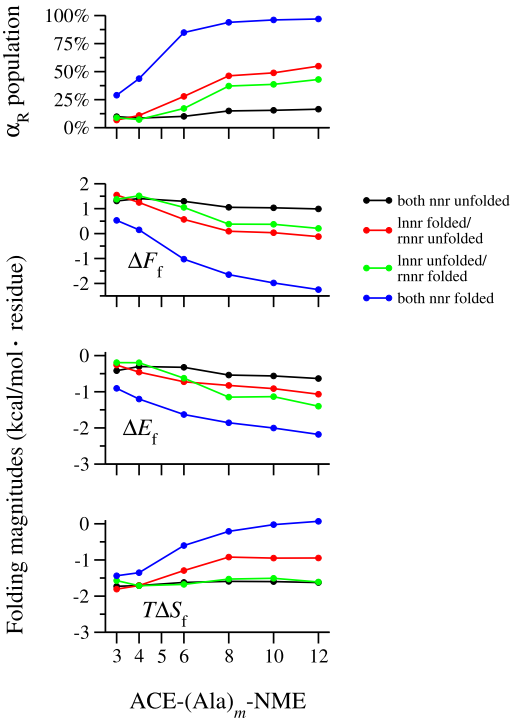


Figure 5. Average α_R populations, folding Helmholtz free energy (ΔF_f), energy (ΔE_f), and entropic contribution at 25° ($T\Delta S_f$) of the inner residues of the polyalanine molecules as a function of the folding state of the nnrs: both are unfolded (black), only the left one is folded (red), only the right one is folded (green), and both are folded (blue).

folding state of its nnrs which provides the corresponding thermodynamic folding magnitudes also shown in Figure 5. As a general trend, we observe that the α_R-folding of an alanine residue is both energetically and entropically favored when its nnrs are already α_R-folded and that the magnitudes of both contributions are similar. Although the mechanism of action, at the molecular level, behind this effect is not evident, the sterical effects[49] can play an important role also in the folding of a residue by modifying the conformational space to be accessed and can also alter the interaction of the CO and NH peptide groups with the water molecules.

This cooperative effect is not exclusive to the α_R conformations. Previous theoretical[50,51] and experimental[14] works have shown that the pPII conformation of a given alanine residue is favored when its nnrs are already in the pPII conformation, so there is a pPII propagation free energy that enhances the formation of segments of pPII structure.[51] In order to check if our MD simulations properly account for this effect, we have extended our analysis to the pPII conformations, as shown in Figure 6. The results show that the presence of nnrs in the pPII conformation also favors the pPII-folding of a given alanine residue, similarly to the α_R conformations. However, our results show two significant differences between the α_R and pPII cooperativity. First, we see in Figure 6 that in the pPII-folding, the origin of the effect is almost completely energetic, since the entropic contribution is mostly independent of the pPII-folding state of the nnrs. Second, the evolution of the pPII populations with the length of the polyalanine molecules shows that the percentage obtained when both nnrs are already pPII-folded is basically constant (∼60%), while the equivalent results for the α_R-

https://dx.doi.org/10.1021/acs.jcim.0c01177
J. Chem. Inf. Model. 2020, 60, 6523−6531

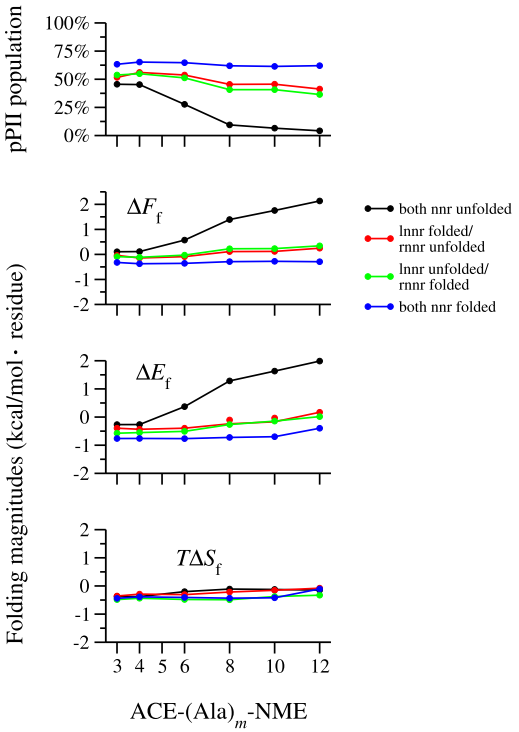


Figure 6. Average pPII populations, folding free energy (ΔF_f), energy (ΔE_f), and entropic contribution at 25° ($T\Delta S_f$) of the inner residues of the polyalanine molecules as a function of the folding state of the nnrs: both are unfolded (black), only the left one is folded (red), only the right one is folded (green), and both are folded (blue).

folding cooperativity increased with the chain length. This fact demonstrates that the α_R-cooperativity is enhanced by the formation of the 3D helical structure, while the pPII-folding cooperativity is basically a local effect non-dependent on the secondary structure of the peptide. These results would agree well with the interpretation proposed by Gnanakaran and Garcia[50,51] that an energetic mechanism in which the propagation of the pPII structure is helped by the formation of a delocalized water channel around the backbone, which favors the pPII conformations.

4. CONCLUSIONS AND PERSPECTIVES

In this work, we have investigated the thermodynamics of the folding of polyalanine chains, known for their propensity to fold in α_R-helix conformation, by analyzing the data extracted from MD simulations at the conformational equilibrium of polyalanine molecules of different lengths, with 1 to 12 residues, dissolved in water.

Average thermodynamic magnitudes of folding, including both energetic and entropic contributions, are obtained from the MD trajectories by combining standard tools to evaluate the free energy with an efficient and accurate method recently developed by our group (the K2V method) to obtain the potential-energy landscapes in terms of residue conformation. The protocol used is rooted in the availability of free- and potential-energy landscapes and rigorously defines the folded and unfolded states by selecting regions that ensure consistency, along the simulation, between the formation of $(i, i + 4)$ HBs and the α_R conformation adopted by all residues participating in the helical fragment.

Our predictions for the free energy, the energy, and the entropy of folding of the largest peptides agree reasonably well with the available experimental data. In addition, the variations of the thermodynamic quantities with the peptide length clearly support the remarkable role that both energy and entropy play in facilitating the process. In particular, a decrease of the energetic term combined with an increase of the entropic term is observed when going from the smaller (1 to 4 residues) to the longer peptides. The energetic and entropic components per residue along the chain for the longest peptides also reveals that folding is energetically and entropically favored for inner residues. The analysis of the thermodynamic folding magnitudes in terms of the folding state of the nnrs reveals that the folding of a residue is both energetically and entropically favored when its nnrs are already folded. More importantly, it indicates that the folding propensity of a given residue is tuned by its neighbor residues, implying that predictions about folding should be made on the basis of residue triads, in agreement with the findings of previous studies. Similar nnr effects are also found for extended pPII-folding, although in this case, the entropic term does not participate in the cooperativity, which is solely ruled by the energetic contribution.

Our results offer a unique way to rationalize the contribution of solvent degrees of freedom to the entropic stabilization of α-helices. Globally, the entropic contribution of the solvent is estimated in ∼0.8 kcal/mol per residue at room temperature, which partially compensates for the decrease of the conformational entropic contribution upon folding. Our analysis reveals also that this compensation is more effective for the inner residues in the α-helix, where the formation of HBs between the peptide and the solvent is hindered. This means that the available sites for HB are being used in intramolecular interactions which facilitates the mobility of the water molecules. Interestingly, such a mechanism would be absent in the case of folding in extended secondary structures, with solvent contribution to entropy being expected to be negligible, which is consistent with our analysis of the entropic term for such conformations.

In summary, we think that this work provides new insights into the thermodynamic factors that trigger the peptide folding processes, providing evidence of the actual role played by the energetic and entropic terms, the latter mostly due to the solvent, in the stabilization of the helical structures of polyalanines. In perspective, the extension of this analysis to peptides with different compositions promises to clarify the propensity of residue triads to drive specific conformations and to be therefore of great help in the prediction of folding patterns from residues sequences.

AUTHOR INFORMATION

Corresponding Authors

Adolfo Bastida − *Departamento de Química Física, Universidad de Murcia, 30100 Murcia, Spain;* orcid.org/0000-0002-8193-9233; Email: bastida@um.es

Javier Cerezo − *Departamento de Química, Universidad Autónoma de Madrid, 28049 Madrid, Spain;* orcid.org/0000-0003-4820-4371; Email: javier.cerezo@uam.es

Authors

José Zúñiga − *Departamento de Química Física, Universidad de Murcia, 30100 Murcia, Spain*

Alberto Requena − *Departamento de Química Física, Universidad de Murcia, 30100 Murcia, Spain;* orcid.org/0000-0002-9408-9493

https://dx.doi.org/10.1021/acs.jcim.0c01177
J. Chem. Inf. Model. 2020, 60, 6523−6531

Beatriz Miguel − *Departamento de Ingeniería Química y Ambiental, Universidad Politécnica de Cartagena, 30203 Cartagena, Spain*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.0c01177

Author Contributions

A.B., J.Z., A.R., B.M., and J.C. contributed to the design and implementation of the research, to the analysis of the results, and to the writing of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was partially supported by the Spanish Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional (FEDER, UE) under Project CTQ2016-79345-P and by the Fundación Séneca under Project 20789/PI/18. We thank the computational assistance provided by J.F. Hidalgo of the Servicio de Infraestructuras TIC de ATICA.

■ REFERENCES

(1) Liu, L.; Yang, C.; Guo, Q.-X. A Study on the Enthalpy-Entropy Compensation in Protein Unfolding. *Biophys. Chem.* **2000**, *84*, 239−251.
(2) Rohl, C. A.; Baldwin, R. L. Comparison of NH Exchange and Circular Dichroism as Techniques for Measuring the Parameters of the Helix-Coil Transition in Peptides. *Biochem* **1997**, *36*, 8435−8442.
(3) Luo, P.; Baldwin, R. L. Interaction Between Water and Polar Groups of the Helix Backbone: An Important Determinant of Helix Propensities. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4930−4935.
(4) Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. Polyproline II Structure in a Sequence of Seven Alanine Residues. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9190−9195.
(5) Goch, G.; Maciejczyk, M.; Oleszczuk, M.; Stachowiak, D.; Malicka, J.; Bierzyński, A. Experimental Investigation of Initial Steps of Helix Propagation in Model Peptides. *Biochemistry* **2003**, *42*, 6840−6847.
(6) Lopez, M. M.; Chin, D.-H.; Baldwin, R. L.; Makhatadze, G. I. The Enthalpy of the Alanine Peptide Helix Measured by Isothermal Titration Calorimetry Using Metal-Binding to Induce Helix Formation. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1298−1302.
(7) Moreau, R. J.; Schubert, C. R.; Nasr, K. A.; Török, M.; Miller, J. S.; Kennedy, R. J.; Kemp, D. S. Context-Independent, Temperature-Dependent Helical Propensities for Amino Acid Residues. *J. Am. Chem. Soc.* **2009**, *131*, 13107−13116.
(8) Jas, G. S.; Kuczera, K. Equilibrium Structure and Folding of a Helix-Forming Peptide: Circular Dichroism Measurements and Replica-Exchange Molecular Dynamics Simulations. *Biophys. J.* **2004**, *87*, 3786−3798.
(9) Penkett, C. J.; Redfield, C.; Dodd, I.; Hubbard, J.; McBay, D. L.; Mossakowska, D. E.; Smith, R. A. G.; Dobson, C. M.; Smith, L. J. NMR Analysis of Main-Chain Conformational Preferences in An Unfolded Fibronectin-Binding Protein. *J. Mol. Biol.* **1997**, *274*, 152−159.
(10) Zaman, M. H.; Shen, M.-Y.; Berry, R. S.; Freed, K. F.; Sosnick, T. R. Investigations Into Sequence and Conformational Dependence of Backbone Entropy, Inter-Basin Dynamics and the Flory Isolated-Pair Hypothesis for Peptides. *J. Mol. Biol.* **2003**, *331*, 693−711.
(11) Chen, K.; Liu, Z.; Zhou, C.; Shi, Z.; Kallenbach, N. R. Neighbor Effect on PPII Conformation in Alanine Peptides. *J. Am. Chem. Soc.* **2005**, *127*, 10146−10147.
(12) Avbelj, F.; Baldwin, R. Limited Validity of Group Additivity for the Folding Energetics of the Peptide Group. *Proteins* **2006**, *63*, 283−289.
(13) Baruah, A.; Rani, P.; Biswas, P. Conformational Entropy of Intrinsically Disordered Proteins from Amino Acid Triads. *Sci. Rep.* **2015**, *5*, 11740.
(14) Toal, S. E.; Kubatova, N.; Richter, C.; Linhard, V.; Schwalbe, H.; Schweitzer-Stenner, R. Randomizing the Unfolded State of Peptides (and Proteins) by Nearest Neighbor Interactions between Unlike Residues. *Chem.—Eur. J.* **2015**, *21*, 5173−5192.
(15) Schweitzer-Stenner, R.; Toal, S. E. Construction and Comparison of the Statistical Coil States of Unfolded and Intrinsically Disordered Proteins From Nearest-Neighbor Corrected Conformational Propensities of Short Peptides. *Mol. BioSyst.* **2016**, *12*, 3294−3306.
(16) Bastida, A.; Zúñiga, J.; Requena, A.; Cerezo, J. Energetic Self-Folding Mechanism in alpha-Helices. *J. Phys. Chem. B* **2019**, *123*, 8186−8194.
(17) Bastida, A.; Carmona-García, J.; Zúñiga, J.; Requena, A.; Cerezo, J. Intraresidual Correlated Motions in Peptide Chains. *J. Chem. Inf. Model.* **2019**, *59*, 4524−4527.
(18) Dill, K. A.; MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **2012**, *338*, 1042−1046.
(19) Rao, V. V. H. G.; Gosavi, S. Using the Folding Landscapes of Proteins to Understand Protein Function. *Curr. Opin. Struct. Biol.* **2016**, *36*, 67−74.
(20) Meirovitch, H.; Cheluvaraja, S.; White, R. Methods for Calculating the Entropy and Free Energy and Their Application to Problems Involving Protein Flexibility and Ligand Binding. *Curr. Protein Pept. Sci.* **2009**, *10*, 229−243.
(21) Dragan, A. I.; Read, C. M.; Crane-Robinson, C. Enthalpy-Entropy Compensation: The Role of Solvation. *Eur. Biophys. J.* **2017**, *46*, 301−308.
(22) Lin, M. M.; Shorokhov, D.; Zewail, A. H. Dominance of Misfolded Intermediates in the Dynamics of Alpha-Helix Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 14424−14429.
(23) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(24) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(25) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D., Jr. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14*, 71−73.
(26) Zhang, S.; Schweitzer-Stenner, R.; Urbanc, B. Do Molecular Dynamics Force Fields Capture Conformational Dynamics of Alanine in Water? *J. Chem. Theory Comput.* **2020**, *16*, 510−527.
(27) Best, R. B.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr. Inclusion of Many-Body Effects in the Additive CHARMM Protein CMAP Potential Results in Enhanced Cooperativity of alpha-Helix and beta-Hairpin Formation. *Biophys. J.* **2012**, *103*, 1045−1051.
(28) Lifson, S.; Roig, A. On the Theory of Helix-Coil Transition in Polypeptides. *J. Chem. Phys.* **1961**, *34*, 1963−1974.
(29) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004−9015.
(30) Rohl, C. A.; Chakrabartty, A.; Baldwin, R. L. Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci.* **1996**, *5*, 2623−2637.
(31) Soler, M. A.; Zúñiga, J.; Requena, A.; Bastida, A. Understanding the Connection Between Conformational Changes of Peptides and Equilibrium Thermal Fluctuations. *Phys. Chem. Chem. Phys.* **2017**, *19*, 3459−3463.
(32) Richardson, J. M.; Lopez, M. M.; Makhatadze, G. I. Enthalpy of Helix-Coil Transition: Missing Link in Rationalizing the Thermody-

namics of Helix-Forming Propensities of the Amino Acid Residues. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1413−1418.
(33) Toal, S.; Meral, D.; Verbaro, D.; Urbanc, B.; Schweitzer-Stenner, R. pH-Independence of Trialanine and the Effects of Termini Blocking in Short Peptides: A Combined Vibrational, NMR, UVCD, and Molecular Dynamics Study. *J. Phys. Chem. B* **2013**, *117*, 3689−3706.
(34) Baxa, M. C.; Haddadian, E. J.; Jumper, J. M.; Freed, K. F.; Sosnick, T. R. Loss of Conformational Entropy in Protein Folding Calculated Using Realistic Ensembles and Its Implications for NMR-Based Calculations. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 15396−15401.
(35) Baldwin, R. L. In Search of the Energetic Role of Peptide Hydrogen Bonds. *J. Biol. Chem.* **2003**, *278*, 17581−17588.
(36) Baldwin, R. L. Energetics of Protein Folding. *J. Mol. Biol.* **2007**, *371*, 283−301.
(37) Bolen, D. W.; Rose, G. D. Structure and Energetics of the Hydrogen-Bonded Backbone in Protein Folding. *Annu. Rev. Biochem.* **2008**, *77*, 339−362.
(38) Gong, H.; Porter, L. L.; Rose, G. D. Counting Peptide-Water Hydrogen Bonds in Unfolded Proteins. *Protein Sci.* **2011**, *20*, 417−427.
(39) Makhatadze, G. I. In *Peptide Solvation and H-Bonds*; Baldwin, R. L., Baker, D., Eds.; Advances in Protein Chemistry; Academic Press, 2005; Vol. *72*, pp 199−226.
(40) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. Structure and Dynamics of Homologous Series of Alanine Peptides: A Joint Molecular Dynamics/NMR Study. *J. Am. Chem. Soc.* **2007**, *129*, 1179−1189.
(41) Verbaro, D.; Ghosh, I.; Nau, W. M.; Schweitzer-Stenner, R. Discrepancies between Conformational Distributions of a Polyalanine Peptide in Solution Obtained from Molecular Dynamics Force Fields and Amide I′ Band Profiles. *J. Phys. Chem. B* **2010**, *114*, 17201−17208.
(42) Baxa, M. C.; Haddadian, E. J.; Jha, A. K.; Freed, K. F.; Sosnick, T. R. Context and Force Field Dependence of the Loss of Protein Backbone Entropy upon Folding Using Realistic Denatured and Native State Ensembles. *J. Am. Chem. Soc.* **2012**, *134*, 15929−15936.
(43) Rossi, M.; Scheffler, M.; Blum, V. Impact of Vibrational Entropy on the Stability of Unsolvated Peptide Helices with Increasing Length. *J. Phys. Chem. B* **2013**, *117*, 5574−5584.
(44) Silva, R. A. G. D.; Kubelka, J.; Bour, P.; Decatur, S. M.; Keiderling, T. A. Site-Specific Conformational Determination in Thermal Unfolding Studies of Helical Peptides Using Vibrational Circular Dichroism with Isotopic Substitution. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8318−8323.
(45) Decatur, S. M. IR Spectroscopy of Isotope-Labeled Helical Peptides: Probing the Effect of N-Acetylation on Helix Stability. *Biopolymers* **2000**, *54*, 180−185.
(46) Ianoul, A.; Mikhonin, A.; Lednev, I. K.; Asher, S. A. UV Resonance Raman Study of the Spatial Dependence of Alpha-Helix Unfolding. *J. Phys. Chem. A* **2002**, *106*, 3621−3624.
(47) Shalongo, W.; Dugad, L.; Stellwagen, E. Distribution of Helicity within the Model Peptide Acetyl$(AAQAA)_3$Amide. *J. Am. Chem. Soc.* **1994**, *116*, 8288−8293.
(48) Hazel, A.; Chipot, C.; Gumbart, J. C. Thermodynamics of Deca-Alanine Folding in Water. *J. Chem. Theory Comput.* **2014**, *10*, 2836−2844.
(49) Pappu, R. V.; Srinivasan, R.; Rose, G. D. The Flory Isolated-Pair Hypothesis is Not Valid for Polypeptide Chains: Implications for Protein Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12565−12570.
(50) Gnanakaran, S.; Garcia, A. E. Validation of An All-Atom Protein Force Field: From Dipeptides to Larger Peptides. *J. Phys. Chem. B* **2003**, *107*, 12555−12557.
(51) Garcia, A. E. Characterization of Non-Alpha Helical Conformations in Ala Peptides. *Polymer* **2004**, *45*, 669−676.

https://dx.doi.org/10.1021/acs.jcim.0c01177
J. Chem. Inf. Model. 2020, 60, 6523−6531