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

Adolfo Bastida,[*,†] José Zúñiga,[†] Alberto Requena,[†] Beatriz Miguel,[‡] and Javier Cerezo[*,¶]

† *Departamento de Química Física. Universidad de Murcia, 30100 Murcia, Spain*

‡*Departamento de Ingeniería Química y Ambiental, Universidad Politécnica de Cartagena, 30203 Cartagena, Spain*

¶ *Departamento de Química. Universidad Autónoma de Madrid. 28049 Madrid, Spain*

E-mail: bastida@um.es; javier.cerezo@uam.es

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 the free energy upon reaching the folded structure. A complete understanding of this process requires, therefore, a deep insight into both contributions to the 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 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 mostly by an increase of the entropic contribution of the solvent, which compensates the decrease of conformational entropy of the polypeptide, thus unveiling a

1

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 entropic stabilization is basically governed by the nearest neighbor residues 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 the 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 form the sizes of the 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] (nnr), 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 the 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 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 presente work, we extend our study to the evaluation of the folding thermodynamic magnitudes of peptides after equilibrium is reached, revealing the fundamental role that entropy plays in the stabilization of the α-helices. We have chosen polyalanine peptides as test modles 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 =$1-6, 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 modestly with the total number of alanine residues, while the entropic contribution significantly 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 nearest neighbor residues.

2   Methods

2.1   MD Simulations.

MD simulations of the ACE-(Ala)$_m$-NME molecules with $m =$1-6, 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 the polyalanine molecules were described using the CHARMM27[25,26] force field and the flexible TIP3P model was used for the solvent water molecules. 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 5 ns followed. The last step was repeated 448 times with different sets of velocities. Each of these 448 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 free energy through $\Delta F_f = -k_B T \ln w$. By performing MD simulations at different temperatures and fitting the results to the 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, computational 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 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 folding process of

5

polypeptides, the folding free energy is given by

$$\Delta F_{\mathrm{f}} = -k_{\mathrm{B}}T \ln \frac{p_{\Omega_{\mathrm{f}}}}{1 - p_{\Omega_{\mathrm{f}}}} \tag{3}$$

where $p_{\Omega_{\mathrm{f}}}$ is the total population of 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 in 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 when three consecutive residues $(i + 1, i + 3)$ are in the helical conformation, an intramolecular hydrogen bond (HB) is formed between the peptide CO of residue i and the peptide NH of residue $i + 4$. Accordingly, the average number of HBs is expected to be equal to the number of helical segments formed by three consecutive residues in the helical conformation, and the optimal δ interval is that providing a definition of the α_{R} region which yields the best estimate of the number of HBs. In Figure 2 we show the related results obtained from our simulations. As observed, the number of HBs is low for the shorter, $m \leq 5$, polyalanine molecules and sharply increases for the $m = 6$ molecule. For larger molecules, the number of HBs shows a notable proportionality with the length of the molecule, which makes it clear that the intramolecular HB net is well established for these molecules. We also observe in Figure 2 that the average number of three consecutive helical residues is highly sensitive to the choice of the δ interval, thus allowing its optimization. We note that the resulting

optimal value, $\delta = 12°$, is substantially lower that the usual intervals employed to define the helical region.

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 the 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 vanishes at thermal equilibrium, in which the average kinetic energy per degree of freedom must be equal to $\frac{1}{2} k_{\mathrm{B}} T$. In Eq. (4), the averaged kinetic (ΔT) and potential (ΔV) energy differences run over all the $i \to j$ 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-$\mathrm{C}_{\alpha,i}$-C_i-N_{i+1}) plus the $\mathrm{H}_{\alpha,i}$ and $\mathrm{H}_{\mathrm{N},i}$ 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
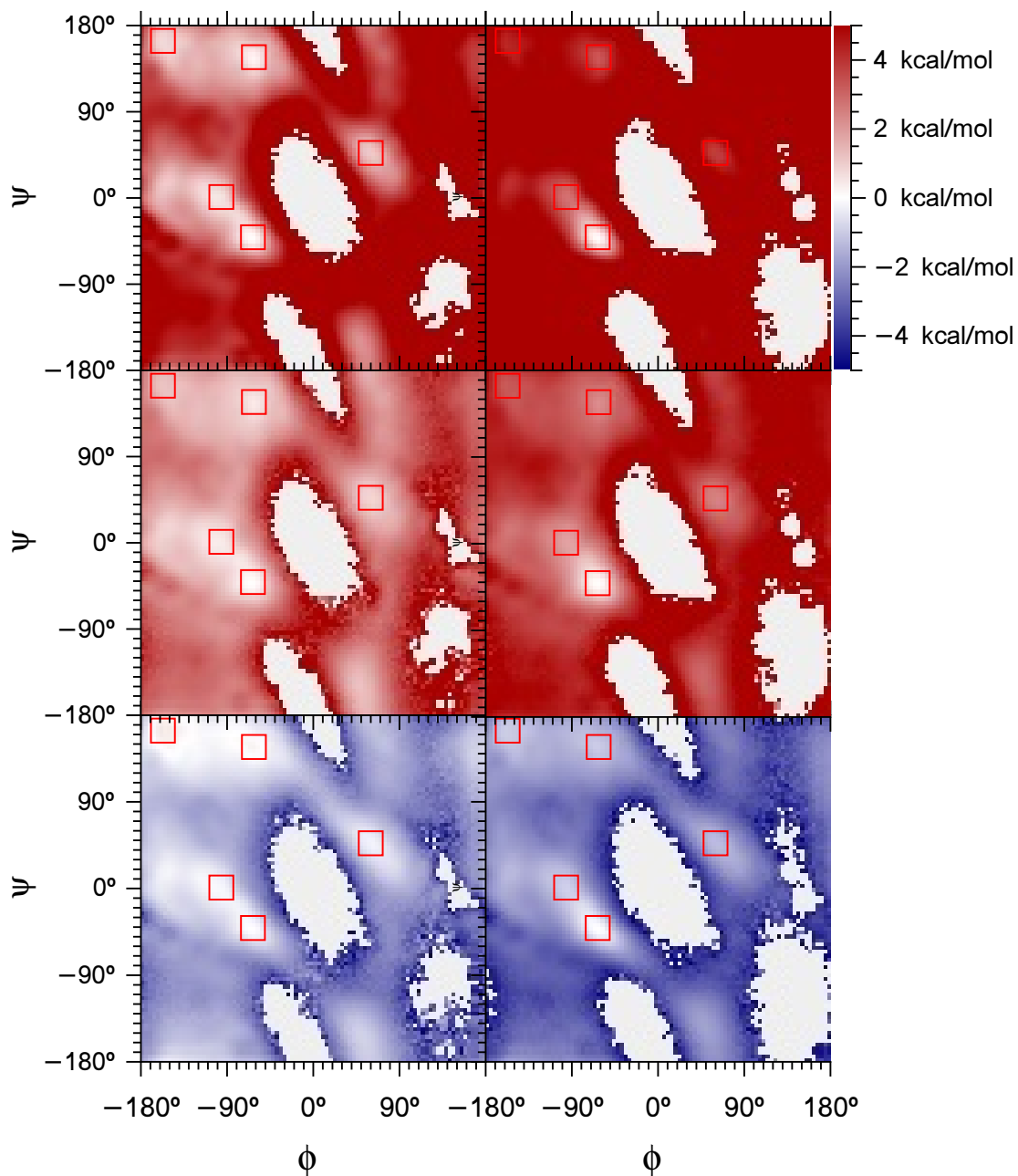
Figure 1: Free energy (top), energy (medium) and entropy ($T\Delta S$) (bottom) maps for the $m = 1$ (left) and $m = 12$ (right) polyalanine molecules. The five regions corresponding to the α_{R}, $\alpha\prime$, α_{L}, β, and pPII conformations are plotted with $\pm 12°$ intervals around each maxima. All maps were calculated using a 90x90 grid. Zero point energies were arbitrarily set at the maximum of the α_{R} conformation.
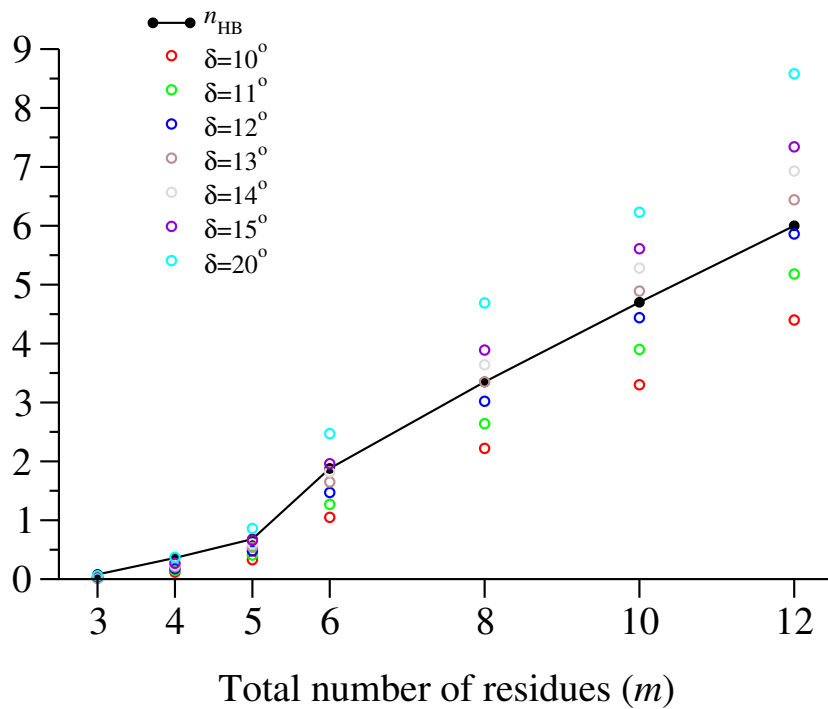
Figure 2: Average number of hydrogen bonds (black line) obtained using the definition of Zewail *et al*,[22] and the average number of helical segments consisting of three consecutive residues in the α_R conformation defined as centered at $(-63°, -41°)$ with $\pm\delta$ intervals for the polyalanine molecules.

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 (on) 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:

$$\Delta E_{\mathrm{f}} = \frac{\sum\limits_{i \in \Omega_{\mathrm{f}}} p_i \, V_i}{\sum\limits_{i \in \Omega_{\mathrm{f}}} p_i} - \frac{\sum\limits_{i \notin \Omega_{\mathrm{f}}} p_i \, V_i}{\sum\limits_{i \notin \Omega_{\mathrm{f}}} p_i} \tag{5}$$

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

3   Results and Discussion

3.1   Folding magnitudes 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 tool 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 later over the twelve residues. As observed, the energy barriers between the different conformers are typically

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 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. Interestingly, this does not necessarily modify the average folding energy at equilibrium, since it depends not only on the relative energy of the folded and unfolded regions, but also on how these regions are populated, with the conformational populations being much more concentrated around the conformational maxima for the longer ($m = 12$) polyalanine molecule than for the shortest ($m = 1$) molecule.

The availability of free energy and potential energy maps with a similar, and high, accuracy make 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 there are small entropic changes between the more stable conformations, while for the longer $m = 12$ molecule, the entropy is maximum in the folded conformation. The larger value obtained at the folded α_R conformation of the longer $m = 12$ polyalanine 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. For the shortest $m = 1$ polyalanine molecule, the absence of intramolecular HBs makes the intermolecular interactions with the water molecules more similar for the different conformations of the solute. 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 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 Equations (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 of 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] 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), $T\Delta S_f = -0.80$ kcal/mol and $\Delta G_f = -0.20$ kcal/mol per alanine residue. 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] Secondly, 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 the 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 12.9% and 80.1% respectively. We see also that the ΔF_f values calculated for the largest polyalanine molecules underestimate the experimental values, that is, the α_R population is overestimated, in line with previous studies[27] reporting that the CHARMM27 force field tends to overestimate the α_R conformation.

As for the evolution of the folding energetic and entropic contributions with the chain length, we observe in Figure 3 quite different variations. The folding energies values are more similar for all polyalanine molecules studied, with a decrease of -0.4 kcal/mol·residue from the shorter to the longer chains, and a mean value of -1.05 ± 0.1 kcal/mol per residue, in agreement with the experimental values, which also show that the folding energy is practically independent of the chain length in peptides enriched in alanine residues.[6] Interestingly, the small fluctuations of the average folding energies arise from the balance between the different depth of the α_R potential energy wells and their relative populations, as anticipated from the analysis of the maps in the previous section. These results indicate that the formation of the intramolecular $(i, i + 4)$ HBs in the larger polyalanine molecules hardly alters the energetic

balance between the folded and unfolded conformations. Consequently, the decrease in the folding free energy with the chain length comes mainly from the decrease of the entropic term.

Actually, the behavior of the entropic term shown in Figure 3 is rather different than that for the folding energy, as it significantly changes when going from the shortest to the longest polyalanine chains. 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 $et\ al$,[40] who obtained an average value of -1.25 kcal/mol for the conformational entropic contribution per alanine residue based on MD simulations carried out with the same force field used in this work. 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 are close to the conformational entropy calculated by Sosnick $et\ al$[40] with the solvent playing just a minor role. In contrast, comparison of the protein backbone entropy with the total folding entropy (~ -0.2 kcal/mol) for the larger chains reveals an entropic contribution from the water solvent of $\sim$1 kcal/mol, 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.[41]
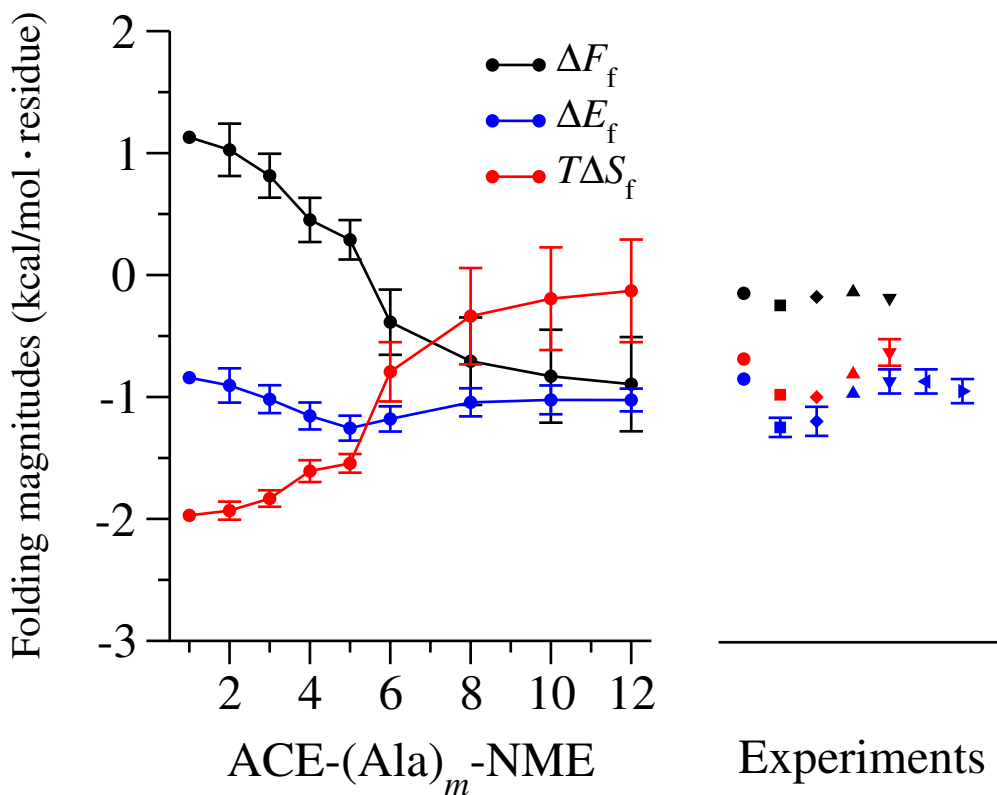
Figure 3: Folding free energy (ΔF_f), energy (ΔE_f) and entropic contribution at $25°$ ($T\Delta S_\mathrm{f}$) for the polyalanine molecules obtained from MD simulations. Error bars correspond to the standard deviations of the magnitudes calculated for every residue in the molecule. Experimental results from references 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).

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 variations of the entropic term between different polypeptides, and also within each chain, are more significant than the enthalpic counterpart, with the overall profile of free energy changes being mainly dictated by the entropic contribution. We observe in Figure 4 that the evolution of the folding magnitudes of the molecules shows a common pattern. There are subtantial variations of these for the shorter ($m = 1-6$) peptides, while the differences between them for the longer ($m = 8-12$) peptides come from the increasing number of inner residues, with similar results for the extreme residues. This is a direct consequence of the different helicoidal structure of the molecules previously noted in the discussion of the number of HBs. For the shorter peptides, the population of the α_R region of all the residues increases as the chain is lengthened, while for the largest peptides the only effect is the addition of more central residues, with no change in the population of the remaining residues.

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,[42–44] and that the folding in the N-terminus residues is higher than the folding in the C-terminus residues, again in agreement with previous observations.[45] The fact that the residues in larger peptides have different folding magnitudes, depending on their relative positions in the chain, explains why the global thermodynamic magnitudes shown 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 folding magnitudes get closer to the values obtained for the inner residues.

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
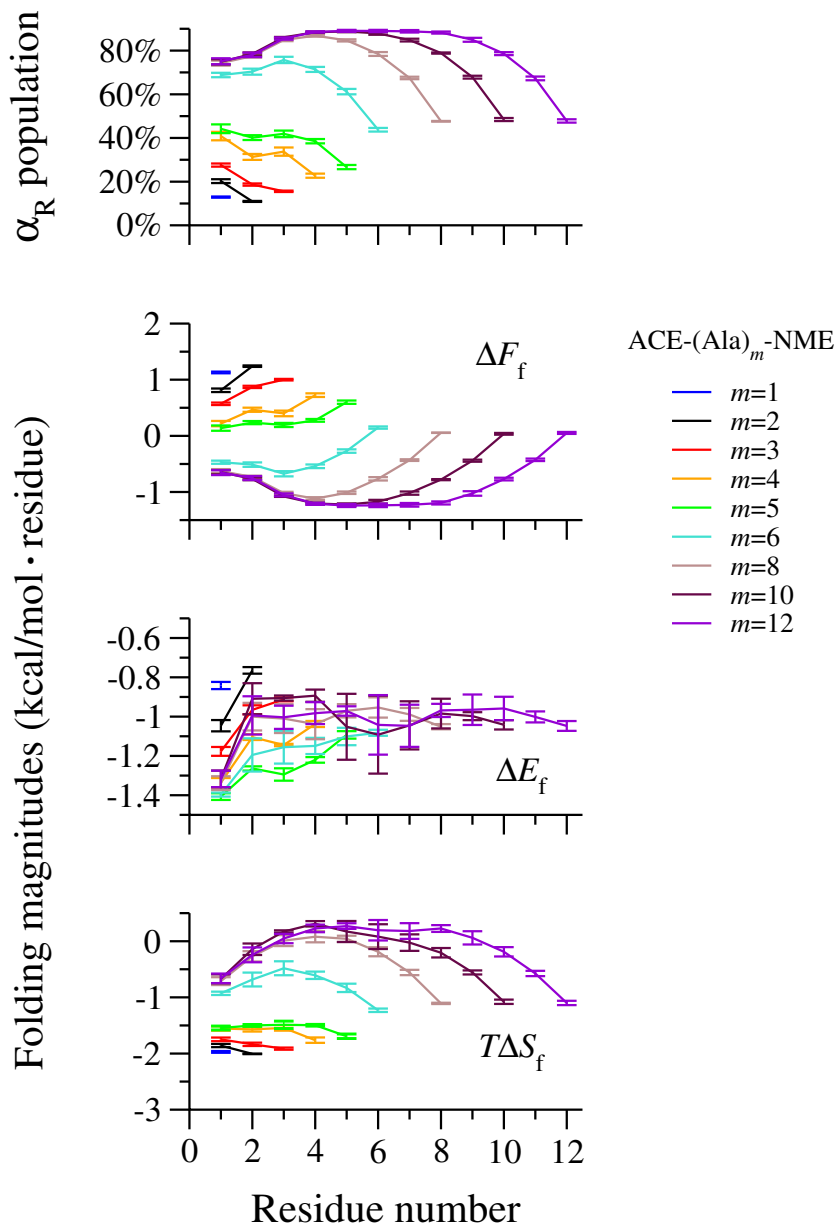
Figure 4: α_{R} populations, folding free energy (ΔF_{f}), energy (ΔE_{f}) and entropic contribution at 25° ($T\Delta S_{\mathrm{f}}$) for every residue in the polyalanine molecules obtained from MD simulations. We note that the energy scale for the folding energies is three and a half times smaller than that for the free energies and entropies. Error bars correspond to the standard deviations of the results obtained by block analysis by performing the statistics in four independent groups of trajectories.

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 are increasing evidences[9–17] supporting the idea that the nearest neighbor residues conformation affects the dynamics and energetics of a given residue, and that the secondary structure of polypeptides cannot be interpreted using the Flory's isolated pair hypothesis, but in terms of the preferential conformations of amino acids triads. In order to test this possibility, we show in Figure 5 the average folding percentage of the inner residues of the polyalanine molecules as a function of the folding state of their nearest neighbor residues. As seen, the general trend is that a given residue has a higher percentage of folding as more nearest neighbor residues are already folded. Therefore, the folding of the nearest neighbor residues of a given residue increases its own folding propensity. This behavior is observed for all polyalanine peptides analyzed, regardless of their length. i.e., also including those that are not able to form strong intramolecular H bonds, which indicates that it is not an effect of the formation of H bonds but, rather is intrinsic to the polyalanine chains. Moreover, this effect is slightly more pronounced when the nearest neighbor residue is placed on the rigth side with respec to the left side. The only exception is obtained for the shortest

($m=3$) polyalanine molecule where the folding percentages when one or two nearest neighbor residues are folded are similar. Overall these results prove that the folding energetics of a residue is greatly influenced by its nearest neighbor residues.

Therefore we applied the K2V method to calculate mean potential energy maps for every residue conditioned to the folding state of its nearest neighbor residues what provides the corresponding thermodynamic folding magnitudes also shown in Figure 5. Concerning the folding energies we observe that folding of an alanine residue is less exoenergetic, that is, it is energetically hindered if any or both nearest neighbor residues are already in the α_R conformation. This effect was already noted in our previous study about the initial folding process of the decalanine molecule[16] but the present results confirm that it is an intrinsic property of the polyalanine chains regardless their length. This proximity shifts have been previously attributed to sterical[46] and peptide electrostatic effects.[12] We also see that this energetic effect is slightly more pronounced (≤ 0.15 kcal/mol) for the nearest neighbor residues placed on the left side in the larger ($m \geq 6$) molecules.

Concerning the entropic contribution we observe that this term always increases when the nearest neighbor residues are already in the α_R conformation. Consequently the folding of an alanine residue is entropically facilitated if any or both nearest neighbor residues are already folded. While the energetic and entropic contributions follow opposite trends we note that the magnitude of the entropical stabilization is substantially larger that the energetic hindrances. As a consequence the trends observed for the folding free energies and α_R populations are entropically driven. It is also interesting that the entropic stabilization is similar for the folding of the right and left nearest neighbor residues so that the folding asymmetry of the two extremes observed in Figure 4 comes mainly from the energetic term. Moreover, compared to the energetic term, left and right entropic contributions are much closer to be additive when both nearest neighbors are folded.

On the whole, the stabilization of the α-helix mainly arises from an entropic effect induced by the nearest neighbor residues already folded. Although the mechanism of action, at
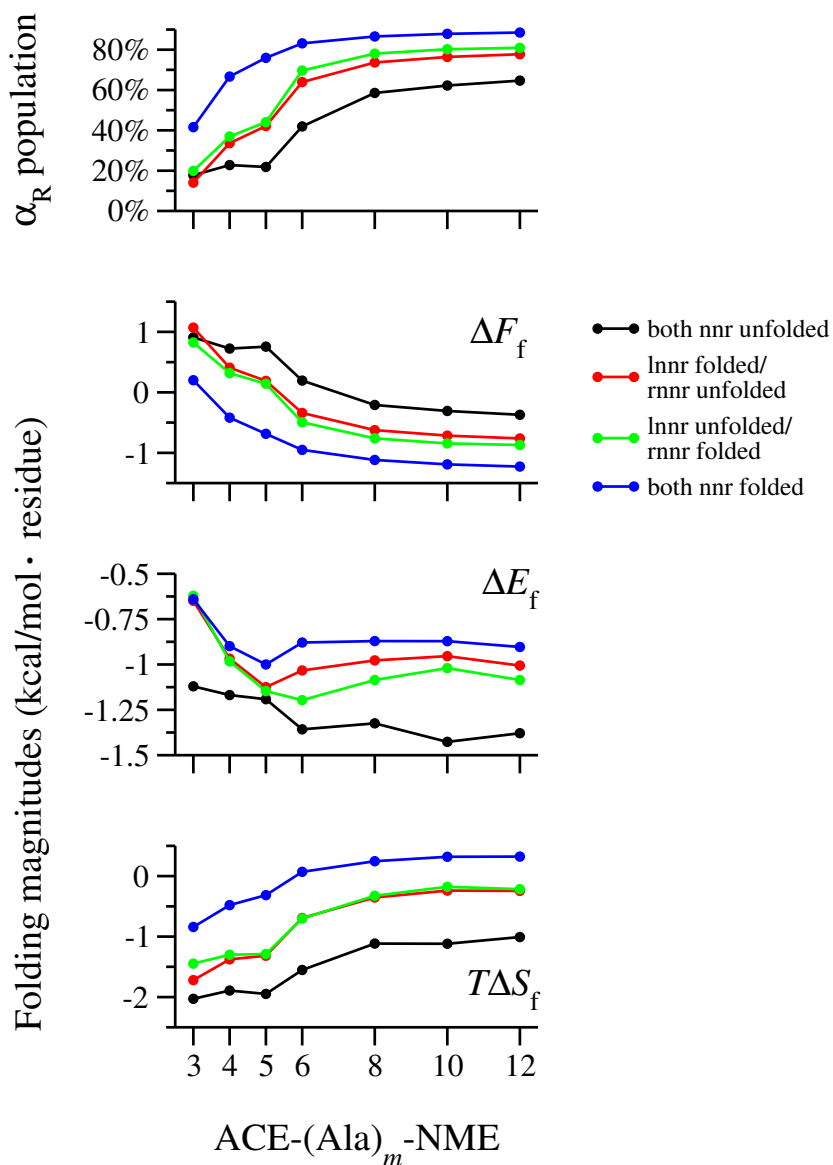
Figure 5: Average α_R populations, folding free energy (ΔF_f), energy (ΔE_f) and entropic contribution at 25° ($T\Delta S_f$) of the inner residues of the polyalanine molecules as a function of the folding state of the nearest neighbor residues: both are unfolded (black), only the left one is folded (red), only the right one is folded (green), and both are folded (blue). We note that the energy scale for the folding energies is three times smaller than that for the free energies and entropies.

molecular level, behind this effect is not evident, the sterical effects[46] 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.

4   Conclusions and perspectives

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

Average thermodynamic magnitudes of folding, including 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 avoids arbitrary choices of parameters by defining rigorously the folded and unfolded states by selecting regions that ensure consistency between the average number of HBs and the helical fragments along the simulation.

Our predictions for the free 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 entropy plays in facilitating the process. In particular, a drastic increase of the entropic term is observed when going from the smaler (1 to 5 residues) to the longer peptides, while the enthalpy contributions display much smaller fluctuations. The entropic component per residue along the chain for the longest peptides also reveals that folding is entropically favored for inner residues, while the enthalpic contribution changes more smoothly among residues. The analysis of the thermodynamic folding magnitudes in terms of the folding state

of the nearest neighbor residues reveals that the folding of a residue is energetically hindered and entropically favored when its nearest neighbor residues are already folded and that the entropic stabilization is comparatively larger. These effects are more pronounced for the nearest neighbor residue placed on the right side what explains the folding asymmetry of the two extremes experimentally observed. 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 residues triads.

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 $\sim 1\,\text{kcal/mol}$ per residue at room temperature, which partially compensates 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 what facilitates the mobility of the water molecules.

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 entropic term, 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 composition 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.

Acknowledgements

References

(1) Liu, L.; Yang, C.; Guo, Q. A Study on the Enthalpy-Entropy Compensation in Protein Unfolding. *Biophys. Chem.* **2000**, *84*, 239–251.

(2) Rohl, C.; Baldwin, R. 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. Interaction Between Water and Polar Groups of the Helix Backbone: An Important Determinant of Helix Propensities. *PNAS* **1999**, *96*, 4930–4935.

(4) Shi, Z.; Olson, C.; Rose, G.; Baldwin, R.; Kallenbach, N. Polyproline II Structure in a Sequence of Seven Alanine Residues. *PNAS* **2002**, *99*, 9190–9195.

(5) Goch, G.; Maciejczyk, M.; Oleszczuk, M.; Stachowiak, D.; Malicka, J.; Bierzynski, A. Experimental Investigation of Initial Steps of Helix Propagation in Model Peptides. *Biochemistry* **2003**, *42*, 6840–6847.

(6) Lopez, M.; Chin, D.; Baldwin, R.; Makhatadze, G. The Enthalpy of the Alanine Peptide Helix Measured by Isothermal Titration Calorimetry Using Metal-Binding to Induce Helix Formation. *PNAS* **2002**, *99*, 1298–1302.

(7) Moreau, R. J.; Schubert, C. R.; Nasr, K. A.; Toeroek, M.; Miller, J. S.; Kennedy, R. J.; Kemp, D. S. Context-Independent, Temperature-Dependent Helical Propensities for Amino Acid Residues. *JACS* **2009**, *131*, 13107–13116.

(8) Jas, G.; 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.; Redfield, C.; Dodd, I.; Hubbard, J.; McBay, D.; Mossakowska, D.; Smith, R.; Dobson, C.; Smith, L. NMR Analysis of Main-Chain Conformational Preferences in An Unfolded Fibronectin-Binding Protein. *J. Mol. Bio.* **1997**, *274*, 152–159.

(10) Zaman, M.; Shen, M.; Berry, R.; Freed, K.; Sosnick, T. 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. Neighbor Effect on PPII Conformation in Alanine Peptides. *JACS* **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. *Scientific Reports* **2015**, *5*.

(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. *Chemistry AEJ* **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. Biosystems* **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. P. Methods for Calculating the Entropy and Free Energy and Their Application to Problems Involving Protein Flexibility and Ligand Binding. *Curr. Prot. & 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. *PNAS* **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. Theor. Comp.* **2008**, *4*, 435–447.

(24) Pronk, S.; Pll, 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) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(26) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Re-

producing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comp. Chem.* **2004**, *25*, 1400–1415.

(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. Theory of Helix-Coil Transition in Poplypeptides. *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.; Chakrabartty, A.; Baldwin, R. Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Prot. Science* **1996**, *5*, 2623–2637.

(31) Soler, M.; 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.; Lopez, M.; Makhatadze, G. Enthalpy of Helix-Coil Transition: Missing Link in Rationalizing the Thermodynamics of Helix-Forming Propensities of the Amino Acid Residues. *PNAS* **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. *PNAS* **2014**, *111*, 15396–15401.

(35) Baldwin, R. 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. *Ann. Rev. Biochem.* **2008**, *77*, 339–362.

(38) Gong, H.; Porter, L. L.; Rose, G. D. Counting Peptide-Water Hydrogen Bonds in Unfolded Proteins. *Prot. Sci.* **2011**, *20*, 417–427.

(39) Makhatadze, G. I. In *Peptide Solvation and H-Bonds*; Baldwin, RL and Baker, D,, Ed.; Advances in Protein Chemistry; 2005; Vol. 72; pp 199–226.

(40) 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. *JACS* **2012**, *134*, 15929–15936.

(41) 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.

(42) Silva, R.; Kubelka, J.; Bour, P.; Decatur, S.; Keiderling, T. Site-Specific Conformational Determination in Thermal Unfolding Studies of Helical Peptides Using Vibrational Circular Dichroism with Isotopic Substitution. *PNAS* **2000**, *97*, 8318–8323.

(43) Decatur, S. IR Spectroscopy of Isotope-Labeled Helical Peptides: Probing the Effect of N-Acetylation on Helix Stability. *Biopolymers* **2000**, *54*, 180–185.
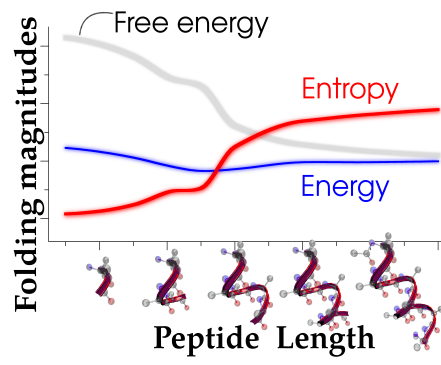
(44) Ianoul, A.; Mikhonin, A.; Lednev, I.; Asher, S. UV Resonance Raman Study of the Spatial Dependence of Alpha-Helix Unfolding. *J. Phys. Chem. A* **2002**, *106*, 3621–3624.

(45) Hazel, A.; Chipot, C.; Gumbart, J. C. Thermodynamics of Deca-Alanine Folding in Water. *J. Chem. Theor. Comp.* **2014**, *10*, 2836–2844.

(46) Pappu, R.; Srinivasan, R.; Rose, G. The Flory Isolated-Pair Hypothesis is Not Valid for Polypeptide Chains: Implications for Protein Folding. *PNAS* **2000**, *97*, 12565–12570.

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.

Competing financial interests

The authors declare no competing financial interests.

TOC Graphic