Side-chain–side-chain interactions and stability of the helical state

Ronen Zangi*

Department of Organic Chemistry I, University of the Basque Country UPV/EHU, Avenida de Tolosa 72, 20018, San Sebastian, Spain and IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain

(Received 22 October 2013; revised manuscript received 12 December 2013; published 30 January 2014)

Understanding the driving forces that lead to the stability of the secondary motifs found in proteins, namely α-helix and β-sheet, is a major goal in structural biology. The thermodynamic stability of these repetitive units is a result of a delicate balance between many factors, which in addition to the peptide chain involves also the solvent. Despite the fact that the backbones of all amino acids are the same (except of that of proline), there are large differences in the propensity of the different amino acids to promote the helical structure. In this paper, we investigate by explicit-solvent molecular dynamics simulations the role of the side chains (modeled as coarse-grained single sites) in stabilizing α helices in an aqueous solution. Our model systems include four (six-mer–nine-mer) peptide lengths in which the magnitude of the effective attraction between the side chains is systematically increased. We find that these interactions between the side chains can induce (for the nine-mer almost completely) a transition from a coil to a helical state. This transition is found to be characterized by three states in which the intermediate state is a partially folded α-helical conformation. In the absence of any interactions between the side chains the free energy change for helix formation has a small positive value indicating that favorable contributions from the side chains are necessary to stabilize the helical conformation. Thus, the helix-coil transition is controlled by the effective potentials between the side-chain residues and the magnitude of the required attraction per residue, which is on the order of the thermal energy, reduces with the length of the peptide. Surprisingly, the plots of the population of the helical state (or the change in the free energy for helix formation) as a function of the total effective interactions between the side chains in the helical state for all peptide lengths fall on the same curve.

DOI: 10.1103/PhysRevE.89.012723 PACS number(s): 87.14.ef, 87.15.ap, 87.15.bd, 87.15.hp

I. INTRODUCTION

The tertiary structure of proteins is composed of segments of secondary structures, such as α helices and β sheets. These secondary motifs are believed to form early, and independently, in the protein folding process and then assemble to form the tertiary structure [1,2]. If this is true, then understanding the folding process of proteins necessitates the comprehension of the driving forces responsible for the formation of these secondary structures. However, despite their relative structural simplicity and decades of experimental and theoretical investigations, the dynamics and thermodynamics of forming α helices and β sheets are still not completely understood.

A segment in a peptide or protein chain adopts an α-helical conformation when the carbonyl group of residue i forms a hydrogen bond with the amide hydrogen of the residue $i+4$, yielding a right-handed helical turn of 3.6 residues. The formation of this hydrogen bond restricts the ϕ and ψ backbone dihedral angles of the three intervening amino acids. Thus, beside the formation of one intrapeptide hydrogen bond, the process of forming the first helical turn (nucleation) is associated with an entropic penalty of fixing six dihedral angles. The process of extending this helical segment is thermodynamically less costly because it involves the formation of an additional intrabackbone (amide-carbonyl) hydrogen bond and the fixation of only two dihedral angles. If helix formation is associated with an entropic penalty, there must be a favorable gain in another component of the free energy. Already at the time the α-helix structure was proposed it was assumed that the formation of the backbone interamide hydrogen bond in aqueous solution has a net enthalpic gain rendering the formation of this secondary structure possible [3]. This was adopted by several theoreticians who proposed analytical models, e.g., Zimm-Bragg [4] and Lifson-Roig [5], for the helix formation. Nevertheless, in the formulation of these theoretical models the identity of the favorable enthalpic term does not need to be restricted to the energetic gain in forming the backbone interamide hydrogen bonds. Alternatively, it can be identified with any favorable interaction of the system (including the surrounding solvent) associated with adopting the helical conformation between the i and $i+4$ residues.

That intraprotein, or intrapeptide, hydrogen bonds are not the driving force for the native folded, or secondary, structure is the idea behind the influential paper by Kauzmann in 1959 [6]. He argued that the enthalpic gain obtained from the formation of these intraprotein hydrogen bonds in the native structure, including the formation of the associated water-water hydrogen bonds, cancels out by the enthalpy of breaking the hydrogen bonds of the carbonyl and amide groups with the water molecules. Instead, it is proposed that because proteins or peptides are, to some extent, hydrophobic in character they tend to minimize their solvent (water) exposed surface area. Thus, folding is driven by hydrophobic forces. Nevertheless, intraprotein hydrogen bonds, although not contributing to the net stability of the native structure, are formed to avoid large loss of enthalpy. It is these intraprotein hydrogen bonds that shape the structure of the secondary motifs and, consequently, the native fold. However, attempts to confirm or refute this picture have not been successful yet. This is because when

*r.zangi@ikerbasque.org

considering the stability of the native structure, it is not so simple to isolate the contribution of the hydrophobic effect from that of the backbone interamide hydrogen bonds and as a result this issue is far from being settled [7–14].

Assuming no significant favorable change in the entropy of the surrounding water molecules, the dominant factor responsible for forming the structure of an α helix must be enthalpic. Indeed, in this case, the thermodynamic reports in the literature agree with each others and the enthalpy change associated with the coil to helix transition is reported to be in the range of -3.3 to -6.3 kJ/mol per residue [15–21]. Another characteristic of the helix-coil transition is that helix formation is easier for larger peptides. This can be explained by the decrease in the relative contribution of the nucleation penalty with the increase in length. However, this does not proceed indefinitely, and the fractional helicity of helical peptides in solution, as a function of the peptide length, starts to plateau after approximately 30 residues [22]. In protein tertiary structures, the distribution of the lengths of α helices is peaked around 14 and the probability to observe longer helices decreases sharply [23]. Note that although the free energy per residue of forming an α helix depends on the peptide length, the associated change in enthalpy does not [20,21]. In addition, it is found that the stability of the helix is larger in the middle of the peptide chain [24]. It is easier to form a helical turn from the N-terminal side and unfolding normally starts at the C terminal [25–28].

In most of the theoretical models, the role of the side-chain residues to the stability of the helix is ignored. It is clear, however, that there are large differences in the propensity of different amino acids to form an α-helical structure, simply because their relative frequencies of occurrence in helices in proteins vary significantly [29]. The same conclusion is also observed in systematic mutational studies of replacing an amino acid, which participates in an α-helical structure within a folded protein, with all other amino acid residues [30,31], as well as, by measuring the helical content of synthetic model peptides in solution [32–35]. These analyses and experiments gave rise to various helix propensity scales that are quite similar to each other [36,37]. For example, alanine is generally ranked as the highest helix promoter amino acid residue whereas glycine as the most disruptive one (except of proline). In fact, short polyalanine peptides can form an α-helical structure in aqueous solutions, however, in order to solubilize these hydrophobic peptides insertion of charged amino acids is necessary [38–41]. Alternatively, promotion of the α-helical conformation of short peptides (which are not polyalanine) can also be realized by insertion of charged residues that form i to $i+4$ salt bridge [42–45].

Initially, the variations to promote differently the α-helical structure was attributed to the restriction of the configurational entropy of the amino acid side chains [34,35]. It is argued that, except for alanine, the rotational degrees of freedom of the side chain in the α-helix conformation, relative to the unfolded state, are suppressed rendering the effect of these side chains helix destabilizers. For different amino acids, this loss in side-chain entropy is different [46]. However, a poor correlation between the reduction in the side-chain entropy and helix propensity is found [31] putting into question the validity of this argument [47].

Another possibility is that helix propensities are driven by enthalpy. This was first demonstrated by Luo and Baldwin who used thermal unfolding curves of five nonpolar amino acids in water/trifluoroethanol mixtures [48]. The physical mechanism inducing the different propensities was argued to be the extent by which the side-chain residues shield (or desolvate) the intrapeptide backbone hydrogen bonds from the solvent. In this picture, shielding of the backbone hydrogen bonds opposes folding. Thus, bulky nonpolar side chains reduce the interaction energy between the solvent waters and the dipoles of the hydrogen bonds, whereas alanine does not. A more extensive study of the importance of the enthalpy of the side chains was performed by Makhatadze and coworkers [47,49]. They used calorimetric measurements of folding a model host peptide in which the helix formation is induced by metal binding [50,51]. The guest amino acid residues were of different size, shape, hydrophobicity, and hydrogen bonding potential. Different enthalpies of the helix-coil transition were found for the different amino acid residues. Furthermore, it was demonstrated that the correlation to the helical propensities increases significantly by combining the enthalpy change with the configurational entropy change of the side-chain residues. Other studies aiming to predict α-helix formation utilized experimental data and parameterized empirically a set of energy contributions, including side-chain–side-chain interactions [52,53]. The resulting algorithms are able to calculate correctly the average helical behavior of many peptides.

In this paper we study the transition from a coil state to an α-helical state by a series of molecular dynamics simulations in which the attractions between the side-chain residues are systematically increased. We find that these side-chain–side-chain interactions are able to control almost completely the helix-coil transition. This demonstrates that the changes associated only with the backbone of the peptide are insufficient and contribute a small positive value to the free energy difference of forming the helical conformation. Our model systems include different peptide lengths ranging from six-mer to nine-mer chains. Interestingly, we find that a plot of the fraction of time the helical state is observed, or the free energy change for helix formation, as a function of the total effective energy between the side chains in the helical state is the same for all peptide lengths.

II. METHODS

We modeled six-, seven-, eight-, and nine-residue homogeneous (uncharged) peptides deprotonated at the N terminal and protonated at the C terminal. For simplicity we represent the side chain of each residue by a single coarse-grained site, thus, from a structural point of view the resulting peptides can be viewed as perturbed polyalanine chains. The interactions between these side-chain sites were systematically varied. In particular, the side-chain–side-chain Lennard-Jones (LJ) σ parameter was equal to 0.47 nm (in this case the distance at which the potential exhibits its minimum $r_m = 2^{1/6}\sigma$ is equal to the distance between the side chains of adjacent neighbors in an ideal α-helical conformation) and the value of ϵ increased from 0.0 to 4.0 kJ/mol in steps of 1.0 kJ/mol, yielding five simulations for each of the four peptide lengths. All other

atoms of the peptides were represented by the OPLSAA force field [54–57]. Nevertheless, the interaction between a side-chain site and any other atom in the system was the same for all values of ϵ. In particular, it was calculated by the OPLSAA combination rule (geometric averages) in which the former is represented by a methyl group. This was done so that in calculating the energy difference of the helical state between the different ϵ's, only the direct side-chain–side-chain interactions are needed because the other interaction energies (in particular those involving the solvent molecules which give rise to errors with large magnitudes) are the same. The peptides were solvated by 1535 water molecules described by the TIP4P model [58]. Water bond distances and angles were constrained using the SETTLE algorithm [59] whereas the peptides bond distances were constrained using the LINCS algorithm [60].

The starting conformations for the simulations were relaxed structures of fully extended conformations constructed using the program WHATIF [61]. The fully extended conformations were relaxed for 100 ps molecular dynamics simulations in which the side-chain–side-chain LJ ϵ parameter was equal to 0.0 kJ/mol. This relaxation step was performed in order to minimize the required cubic box sizes subject to the condition that the peptide did not interact with its periodic image. This procedure yielded extended random coils conformations for the four different peptide lengths (see Fig. S1 in the Supplemental Material [62]). Within each peptide length, we used the same starting configurations for the simulations at different ϵ.

Each of the systems was propagated for 400 ns (thus, a combined trajectory of 8 μs for the 20 simulations). The peptides' coordinates were saved every 10 ps and were used in all analyses. Note that a recent computational study of hepta-alanine modeled by the OPLSAA force field and compared with NMR-derived J-coupling constants reported convergence of the value of χ^2 within the first 250 ns of the trajectory [63]. The root mean square deviation (RMSD) were computed by fitting the peptide structure, by the least-squares procedure of all the peptide atoms excluding hydrogens, to a perfect α-helix conformation (built with the WHATIF program). Then, the deviations of these heavy atoms from the ideal α-helix conformation were calculated. In analyzing the trajectories and calculating averages we skipped the first 10 ns to eliminate the biasing towards the initial conformation.

The molecular dynamics package GROMACS version 4.5.5 [64] was used to perform all simulations with a time step of 0.002 ps. The electrostatic forces were evaluated by the Particle-Mesh Ewald method [65] (with real-space cutoff of 1.2 nm, grid spacing of 0.12 nm, and quadratic interpolation) and the Lennard-Jones (LJ) forces by a cutoff of 1.2 nm (with long-range dispersion correction for the energy and pressure). The entire system was maintained at a constant temperature of 300 K by the velocity rescaling thermostat [66] with a coupling time of 0.1 ps, and at a pressure of 1.0 bar by the Berendsen thermostat [67] with a compressibility of 5×10^{-5} 1/bar and a coupling time of 1.0 ps.

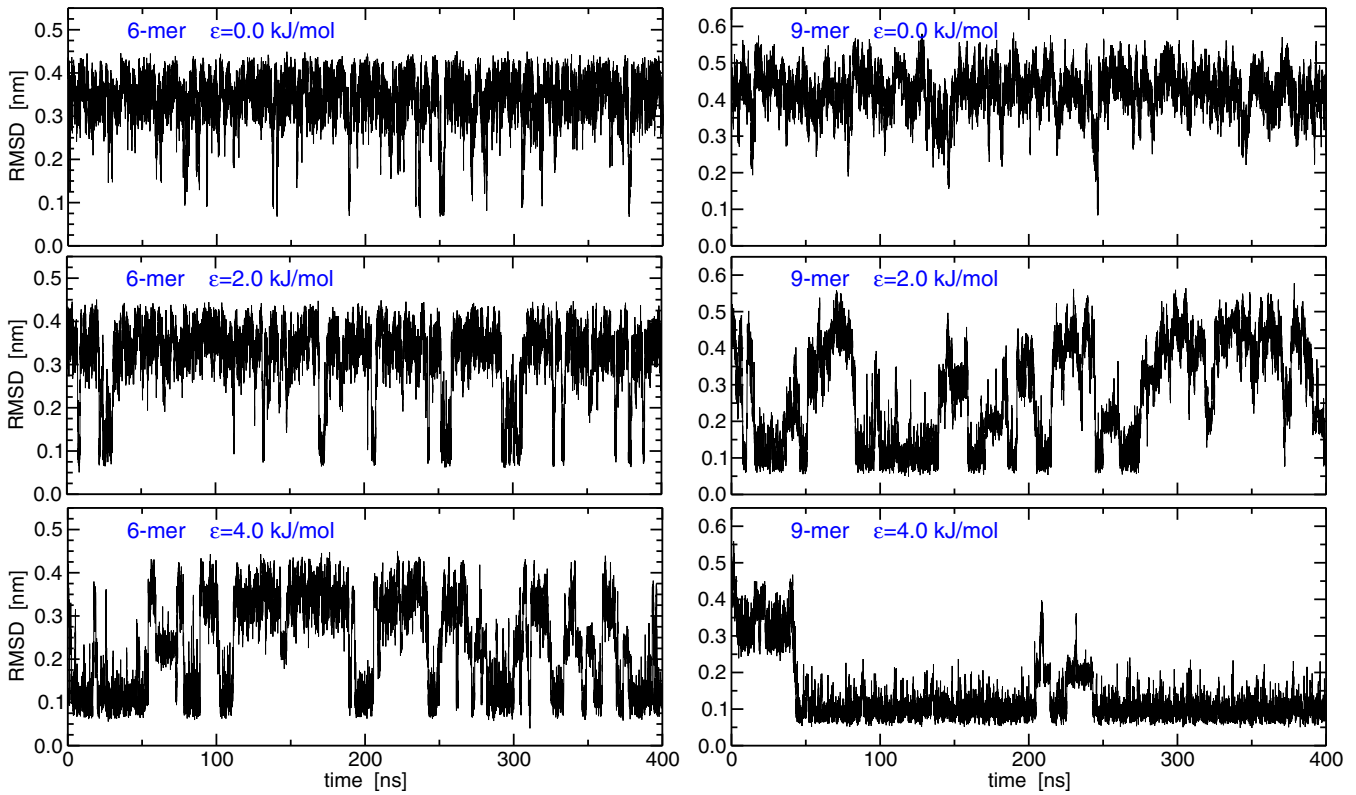

FIG. 1. (Color online) The root mean squared deviations of the peptide heavy atoms obtained by fitting to an ideal α-helix structure. The left column is for the six-mer, whereas the right column is for the nine-mer, residue peptides. The top, middle, and bottom panels correspond to an increasing strength of the side-chain–side-chain interactions.

III. RESULTS AND DISCUSSION

Figure 1 exhibits the RMSD of the peptide with respect to an ideal α-helical structure for the shortest (six-mer) and longest (nine-mer) chains at three different strengths of the interaction between the side-chain residues. In both cases, the instantaneous values of the RMSD for $\epsilon = 0.0$ kJ/mol are large indicating that the peptide hardly samples the α-helical conformation. However, with the increase in the side-chain–side-chain attraction the peptide also samples, via sharp transitions, conformations that are represented by smaller values of the RMSD. The stability of these small RMSD states increases with increasing the strength of this interaction (ϵ). More specifically, at $\epsilon = 4.0$ kJ/mol, the state at the smallest RMSD value ($\sim$0.1 nm) is sampled almost half of the time along the trajectory of the six-mer peptide and is by far the most stable state for the nine-mer peptide. Very similar behavior, which is interpolated between these two peptide lengths, is also exhibited by the seven-mer and eight-mer peptides (see Fig. S2 in the Supplemental Material [62]).

Histograms of the RMSD plots (excluding the first 10 ns) are shown in Fig. 2 for all four peptide lengths. To a very good approximation, all four systems display three distinct states. The locations of the minima separating the maxima identifying these states were used as the cutoff values and are depicted in Fig. 2. Visual inspections of the different states clearly confirm that the state for which the RMSD is smaller than 0.15 nm corresponds to that in which the entire peptide adopt an α-helical conformation. This is shown in Fig. 3 for the six-mer and nine-mer peptides. In Fig. S3 in the Supplemental Material [62] we plot the number of intrapeptide hydrogen bonds for two trajectories and compare these curves with the corresponding RMSD. The comparisons indicate a strong correlation between the number of intrapeptide hydrogen bonds and the RMSD values; frames with small RMSD values are characterized by large numbers of hydrogen bonds and vice versa. The state in which the RMSD value is in the range, 0.15 nm $<$ RMSD $<$0.25/0.26 nm, corresponds to a partial α-helical structure. For the peptide lengths larger than 6 a clear segment of an α-helical structure can be identified along the chain. Finally, the state with RMSD values larger than 0.25 nm (0.26 nm for the six-mer peptides) corresponds to an unfolded state. Note that in few of the distributions shown in Fig. 2 (e.g., for the six-mer and nine-mer peptides at $\epsilon = 2.0$ kJ/mol) the curve in the unfolded regime displays a shoulder or a second peak. In Fig. 3 we present snapshots of both substates, which indicate that the one with the larger RMSD corresponds to an extended conformation. The probability of observing this extended conformation reduces for larger ϵ and it actually disappears in the nine-mer peptide for $\epsilon \geq 3.0$ kJ/mol. Indeed, it is known that for long peptide chains the extended conformation is not stable and even in absence of any secondary structure the peptide will collapse to a globular shape because of its hydrophobic character. Note that the existence of extended conformations in the ensemble of the unfolded states does not necessarily mean the existence of multiple maxima in the distribution. However it was shown that in the case of short peptides the unfolded state does not randomly sample its conformational space and the transitions between distinct conformers can be abrupt [68].

Using the definitions of the cutoff values marked in Fig. 2 we counted the number of times each state was visited throughout the trajectory. In Fig. 4(a) we plot the fraction

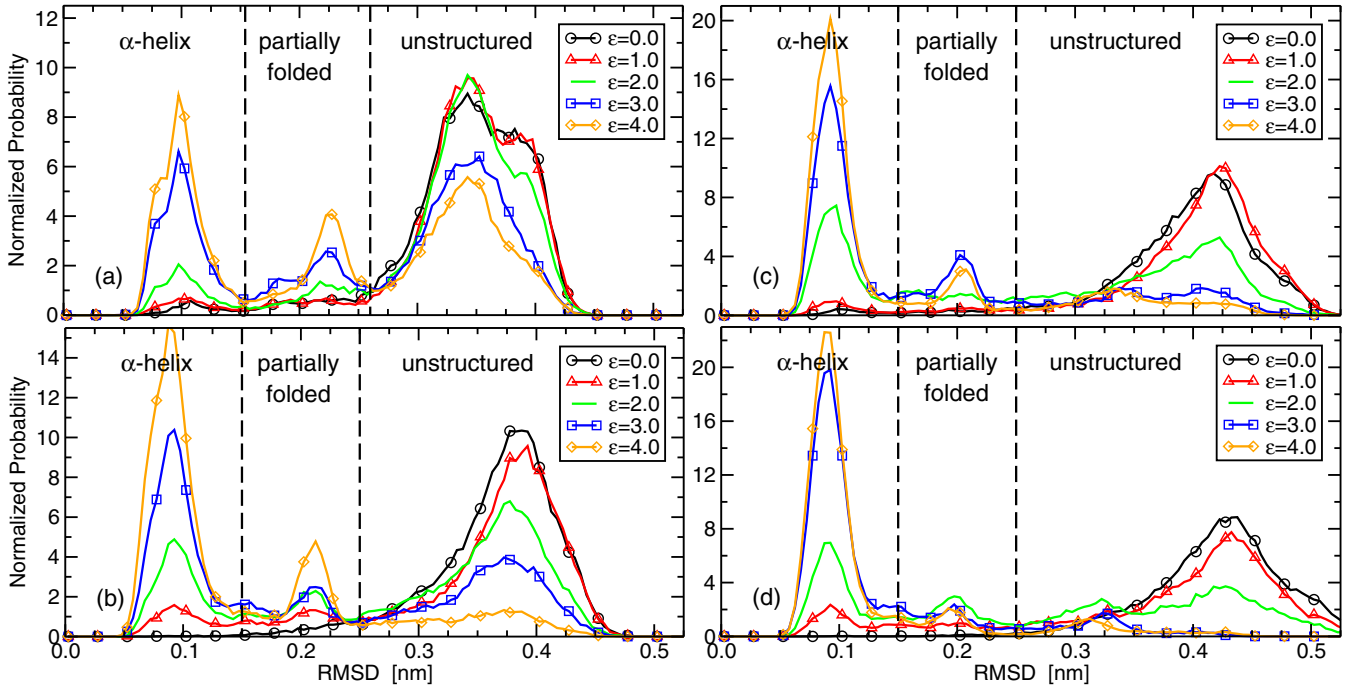

FIG. 2. (Color online) Histograms of the RMSD (from an ideal α-helical structure) of the peptides for different strengths of interactions between the side-chain residues. (a), (b), (c), and (d) are for six-mer, seven-mer, eight-mer, and nine-mer residue oligomers, respectively. The vertical dashed lines denote the RMSD cutoff values defining the three different states. These cutoff values were taken to be 0.150 nm and 0.250 nm in all cases except for the six-residue peptide in which case they were assigned the values 0.154 nm and 0.260 nm.

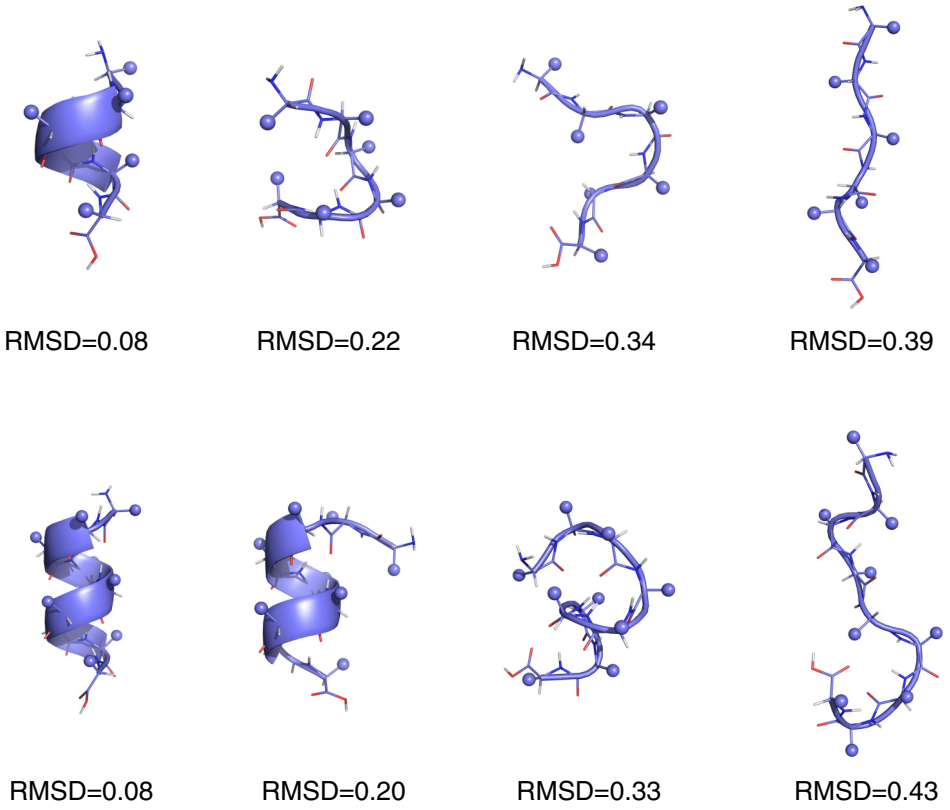

FIG. 3. (Color online) Snapshots of the peptide for four conformations with different RMSD (relative to an ideal α-helix) values. The upper and lower panels correspond to six-residue and nine-residue peptides, both at $\epsilon = 2.0$ kJ/mol. For each peptide length, the different conformations correspond to three different states: α-helical, partial-helical (partially folded), and unfolded. The latter (represented by the snapshots with the two largest RMSD values) can adopt either an extended or a more compact conformation.

of time the entire peptide adopts an α-helical conformation as a function of ϵ for all four chain lengths. Considering the process in which the unfolded (U) and partially folded (P) states change their conformation to the helical (H) state,

$$U + P \leftrightarrows 2H, \tag{1}$$

the equilibrium constant of this process can be defined by the probabilities of observing the different states, $K = \rho_H^2/(\rho_U \rho_P)$. Figure 4(b) displays the corresponding Gibbs (free) energy change $\Delta G = -RT \ln K$. For all peptide lengths, the content of the helical state increases with increasing the attraction between the side-chain residues. However, the magnitude of the increase (i.e., the slope of the curve) is different for the different oligomer lengths. The increase is larger for longer chains. This is not surprising because, as mentioned in the introduction, it is experimentally observed that the propensity for the helical state is larger for longer peptides (up to a certain peptide length) due to the nucleation penalty. Note that in the process written in Eq. (1) we considered only the helical state in the products side simply because this is the state we are interested in.

At $\epsilon = 0.0$ kJ/mol all peptides display a negligible amount of helical content, which indicate that the backbone alone can not induce the transition to a helical conformation. It is then the role of the side chains to contribute a favorable free energy term to render this transition possible. In other words, it is a character of the peptide backbone to permit the stability of the helical state to be sensitive to contributions from the side chains. The range of the energy required to observe a substantial population of helical conformation in the short peptides we modeled are on the order of the thermal energy, kT. Note that experimentally, seven-mer polyalanine based peptides are too short to adopt a stable helical structure in aqueous solutions [69]. Indeed, this is reproduced in the simulations. The value of ϵ that would correspond to a 40% population of the helical state in the six-mer peptide is about 4 kJ/mol, which is much larger than a common value of ϵ for a methyl group in all force fields. Nevertheless, helical content in a five-mer peptide, $WAAAH^+$, in aqueous solution is observed experimentally [70] and computationally [71]. The reason for the enhanced stability of the helical state in such a short peptide is the strong cation-π interaction between the side chain of the positively charged histidine and that of tryptophan [72]. In our analysis we do not count the fractional helicity of the peptide, that means the fraction of residues that adopt a helical conformation (e.g., based on the values of their ϕ and ψ dihedral angles) but instead count the fraction of time the entire peptide (or almost entire) adopt an α-helical structure.

The (Lennard-Jones) energy between the n side-chain residues in the helical state can be calculated by,

$$U_{\text{sc-sc}}^{\text{helix}} = \epsilon \sum_{i=1}^{n} \sum_{j=i+1}^{n} \xi_{ij}. \tag{2}$$

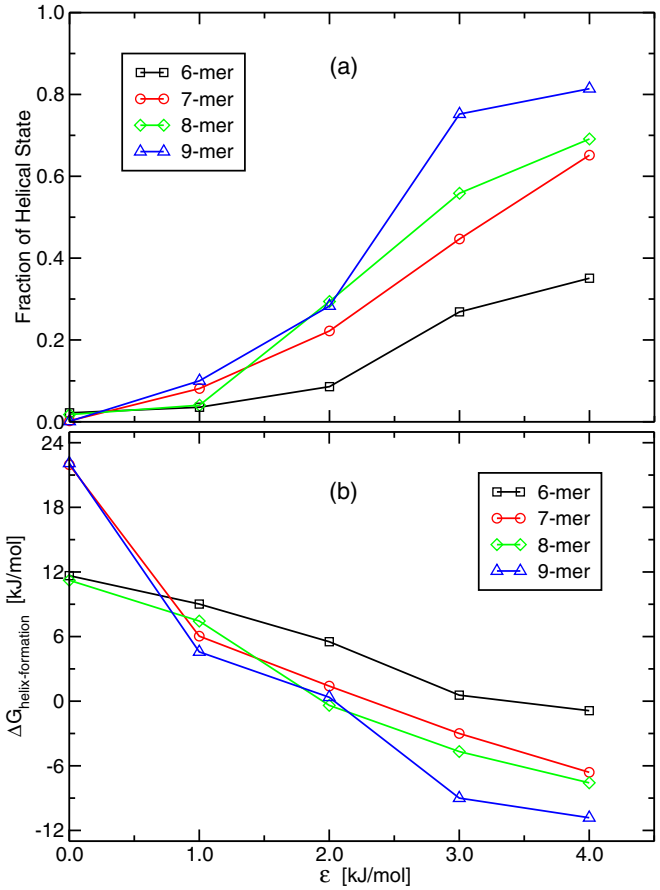

FIG. 4. (Color online) The probability of the peptides to adopt an α-helical conformation (a), as well as the free energy change of forming this helical state (b), as a function of the side-chain–side-chain Lennard-Jones ϵ parameter.

The term ξ_{ij} is the LJ potential between side-chain residue i and j,

$$\xi_{ij} = \left(\frac{r_m}{r_{ij}}\right)^{12} - 2\left(\frac{r_m}{r_{ij}}\right)^{6}. \tag{3}$$

Due to the periodicity of an ideal helical conformation, ξ_{ij} depends only on the difference of the positions along the chain of these residues, $|i-j|$. These side-chain–side-chain interactions are attractive and because of the $1/r^6$ dependence they approach zero quickly. For example, the strength of interactions of side chain i with its side-chain neighbors decreases in the following order: $i \pm 1$ ($r = 0.53$ nm) $> i \pm 4$ ($r = 0.60$ nm) $> i \pm 3$ ($r = 0.62$ nm) $> i \pm 2$ ($r = 0.75$ nm) $> i \pm 5$ ($r = 0.94$ nm) $> i \pm 6$ ($r = 1.09$ nm) and so on. In order to prevent double counting, the sum in Eq. (2) considers the pair interactions only towards one end of the helix, resulting in a decreasing number of pair-interaction terms in the second summation (over j) as the value of i increases. An alternative way to calculate the interactions in Eq. (2) is to consider a periodic image of the peptide only on one side of the termini and from that to subtract the interactions involving the periodic

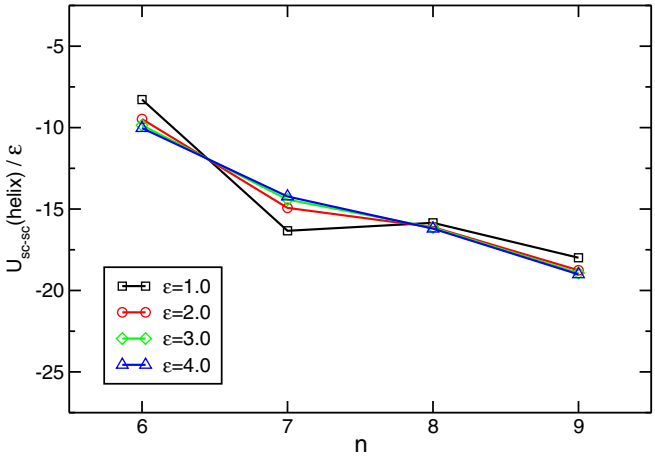

FIG. 5. (Color online) The side-chain–side-chain LJ energy of the helical state as a function of the number of residues for four different values of ϵ. This energetic term is represented by $U^{\text{helix}}_{\text{sc-sc}}/\epsilon = -\nu n + \Gamma$. The average values of the ν and Γ parameters obtained from linear regression fittings equal 2.9 and 6.9, respectively (see Table I).

image,

$$U^{\text{helix}}_{\text{sc-sc}} = \epsilon\left[n\sum_{k=1}^{n-1}\xi_k - \sum_{i=2}^{n}\sum_{k=1}^{i-1}\xi_{n-k}\right]. \tag{4}$$

The first term is the energy, summed over all residues i, of the interactions between this i residue and all $k = |i-j|$ differences in the position of the residues including those within the onesided periodic image. Because this interaction is the same for all i residues, the summation over i is substituted with multiplication by n. The second term in Eq. (4) is the energy of the interactions with the onesided periodic image that has to be removed. Because of the short-range nature of the interactions between the side chains, in practice the sum of the ξ_k (for long enough n) is independent of n and we let $\nu = -\sum_{k=1}^{n-1}\xi_k$. A similar argument is made for the missing interactions and we assign $\Gamma = -\sum_{i=2}^{n}\sum_{k=1}^{i-1}\xi_{n-k}$. Equation (4) then becomes,

$$U^{\text{helix}}_{\text{sc-sc}}/\epsilon = -\nu n + \Gamma. \tag{5}$$

Thus, a plot of $U^{\text{helix}}_{\text{sc-sc}}/\epsilon$ vs n should yield a straight line with a slope of $-\nu$ and and intercept of Γ for all (nonzero) values of ϵ. We applied this relation to all four nonzero values of ϵ and the results are shown in Fig. 5 and presented in Table I.

TABLE I. The values of the parameters ν and Γ, in the equation: $U^{\text{helix}}_{\text{sc-sc}}/\epsilon = -\nu n + \Gamma$ obtained by linear regressions of the curves shown in Fig. 5. The absolute values of the correlation coefficient R are also indicated.

	ν	Γ	$\lvert R\rvert$
$\epsilon = 1.0$	2.87	6.87	0.856
$\epsilon = 2.0$	2.90	6.96	0.960
$\epsilon = 3.0$	2.89	6.88	0.982
$\epsilon = 4.0$	2.89	6.82	0.989
Average	2.89	6.88	0.947

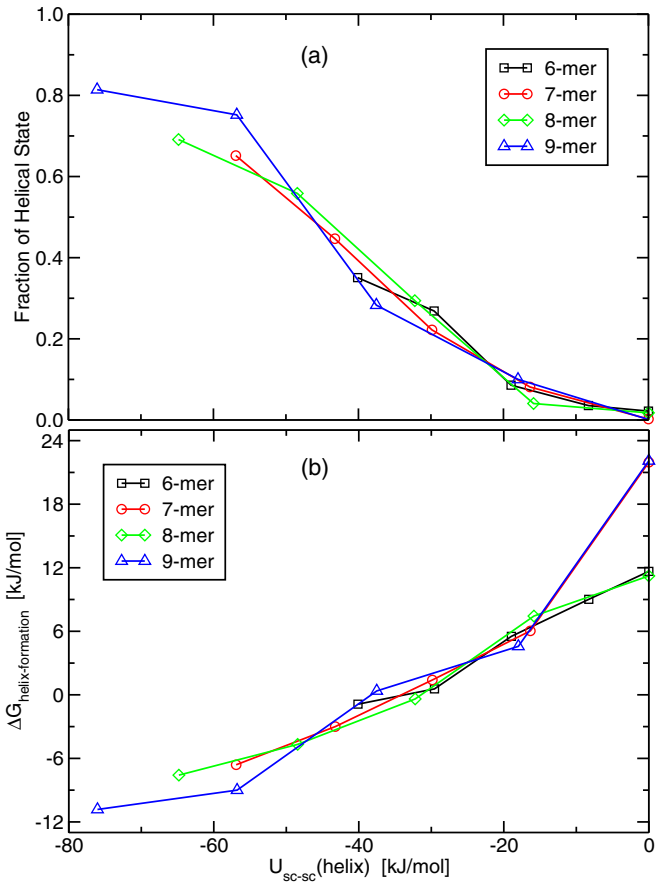

FIG. 6. (Color online) The probability of the peptides to adopt an α-helical conformation (a), as well as the free energy change of forming the helix (b) as a function of the energy between the side-chain residues in the helical state.

Except for $\epsilon = 1.0$ kJ/mol, there is a very good fit to a straight line and the values of ν and Γ are independent of ϵ with average values of 2.89 and 6.88, respectively. For a comparison, we obtained analytically the value of ν by summing ξ_i over the first six closest i neighbors in an ideal helix conformation which yielded a value of 2.66. The corresponding calculation for the theoretical value of Γ yield a value of 6.68. Thus, there is a discrepancy of about 9% and 3% from the hypothetical values of an ideal α helix.

Plots of the fraction of time the helical state is observed throughout the trajectory and the corresponding free energy change as a function of the energy between the side-chain residues in the helical conformation are shown in Fig. 6. The plots for all n fall on the same curve. Two major deviations are observed. Both are in the change of the free energy at $\epsilon = 0.0$ kJ/mol, one for the seven-mer and the other for nine-mer. As shown in Fig. 2, at $\epsilon = 0.0$ kJ/mol the populations of the helical state (especially at these two peptide lengths) are very small, and therefore, small absolute inaccuracies in their values will produce large errors in the calculation of the equilibrium constant. If we approximate the shape of the curves to a linear line [although Fig. 6(a) resembles more of a saturation curve] the average slope over the four different lengths is -0.011 mol/kJ. This means that in order to increase the probability of the helical state of a peptide from X% to Y% there is a need to increase the *total* interaction energy between the side-chains residues (in experimental realizations by, for example, direct or solvent-induced interactions, salt bridges, or cations binding) by about 0.9(Y–X) kJ/mol, irrespective of the length of the peptide (within the range considered here). If we extrapolate this behavior to a 22-mer peptide, the results indicate that in order to observe the helical state 40% of the time ($U_{\text{sc-sc}}^{\text{helix}} = -40$ kJ/mol), the value of ϵ between the side chains [Eq. (5)] should be about 0.7 kJ/mol.

The scaled behavior with respect to n described above is not trivial because the increase in the value of ϵ (or alternatively in the value of $U_{\text{sc-sc}}^{\text{helix}}$) changes also the interaction energy of the partially folded and coil states. Consider, for example, the process in which several unfolded states, u_i, fold into an α-helical structure, h,

$$\sum_{i=1}^{m} u_i \leftrightarrows mh. \tag{6}$$

In this case,

$$\begin{aligned}\Delta H &= mH_h - \sum_{i=1}^{m} H_{u_i} \\ &= m[-\epsilon\nu n + \Gamma\epsilon + C_h(n)] - \sum_{i=1}^{m} H_{u_i}(n,\epsilon).\end{aligned} \tag{7}$$

The enthalpy (energy) of the helical state is decomposed into the side-chain–side-chain contribution ($-\epsilon\nu n + \Gamma\epsilon$) and the remaining interactions in the system $C_h(n)$. The latter, which in addition to the peptide-peptide interactions includes also the peptide-solvent and solvent-solvent interactions is a function of the length of the peptide n but not of ϵ (simply because in our model a change in ϵ changes only the LJ energy *between* the side chains).

Assuming that the entropies of the different states do not depend on the value of ϵ and that the entropy of the helical state is much smaller than that of the unfolded states we can write,

$$\Delta S = mS_h(n) - \sum_{i=1}^{m} S_{u_i}(n) \approx -\sum_{i=1}^{m} S_{u_i}(n) = -\sum_{i=1}^{m} ns_{u_i}, \tag{8}$$

where s_{u_i} is the average entropy per residue of the unfolded state i (for the partially folded states this may not be an adequate representation because the entropy of the different residues are not likely to be homogeneous along the chain). The change in the free energy for helix formation is then,

$$\begin{aligned}\Delta G &= \Delta H - T\Delta S \\ &= m[-\nu\epsilon n + \Gamma\epsilon + C_h(n)] - \sum_{i=1}^{m} \left[H_{u_i}(n,\epsilon) - Tns_{u_i}\right].\end{aligned} \tag{9}$$

From Eq. (9) it is not clear at all why the value of ΔG as a function of $U_{\text{sc-sc}}^{\text{helix}} = -\nu\epsilon n + \Gamma\epsilon$ would have the same dependency for all peptide lengths (n). Obviously, more information about the last three terms are needed. We attempted to obtain

some knowledge of the behavior of $C_h(n)$ and $\sum_{i=1}^{m} H_{u_i}(n,\epsilon)$ as a function of n, ϵ, and $U_{\text{sc-sc}}^{\text{helix}}$. However, unfortunately, due to the large errors characterizing the potential energy of these terms (which is mainly due to the solvent-solvent interactions) we were unable to draw any further conclusions.

We would like to emphasize that in this study the controlled parameter of the interaction between the side chains was chosen to be ϵ because we could then design all other interactions to be invariant. However, in experimental realization changing the side-chain–side-chain interactions will modify the interactions with all other particles as well. The controlled parameter in this case will be the *effective* interactions (or the potential of mean force) between the side-chain residues.

IV. CONCLUSIONS

All amino acids, except for proline, have exactly the same backbone chemical structure. Nevertheless, large differences exist in the propensity of these different amino acids to form an α-helical conformation. Although initially it was proposed that the difference in this propensity originates from the entropic penalty of the side chains in the helical conformation, it is now well acknowledged that a gain in enthalpy, due to the interactions between the side chains, also contributes to the difference in the ability to promote the helical conformation. In this paper we investigated the effect of the magnitude of these side-chain–side-chain interactions on the stability of the helical state in six- to nine-mer homogeneous peptides. The results indicate that the transformation from a coil state to a helical state can be completely induced by augmenting the effective interactions between the side-chains residues. This indicates that the contribution of the backbone alone to the change in the free energy of the helix-coil transition is unfavorable but small, and it therefore allows the contribution of the side chains to determine whether the peptide adopts a coil or a helical conformation. Although, for short peptides the effective interaction (per residue) between the side chains has to be large (larger, for example, than that of a methyl group) to induce a significant population of a helical state, it is still on the order of a thermal energy kT. With increasing the length of the chain, this required interaction per residue reduces substantially due to the relative decrease in the importance of the nucleation penalty. Additionally, we find that a plot of the fraction of time the helical state is observed, or alternatively the corresponding free energy change for helix formation, as a function of the total effective energy between the side-chains residues in the helical state is the same for all peptide lengths. This behavior is not trivial and we could not explain it. Nevertheless, from the slope of the curve, it is predicted that in order to increase the probability of the helical state from X% to Y%, the total effective interactions between the side chains need to increase by about 0.9(Y–X) kJ/mol. Thus, in order to predict the probability of the helical state of a peptide it is necessary to evaluate the effective energy between all side chains in this helical conformation. It is yet to be studied whether the extension of the system to more complicated models (for example, inhomogeneous peptides in which the side-chain–side-chain interactions along the chain is not the same) will preserve the conclusions obtained in this study.

ACKNOWLEDGMENTS

I would like to thank Professor Robert L. Baldwin for stimulating discussions. This work has been funded with support from the Spanish Ministry of Science and Innovation, MICINN (Grant No. CTQ2010-20297) and the European Commission, Marie Curie International Reintegration Grant, Project No. 247485. Technical and human support provided by SGIker (USED SERVICES) (UPV/EHU, MICINN, GV/EJ, ESF) is gratefully acknowledged.

[1] M. Karplus and D. L. Weaver, Prot. Sci. **3**, 650 (1994).
[2] R. L. Baldwin and G. D. Rose, Trends Biochem. Sci. **24**, 26 (1999).
[3] L. Pauling and R. B. Corey, J. Am. Chem. Soc. **72**, 5349 (1950).
[4] B. H. Zimm and J. K. Bragg, J. Chem. Phys. **31**, 526 (1959).
[5] S. Lifson and A. Roig, J. Chem. Phys. **34**, 1963 (1961).
[6] W. Kauzmann, Adv. Prot. Chem. **14**, 1 (1959).
[7] I. M. Klotz and J. S. Franzen, J. Am. Chem. Soc. **84**, 3461 (1962).
[8] K. A. Dill, Biochemistry **29**, 7133 (1990).
[9] E. S. Eberhardt and R. T. Raines, J. Am. Chem. Soc. **116**, 2149 (1994).
[10] Z. Shi, B. A. Krantz, N. Kallenbach, and T. R. Sosnick, Biochemistry **41**, 2120 (2002).
[11] B. A. Krantz *et al.*, Nature Struct. Biol. **9**, 458 (2002).
[12] R. L. Baldwin, J. Biol. Chem. **278**, 17581 (2003).
[13] J. Gao and J. W. Kelly, Prot. Sci. **17**, 1096 (2008).
[14] D. W. Bolen and G. D. Rose, Annu. Rev. Biochem. **77**, 339 (2008).
[15] J. A. Schellman, C. R. Trav. Lab. Carlsberg Chim. **29**, 223 (1955).
[16] J. A. Schellman, C. R. Trav. Lab. Carlsberg Chim. **29**, 230 (1955).
[17] G. Rialdi and J. Hermans, J. Am. Chem. Soc. **88**, 5719 (1966).
[18] P. Y. Chou and H. A. Scheraga, Biopolymers **10**, 657 (1971).
[19] J. M. Scholtz and R. L. Baldwin, Annu. Rev. Biophys. Biomol. Struct. **21**, 95 (1992).
[20] M. M. Lopez, D.-H. Chin, R. L. Baldwin, and G. I. Makhatadze, Proc. Natl. Acad. Sci. USA **99**, 1298 (2002).
[21] G. Goch *et al.*, Biochemistry **42**, 6840 (2003).
[22] J. M. Scholtz, H. Qian, E. J. York, J. M. Stewart, and R. L. Baldwin, Biopolymers **31**, 1463 (1991).
[23] S. Kumar and M. Bansal, Biophys. J. **71**, 1574 (1996).
[24] M. I. Liff, P. C. Lyu, and N. R. Kallenbach, J. Am. Chem. Soc. **113**, 1014 (1991).
[25] W. S. Young and C. L. Brooks, III, J. Mol. Biol. **259**, 560 (1996).
[26] G. Hummer, A. E. García, and S. Garde, Phys. Rev. Lett. **85**, 2637 (2000).
[27] R. Zangi, H. Kovacs, W. F. van Gunsteren, J. Johansson, and A. E. Mark, Proteins **43**, 395 (2001).
[28] L. Monticelli, D. P. Tieleman, and G. Colombo, J. Phys. Chem. B **109**, 20064 (2005).

[29] T. E. Creighton, *Proteins: Structures and Molecular Properties* (W. H. Freeman & Co., New York, 1992).
[30] A. Horovitz, J. M. Matthews, and A. R. Fersht, J. Mol. Biol. **227**, 560 (1992).
[31] M. Blaber *et al.*, J. Mol. Biol. **235**, 600 (1994).
[32] K.-H. A. J. Wójcik and H. A. Scheraga, Biopolymers **30**, 121 (1990).
[33] P. C. Lyu, M. I. Liff, L. A. Marky, and N. R. Kallenbach, Science **250**, 669 (1990).
[34] S. Padmanabhan, S. Marqusee, T. Ridgeway, T. M. Laue, and R. L. Baldwin, Nature (London) **344**, 268 (1990).
[35] K. T. O'Neil and W. F. DeGrado, Science **250**, 646 (1990).
[36] J. K. Myers, C. N. Pace, and J. M. Scholtz, Biochemistry **36**, 10923 (1997).
[37] C. N. Pace and J. M. Scholtz, Biophys. J. **75**, 422 (1998).
[38] S. Marqusee, V. H. Robbins, and R. L. Baldwin, Proc. Natl. Acad. Sci. USA **86**, 5286 (1989).
[39] J. A. Vila, D. R. Ripoll, and H. A. Scheraga, Proc. Natl. Acad. Sci. USA **97**, 13075 (2000).
[40] A. E. García and K. Y. Sanbonmatsu, Proc. Natl. Acad. Sci. USA **99**, 2782 (2002).
[41] S. Gnanakaran and A. E. García, Proteins **59**, 773 (2005).
[42] S. Marqusee and R. L. Baldwin, Proc. Natl. Acad. Sci. USA **84**, 8898 (1987).
[43] W.-Z. Wang, T. Lin, and Y.-C. Sun, J. Phys. Chem. B **111**, 3508 (2007).
[44] J. Dzubiella, J. Am. Chem. Soc. **130**, 14000 (2008).
[45] M. Hinczewski, Y. von Hansen, J. Dzubiella, and R. R. Netz, J. Chem. Phys. **132**, 245103 (2010).
[46] T. P. Creamer and G. D. Rose, Proc. Natl. Acad. Sci. USA **89**, 5937 (1992).
[47] G. I. Makhatadze, Adv. Prot. Chem. **72**, 199 (2005).
[48] P. Luo and R. L. Baldwin, Proc. Natl. Acad. Sci. USA **96**, 4930 (1999).
[49] J. M. Richardson, M. M. Lopez, and G. I. Makhatadze, Proc. Natl. Acad. Sci. USA **102**, 1413 (2005).
[50] J. Wójcik, J. Góral, K. Pawłowski, and A. Bierzyński, Biochemistry **36**, 680 (1997).
[51] M. Siedlecka, G. Goch, A. Ejchart, H. Sticht, and A. Bierzyńsk, Proc. Natl. Acad. Sci. USA **96**, 903 (1999).
[52] V. Muñoz and L. Serrano, Nature Struct. Biol. **1**, 399 (1994).
[53] P. Cossio, D. Granata, A. Laio, F. Seno, and A. Trovato, Sci. Rep. **2**, 351 (2012).
[54] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).
[55] N. A. McDonald and W. L. Jorgensen, J. Phys. Chem. B **102**, 8049 (1998).
[56] R. C. Rizzo and W. L. Jorgensen, J. Am. Chem. Soc. **121**, 4827 (1999).
[57] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, J. Phys. Chem. B **105**, 6474 (2001).
[58] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[59] S. Miyamoto and P. A. Kollman, J. Comp. Chem. **13**, 952 (1992).
[60] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, J. Comp. Chem. **18**, 1463 (1997).
[61] G. Vriend, J. Mol. Graph. **8**, 52 (1990).
[62] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.89.012723 for additional figures (Figs. S1–S3) associated with this manuscript.
[63] P. S. Georgoulia and N. M. Glykos, J. Phys. Chem. B **115**, 15221 (2011).
[64] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, J. Chem. Theory Comput. **4**, 435 (2008).
[65] T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[66] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).
[67] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[68] X. Daura, B. Jaun, D. Seebach, and W. F. van Gunsteren, J. Mol. Biol. **280**, 925 (1998).
[69] P. Luo and R. L. Baldwin, Biochemistry **36**, 8413 (1997).
[70] M. M. Lin, O. F. Mohammed, G. S. Jas, and A. H. Zewail, Proc. Natl. Acad. Sci. USA **108**, 16622 (2011).
[71] D. De Sancho and R. B. Best, J. Am. Chem. Soc. **133**, 6809 (2011).
[72] M. M. Torrice, K. S. Bower, H. A. Lester, and D. A. Dougherty, Proc. Natl. Acad. Sci. USA **106**, 11919 (2009).

Supplementary Material:

Side-Chain–Side-Chain Interactions and Stability of the Helical State

Ronen Zangi[1,2,‡]

1. Department of Organic Chemistry I, University of the Basque Country UPV/EHU, Avenida de Tolosa 72, 20018, San Sebastian, Spain

2. IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain

January 14, 2014

‡ Email: r.zangi@ikerbasque.org

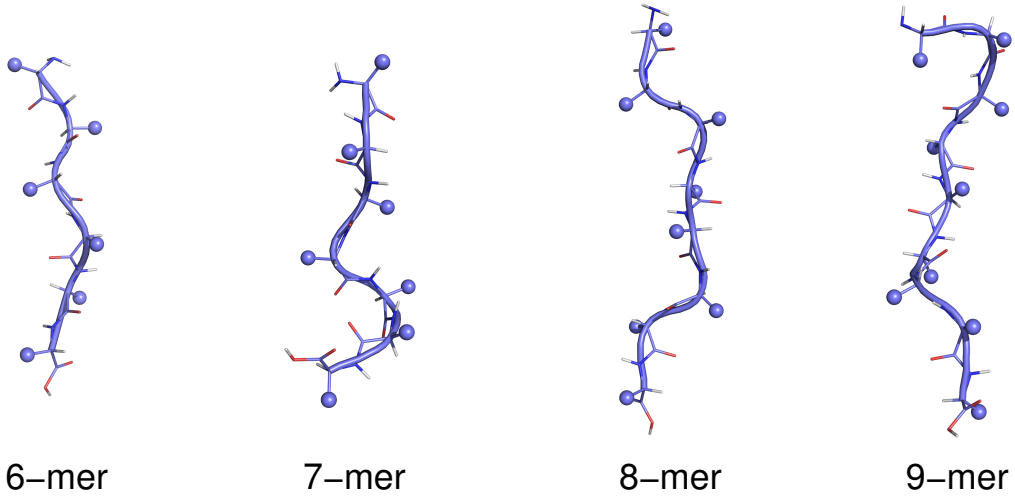

Figure S1: The starting random-coil conformations for the simulations of the four different peptide lengths. The side-chain of each residue is shown by a blue solid sphere.

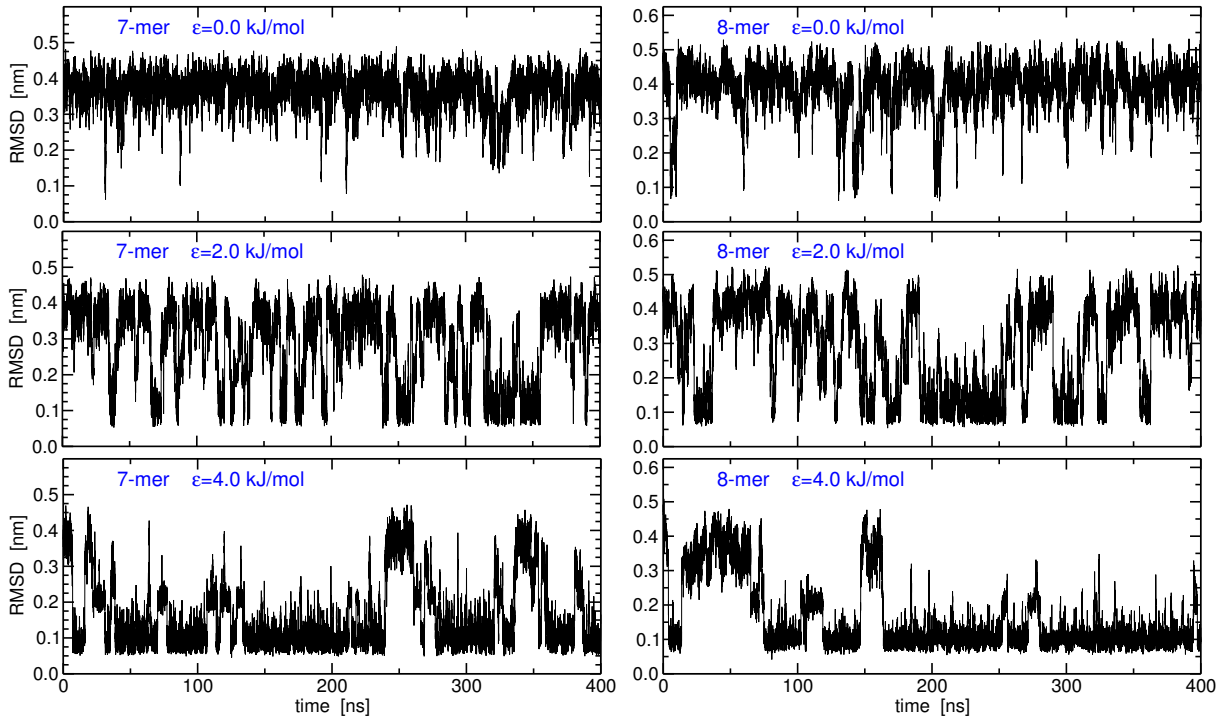

Figure S2: The root mean squared deviations of the peptide heavy atoms obtained by fitting to an ideal α-helix structure. The left column is for the 7-mer, whereas the right column is for the 8-mer, residue peptides. The top, middle, and bottom panels correspond to an increasing strength of the side-chain–side-chain interactions.

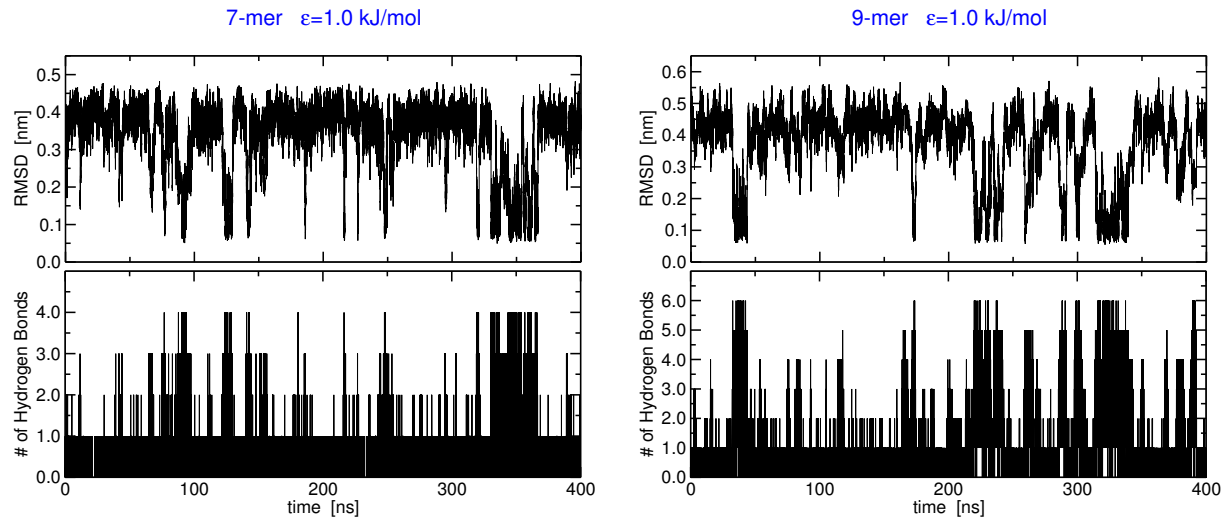

Figure S3: The number of the intra-peptide (backbone) hydrogen bonds (lower panels) as a function of time for the 7-mer and 9-mer peptides with side-chain–side-chain strength of 1.0 kJ/mol. The corresponding RMSDs are also plotted (upper panels) demonstrating a very strong correlation to the hydrogen bonds plots. A hydrogen bond is defined by a donor–acceptor cutoff distance of 0.35 nm and a donor–hydrogen–acceptor angle larger than 150°. Note that for a perfect α-helical structure, the number of $i \cdots i+4$ hydrogen bonds is 3 and 5 for the 7-mer and 9-mer, respectively. However, in the plots a negligible population of 4 (0.36%) and 6 (0.21%) hydrogen bonds are present for the 7-mer and 9-mer, respectively. These 'extra' hydrogen bonds are due to instantaneous additional participation of one of the $i \cdots i+3$ hydrogen bonds that fall within the cut-offs defining the hydrogen bonds.