doi:10.1016/j.jmb.2005.08.053 J. Mol. Biol. (2005) 353, 232–241

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

How Large is an α-Helix? Studies of the Radii of Gyration of Helical Peptides by Small-angle X-ray Scattering and Molecular Dynamics

Bojan Zagrovic[1], Guha Jayachandran[2], Ian S. Millett[3,4] Sebastian Doniach[3,4] and Vijay S. Pande[2]*

[1]*Department of Chemistry and Applied Biosciences, Laboratory of Physical Chemistry, ETH Hönggerberg, Zürich 8093 Switzerland*

[2]*Department of Chemistry Stanford University, Stanford CA 94305, USA*

[3]*Department of Physics, Stanford University, Stanford CA 94305, USA*

[4]*Department of Applied Physics Stanford University, Stanford CA 94305, USA*

**Corresponding author*

Using synchrotron radiation and the small-angle X-ray scattering technique we have measured the radii of gyration of a series of alanine-based α-helix-forming peptides of the composition Ace-$(AAKAA)_n$-GY-NH_2, $n=2$–7, in aqueous solvent at 10(±1) °C. In contrast to other techniques typically used to study α-helices in isolation (such as nuclear magnetic resonance and circular dichroism), small-angle X-ray scattering reports on the global structure of a molecule and, as such, provides complementary information to these other, more sequence-local measuring techniques. The radii of gyration that we measure are, except for the 12-mer, lower than the radii of gyration of ideal α-helices or helices with frayed ends of the equivalent sequence-length. For example, the measured radius of gyration of the 37-mer is 14.2(±0.6) Å, which is to be compared with the radius of gyration of an ideal 37-mer α-helix of 17.6 Å. Attempts are made to analyze the origin of this discrepancy in terms of the analytical Zimm–Bragg–Nagai (ZBN) theory, as well as distributed computing explicit solvent molecular dynamics simulations using two variants of the AMBER force-field. The ZBN theory, which treats helices as cylinders connected by random walk segments, predicts markedly larger radii of gyration than those measured. This is true even when the persistence length of the random walk parts is taken to be extremely short (about one residue). Similarly, the molecular dynamics simulations, at the level of sampling available to us, give inaccurate values of the radii of gyration of the molecules (by overestimating them by around 25% for longer peptides) and/or their helical content. We conclude that even at the short sequences examined here (≤37 amino acid residues), these α-helical peptides behave as fluctuating semi-broken rods rather than straight cylinders with frayed ends.

Keywords: α-helix; SAXS; radius of gyration; molecular dynamics

Introduction

The α-helix is the most prevalent secondary structural element in proteins.[1] However, in the absence of stabilizing tertiary interactions, α-helices rarely persist in isolation. One of the notable exceptions is a series of alanine-based, helix-forming peptides introduced by Robert Baldwin and his group.[2–4] The study of these peptides has significantly furthered our understanding of the stability of α-helices,[4–6] helical preference of individual amino acids,[7–9] the role of helix-termination signals[6,10] and the kinetics of α-helix folding.[11] In addition, these peptides have been studied extensively by computer simulations, which lead to a more detailed microscopic picture of α-helix folding and dynamics in general.[12–20]

The two most commonly used techniques to study α-helices in isolation, CD and NMR, are predominantly short distance-range techniques. They report on the local structure and dynamics of

Abbreviations used: SAXS, small-angle X-ray scattering technique; ZBN, Zimm–Bragg–Nagai; MD, molecular dynamics.

E-mail address of the corresponding author: pande@stanford.edu

polypeptides while giving limited information on the long-range ordering and correlations. On the other hand, the small angle X-ray scattering technique (SAXS),[21] with molecular radius of gyration as its principal readout, is a long-range technique. It reports on the global structure of a molecule and, as such, gives information complementary to that obtained by CD, NMR and other local techniques. This ability of SAXS to probe the long-range structure in polypeptides is shared by electron paramagnetic resonance[22] and fluorescence resonance energy transfer[23,24] techniques, but with greater accuracy compared to these methods. In this study, we have used SAXS to measure the radii of gyration of six α-helix-forming peptides with composition Ace-(AAKAA)$_n$-GY-NH$_2$ (n=2–7) in aqueous solvent. The measured radii of gyration are smaller than those of the equivalent ideal α-helices or α-helices with frayed ends. This difference is analyzed in terms of the Zimm–Bragg–Nagai (ZBN) analytical theory[25] and explicit solvent molecular dynamics (MD) simulations using world-wide distributed computing. The ZBN theory has been very successful in describing the behavior of long homopolymeric helices under various conditions,[26] but fails to account for the low radii of gyration measured here. The ZBN theory describes helices as cylinders connected by random walk segments. It combines the standard Zimm–Bragg helix-coil theory[27] with the statistics of ideal random walks. However, even if the persistence length of the random walk segments is taken to be very short (one amino acid), the ZBN theory predicts radii of gyration that are significantly larger than the values obtained by SAXS. The explicit solvent MD simulations described here were performed using distributed computing techniques and two variants of the AMBER force-field with a hope of finding a microscopic explanation for the low radii of gyration observed in the experiment. However, at the level of sampling used in this study (50 independent 50 ns long trajectories for each peptide and each force field), the radii of gyration of the simulated peptides deviate from the measured values, especially for the longest peptides where they are overestimated significantly. In addition, the secondary structure content from simulations deviates from the estimates based on CD. This suggests that capturing the behavior of flexible peptides is an area where modern atomistic force-fields could be improved. Our SAXS results are most consistent with alanine-based polypeptides adopting fluctuating semi-broken helix conformations, and are somewhat at odds with the picture of helices as straight cylinders with frayed ends.

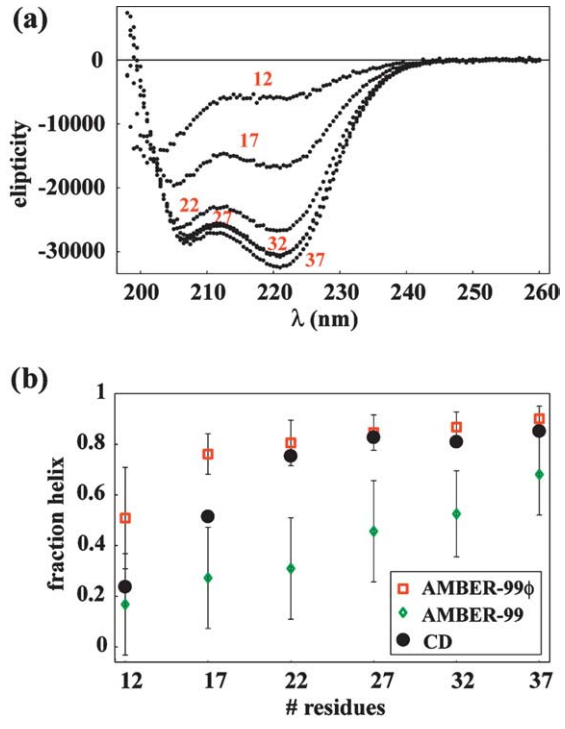

Figure 1. (a) Circular dichroism spectra of the six peptides in aqueous solvent at 10 °C. Peptide length is indicated in red. (b) Helical fraction for each of the peptides (black dots) calculated based on CD absorption at 222 nm and using normalization as described.[28] We also show the predicted helical fraction based on the simulations using DSSP to define helical residues (a residue is defined as helical if it falls in the G-H-I category in DSSP; this includes α-helices, 3_{10}-helices and π-helices; see Materials and Methods for details).

Results

How helical are the six peptides that we studied? Figure 1 (a) shows the CD spectra of the molecules. All peptides other than AK12 give rise to standard α-helical profiles with double minima at around 208 nm and 222 nm. The 12-mer is known to be too short to form a sizable helix, and its CD spectrum is consistent with random-coil behavior with some small level of helix present (~20%, see below). By measuring the mean molar ellipticity at 222 nm and using the standard length-dependent baseline,[28] we have estimated the mean fractional helical content (f_H) of the molecules, which is given in Figure 1(b). As expected, f_H increases with sequence length and reaches about 86% for AK37. We also show f_H and the standard deviation for the ensembles simulated using two AMBER force-field variants. Except for AK12 and AK17, the AMBER-99φ force-field matches the experimental results very well. The AMBER-99 force-field, on the other hand, significantly underestimates the average helical content, except for the very shortest peptide.

The scattering curves for the six peptides are given in Figure 2 in Guinier representation. We show the linear fits to the Guinier curves in the small-angle scattering limit that were used to determine the radii of gyration of the six molecules. One thing to note is that, as the molecules are elongated, the region over which the Guinier curves are linear is significantly narrower compared to the equivalent region in the case of spherical molecules

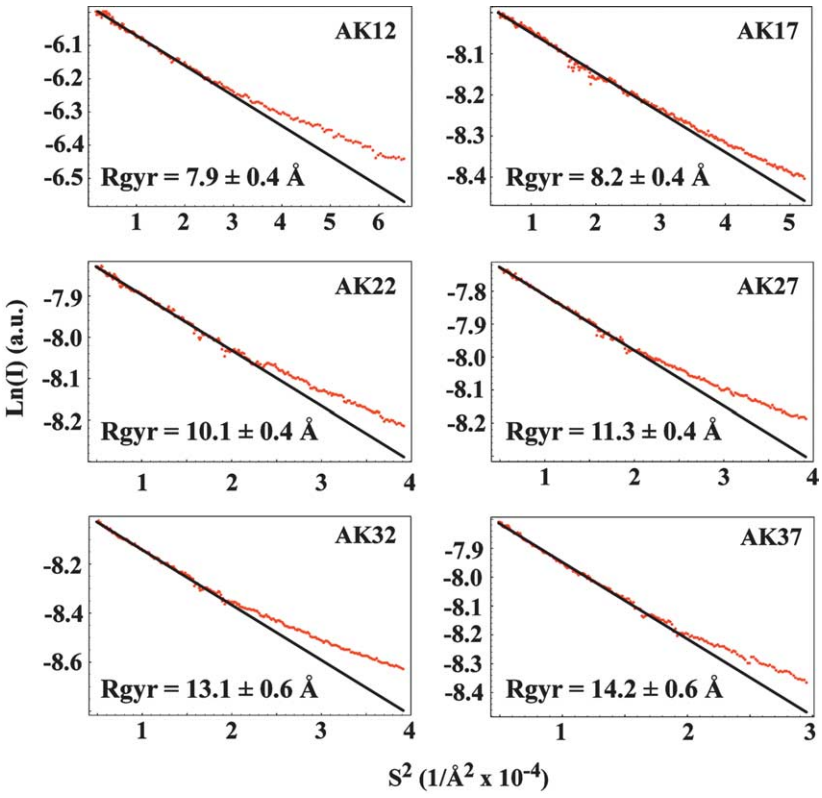

Figure 2. Experimentally determined scattering curves in the Guinier representation (in red) together with the linear fits that were used to determine the molecular radii of gyration (in black).

of the same molecular mass. Attempts were made to determine the radii of gyration of the molecules based on entire scattering curves and without resort to Guinier approximation, but this approach lead to inconsistent results, as the data were collected out to only a relatively small upper limit in scattering vector ($S_{max} = 0.04$ Å^{-1}).

Figure 3 exhibits the central result of the study: the radii of gyration of the six peptides as determined by SAXS (black dots). For all peptides except the 12-mer, the measured radii of gyration are significantly lower than the expected values for ideal helices of the same length (black line). This discrepancy is close to 25% for the 37-mer. We show also the average radii of gyration from our simulations using the AMBER-99 (green diamonds) and AMBER-99φ (red boxes) force-fields calculated for the ensembles between 30 ns and 50 ns. The radii of gyration from simulations using the AMBER-99φ force-field are relatively close to the expected ideal α-helix values throughout the sequence range studied. On the other hand, the radii of gyration from simulations using AMBER-99 are significantly lower, consistent with their unnaturally low helical content (Figure 1). The simulations were performed with a hope that they would shed light on the apparent discrepancy between the measured radii of gyration and what is expected for ideal α-helices or even α-helices with frayed ends. However, at the level of sampling used here, the simulations do not reproduce the measured radii of gyration and/or helical fraction close enough to justify such analysis.

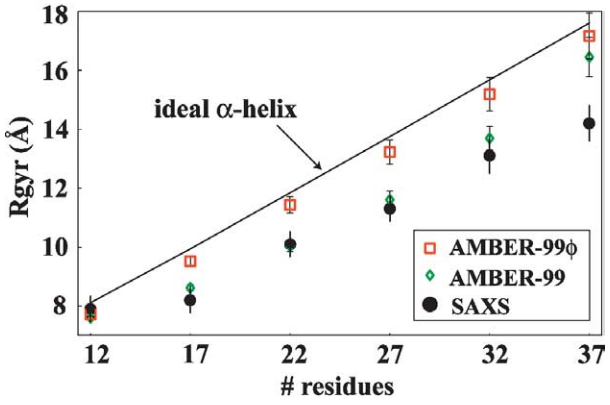

Figure 3. Radii of gyration of the six peptides as determined by SAXS (black dots) and molecular dynamics (red squares, AMBER-99φ; green diamonds, AMBER-99). The continuous line captures the chain-length dependence of the radii of gyration of ideal α-helices. The theoretical radii of gyration were calculated using Crysol (see Materials and Methods for details).

Discussion

The radii of gyration of the peptides studied here are significantly smaller than the radii of gyration of ideal α-helices with the same sequence length. The presence of 3_{10} or π-helices, the occurrence of fraying at the ends and the potentially inaccurate ellipticity of the CD coil-state baseline do not

provide a satisfactory explanation for this discrepancy. First, the 3_{10}-helix, which has been suggested as an important structural element in polyalanine peptides,[29,30] is more elongated compared to the ideal α-helix.[1] There are three residues per turn in a 3_{10}-helix with 2 Å rise per residue compared with 1.5 Å rise per residue in an α-helix. Therefore, any presence of the contiguous, unbroken 3_{10}-helix in the peptides studied would necessarily only increase their radii of gyration compared to ideal α-helices. For example, the radius of gyration of the AK37 structure in which three residues on each terminus are 3_{10}-helical, while the rest is α-helical, is 17.8 Å. This should be compared with 17.6 Å for an ideal α-helix and 14.2(±0.6) Å determined in our experiment. In light of the discussion below, we emphasize that this is true only if the 3_{10} segments are unbroken and adopt straight helical structure.

Second, the contracted radii of gyration are qualitatively consistent with the presence of π-helix. This structure, which has been observed in some computer simulations of short peptides,[31,32] is more compact (1.1 Å rise per residue) than the classical α-helix and its presence could lead to the measured radius of gyration (R_{gyr}) values. However, the π-helix has been detected only rarely in the experiment, and even then only in the context of large, complex folds that could stabilize its structure.[33] Moreover, only trace quantities of π-helix have been seen in our present simulations.

Third, the discrepancy cannot be explained by fraying at the ends. For instance, even if up to three residues at each end of the 37-mer are assumed to be frayed and random-walk-like with a persistence length of only one amino acid, the radius of gyration of the molecule remains roughly the same as for an ideal α-helix of the equivalent length. For longer persistence lengths of the frayed ends, the overall radius of gyration of the molecule actually increases compared to the ideal α-helix.

Finally, the low radii of gyration could potentially be explained if one assumes that the helicity of the peptides studied is actually not as high as CD measurements suggest, primarily due to the inaccuracy of the ellipticity of the coil baseline. The coil baseline (equation (3) in Materials and Methods) was determined originally by analyzing the CD spectra of an alanine heptamer,[28] and it is possible that this model for the unfolded state is not fully appropriate. This is particularly important because the exact contribution of the polyproline type II helix (PII)[34] to the coil state and its CD spectrum is not known. Shellwagen and co-workers have proposed a formula for estimating the ellipticity of the coil state as a function of the relative contribution of the PII helix:[35]

$$\Theta_{222} = 9580 f_{PII} - 5560 f_u \quad (1)$$

where f_{PII} and f_u are the fractions of PII and truly unordered conformations in the two-state equilibrium, respectively. Using this approach, we have examined what is the influence of varying the PPII content of the coil state on our estimates of the helicity of AK peptides, and the results are twofold. While for shorter peptides the effects are significant (the helix content of AK12 increases by 15% in absolute units if one assumes that the coil state is 100% PII), for peptides of 22 or more amino acid residues (those ones where the low R_{gyr} is most pronounced), the effects are negligible. For instance, changing the PII contribution of the coil state from 0% to 100%, changes the helix content of AK37 from 82% to 88%, which is within error from the value of 86% reported here. How then can we explain the low radii of gyration measured?

To fully account for the experimental data presented here and what is known about these peptides from previous work, it seems necessary to view the structures of these molecules as fluctuating broken rods. Namely, to reduce the radius of gyration of a mostly helical molecule compared to a straight ideal α-helix, the molecule needs to be bent somewhere in the middle. Nagai has derived an analytical theory that treats long α-helices as semi-broken rods connected by ideal random-walk-like segments.[25] This theory is based on the helix-coil formalism described by Zimm & Bragg,[27] combined with the standard statistics of ideal random walks, and it gives predictions for the radii of gyration of helical molecules in the helix-coil transition region. In the ZBN theory, the radius of gyration of a semi-broken helix depends on: (1) the number of residues; (2) the rise per residue of an ideal α-helix; (3) the Zimm–Bragg parameters s and σ; and (4) the persistence length of the random-walk-like segments connecting the α-helical stretches, a_0. The number of residues and the rise of the α-helix (1.5 Å) are known, while the Zimm–Bragg parameters can be calculated from the analogous Lifson–Roig parameters for similar peptides already present in the literature ($s=1.6$, $\sigma=0.001$).[8,36]

It is important to note that these Zimm–Bragg parameters reproduce the fractional helicity of our molecules very well (to rms deviation of 0.06) as shown in Figure 4(a). The only missing parameter then is the persistence length of the random walk segments. We have calculated the predicted values of the radii of gyration for several different values of the persistence length and the results are shown in Figure 4(b). Even for the shortest value of persistence length of 3.8 Å (equivalent to the length of one amino acid), the ZBN theory significantly overestimates the helical radii of gyration as compared with the experiment. For longer values of the persistence length, the disagreement gets markedly worse, while for even shorter values the improvement is only marginal. For instance, in the limit that a_0 goes to zero, the predicted radius of gyration of the 37-mer becomes 17.6 Å compared with 17.9 Å for $a_0=3.8$ Å. The ZBN theory was originally developed for very long homopolymeric broken helices with sizeable random-walk-like stretches.[26] It is possible that treating the short disordered parts of the AK helices as ideal random-flight chains (with no sterics or excluded volume) is the main

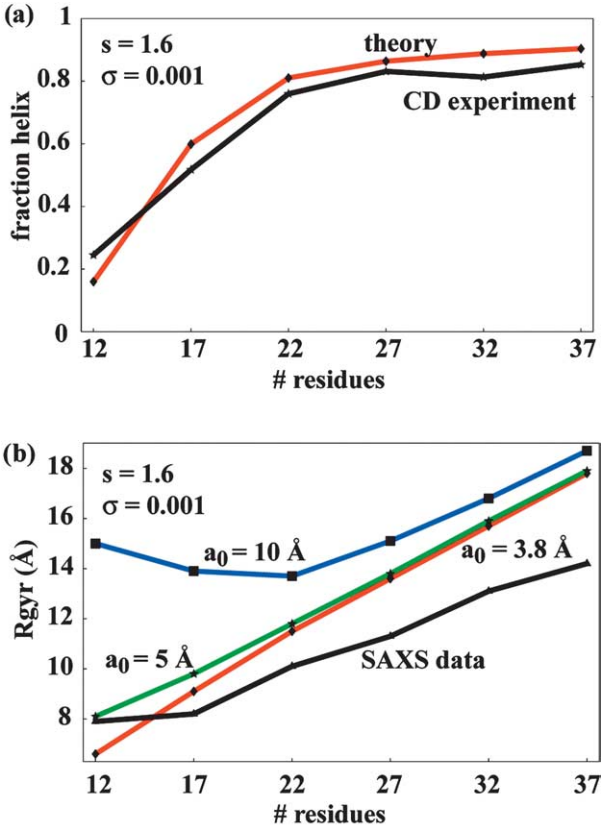

Figure 4. (a) Predictions of the ZBN theory *versus* the measured α-helical fraction. (b) Predictions of the ZBN theory *versus* the measured radii of gyration. We show the predictions of the theory for three different values of the persistence length, a_0, of the random walk segments connecting the α-helical stretches. The Zimm–Bragg parameters in (a) and (b) are the same ($s=1.6$, $\sigma=0.001$) and they correspond to the Lifson–Roig parameters of $w=1.7$, $v=0.036$ (adopted from the literature).

culprit for the failure of the theory to account for the measured radii of gyration. Finally, it should be mentioned that the ZBN theory does not take into account the ordered layer of surface water surrounding the peptide and contributing to the measured scattering. If this were taken into account, the ZBN predictions would increase even further by an estimated 0.5–1 Å.

The MD simulations described here also deviate from the experimental data. A potential reason for this, apart from the ever-present sampling issue, could be the fact that, when force-fields are parameterized and tested, significantly more emphasis is placed on reproducing the local geometries and energies, such as those of dihedral angles, than on long-range structure. However, small inaccuracies in local details can lead to large errors when it comes to global structure, especially in the case of flexible systems such as the ones studied here. Interestingly, the AMBER-99 force-field, which severely underestimates the helical fraction of the molecules studied here, agrees better with the experimental radii of gyration (Figures 2 and 3) compared to AMBER-99ϕ. On the other hand, AMBER-99ϕ, which was shown to match the Lifson–Roig parameters of a 21-residue helix better than other parameterizations of AMBER,[37] matches the helical content better while overestimating the experimental radii of gyration, especially for longer peptides. This suggests that AMBER-99ϕ works reasonably well in the range it was parameterized for (∼20 residues), but not much smaller or much longer. We hope that the experimental results presented here will provide a useful benchmark for further simulations and force-field development efforts.

The SAXS results presented here are most consistent with the picture of the α-helix as a semi-broken rod in which cylindrical segments are interspersed with more random-walk-like sections. Only by bending somewhere in the middle of the sequence, can the helix achieve the low radii of gyration measured. In order to illustrate this point, we have carried out excluded volume Monte Carlo simulations in which a given number of randomly chosen residues are placed at random in the β-sheet region of the Ramachandran plot while the rest is kept α-helical (see Materials and Methods). The number of α-helical residues is chosen such that it matches exactly the experimentally determined α-helical fraction for a given peptide. The structures generated in such a way are screened for steric clash, and, in the absence thereof, accepted. A similar procedure has been used recently by Fitzkee and Rose to study the long-range properties of unfolded proteins.[38] However, the ensembles generated in this way significantly overestimate the measured R_{gyr} values for all peptides (Figure 5(a)). The potential reasons for this discrepancy are multifaceted. First, the above procedure results in non-Boltzmann-weighted ensembles and it essentially assumes that all the configurations generated have the same free energy. Second, the effects of solvation on the stability of different structures are not taken into account. Finally, fixing the number of α-helical residues to the average level determined by CD precludes the possibility that this level fluctuates with time.

Despite its shortcomings, the Monte Carlo procedure described above allowed us to select and analyze the subset of structures from the generated ensembles that match the experimental R_{gyr} values (the α-helical level from CD is matched by default). For this purpose, all the structures that agree with the average experimental R_{gyr} to within the experimental error were chosen from the Monte Carlo ensembles. Nine representative structures chosen in this way for AK37 are shown in Figure 5(b) and they are meant to illustrate how broken and compact the peptide would have to be to match the experimental R_{gyr} and helical fraction. We do not suggest that any one structure shown in Figure 5(b) represents alone the geometry of the AK peptides in solution: it is likely that these molecules adopt flexible helical structures whose complete

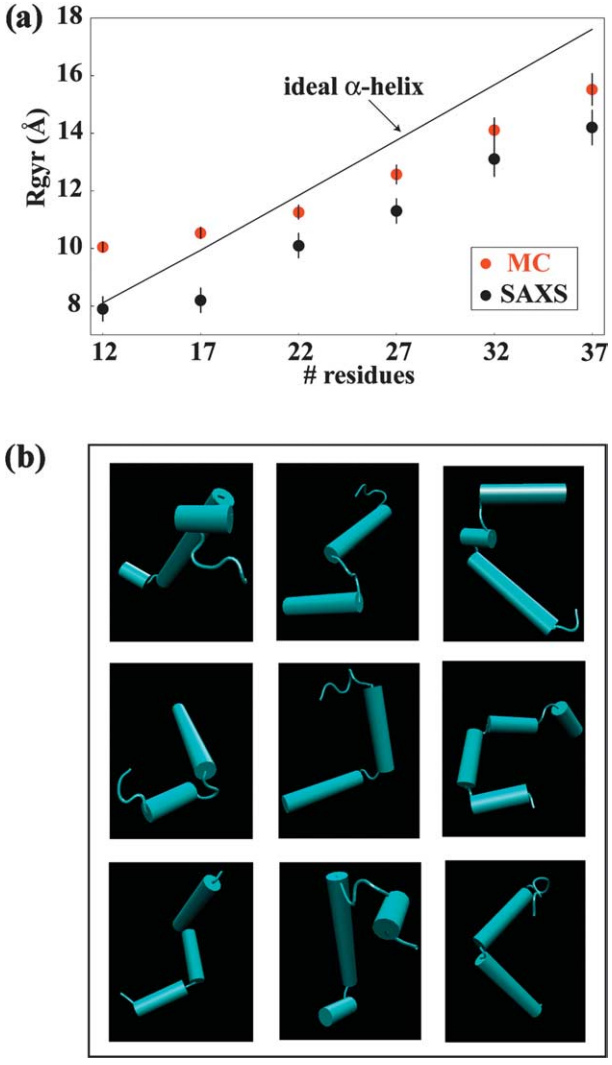

Figure 5. (a) Ensemble average radii of gyration from the excluded volume Monte Carlo simulations (MC) compared with the experimentally determined values (SAXS). The radii of gyration from simulation were determined based on theoretically calculated Guinier plots. (b) Representative structures from the excluded volume Monte Carlo simulations for AK37 which agree exactly with the experimentally determined α-helical fraction and agree to within experimental error (±0.6 Å) with the experimentally determined radius of gyration. These structures are meant solely to illustrate how broken helices would need to be in order to match the experimental α-helical fraction and radius of gyration.

description can be done only on the level of ensemble. The points of bending likely fluctuate and are not stabilized in any way. What is clear is that, to match the SAXS data, the members of the ensemble would need to be broken, on average, once or twice somewhere in the middle of the sequence, just like the structures in Figure 5(b). It is not hard to see how this picture could be reconciled with the structural characterization of these molecules coming from mostly sequence-local observables such as CD spectra or NMR nuclear Overhauser effects (NOEs) and ^3J-coupling. It is somewhat harder to see how this picture fits with the results coming from the hydrogen-exchange experiments, which suggest larger fluctuations at the ends of the α-helical rods than in the middle.[39] We hope future research will bring answers to this and related questions.

There is one puzzling aspect about the model ensembles of structures with experimental R_{gyr} values and helical fractions from our Monte Carlo simulations such as the structures shown in Figure 5(b). Namely, if one calculates complete scattering Kratky profiles based on these ensembles, one sees significant deviation from the experimentally determined Kratky profiles (Figure 6). While these structures match the experimental radii of gyration and helical fractions, they are clearly not completely representative of the real ensembles. What is particularly intriguing is the fact that the experimental ensembles exhibit approximately linearly rising Kratky plots (Figure 6). This, specifically, is a hallmark of random-walk-like chains,[21] and it provides additional evidence for the flexible and fluid nature of these structures. The fact that the simulated ensembles with experimental R_{gyr} values and helical fractions do not reproduce the linear Kratky plots could be due to incomplete sampling. Another possibility is that, in fact, the helical stretches, which in the simulation are kept fixed in the canonical α-helical configuration, actually also exhibit significant fluctuations around the ideal values, contributing in such a way to the linear Kratky plots observed.

Recently, we have shown how the inter-residue distances in the α-helix over short stretches correspond closely to the average distances between the residues in an ideal random-flight chain with link length of 3.8 Å.[40] On the basis of this observation, we have speculated that, under certain circumstances, the geometry of a highly averaged random-flight chain might be interpreted as an α-helix. The present SAXS results argue against this hypothesis, as the dependence of the radius of gyration on sequence length does not follow $N^{1/2}$ dependence expected for ideal random-flight chains. Indeed, the dependence that we observe (Figure 3) falls between the linear relationship expected for ideal α-helices and the $N^{1/2}$ relationship characterizing the ideal random-flight chains. For example, while for AK37 the random-flight chain model (with link length of 3.8 Å) predicts the radius of gyration of 9.4 Å and the ideal α-helix value is 17.6 Å, the measured value is 14.2(±0.6) Å. However, it should be pointed out that the random-flight chain value of 9.4 Å quoted here does not include the contribution of the ordered water layer and is, in reality, likely to be larger by an estimated 0.5–1 Å. Finally, an intriguing counterpoint is provided by the fact that the Kratky plots observed here (Figure 6) are actually consistent with random-walk-like behavior. We hope that future research will bring an answer to this seeming puzzle.

The canonical view of breaks in the α-helix is that

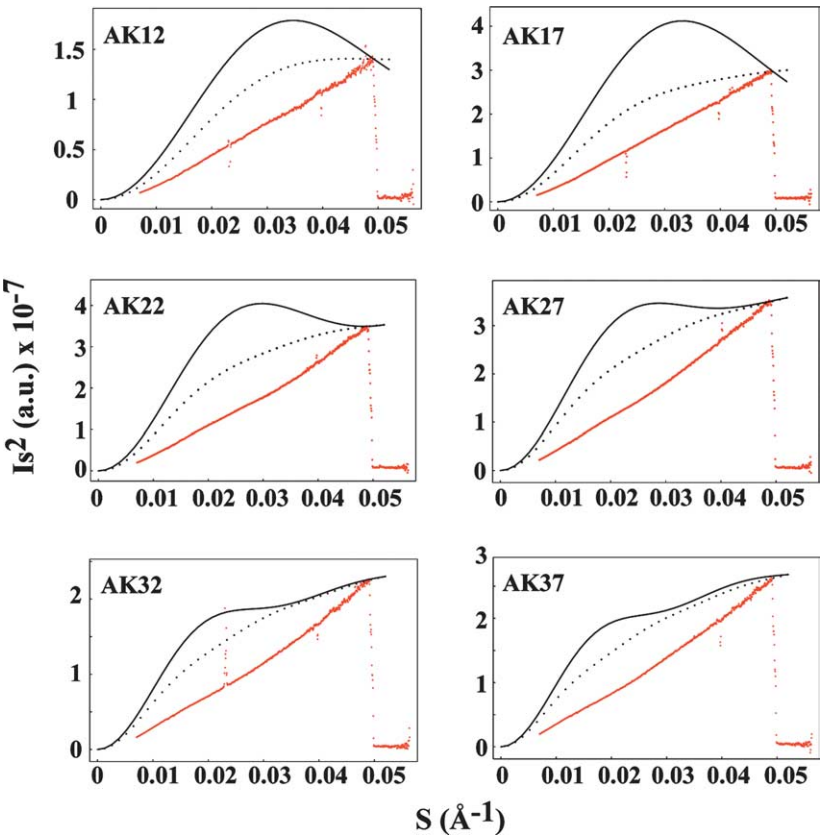

Figure 6. Experimentally determined Kratky plots (in red) together with the Kratky plots calculated for ideal α-helices of equivalent length (dotted lines) and the average Kratky plots (continuous black lines) based on all of the structures from the excluded volume Monte Carlo simulations that agree exactly with the experimentally determined helical fraction and agree with the experimentally determined radii of gyration to within experimental error (±0.6 Å for AK32 and AK37 and ±0.4 Å for the rest). The theoretical Kratky plots were scaled such that they match the experimental Kratky plot at its maximum.

these are very unlikely, since three consecutive peptide H-bonds must be broken simultaneously. This was discussed for the first time by Gibbs & DiMarzio,[41] following the original work by Zimm & Bragg. On the other hand, these papers were written before the data on the strength of peptide hydrogen bonds in water was available. The ΔG for the propagation reaction of the alanine peptide α-helix is only −0.27 kcal/mol at 0 °C in water.[42] This is significantly less than thermal energy (kT), and breaks in the helix may not be so surprising after all. At this point, one should mention the striking result found by A. Wada for the absence of helix breaks in non-polar solvents.[43] He measured dipole moments of an uncharged peptide helix (poly-benzyl-L-glutamate) in two non-polar solvents as a function of chain length, and found exact agreement with the predicted values for a rigid helix.[43] This can be explained by the fact that the solvents used in Wada's studies were highly helicogenic, i.e. the ΔG for helix propagation in these solvents is significantly more negative than in water.[26,42,43]

To conclude, our study demonstrates the utility of the SAXS technique to study the long-range structure of even the shortest peptides in aqueous solvent. We have shown how certain features of flexible molecules, such as radius of gyration, may differ from what could be deduced from the results of local, short-range measurements such as NMR or CD. Finally, we have shown that simulating the behavior of flexible peptides, even when predominantly α-helical, may present significant challenges for modern-day atomistic force-fields.

Materials and Methods

Peptide synthesis

The peptides (sequence composition Ace-$(AAKAA)_n$-GY-NH_2, with n=2–7) were synthesized at the Stanford University PAN facility using solid-phase synthesis using standard Fmoc synthetic chemistry. The final products of the synthesis and their purity were analyzed by reverse-phase HPLC and MALDI-TOF mass spectrometry. In the text, the peptides are referred to as AKxx, where xx corresponds to the number of amino acid residues in a given molecule.

Circular dichroism

Peptide stock solutions for CD measurements were made at 0.5–1.0 mM in water. Concentrations of stock solutions were determined by measuring the tyrosine absorbance at 275 nm in 6.5 M guanidine hydrochloride, 20 mM potassium phosphate (pH 6.5) using $\varepsilon_{275\,nm}$ = 1450 M^{-1} cm^{-1}.[44] CD measurements were made on an Aviv 60DS spectrometer equipped with a Hewlett-Packard 89100A temperature-control unit using quartz cuvettes with 1.0 mm pathlengths. Ellipticity was calibrated with (+)-10-camphorsulfonic acid. The spectra were recorded at 10 °C at pH 7.0 in 1 M sodium chloride, 3 mM potassium phosphate. Data were collected in 0.5 nm increments from 260 nm to 200 nm with triple averaging at each point.

To estimate fractional helicity of the molecules, the ellipticity at 222 nm was assumed to be linearly related to mean helix content, f_H, which can be calculated from the Lifson–Roig-based helix-coil model (see below).[45] The conversion of the ellipticity at 222 nm to f_H depends on the knowledge of the baseline ellipticities of both the random coil, Θ_C, and the complete helix, Θ_H:

$$f_H = (\Theta_{222} - \Theta_C)/(\Theta_H - \Theta_C) \quad (2)$$

The values of Θ_C and Θ_H depend on temperature and are given by the following expressions:

$$\Theta_C = 2220 - 53T \quad (3)$$

$$\Theta_H = (-44000 + 250T)(1 - 3/N_r) \quad (4)$$

where T is the temperature in °C, and N_r is the chain length in residues. These equations were given by Luo & Baldwin,[28] and they were determined by fitting the Lifson–Roig model[45] for the helix-coil transition to the ellipticity as a function of temperature and concentration of trifluoroethanol.

Small-angle X-ray scattering

SAXS experiments were carried out on the BESSRC-CAT beamline at the Advanced Photon Source, Chicago, IL. The molecules were studied at 10(±1) °C using a flow-cell setup. The concentration of the molecules was (for duplicated measurement, two values are given): AK12, 14 mg/ml, 10 mg/ml; AK17, 25 mg/ml, 22 mg/ml; AK22, 20 mg/ml, 22 mg/ml; AK27, 20 mg/ml, 20 mg/ml; AK32, 15 mg/ml; and AK37, 15 mg/ml. The molecules were dissolved in 3 mM phosphate buffer and 1 M NaCl, in the presence of 5 mM radical scavenger *N*-tert-butyl-α-(4-pyridyl)nitrone *N*′-oxide, at the final pH of 3.5±0.5. The radii of gyration were determined by Guinier analysis. The zero-angle forward scattering was monitored and no indication of aggregation was observed (data not shown).

The reported confidence intervals for the experimentally determined R_{gyr} values represent estimated sample standard deviations for the fitted parameters based on a single experiment. When duplicate measurements have been carried out, this value has been adjusted accordingly.

Calculating theoretical scattering profiles

The effective R_{gyr} values measured by SAXS differ from the ideal, geometry-based values due to the surface layer of ordered water surrounding the molecule. The theoretical expected values for the radii of gyration in solution of ideal α-helices and the simulated structures (Figures 3 and 5) were calculated using Guinier analysis on the predicted scattering profiles generated using Crysol software,[46] with the following input parameters. The maximum order of harmonics was set to 15 while the order of the Fibonacci grid was taken as 17. The contrast of the solvation shell was taken as 3×10^{-2}, while the structural parameters for the solvation shell were taken at their default Crysol values. The errors were calculated from the standard errors of the linear Guinier fits.

MD simulations

The simulations were carried out on the Folding@Home distributed computing cluster[47–49] using GROMACS simulation package[50,51] and two variants of the AMBER force-field: the original AMBER-99 force-field[52] and a variant of this force-field AMBER-99φ.[37] In this modification, the torsion potential for the φ dihedral angle is set to zero. It was demonstrated recently that this alteration leads to better agreement with experiment when it comes to folding kinetics and Lifson–Roig parameters.[37] The basic side-chains were protonated by assuming low pH. Sodium and chloride ions were added at experimental concentrations (1 M). More chloride atoms were added to exactly balance the positive charges on the peptides, thus assuring charge neutrality. Simulations were performed at constant temperature and pressure (283 K, 101.3 kPa), using the explicit TIP3P water model[53] and periodic boundary conditions in a triclinic box. The starting dimensions of the simulation boxes were (in Å): AK12, 27.4, 39.4, 43.3; AK17, 37.1, 25.5, 33.8; AK22, 37.1, 25.9, 43.7; AK27, 42.4, 30.5, 45.9; AK32, 52.0, 27.5, 50.9; AK37, 51.7, 28.6, 61.3; and AK42, 62.5, 33.7, 61.4. The temperature and the pressure were controlled by coupling the system to external heat-baths with a relaxation time of 0.5 ps.[54] The electrostatics were treated by using the reaction-field method with a cutoff of 10 Å, while for Lennard–Jones interactions, 10 Å cutoffs with 8 Å tapers were used. Non-bonded pair lists were updated every ten steps of molecular dynamics, and the integration step size was 2 fs in all simulations. The bonds involving hydrogen atoms were constrained by using the LINCS algorithm. The sizes of the simulation systems were: 2046 atoms, 621 water molecules (AK12); 3271 atoms, 1043 water molecules (AK17); 4303 atoms, 1322 water molecules (AK22); 6121 atoms, 1901 water molecules (AK27); 7562 atoms, 2355 water molecules (AK32); 9484 atoms, 2968 water molecules (AK37). Starting from the polyproline II conformation (all $\phi = -78°$, $\psi = 149°$) for AK12, or from the ideal α-helical conformations (all $\phi = -62°$, $\psi = -41°$) for all other peptides, a total of 50 independent, 50 ns long molecular dynamics simulations was performed for each peptide. All of the analysis presented here was carried out on composite ensembles containing structures from all of the simulated trajectories between 30 ns and 50 ns sampled every 2 ns (a total of 500 structures for each of the peptides). The simulations were run for an aggregate time of 30 μs.

Excluded volume Monte Carlo simulations

In the excluded volume Monte Carlo simulations, a given number of randomly chosen residues were placed at random in the β-sheet region of the Ramachandran plot (defined as $-180° \le \phi \le -50°$ and $30° \le \psi \le 180°$), while the rest were kept α-helical ($\phi = -62°$ and $\psi = -41°$). The number of α-helical residues that were retained was chosen such that it exactly matches the experimentally determined helical fraction for a given peptide. In all of the structures, the side-chains of tyrosine and lysine were placed in their most frequent rotameric state.[55] The structures generated in such a way were screened for steric clash, and, in the absence thereof, accepted. This procedure assumes that all of the accepted structures have essentially the same free energy and that the steric repulsion is the only relevant energetic contribution.[38] The generated ensembles, described in Figure 5(a), contained anywhere between 300 and 3000 structures. From these, we have chosen subsets of structures that match the average experimental R_{gyr} to within the experimental error (Figures 5(b) and 6). These small ensembles contained 52, 21, 50, 39, 73 and 58 structures for AK12 through AK37, respectively, and were used for

the calculation of the theoretical Kratky plots shown in Figure 6.

Secondary structure in simulation

The secondary structure assignment of the simulated structures (Figure 1(b)) was based on the DSSP classification.[56] A given residue was defined as helical if it fell in the G-H-I category of DSSP. This includes the α-helix, 3_{10}-helix and π-helix. The fraction of the residues defined as helical that were α-helical was: 0.51, 0.52, 0.62, 0.79, 0.87 and 0.92 for AMBER-99, or 0.86, 0.98, 0.97, 0.98, 0.98 and 0.99 for AMBER-99φ for AK12 through AK37, respectively. The rest of the helical residues were predominantly 3_{10}-helical. Appreciable amounts of π-helix were observed only in AK12 with AMBER-99 (4% of helical residues).

Acknowledgements

We especially thank the thousands of Folding@Home contributors, without whom this work would not be possible. A complete list of contributors can be found at http://folding.stanford.edu. We thank Soenke Seifert and staff at APS for help with data collection. We thank Buzz Baldwin and Eric J. Sorin for useful discussions, and the anonymous referee for suggestions on how to improve the manuscript. This work was supported by grants from the NIH (R01GM62868), a postdoctoral fellowship by EMBO (to B.Z.) and gifts from Google.

References

1. Creighton, T. E. (1993). *Proteins: Structure and Molecular Properties*, W.H. Freeman, New York.
2. Marqusee, S. & Baldwin, R. L. (1987). Helix stabilization by Glu–Lys+ salt bridges in short peptides of *de novo* design. *Proc. Natl Acad. Sci. USA*, **84**, 8898–8902.
3. Marqusee, S., Robbins, V. H. & Baldwin, R. L. (1989). Unusually stable helix formation in short alanine-based peptides. *Proc. Natl Acad. Sci. USA*, **86**, 5286–5290.
4. Rohl, C. A. & Baldwin, R. L. (1998). Deciphering rules of helix stability in peptides. *Methods Enzymol.* **295**, 1–26.
5. Creamer, T. P. & Rose, G. D. (1992). Side-chain entropy opposes alpha-helix formation but rationalizes experimentally determined helix-forming propensities. *Proc. Natl Acad. Sci. USA*, **89**, 5937–5941.
6. Rohl, C. A., Fiori, W. & Baldwin, R. L. (1999). Alanine is helix-stabilizing in both template-nucleated and standard peptide helices. *Proc. Natl Acad. Sci. USA*, **96**, 3682–3687.
7. Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T. M. & Baldwin, R. L. (1990). Relative helix-forming tendencies of nonpolar amino acids. *Nature*, **344**, 268–270.
8. Chakrabartty, A., Kortemme, T. & Baldwin, R. L. (1994). Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci.* **3**, 843–852.
9. Creamer, T. P. & Rose, G. D. (1994). Alpha-helix-forming propensities in peptides and proteins. *Proteins: Struct. Funct. Genet.* **19**, 85–97.
10. Baldwin, R. L. & Rose, G. D. (1999). Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem. Sci.* **24**, 26–33.
11. Eaton, W. A., Munoz, V., Hagen, S. J., Jas, G. S., Lapidus, L. J., Henry, E. R. & Hofrichter, J. (2000). Fast kinetics and mechanisms in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 327–359.
12. Daggett, V., Kollman, P. A. & Kuntz, I. D. (1991). A molecular dynamics simulation of polyalanine: an analysis of equilibrium motions and helix-coil transitions. *Biopolymers*, **31**, 1115–1134.
13. Daggett, V. & Levitt, M. (1992). Molecular dynamics simulations of helix denaturation. *J. Mol. Biol.* **223**, 1121–1138.
14. Huo, S. & Straub, J. E. (1999). Direct computation of long time processes in peptides and proteins: reaction path study of the coil-to-helix transition in polyalanine. *Proteins: Struct. Funct. Genet.* **36**, 249–261.
15. Garcia, A. E. & Sanbonmatsu, K. Y. (2002). Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl Acad. Sci. USA*, **99**, 2782–2787.
16. Chowdhury, S., Zhang, W., Wu, C., Xiong, G. & Duan, Y. (2003). Breaking non-native hydrophobic clusters is the rate-limiting step in the folding of an alanine-based peptide. *Biopolymers*, **68**, 63–75.
17. Ghosh, T., Garde, S. & Garcia, A. E. (2003). Role of backbone hydration and salt-bridge formation in stability of alpha-helix in solution. *Biophys. J.* **85**, 3187–3193.
18. Gnanakaran, S., Nymeyer, H., Portman, J., Sanbonmatsu, K. Y. & Garcia, A. E. (2003). Peptide folding simulations. *Curr. Opin. Struct. Biol.* **13**, 168–174.
19. Nymeyer, H. & Garcia, A. E. (2003). Simulation of the folding equilibrium of alpha-helical peptides: a comparison of the generalized Born approximation with explicit solvent. *Proc. Natl Acad. Sci. USA*, **100**, 13934–13939.
20. van Giessen, A. E. & Straub, J. E. (2005). Monte Carlo simulations of polyalanine using a reduced model and statistics-based interaction potentials. *J. Chem. Phys.* **122**, 024904.
21. Doniach, S. (2001). Changes in biomolecular conformation seen by small angle X-ray scattering. *Chem. Rev.* **101**, 1763–1778.
22. Rabenstein, M. D. & Shin, Y. K. (1995). Determination of the distance between two spin labels attached to a macromolecule. *Proc. Natl Acad. Sci. USA*, **92**, 8239–8243.
23. Ratner, V., Sinev, M. & Haas, E. (2000). Determination of intramolecular distance distribution during protein folding on the millisecond timescale. *J. Mol. Biol.* **299**, 1363–1371.
24. Jares-Erijman, E. A. & Jovin, T. M. (2003). FRET imaging. *Nature Biotechnol.* **21**, 1387–1395.
25. Nagai, K. (1961). Dimensional change of polypeptide molecules in the helix-coil transition region II. *J. Chem. Phys.* **34**, 887–904.
26. Teramoto, A. & Fujita, H. (1975). Macroconformation of polymers. *Advan. Polymer Sci.* **18**, 65–149.
27. Zimm, B. H. & Bragg, J. K. (1959). Theory of the phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.* **31**, 526–535.
28. Luo, P. & Baldwin, R. L. (1997). Mechanism of helix induction by trifluoroethanol: a framework for

extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry*, **36**, 8413–8421.
29. Miick, S. M., Martinez, G. V., Fiori, W. R., Todd, A. P. & Millhauser, G. L. (1992). Short alanine-based peptides may form 3(10)-helices and not alpha-helices in aqueous solution. *Nature*, **359**, 653–655.
30. McNulty, J. C., Silapie, J. L., Carnevali, M., Farrar, C. T., Griffin, R. G., Formaggio, F. *et al*. (2000). Electron spin resonance of TOAC labeled peptides: folding transitions and high frequency spectroscopy. *Biopolymers*, **55**, 479–485.
31. Shirley, W. A. & Brooks, C. L., 3rd (1997). Curious structure in "canonical" alanine-based peptides. *Proteins: Struct. Funct. Genet.* **28**, 59–71.
32. Lee, K. H., Benson, D. R. & Kuczera, K. (2000). Transitions from alpha to pi helix observed in molecular dynamics simulations of synthetic peptides. *Biochemistry*, **39**, 13737–13747.
33. Weaver, T. M. (2000). The pi-helix translates structure into function. *Protein Sci.* **9**, 201–206.
34. Shi, Z., Woody, R. W. & Kallenbach, N. R. (2002). Is polyproline II a major backbone conformation in unfolded proteins? *Advan. Protein Chem.* **62**, 163–240.
35. Park, S. H., Shalongo, W. & Stellwagen, E. (1997). The role of PII conformations in the calculation of peptide fractional helix content. *Protein Sci.* **6**, 1694–1700.
36. Doig, A. J. (2002). Recent advances in helix-coil theory. *Biophys. Chem.* **101–102**, 281–293.
37. Sorin, E. J. & Pande, V. S. (2005). Exploring the helix-coil transition *via* all-atom equilibrium ensemble simulations. *Biophys. J.* **88**, 2472–2493.
38. Fitzkee, N. C. & Rose, G. D. (2004). Reassessing random-coil statistics in unfolded proteins. *Proc. Natl Acad. Sci. USA*, **101**, 12497–12502.
39. Rohl, C. A. & Baldwin, R. L. (1994). Exchange kinetics of individual amide protons in ^{15}N-labeled helical peptides measured by isotope-edited NMR. *Biochemistry*, **33**, 7760–7767.
40. Zagrovic, B. & Pande, V. S. (2003). Structural correspondence between the alpha-helix and the random-flight chain resolves how unfolded proteins can have native-like properties. *Nature Struct. Biol.* **10**, 955–961.
41. Gibbs, J. H. & DiMarzio, E. A. (1958). Theory of helix-coil transitions in polypeptides. *J. Chem. Phys.* **28**, 1247–1248.
42. Rohl, C. A., Chakrabartty, A. & Baldwin, R. L. (1996). Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci.* **5**, 2623–2637.
43. Wada, A. (1976). The alpha-helix as an electric macrodipole. *Advan. Biophys.* **9**, 1–63.
44. Brandts, J. F. & Kaplan, L. J. (1973). Derivative sspectroscopy applied to tyrosyl chromophores. Studies on ribonuclease, lima bean inhibitors, insulin, and pancreatic trypsin inhibitor. *Biochemistry*, **12**, 2011–2024.
45. Lifson, S. & Roig, A. (1961). On the theory of helix-coil transition in polypeptides. *J. Chem. Phys.* **34**, 1963–1974.
46. Svergun, D. I., Barberato, C. & Koch, M. H. J. (1995). CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallog.* **28**, 768–773.
47. Zagrovic, B., Sorin, E. J. & Pande, V. (2001). Beta-hairpin folding simulations in atomistic detail using an implicit solvent model. *J. Mol. Biol.* **313**, 151–169.
48. Snow, C. D., Nguyen, H., Pande, V. S. & Gruebele, M. (2002). Absolute comparison of simulated and experimental protein-folding dynamics. *Nature*, **420**, 102–106.
49. Pande, V. S., Baker, I., Chapman, J., Elmer, S. P., Khaliq, S., Larson, S. M. *et al*. (2003). Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. *Biopolymers*, **68**, 91–109.
50. Berendsen, H. J. C., van der Spoel, D. & Drunen, R. (1995). GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56.
51. Lindahl, E., Hess, B. & van der Spoel, D. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Mod.* **7**, 306–317.
52. Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W. *et al*. (2003). A point-charge force-field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012.
53. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
54. Berendsen, H. J. C., Postma, J. P., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690.
55. Dunbrack, R. L., Jr & Cohen, F. E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **6**, 1661–1681.
56. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Edited by C. R. Matthews

(Received 10 June 2005; received in revised form 18 August 2005; accepted 23 August 2005)
Available online 8 September 2005