

Secondary Structure Assignment of Amyloid- β Peptide Using Chemical Shifts

Geoffrey P. F. Wood* and Ursula Rothlisberger*

Laboratory of Computational Chemistry and Biochemistry, BCH 4107 EPF Lausanne, CH-1015 Lausanne, Switzerland

 Supporting Information

ABSTRACT: The distinct conformational dependence of chemical shifts caused by α -helices and β -sheets renders NMR chemical shift analysis a powerful tool for the structural determination of proteins. However, the time scale of NMR experiments can make a secondary structure assignment of highly flexible peptides or proteins, which may be converting between conformational substates, problematic. For instance the amyloid- β monomer, according to NMR chemical shifts, adopts a predominately random coil structure in aqueous solution (with <3% α -helical content). Molecular dynamics simulations, on the other hand, suggest that α -helical content can be significant (10–25%). In this paper, we explore the possible reasons for this discrepancy and show that the different results from experiments and theory are not necessarily mutually exclusive but may reflect a general problem of secondary structure assignment of conformationally flexible biomolecules.

1. INTRODUCTION

One of the hallmarks of Alzheimer's disease is the deposition of fibrils containing the amyloid- β peptide ($A\beta$) in the extracellular space of the limbic and association cortices (for recent reviews see refs 1 and 2). $A\beta$ is a 39 to 43 amino acid protein derived from the normal metabolism of the transmembrane amyloid precursor protein (APP).^{3–5} The 43 amino acid fragment is characterized by the sequence: DAEFRHDSGYEVHH-QKLVFFAEDVGSNKGAIIGLMVGGVVIAT. The remaining alloforms are derived by deleting residues from the C-terminal end.

The most abundant form is the 40 amino acid alloform denoted $A\beta(1–40)$.⁶ However, the most toxic form has been identified as the 42 amino acid peptide $A\beta(1–42)$.⁶ There is growing evidence that the toxic agents in Alzheimer's disease are small soluble oligomeric structures of $A\beta$; in this model, the fibril is a symptom of the disease rather than a causative agent itself.^{7–9} In order to understand the pathology of Alzheimer's disease and to be in a position to effectively develop drugs, a detailed atomistic knowledge of the conformational transitions connecting the native form (transmembrane and APP-incorporated) to the fibril is desirable.

Under normal physiological conditions, $A\beta$ aggregates quickly, rendering standard experimental tools of biochemistry ineffective for characterizing the intermediate species. As a consequence, the bulk of our structural knowledge refers to the two end-points of the conformational transition. Although no direct structure of APP-incorporated $A\beta$ is available, a number of experimental model studies,^{10–14} which simulate transmembrane conditions, have shown that $A\beta$ exists primarily as two α -helices connected through a kink spanning residues Gly25–Asn27. A longer helix contains residues 10–25, and a shorter one contains residues 27–35 (see Figure 1). The N-terminal end (residues 1–9) adopts variable conformations depending on the exact experimental conditions (i.e., pH, temperature, and stabilizing agents) but is described mainly as "unstructured".

The fibril containing $A\beta(1–40)$ monomers has been characterized by solid-state NMR¹⁵ and is present as a pair of

antiparallel β -sheets spanning residues 10 to 40 with a turn located between Glu22 and Lys28. Again, the N-terminal end (residues 1–9) is unstructured (not shown in Figure 1). In addition, an important turn-stabilizing salt bridge between Asp23 and Lys28 has also been identified.

The $A\beta$ monomer in aqueous solution at a pH of ~7 can be considered as the first stage of the transition from membrane bound $A\beta$ to the fibril, making a structural analysis of this situation highly desirable. However, the rapid *in vitro* aggregation of $A\beta$ makes a direct experimental evaluation of the monomeric structure difficult to obtain. Despite this difficulty, a number of CD spectra and solution NMR studies of the full-length species, i.e., $A\beta(1–40)$ and $A\beta(1–42)$, have been carried out.^{16–19} In addition, the solution structure of the fragment containing residues 10–35 has been characterized by NMR.^{20,21} This fragment has a higher solubility in aqueous solution and has been suggested as a good structural model for the full-length peptides. The consensus of these studies is that the $A\beta$ monomer adopts predominately a random coil structure in solution with little or no well-defined secondary structure motifs. One particular spectrum based on CD suggests, for $A\beta(1–42)$, 79% random coil, 13% β -turn/ β -strand, and 3% α -helix.¹⁸ An NMR study of $A\beta(1–42)$,¹⁶ which analyzed the secondary structure using chemical shifts, indicated that the α -helical content is 0% (derived from $C\alpha$ chemical shifts) and the β -strand content is 20% (derived from $H\alpha$ chemical shifts), in good agreement with the CD spectrum. This study employed the use of the chemical shift index (CSI) method,²² which is an important tool for identifying secondary structure elements of large biomolecules. The method is based on the observation made in the early 1980s²³ that $C\alpha$ protons experience a relative upfield shift when the residue in question is incorporated into an α -helix and a relative downfield shift when incorporated into a β -sheet.

Received: March 4, 2011

Published: April 22, 2011

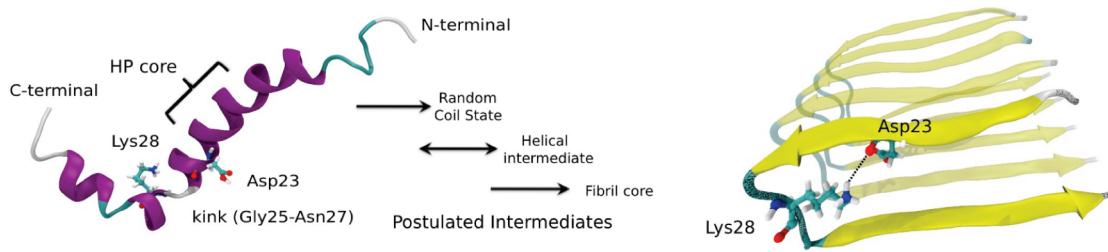


Figure 1. Key features of the conformations associated with transmembrane amyloid- β and when incorporated in a fibril.

Since then, similar chemical shift-to-structure correlations have been reported for other nuclear centers, including $C\alpha$ s, $C\beta$ s, and C's, (for reviews, see refs 24 and 25); however, the correlation may be reversed. The identification process requires the determination of a set of "random coil" (rc) values that lie in the middle of the α -helix and β -sheet extremes. The reference values give the conformational dependence on the chemical shift as

$$\Delta\delta_s^i = \delta_{\text{obs}}^i - \delta_{\text{rc}}^i \quad (1)$$

Secondary structure is then assigned through a set of well-defined rules.²⁵

Molecular dynamics (MD) simulations do not suffer from the same difficulties as experimental studies in that the solution structure of the monomer can be probed directly. However, caution needs to be exercised with MD simulations of $A\beta$ because it has been shown that kinetic trapping may occur,²⁶ therefore, long time scales and/or enhanced sampling is required in order to effectively sample the conformational space. From the plethora of MD studies in the literature, only a handful satisfy either or both of these criteria.^{26–30} The general consensus of these studies is that $A\beta$ has a predominately random coil and turn structure, in agreement with experimental results, but the α -helical content differs significantly from the experimental findings. Depending on the MD protocol used, the total helical content ($\alpha + 3_{10} + \pi$) is at least 10% (using the AMBER99SB³¹ force field) and can be as high as 25% (using the CHARMM19³² force field). However, a number of force fields have been shown to overstabilize helical content by up to a factor of 2.³¹ Taking the possible overstabilization into account, a helical content of no less than 10–12% is predicted by theory, which is at least 4 times greater than the experimental findings.

In the present work, we investigate this discrepancy by carrying out enhanced sampling molecular dynamics simulations and evaluating the secondary structure content of $A\beta(1–42)$ via the conformational dependencies of chemical shifts. Our results suggest that because both downfield and upfield shifts may occur, conformationally flexible biomolecules may on average give NMR signals that can be interpreted as random coil or turn elements. However, in certain extreme cases, the average chemical shift of a particular nuclear center may be sampling from both helical and β structures with little or no "random coil" structure present, even though the average NMR signal suggests that a random coil structure dominates. In other less extreme cases, the average chemical shift is sampling from any of a number different motifs during the course of the simulation, but one single averaged chemical shift does not capture this complexity. We then show how the helical content depends on its definition, and for a flexible biomolecule, a single "static" content is not the appropriate measure for helicity.

The idea that a single dominant configuration may lead to inconsistencies in the interpretation of spectroscopic data for

flexible biomolecules is not new. In particular, van Gunsteren and co-workers have shown how the correct interpretation of NOE data and J -coupling constants requires the incorporation of additional conformations along with the dominant structure.^{33,34} Further examples where a conformational distribution is important for the correct interpretation of spectroscopic quantities have been reviewed elsewhere.³⁵

2. COMPUTATIONAL DETAILS

Replica exchange molecular dynamics (REMD)³⁶ has emerged as a method that explores a significantly larger portion of phase space than what a single-temperature simulation of the same (aggregate) length can achieve.³⁷ However, a major drawback of conventional (temperature) replica exchange molecular dynamics is that the number of replicas needed to span a given temperature range increases as the square root of the number of degrees of freedom in the system. Numerous techniques have been developed in order to deal with this problem, which include optimizing the allocation of replicas,³⁸ using a perturbation on the Hamiltonian (rather than temperature) when defining each replica (H-REMD),^{39–41} and coupling replicas to reservoirs of pregenerated ensembles.^{42,43}

Another novel approach is to retain the conventional (temperature) REMD protocol but use a modified Hamiltonian when attempting exchanges. In the formulation by Simmerling and co-workers,^{44,45} the exchange attempts are made with a Hamiltonian that reduces the total number of degrees of freedom of the system. Instead of using the entire periodic box for exchange attempts, a predetermined subset of explicit water molecules is retained around the solute of interest; this subsystem is then immersed in a dielectric continuum. We emphasize that only the exchange attempts use this hybrid Hamiltonian; the dynamics are propagated with a full periodic box of explicit waters. With a reduced number of degrees of freedom for the exchange attempts, the "hybrid-REMD" approach uses fewer replicas for a given temperature range. Promising results have been published showing that this scheme can reproduce properties derived from full REMD simulations.⁴⁴

There is no recipe for choosing the number of explicit water molecules when swaps are carried out. It has been recommended that the first solvation shell be used because in some instances increasing the number of waters to include the second solvation shell gave less satisfactory results. This indicates that convergence with respect to the number of explicit water molecules may not be a monotonically decreasing function. In light of this, we chose to saturate the system with as many explicit water molecules as we could computationally afford and thus settled on 5000. This number may be compared with approximately 350 and 600 water molecules in the first and second solvation shells,

respectively, of $\text{A}\beta(1-42)$. To determine the temperature spacing between replicas, we chose initial temperatures using an online predicting algorithm.^{46,47} After propagating the dynamics for approximately 2 ns, energetic data from the exchange attempts using the hybrid Hamiltonian were plotted against the temperature. An exponential curve was then fitted to the data to give an expression for the temperature as a function of the energy. The resultant curves were used to solve iteratively the Monte Carlo swap condition for a given acceptance ratio (p), i.e.,

$$p = \exp[(E_2 - E_1)(\beta_2 - \beta_1)] \quad (2)$$

and thus obtain an approximation to the ideal temperature spacing.

The final numbers of replicas, theoretical swapping probability, and temperature distributions were set to the following: 16 replicas, 8.5% probability, with temperatures set to 282.0, 291.0, 300.0, 310.0, 320.0, 331.0, 342.0, 354.0, 366.0, 380.0, 393.0, 408.0, 424.0, 440.0, 457.0, and 476.0 K. In addition to these parameters, the simulations were carried out using the AMBER 9 software package⁴⁸ using the ff03 force field,⁴⁹ which has been shown to give good population distributions for secondary structure elements in biomolecular simulations compared to the other Amber force fields but may overpopulate helical structures.⁵⁰ The replica exchange simulations were set up by first surrounding the $\text{A}\beta(1-42)$ monomer (coordinates were obtained from the PDB structure ID 1Z0Q) in a box of 16 777 TIP3P water molecules. The protonation states of titratable residues were adjusted according to a pH of approximately 7.0 achieved by using empirical structural rules to determine the pK_a of each residue.⁵¹ This resulted in standard protonation states for all residues, giving the protein an overall charge of -3.0. His6, His13, and His14 were found to have pK_a 's of 6.4, 7.0, and 6.4, respectively, and are all surface exposed. It is likely that the protonation states of these residues are in rapid equilibrium; therefore, we arbitrarily chose to protonate the ring in the δ position. To neutralize the -3.0 charge, we chose to use a background jellium rather than explicit counterions to avoid spurious coordination effects. This decision was made because the structure of $\text{A}\beta$ seems to be acutely sensitive to external factors, and we wanted to draw conclusions in the absence of any other extraneous elements that may bias the statistics. Following energy minimization to remove close contacts, NPT simulations were run for a single system using a 2 fs time step and Langevin dynamics with a collision frequency of 1.0 ps^{-1} to couple to the constant target temperature of 300 K and a Berendsen barostat to control pressure using a coupling constant of 2.0 ps to the target pressure of 1 bar. A 10.0 Å cutoff for nonbonded interactions was used in combination with the particle mesh Ewald procedure for long-range electrostatics, while H-X bond lengths were constrained using the SHAKE algorithm.⁵² This single simulation was run until both the pressure and temperature stabilized (~ 100 ps) and was then used as the starting point for each replica of the REMD simulations, which were carried out in the NVT ensemble. Temperatures for each replica were defined in the manner described above, and swaps were attempted every 2 ps. After a further equilibration of approximately 10 ns per replica, structural data were accumulated for 103 ns per replica, with structural data sampled every 2 ps (resulting in $\sim 45\,000$ configurations per replica). This gave an aggregate simulation time of 1.65 μs . The theoretical swapping probability, based on the temperature spacing, is determined to be 8.5%; the actual swapping probability

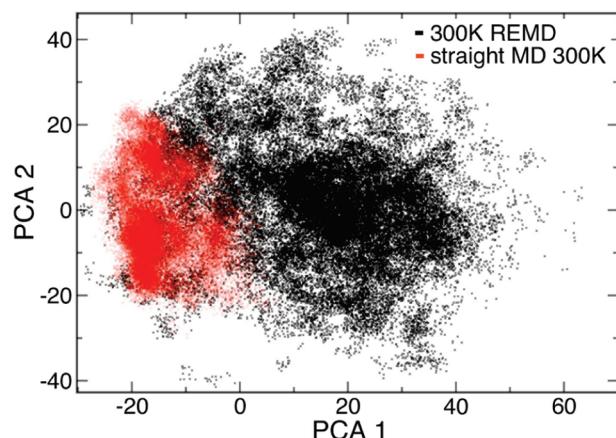


Figure 2. Comparison of replica exchange and constant temperature molecular dynamics at 300 K for $\text{A}\beta(1-42)$. Each point represents one configuration generated either at 300 K from the REMD simulation (black) or at 300 K from a straight constant temperature simulation (red). The configurations are projected onto the plane of the first two principal components (PCA-1, PCA-2) calculated by analyzing the covariance matrix of the combined configurations.

was found to be 6.3%. Although this is a low probability, the frequency of swap trials is high, and also the length of the simulation is long. This means that the number of *successful* swaps for our simulation was relatively high (9857), indicating that the REMD simulation is effective. In order to illustrate the efficiency of the REMD simulation, we ran a straight NPT molecular dynamics simulation at 300 K of the same system for approximately 500 ns, with data sampled every 12 ps ($\sim 40\,000$ configurations), and compared this simulation to the REMD 300 K temperature ensemble. To make the comparison, we projected all configurations from the 300 K REMD and constant temperature ensembles onto their first two principal components, calculated by combining the two trajectories together and analyzing their covariance matrix. The total combined configurations number approximately 80 000. The projections are shown in Figure 2.

The figure clearly shows that the REMD simulation at 300 K covers the PCA-1/PCA-2 plane much more extensively than the constant temperature simulation.⁵³ This is particularly important since the constant temperature simulation is run for effectively 5 times the length of the REMD simulation (500 ns versus 100 ns per replica), indicating that enhanced sampling is required for simulations of $\text{A}\beta(1-42)$ because of kinetic trapping. This has been alluded to in a previous study.²⁶

In addition to the simulations carried out for the Amyloid- β peptide, we also ran a small number of pentapeptide simulations of peptides having the general formula GGXGG. These simulations are discussed in the context of analyzing the change in chemical shift (see eq 1 and section 3.2), as evaluated from our theoretical calculations. Initially, we ran these simulations in 15 Å boxes of TIP3P water using the NVT ensemble; after a short equilibration period (~ 200 ps), we switched to the NPT ensemble to equilibrate the density of water. Statistics were then gathered using the same control parameters given above for the constant temperature simulation of $\text{A}\beta(1-42)$ with the following differences: Acetyl and methyl end-caps were used to prevent charge interactions affecting the equilibrium population of the peptide. In addition, we used the Amber ff99SB force field.³¹ These simulations were not run for a fixed time but rather until

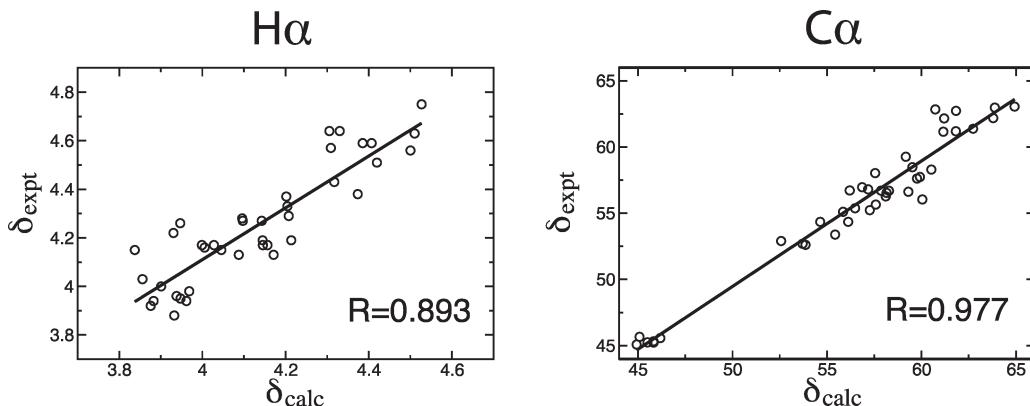


Figure 3. Correlation of average theoretical chemical shifts with experimentally determined chemical shifts. The left panel compares 41 experimentally reported H_α chemical shifts with those determined from the simulation; the right panel compares 41 experimentally reported C_α chemical shifts with those determined from simulation. The lines are linear regressions, and the R values are the corresponding Pearson correlation coefficients.

the difference in the percentage populations of dihedral substates between the entire simulation and of the second half of the simulation was less than 5% (using a sampling frequency of 8 ps). This usually resulted in the simulations being approximately 200-ns-long.

3. RESULTS AND DISCUSSION

3.1. Comparison to NMR Chemical Shift Data. H_α and C_α chemical shifts were calculated using an empirical approach as implemented in the SHIFTS program.⁵⁴ In order to validate the quality of the theoretically determined chemical shifts, we compared calculated average H_α and C_α NMR shifts from our ensemble of structures with the experimentally determined values for A_β(1–42).¹⁶ Figure 3 plots calculated H_α and C_α chemical shifts (δ_{calc} , pH ~ 7, 9 °C) versus experimentally determined chemical shifts (δ_{expt} , pH ~ 7, 5 °C) along with a least-squares line and the corresponding Pearson coefficient (R).

For the H_α chemical shifts, the Pearson coefficient (0.893) is relatively high, suggesting that our simulations reproduce the structural features associated with the chemical shifts. The inherent spread of the data may be due to a number of associated errors including insufficient sampling of configuration space, inaccuracies of the force field, and errors in the theoretical methodology used to calculate the chemical shift. The standard deviation of the theoretical methodology is estimated to be 0.23 ppm,⁵⁴ while the mean residual from the current regression analysis is 0.01 ppm. This indicates that error from the theoretical methodology for calculating the chemical shift eclipses any other source of error, i.e. sampling and force-field inaccuracies. Similar results are found for the calculated versus experimental chemical shifts of α-carbons. In this case, the Pearson coefficient is even higher, 0.977. The estimated standard deviation for the theoretical model is 0.97 ppm,⁵⁵ whereas the mean residual of the current regression analysis is 1.35 ppm, suggesting again that the largest source of error in our analysis comes from the chemical shift calculation, although in this case it seems that other sources also contribute to the error in the regression analysis.

The high Pearson coefficients indicate that the calculated average chemical shifts from our ensemble of structures are in good agreement with experimental results. This result lends itself to a more detailed analysis of the secondary structure of A_β(1–42) based on the change in chemical shifts as given in eq 1.

3.2. Analysis of the Brookhaven PDB. Before undertaking a detailed analysis of A_β(1–42), we wanted to test whether a theoretical protocol can capture the secondary structure shift for globular proteins with well-defined secondary structure. To do this, we analyzed the Brookhaven Protein Data Bank (PDB)⁵⁶ and selected all protein structures with a resolution of better than or equal to 1.50 Å and calculated their C_α and H_α chemical shifts. This resulted in analyzing 4456 structures and producing approximately 8 000 000 chemical shifts. This analysis also removes the possibility that errors are resulting from incorrectly parametrized force fields.

In order to automate the process and minimize errors associated with incorrect placement of protons, we first stripped the structures of all their protons and then re-added them with protons according to residue templates in Amber's LeAP module.⁴⁸ When adding protons, we made the assumption that all histidine residues were protonated at the ε position and all other titratable residues were protonated assuming an environmental pH of 7. Although this may lead to the incorrect placement for some side-chain protons, especially those buried within a protein, we are assuming that these protons are far enough removed from the α-centers that they do not affect their chemical shifts to a large degree. As it turns out, which we note below, this may not be a justifiable assumption in the case of histidine.

To calculate the change in the chemical shifts ($\Delta\delta$), we used random coil reference values taken from experimental results.⁵⁷ However, this presents problems for cysteine and proline residues because they each have two possible random coil references. In the case of cysteine, this is because it is commonly found in two different oxidation states and, for proline, because it may adopt a cis or trans conformation. In order to speed up the automated process, we decided to leave all cysteine residues out of the analysis and assume that all proline residues were in a trans conformation. After calculating the chemical shifts, we accumulated statistics according to what secondary structure element the residue originated from. As there is no unique means to do this, we compared two popular methods, viz., Ramachandran dihedral angles and the DSSP⁵⁸ (define secondary structure of proteins) protocol.

For the Ramachandran analysis, we used the following definitions (in degrees) of secondary structure. Right-handed α-helix (alpha): $-100 \leq \varphi \leq -30$; $-80 \leq \psi \leq -5$. Near right-handed α-helix (alpha N): $-175 \leq \varphi \leq -100$; $-55 \leq \psi \leq -5$. Left-handed

Table 1. Average (av), Standard Deviation (std), and Median Absolute Deviation (MAD) for the Change in the Chemical Shift ($\Delta\delta$) for C α and H α Atoms for Various Secondary Structure Elements as Defined Using Ramachandran Angles and the DSSP Protocol for the Brookhaven PDB

C α	Ramachandran			C α	DSSP		
	av	std	MAD		av	std	MAD
alpha ^a	2.69	1.18	0.61	α -helix ^b	2.87	1.15	0.58
alpha N ^a	0.28	1.00	0.63	π -helix ^b	1.91	1.55	0.91
alpha L ^a	0.72	0.83	0.40	β -sheet (para) ^b	-1.10	1.36	0.83
PPII ^a	-0.01	1.26	0.87	β -sheet (anti) ^b	-1.11	1.34	0.86
beta ^a	-1.66	1.16	0.78	turn ^b	0.84	1.55	1.02
other ^c	-0.70	1.45	0.98	other ^c	-0.08	1.79	1.15

H α	Ramachandran			H α	DSSP		
	av	std	MAD		av	std	MAD
alpha ^a	-0.29	0.29	0.12	α -helix ^b	-0.30	0.27	0.12
alpha N ^a	0.18	0.24	0.13	π -helix ^b	-0.12	0.38	0.26
alpha L ^a	-0.50	0.36	0.10	β -sheet (para) ^b	0.27	0.33	0.19
PPII ^a	0.01	0.38	0.23	β -sheet (anti) ^b	0.27	0.35	0.22
beta ^a	0.26	0.34	0.21	turn ^b	-0.18	0.37	0.21
other ^c	0.00	0.38	0.19	other ^c	-0.08	0.37	0.21

^aThe following definitions for secondary structure were used. Right-handed α -helix (alpha): $-100 \leq \varphi \leq -30$; $-80 \leq \psi \leq -5$. Near right-handed α -helix (alpha N): $-175 \leq \varphi \leq -100$; $-55 \leq \psi \leq -5$. Left-handed α -helix (alpha L): $100 \leq \varphi \leq 30$; $80 \leq \psi \leq 5$. Polyproline II (PPII): $-110 \leq \varphi \leq -50$; $120 \leq \psi \leq 180$. Extended β sheet (beta): $-170 \leq \varphi \leq -110$; $80 \leq \psi \leq 180$ and $-170 \leq \varphi \leq -110$; $-180 \leq \psi \leq -170$. ^bFor DSSP definitions, see ref 58. ^cDefined as being any other element that does not fall within the definitions of the protocol in question, viz, Ramachandran or DSSP.

α -helix (alpha L): $100 \leq \varphi \leq 30$; $80 \leq \psi \leq 5$. Polyproline II (PPII): $-110 \leq \varphi \leq -50$; $120 \leq \psi \leq 180$. Extended β sheet (beta): $-170 \leq \varphi \leq -110$; $80 \leq \psi \leq 180$ and $-170 \leq \varphi \leq -110$; $-180 \leq \psi \leq -170$.

The DSSP algorithm is more sophisticated in that it first identifies hydrogen bonds by using an electrostatic-energy function and then assigns secondary structure either by the topology of hydrogen bonds in a repetitive sequence (for helices and sheets) or by angle restraints (for turns).

After assigning each $\Delta\delta$ to a secondary structure element, either according to Ramachandran angles or the DSSP protocol, we calculated their averages (av), standard deviations (std), and median absolute deviations (MAD). These results are summarized in Table 1. Although there is no one-to-one correspondence between secondary structure elements defined by Ramachandran angles and the DSSP protocol, we can still directly compare the results for α -helices and β -sheets (the β region in the case of Ramachandran angles and the parallel and antiparallel structures in the case of DSSP).

The average $\Delta\delta$ value of α -helical residues for C α atoms was found to be 2.69 ppm and 2.87 ppm when using Ramachandran angles and the DSSP protocol for assigning secondary structure, respectively. Experimentally, the secondary structure shift is usually reported as 2.50 ppm for α -helical residues. This indicates that the combination of either Ramachandran angles or DSSP for secondary structure assignment with the empirical method for

calculating C α chemical shifts is able to capture the $\Delta\delta$ for α -helical residues reasonably well. For β -sheets, the average $\Delta\delta$ for C α chemical shifts was found to be -1.66 ppm when using the Ramachandran angle definitions and -1.10 and -1.11 ppm for parallel and antiparallel β -sheets, respectively, when using the DSSP definitions. Experimentally, the $\Delta\delta$ for C α β -sheets is reported as -2.00 ppm. Again, this indicates that both Ramachandran angles and the DSSP protocol are capturing the $\Delta\delta$ effect. However, using the DSSP definitions moves the average value in the positive direction when compared with experimental results.

A similar analysis of the H α chemical shifts finds the average $\Delta\delta$ of α -helical residues to be -0.29 ppm and -0.30 ppm when using the Ramachandran and DSSP definitions, respectively. These are in good agreement with the experimentally determined value of -0.30 ppm. For β -sheets, the Ramachandran definition gave an average $\Delta\delta$ of 0.26 ppm, while the DSSP definitions of parallel and antiparallel β -sheets gave averages of 0.27 ppm. Again, these results are in good agreement with the experimental value of 0.30 ppm.

To further quantify the distributions of calculated $\Delta\delta$ values, we have also computed the standard deviation (std) and median absolute deviations (MAD) for each secondary structure element; these results are also presented in Table 1. In addition, we have plotted the $\Delta\delta$ distributions by calculating their histograms. Figure 4 presents the $\Delta\delta$ distributions for all C α and H α chemical shifts (upper panels). These distributions are then split according to secondary structure elements using the Ramachandran definitions (middle panels) or the DSSP protocol (lower panels).

The MADs of Table 1 are smaller compared with their respective standard deviations for both C α and H α chemical shifts using either Ramachandran or DSSP definitions. This indicates that the calculated distributions of $\Delta\delta$ have long tails because the MADs give them less weight. Figure 4 supports this conclusion. In addition, the upper panels of Figure 4 show that the total $\Delta\delta$ distributions (independent of secondary structure) are unsymmetrical and, in the case of C α atoms, multimodal.

Comparing the standard deviations of α -helical and β -sheet values with their averages indicates that these distributions overlap more significantly for H α chemical shifts than C α chemical shifts; this is also confirmed by the middle and lower panels of Figure 4 (see for example the overlap between the solid black and dashed red lines in the middle panels and the solid black and solid red lines in the lower panels). More importantly, if other secondary structure elements are included in the analysis, it is difficult to decompose both the C α and H α distributions into areas that do not include contributions from other secondary structure elements; in particular, it would be difficult to assign a β -sheet element to any residue solely on the basis of $\Delta\delta$ if the protein or peptide is dynamically exploring conformational space. This certainly has consequences when using the CSI method to assign secondary structure to flexible biomolecules. That being said, the Ramachandran decomposition of the C α distribution shows that the α -helical area of the $\Delta\delta$ distribution is relatively devoid of contributions from other secondary structure elements. This is not the case for the DSSP decomposition if π -helices are present in addition to α -helices.

The distributions in Figure 4 also highlight another important aspect of the random-coil reference values. Frequently, these references are derived from small unstructured peptides or denatured proteins.^{57,59–61} The “random coil” or more appropriately the statistical coil is not random but based on maintaining predefined equilibrium populations of each substate through

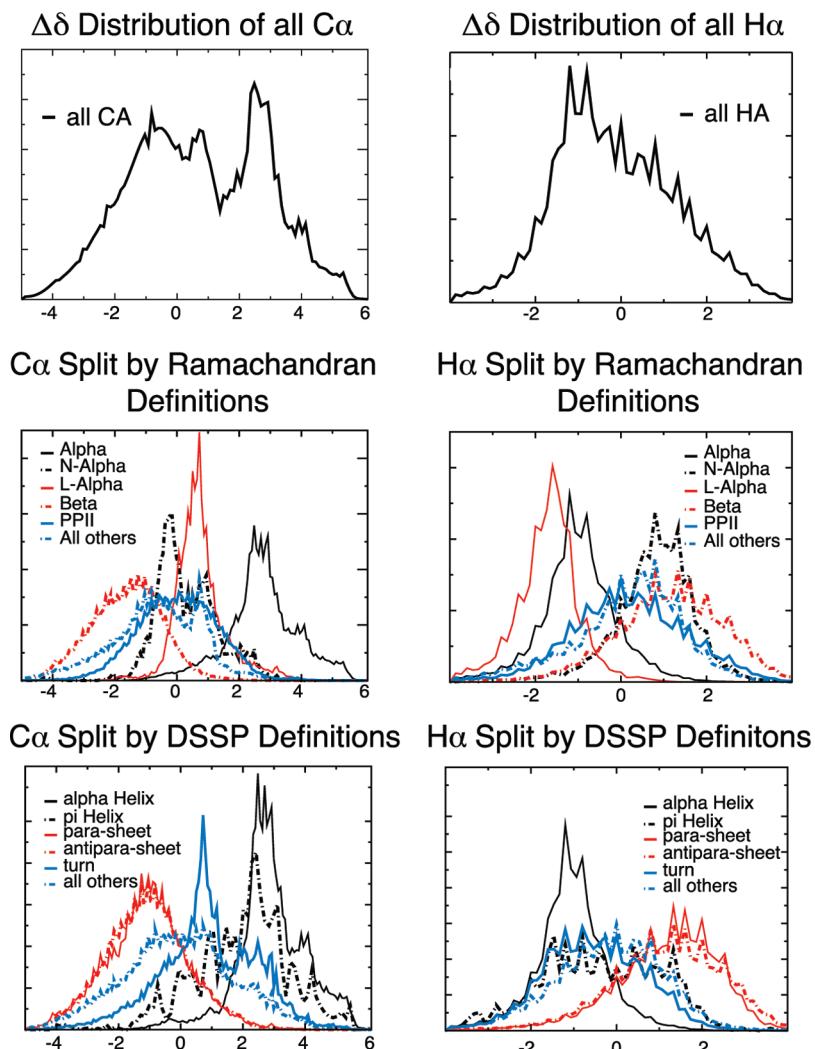


Figure 4. Distributions of $\Delta\delta$ for $C\alpha$ and $H\alpha$ chemical shifts derived from structures in the PDB database with a resolution equal to or better than 1.50 Å. Panels show the full distributions (upper) and then these distributions divided according to secondary structure elements according to either Ramachandran angles (middle) or the DSSP protocol (lower).

rapid conformational switching. The equilibrium populations depend primarily on the residue in question⁶² but also on its sequence-dependent context.⁶³ Small peptides are one means of obtaining random-coil values; another way is to look at nonhomologous protein structures from X-ray experiments.⁶⁴ In essence, this is what we have done in generating Figure 4; therefore, the average $\Delta\delta$ over all residues for both $C\alpha$ and $H\alpha$ should be zero if

- (a) the proteins we selected are sufficiently nonhomologous as to represent the inherent random coil structure
- (b) the empirical model used to calculate the chemical shifts does not contain systematic errors
- (c) the reference random coil values are appropriate

The average of $\Delta\delta$ for all shifts was found to be 0.64 ppm and -0.07 ppm for $C\alpha$ and $H\alpha$ shifts, respectively, indicating that the three conditions are being met to a reasonable degree. By the same arguments above, the residue-by-residue values should also average to zero. These results are given in Table 2.

Table 2 clearly demonstrates that while the overall averages of $\Delta\delta$ for $C\alpha$ and $H\alpha$ chemical shifts of 0.64 ppm and -0.07 ppm, respectively, are reasonably good, their signs are resulting from

systematic errors in the by-residue estimates of $\Delta\delta$. In addition to the systematic errors that are apparent from Table 2, there are a small number of residues for which the total average $\Delta\delta$ deviates significantly from zero. For the $C\alpha$ chemical shifts, these include Ala (1.10 ppm), His (1.71 ppm), and Arg (1.17 ppm). For the $H\alpha$ shifts, the residues are Ala (-0.18 ppm), Glu (-0.18 ppm), and His (-0.16 ppm) and to a lesser degree Lys (-0.12 ppm) and Arg (-0.12 ppm).

In order to probe the origin of the systematic errors and a small number of large deviations, we ran molecular dynamics simulations for a selection of pentapeptides of the general formula GGXGG and calculated the average $\Delta\delta$ for the center residue X. These results are also given in Table 2.

For alanine, the large deviations of 1.10 ppm and -0.18 ppm are reduced to 0.11 ppm and -0.10 ppm for $C\alpha$ and $H\alpha$ centers, respectively, when using molecular dynamics simulations. This indicates that these deviations are originating from a bias in the statistics generated from the PDB analysis. The known prevalence of alanine in α -helical structures and the respective average downfield and upfield shifts of the $H\alpha$ and $C\alpha$ atoms suggests that the PDB structures are over-representing alanine in

Table 2. Per Residue Averages for the Change in the Chemical Shift ($\Delta\delta$) for C α and H α Atoms from an Analysis of the PDB Databank and a Selection of GGXGG Peptides from Molecular Dynamics Simulations

	C α Ave.		C α Ave.		H α Ave.	H α	Ave.
Brookhaven PDB ^a							
Ala	1.10	Met	0.75	Ala	-0.18	Met	-0.07
Asp	0.35	Asn	0.64	Asp	-0.08	Asn	-0.07
Glu	0.71	Gln	0.87	Glu	-0.18	Gln	-0.10
Phe	0.40	Arg	1.17	Phe	-0.05	Arg	-0.12
His	1.71	Ser	0.29	His	-0.16	Ser	-0.06
Pro	0.66	Thr	0.39	Pro	-0.07	Thr	0.10
Ile	0.60	Val	0.65	Ile	0.05	Val	0.05
Lys	0.85	Trp	0.22	Lys	-0.12	Trp	-0.11
Leu	0.79	Tyr	-0.07	Leu	-0.08	Tyr	0.00
Molecular Dynamics ^b							
Ala	0.11	Pro	0.11	Ala	-0.10	Pro ^b	-0.03
Asp	0.11	Lys	-0.32	Asp	-0.12	Lys	-0.06
His	1.30	Thr	-0.02	His	-0.15	Thr	-0.02

^a Statistics obtained from an analysis of the PDB; see text for details.

^b Statistics obtained from molecular dynamics simulations; see text for details.

α -helices compared with the inherent random coil state. In fact, 57.8% of alanine residues from the PDB analyzed structures are found in an α -conformation using Ramachandran angle definitions, whereas the alanine of GGAGG spends only 17.2% of the time in the α region.

The similar large deviations for histidine of 1.71 ppm and -0.16 ppm for C α and H α , respectively, would immediately suggest a similar bias for this residue. However, the molecular dynamics simulations also produce large deviations (1.30 ppm and -0.15 ppm). The small reduction in the $\Delta\delta$ values is because the PDB structures have an α -region population of 33.4%, whereas the molecular dynamics simulations produce an α -region population of 18.2%. The remaining large error is most likely because of the pH effect on C α and H α chemical shifts of histidine, which can be up to 2.5 ppm for the C α atom.^{65,66} Therefore, the large errors seen for histidine are probably a combination of assuming that the side chain is permanently protonated at the ϵ position and the random coil reference values that we have used to generate the $\Delta\delta$ statistics being incorrect for this permanent protonation state. Modest improvements in the average $\Delta\delta$ are also seen for the other residues by using molecular dynamics simulations. This further indicates that even with a large data set from the PDB, biasing is introduced to a small degree. However, because the dynamics simulations do not produce exactly zero deviations, it is suggested that there is also a modest systematic error in the theoretical methodology of on average approximately 0.10 ppm for both C α and H α atoms.

In summary, we have analyzed the empirical methodology for calculating the secondary structure shift ($\Delta\delta$) using a large data set of structures from the PDB. In conjunction with these chemical shifts, we examined two popular means of dividing the conformational subspace into secondary structure elements. We found that the secondary structure distributions of $\Delta\delta$ as divided by Ramachandran angles and the DSSP protocol results in substantial overlap of the distributions for both C α and H α .

chemical shifts. However, if the conformational subspace is divided by Ramachandran angles, then the α region of the C α distribution is not significantly contaminated by other secondary structure elements. In addition, we found that modest systematic errors are introduced into the average $\Delta\delta$'s because of the theoretical methodology. In some cases, however, large errors are introduced because of either incorrect statistics from the data set (Ala) or from misrepresenting the correct dynamic protonation states (His).

3.3. Distributions of Chemical Shifts for A β (1–42). In the previous sections, we have shown that the mean chemical shifts for A β (1–42), as calculated from theory, agree well with experimental results. We also demonstrated by analyzing the PDB that the theoretically determined change in the chemical shift ($\Delta\delta$) is also reproduced well. If the protein in question does not vary significantly from its equilibrium position, then the $\Delta\delta$ distributions for each residue would be expected to have minor fluctuations around their means, making secondary structure assignment unambiguous. However, in the case of a flexible biomolecule like A β (1–42), we will show below that the distributions of $\Delta\delta$ are broad like in Figure 4, where we combined many peptide conformations from the PDB. We will also investigate the consequences this has for assigning its secondary structure. In the discussion below, we will refer to secondary structure elements defined through Ramachandran dihedral angles because the PDB analysis showed that the $\Delta\delta$ effect is in closer agreement with experimental results using these definitions.

It is important to state from the outset the differences between the inherent random-coil state of a residue and the random-coil region of chemical-shift distributions. The random-coil state of a residue is a statistical combination of all of the dihedral angles the backbone explores when it is unconstrained by contextual influences. As the unconstrained residue flips from β to α regions, the H α chemical shift changes from being relatively upfield to being relatively downfield; the reverse is true for C α shifts. The PDB analysis above demonstrated that other structural elements may also move the shift upfield or downfield and in some cases leaves it unchanged. The mean chemical shift of the random-coil state is then the average of these competing shifts, and the change in chemical shift for this state is defined to be zero. The random-coil region of the distribution of changes in chemical shifts is thus a zone centered on zero. When ascribing a random coil structural element to a residue, it is assumed that a zero overall shift is resulting from the residue in question, sampling from the same populations of dihedral angles that originally defined its random coil state.

In order to demonstrate the relationship between the mean chemical shift and the underlying distributions for a flexible peptide, we plotted the $\Delta\delta$ distributions of each residue from our simulations of A β (1–42) and compared these distributions against typical random coil, α -helical, and β -sheet $\Delta\delta$ values. Eight examples of these plots are shown in Figures 5 and 6 (plots for all residues can be found in the Supporting Information, Figures S1 and S2). In addition to the distributions and typical values of $\Delta\delta$, each plot indicates the mean value from our simulations and the observed value from experiments. Random coil values were taken from tabulated data, as were helical and sheet values; the domains of the $\Delta\delta$ zones shown in Figures 5 and 6 were derived from the first standard deviation of $\Delta\delta$ values calculated from the analysis of the PDB described above.

For the H α chemical shifts, we find a number of consistent key features, which are illustrated by the distributions calculated for

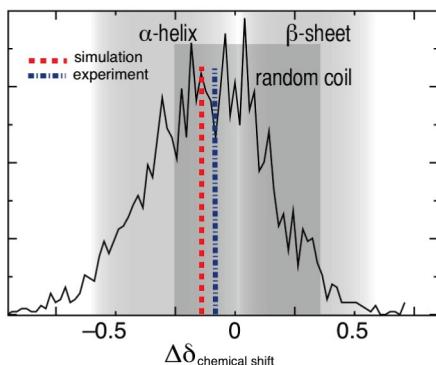
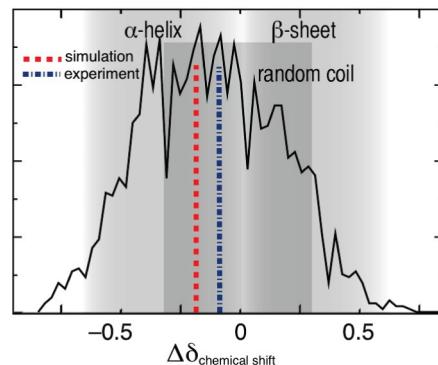
H_α Glu 3**H_α Glu 22**

Figure 5. $\Delta\delta$ Distributions of $\text{H}\alpha$ chemical shifts of Glu3 and Glu22 in $\text{A}\beta(1-42)$ at 300 K. Shaded areas give the domains encompassing typical random coil, helix, and β -sheet values. The sizes of the shaded areas have been determined by using one standard deviation from the mean values found from analyzing the PDB. The mean values from simulations and experiments are indicated with dashed lines. See also Figure S1, Supporting Information.

Glu3 and Glu22 as shown in Figure 5a and b. First, like the analysis of the PDB, the regions spanning the secondary structure domains of $\Delta\delta$ values overlap one another significantly. The significant overlap between different structural elements causes the resultant distributions to be Gaussian-like (often skewed to one side) with the mean value aligned near the distribution's maximum value. This situation is clearly shown in Figure 5. For example, because of the overlap in $\Delta\delta$ values from different structural elements, it is difficult to discern whether or not the similar average $\Delta\delta$ values for Glu3 and Glu22 are resulting from the same combination of structural elements. The mean values for Glu3 and Glu22 from experiments and simulations are located in the random coil region. Naturally, one would infer from this information that both residues are sampling from the random-coil state. However, when the α -content is calculated using Ramachandran angles, we find that Glu3 spends 29.3% of the time in this configuration, while Glu22 spends 61.4% of the time in this configuration. Therefore, Glu3 and Glu22 cannot both be sampling from the inherent random-coil population of glutamic acid.

The large overlap of domains causes 34 of the 38 $\text{H}\alpha$ distributions plotted in this study (see also Figure S1 in the Supporting Information) to have their mean values from simulations and experiments lying in the region combining α -helical and random coil values, even though each residue may be sampling from distinctly different dihedral populations.

The $\text{C}\alpha$ chemical shift distributions show some significant differences compared with the $\text{H}\alpha$ distributions. Unlike the $\text{H}\alpha$ distributions, they are not simply Gaussian-like and centered on the mean value but may be complicated bi- or trimodal distributions, see for example the plots in Figure 6. From this figure, it can be clearly seen that the average value is derived from combinations of three distinct and separated distributions, viz, α -helical, random coil, and β -sheet. Because each zone is clearly observable, the $\text{C}\alpha$ shifts may lend themselves to a more robust means of identifying if distributions with similar averages have distinct populations. To investigate this, we have plotted in Figure 6 the distributions for three pairs of residues with the same side chain from the $\text{A}\beta(1-42)$ sequence.

In the top panels, we have plotted the distributions of Ala2 and Ala21. The mean value from experimental results and theory lie in the random coil region, and the underlying distributions look fairly similar. That is, they have large peaks in the random coil and

α -regions with little signal in the β -region. However, like the situation for glutamic acid above, the secondary structure populations calculated from Ramachandran angles are quite different; for example, Ala2 spends 25.7% of the time in the α -region, whereas Ala21 spends 42.8% of the time in the α -region. The simulation of GGAGG we performed in conjunction with the PDB analysis indicates that alanine spends approximately 26.8% of the time in the α -region when in a random coil configuration, which is supported by another study.⁶² Therefore, even though both distributions average to a random coil value the underlying distribution of Ala2 seems to be much more closely related to the random coil than does the distribution of Ala21.

For Val18 and Val36, the distributions are again divided into distinct areas. Val18 has large peaks in the α - and random coil regions with a smaller peak in the β -region. Val36 on the other hand has large peaks in the β - and random coil regions with little or no signal in the α -region. Again, however, their means from simulations and experiments lie in the random coil region. In this case, the distributions are representative of their secondary structure content. For instance, Val18 spends 35.2% of the time in the α -region, and Val36 spends 6.3% of the time in this region. The sizes of the peaks in the β -regions are commensurate with their calculated β -content of 43.1% and 51.4% for Val18 and Val36, respectively. The large middle peaks are derived from predominately the PII structure, 14.4% in the case of Val18 and 22% in the case of Val36. Although the distributions can be related to the calculated secondary structure and their mean values lie in the random coil region, the random coil distribution of valine has approximately 21.9% helical content and 30.4% β -content;⁶² therefore, neither Val18 nor Val36 seem to be representing this population exactly.

The final pair of distributions we plot in Figure 6 are for Phe19 and Phe20. Like the situation for Ala2 and Ala21, these distributions share similar shapes. Both have large α -peaks and a large peak on the edge of the β -/random coil region. Their mean values lie in the random coil region. Again, the peaks are representative of their calculated secondary structure; Phe19 and Phe20 spend 29.7% of the time in the α -region and 65.5 and 62.3% of the time in the β -region, respectively. This indicates that in the case of phenylalanine the chemical shift for β -sheets is shifted relatively less downfield compared with other residues.

The differences observed between the six residues above highlight a number of important points. First, signals that result

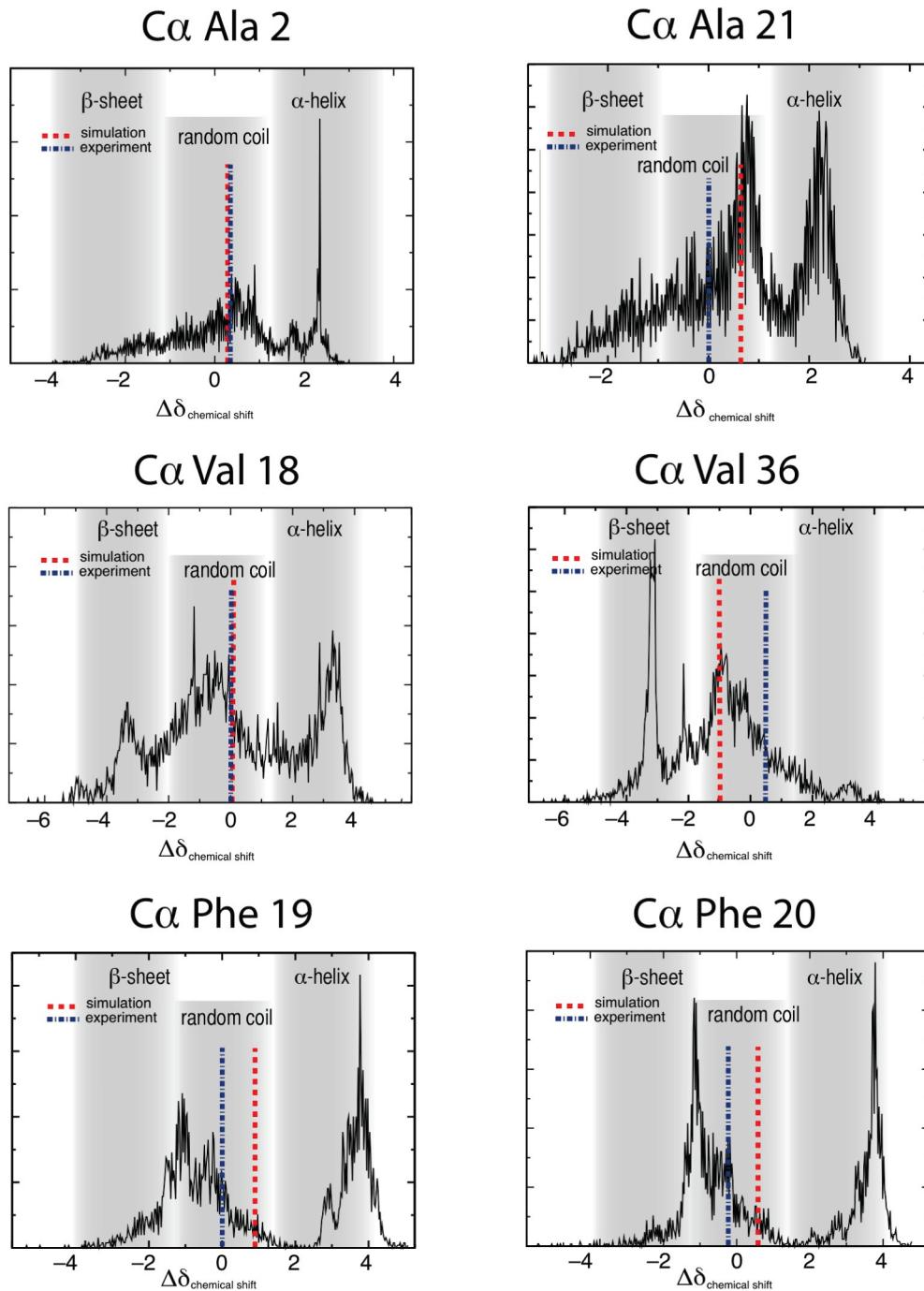


Figure 6. Distributions of C α chemical shifts of Ala2, Ala21, Val18, Val36, Phe19, and Phe20 from A β (1–42) at 300 K. Shaded areas give the domains encompassing typical random coil, helix, and β -sheet values. The sizes of the shaded areas have been determined by using one standard deviation from the mean values found in an analysis of the PDB data bank (see test). The mean values from simulations and experiments are indicated with dashed lines. See also Figure S2, Supporting Information.

in random coil mean values are a result of sampling from combinations of all three regions. Second, the dynamically explored secondary structure elements can change by up to 30% between two residues with the same side chain but still result in random coil averages. Finally, the random coil region of the chemical shift distribution does not necessarily have to contain a signal for the mean value to reside there.

From the distributions given in Figure 6, one might be tempted to estimate the secondary structure content of a particular residue

by integrating the signals in each area and comparing them against the total signal. However, as shown in Figure 4, the β -region, random coil region, and to a lesser degree the α -region may contain a signal from more than one secondary structure element when defined through Ramachandran angles. To demonstrate how this may influence the secondary structure populations from a histogram analysis, we plot in Figure 7 the α - and β -Ramachandran populations for all nonglycine residues of A β (1–42) against the estimated populations by counting the

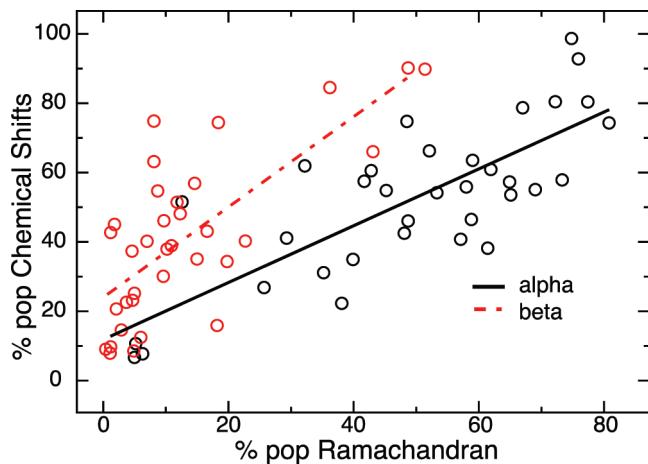


Figure 7. Correlation of % population of α - and β -structure calculated with Ramachandran angles with the % population of α - and β -structure determined by counting the number of chemical shifts found in the α - and β -areas of the chemical shift distributions. Data points are from all nonglycine residues in $A\beta(1-42)$.

total signal within two standard deviations of the mean β - and α -chemical shifts.

The figure shows that both α - and β -Ramachandran populations are correlated to the areas of the chemical shift distributions. However, simply because they correlate does not infer that the model is valid. If it were valid, then the slope of the regression lines would pass through zero and have a slope of one. The line correlating the β -populations has a slope larger than one (1.3) and crosses the y intercept at 22.0. The y intercept indicates that on average the β -area of the chemical shift distribution contains at least 22% additional structure not associated with β dihedral angles. The slope indicates that, as the percentage of β -sheet structure increases, the contamination from other elements in the β -area of the chemical shift plot also increases. The line correlating the size of the α -area to the α -populations is a better model for predicting the α -content with a slope of 0.82 and a y intercept of 11.0. An intercept larger than zero is again an effect associated with contamination from other secondary structure elements, albeit to a lesser degree than the β -area contamination. With a slope smaller than one, this model infers that as the percentage of α -helical structures increases, the number of those that are captured in the α -area of the chemical shift plot decreases.

3.4. Static and Dynamic Helical Content of $A\beta(1-42)$. As stated in the Introduction, molecular dynamics simulations usually report a much higher helical content for $A\beta(1-42)$ than experimental studies. In both types of reporting, the “static” helical content is being calculated, that is, the average over the course of the simulation or over the course of the experiment. In this section, we report the possible underlying reasons for the discrepancy between theory and experimental results. We conclude that it is ambiguous to assign a static content to a peptide that does not constitute one specific conformational fold. This is because the final average helical content depends on how a helix per residue is defined, and as a result, we may calculate a wide range of percent content (from 0% to 45.7% in this case) even when using the same data set.

First, using the change in $C\alpha$ chemical shift as an indicator of helical content, we can calculate the static content in the same way that the CSI method does. Of the 40 $C\alpha$ chemical shift

distributions (see Figure S2 of the Supporting Information), only four of them have mean values from simulation in the α -helical area (Ser8, Tyr10, Leu17, and Ser26); an additional three lie in the β -sheet area (Val36, Val39, and Val40). In order to define an α -helix, the chemical shift index method requires that both the mean chemical shift is in the α -helical area, and at least four consecutive residues are found in this area. Using these rules to interpret our chemical shift data, $A\beta(1-42)$ has 0% α -helical content, reproducing the results found for $A\beta(1-42)$ using the CSI method in ref 16.

Normally, helical content is reported from simulation by averaging the helical content from theoretical calculations; popular methods are to average the content as determined by DSSP, Ramachandran angles, or STRIDE,⁶⁷ which is defined similarly to DSSP. Although for DSSP we found less satisfactory results for the change in chemical shift and, in addition, the DSSP structural decomposition leads to an overlap of α -helices and π -helices, the vast body of literature supports DSSP as a valid method for finding helical structures. Using this methodology, we find an average helical content of 25.9%, in agreement with most molecular dynamics studies. The STRIDE algorithm similarly finds a 29.1% helical content. For both algorithms, the residues contributing most to these numbers are Tyr10–Leu17, which for approximately 50% of the simulation are in a helix, and residues Val18–Ser26, which spend approximately 30% of the simulation as a helix (see Supporting Information Table S1). However, these numbers may overestimate the *static* helical content with respect to experimental results because residues that form part of the boundary between the helical and nonhelical structure are counted even though they may not be part of a persistent helix.

More simply than the DSSP and STRIDE estimates, we could average the Ramachandran angles that fall inside of the α region. For our simulation, this results in a helical content of 45.7%. However, this is a rather naive approach for two reasons: First, a Ramachandran angle pair residing in the α -region does not equate to an α -helix. In essence, this is what the DSSP and STRIDE algorithms are correcting for. Second, the random coil state of an amino acid samples significantly from the α -region, see ref 62. To correct for the random coil populations, we may define the helical “excess” of the species by taking the average of the differences in the helical content per residue from the simulation with the helical population of the random coil state and normalize (see eq 3).

$$\text{helical excess} = \frac{\sum_{i=1}^N (\% \text{helix}_{\text{simulation}}^{\text{res}(i)} - \% \text{helix}_{\text{randomcoil}}^{\text{res}(i)})}{\sum_{i=1}^N (1 - \% \text{helix}_{\text{randomcoil}}^{\text{res}(i)})} \quad (3)$$

If we take the average random coil helical populations⁶⁸ for each residue from ref 62, the helical excess of $A\beta(1-42)$ is found to be 27.5%, which is interestingly in agreement with the DSSP and STRIDE results. However, this formula can be potentially misleading because it may include negative values in the average, much like the chemical shift assignment does.

The preceding discussion demonstrates that it is difficult to ascribe a single static content to $A\beta(1-42)$ even when the data set is the same, as it depends on how the helical content is defined. A more appropriate way of describing the helical content is to look at its dynamic structure. This can be demonstrated through the joint probability of finding a sequence of n residues (starting from

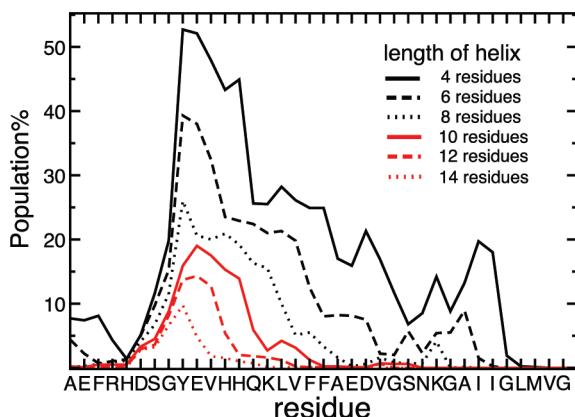


Figure 8. Population of helical structure as a function of helix length and residue number calculated from the joint probability function of Ramachandran angles in the α region, where the joint probability function is $P(i, i+1, \dots, i+n-1)$, i is the starting residue, and n is the length of the helix (see text for further details).

residue i) that reside in the helical basin of the Ramachandran plot. In Figure 8, we have plotted this function for helix lengths of 4, 6, 8, 10, 12, and 14 residues. As expected the probability of finding longer helices is smaller than finding shorter ones. The plot in Figure 8 suggests that there is a strong probability of finding short helices four residues in length starting from Tyr10. This short sequence forms the start of longer helices that extend toward the C-terminal end, which even for a length of 14 residues has a probability of approximately 10% from our simulation. Short helical sequences of four to six residues also have a 10 or 20% probability at the N- and C-terminal ends, respectively.

This description of helicity makes sense when comparing our knowledge of $A\beta$ in transmembrane and apolar environments. If we imagine gradually making the environment around $A\beta(1-42)$ more apolar, then the probability of finding the helix starting at Tyr10 increases. In addition, the probability of finding longer helices starting at this residue increases until the permanent helix spanning residues 10–25 is established, as found from experimental studies in apolar environments. The helix that starts at residue 27 in Figures 1 and 8 behaves in a similar fashion.

4. CONCLUSION

In summary, we have performed a comparison of experimental chemical shift data with theoretically determined chemical shifts for $A\beta(1-42)$. First, we investigated the correlation between chemical shifts and protein secondary structure in the case of structures in the PDB. We found that for $H\alpha$ chemical shifts, the theoretical chemical shift distributions of secondary structure elements overlap one another significantly, making it difficult to discern if the calculated mean was composed of one element or another. For $C\alpha$ chemical shifts, the distributions for each secondary structure element investigated are more well separated, making them in principle more suitable for secondary structure assignment. However, we showed that residues with the same side chain could produce similar means and distributions but could have differences in secondary structure of up to 30%. Finally, we showed that a single static helical content could vary from 0% to 45.7%, depending on the method used and how the helical content is defined, even when the same data set is used.

For flexible molecules, it is thus more appropriate to look at the probability of helical persistence as a function of length.

■ ASSOCIATED CONTENT

S Supporting Information. Plots of the distributions of $H\alpha$ and $C\alpha$ chemical shifts for all residues can be found in Figures S1 and S2, respectively. Percentage populations of secondary structure elements calculated using Ramachandran angles are given in Table S1. This information is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: geoffrey.wood@epfl.ch, ursula.roethlisberger@epfl.ch.

■ ACKNOWLEDGMENT

This work was supported by the Swiss National Science Foundation through grant 200020-116294. We also acknowledge generous computing time from the HPC department at the EPF Lausanne, and we thank Professor M. G. Zagorski for providing us with the experimentally determined chemical shifts of $A\beta(1-42)$.

■ REFERENCES

- (1) Selkoe, D. J. *J. Physiol. Rev.* **2001**, *81*, 741–766.
- (2) Hardy, J.; Selkoe, D. J. *Science* **2002**, *297*, 353–356.
- (3) Shoji, M.; Golde, T. E.; Ghiso, J.; Cheung, T. T.; Estus, S.; Shaffer, L. M.; Cai, X. D.; McKay, D. M.; Tintner, R.; Frangione, B.; Younkin, S. G. *Science* **1992**, *258*, 126–129.
- (4) Selkoe, D. J. *Trends Cell Biol.* **1998**, *8*, 447–453.
- (5) Kang, J.; Lemaire, H. G.; Unterbeck, A.; Salbaum, J. M.; Masters, C. L.; Grzeschik, K. H.; Multhaup, G.; Beyreuther, K.; Mullerhill, B. *Nature* **1987**, *325*, 733–736.
- (6) Gregory, G. C.; Halliday, G. M. *Neurotox. Res.* **2005**, *7*, 29–41.
- (7) Lambert, M. P.; Barlow, A. K.; Chromy, B. A.; Edwards, C.; Freed, R.; Liosatos, M.; Morgan, T. E.; Rozovsky, I.; Trommer, B.; Viola, K. L.; Wals, P.; Zhang, C.; Finch, C. E.; Kraft, G. A.; Klein, W. L. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6448–6453.
- (8) Walsh, D. M.; Hartley, D. M.; Kusumoto, Y.; Fezoui, Y.; Condron, M. M.; Lomakin, A.; Benedek, G. B.; Selkoe, D. J.; Teplow, D. B. *J. Biol. Chem.* **1999**, *274*, 25945–25952.
- (9) Hartley, D. M.; Walsh, D. M.; Ye, C. P.; Deihl, T. D.; Vasquez, S.; Vassilev, P. M.; Teplow, D. B.; Selkoe, D. J. *J. Neurosci.* **1999**, *19*, 8876–8884.
- (10) Barrow, C. J.; Zagorski, M. G. *Science* **1991**, *253*, 179–182.
- (11) Shao, H. Y.; Jao, S. C.; Ma, K.; Zagorski, M. G. *J. Mol. Biol.* **1999**, *285*, 755–773.
- (12) Ma, K.; Clancy, E. L.; Zhang, Y. B.; Ray, D. G.; Wollenberg, K.; Zagorski, M. G. *J. Am. Chem. Soc.* **1999**, *121*, 8698–8706.
- (13) Crescenzi, O.; Tomaselli, S.; Guerrini, R.; Salvadori, S.; D'Ursi, A. M.; Temussi, P. A.; Picone, D. *Eur. J. Biochem.* **2002**, *269*, 5642–5648.
- (14) Tomaselli, S.; Esposito, V.; Vangone, P.; van Nuland, N. A. J.; Bonvin, A. M. J. J.; Guerrini, R.; Tancredi, T.; Temussi, P. A.; Picone, D. *Chembiochem* **2006**, *7*, 257–267.
- (15) Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzutkin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16742–16747.
- (16) Hou, L. M.; Shao, H. Y.; Zhang, Y. B.; Li, H.; Menon, N. K.; Neuhaus, E. B.; Brewer, J. M.; Byeon, I. J. L.; Ray, D. G.; Vitek, M. P.; Iwashita, T.; Makula, R. A.; Przybyla, A. B.; Zagorski, M. G. *J. Am. Chem. Soc.* **2004**, *126*, 1992–2005.

- (17) Danielsson, J.; Jarvet, J.; Damberg, P.; Graslund, A. *FEBS J.* **2005**, *272*, 3938–3949.
- (18) Kirkpatridge, M. D.; Condon, M. M.; Teplow, D. B. *J. Mol. Biol.* **2001**, *312*, 1103–1119.
- (19) Bitan, G.; Kirkpatridge, M. D.; Lomakin, A.; Vinters, S. S.; Benedek, G. B.; Teplow, D. B. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 330–335.
- (20) Lee, J. P.; Stimson, E. R.; Ghilardi, J. R.; Mantyh, P. W.; Lu, Y. A.; Felix, A. M.; Llanos, W.; Behbin, A.; Cummings, M.; Vancriekinge, M.; Timms, W.; Maggio, J. E. *Biochemistry* **1995**, *34*, 5191–5200.
- (21) Zhang, S.; Iwata, K.; Lachenmann, M. J.; Peng, J. W.; Li, S.; Stimson, E. R.; Lu, Y.; Felix, A. M.; Maggio, J. E.; Lee, J. P. *J. Struct. Biol.* **2000**, *130*, 130–141.
- (22) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647–1651.
- (23) Dalgarno, D. C.; Levine, B. A.; Williams, R. J. P. *Biosci. Rep.* **1983**, *3*, 443–452.
- (24) Deslauriers, R.; Smith, I. C. P. In *Topics in Carbon-13 NMR Spectroscopy*; Levy, G. C., Eds.; John Wiley & Sons: New York, 1976; Vol 2, pp 1–80.
- (25) Szilagyi, L. *Proc. Nucl. Magn. Reson. Spectrosc.* **1995**, *27*, 325–443.
- (26) Raffa, D. F.; Rauk, A. *J. Phys. Chem B* **2007**, *111*, 3789–3799.
- (27) Sgourakis, N. G.; Yan, Y.; McCallum, S. A.; Wang, C.; Garcia, A. E. *J. Mol. Biol.* **2007**, *368*, 1448–1457.
- (28) Yang, M. F.; Teplow, D. B. *J. Mol. Biol.* **2008**, *384*, 450–464.
- (29) Takeda, T.; Klimov, D. K. *J. Phys. Chem B* **2009**, *113*, 6692–6702.
- (30) Vitalis, A.; Caflisch, A. *J. Mol. Biol.* **2010**, *403*, 148–165.
- (31) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–725.
- (32) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins: Struct. Funct. Bioinf.* **2002**, *46*, 24–33.
- (33) Trzesniak, D.; Glattli, A.; Bernhard Jaun, B.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **2005**, *127*, 14320–14329.
- (34) Glattli, A.; van Gunsteren, W. F. *Angew. Chem., Int. Ed. Engl.* **2004**, *43*, 6312–6312.
- (35) van Gunsteren, W. F.; Dolenc, J.; Mark, A. E. *Curr. Opin. Struct. Biol.* **2008**, *18*, 149–153.
- (36) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (37) Sanbonmatsu, K. Y.; Garcia, A. E. *Proteins: Struct. Funct. Bioinform.* **2002**, *46*, 225–234.
- (38) Rathore, N.; Chopra, M.; de Pablo, J. J. *J. Chem. Phys.* **2005**, *122*, 024111.
- (39) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058–9067.
- (40) Hritz, J.; Oostenbrink, C. *J. Chem. Phys.* **2008**, *128*, 144121.
- (41) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042–6051.
- (42) Okur, A.; Roe, D. R.; Cui, G. L.; Hornak, V.; Simmerling, C. *J. Chem. Theory Comput.* **2007**, *3*, 557–568.
- (43) Roitberg, A. E.; Okur, A.; Simmerling, C. *J. Phys. Chem. B* **2007**, *111*, 2415–2418.
- (44) Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. *J. Chem. Theory Comput.* **2006**, *2*, 420–433.
- (45) Okur, A.; Wickstrom, L.; Simmerling, C. *J. Chem. Theory Comput.* **2008**, *4*, 488–498.
- (46) Temperature generator for REMD-simulations. <http://folding.bmc.uu.se/remd/index.php> (accessed March, 2008).
- (47) Patriksson, A.; van der Spoel, D. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073–2077.
- (48) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Tsui, V.; Gohlke, H.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. *AMBER 9*; University of California: San Francisco, CA, 2006.
- (49) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (50) Wang, T.; Wade, R. C. *J. Chem. Theory Comput.* **2006**, *2*, 140–148.
- (51) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins* **2005**, *61*, 704–721.
- (52) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 372–341.
- (53) Although the sampling frequencies for the two simulations are different (2 ps versus 12 ps), the figure does not change significantly if the number of configurations is adjusted to reflect equal sampling frequencies (12 ps). The total number of configurations (~80 000) is kept in this analysis for purposes of clarity.
- (54) Osapay, K.; Case, D. A. *J. Am. Chem. Soc.* **1991**, *113*, 9436–9444.
- (55) Xu, X. P.; Case, D. A. *J. Biomol. NMR* **2001**, *21*, 321–333.
- (56) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (57) Wishart, D. S.; Case, D. A. *Methods Enzymol.* **2001**, *338*, 3–34.
- (58) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.
- (59) Braun, D.; Wider, G.; Wuthrich, K. *J. Am. Chem. Soc.* **1994**, *116*, 8466–8469.
- (60) Merutka, G.; Dyson, H. J.; Wright, P. E. *J. Biomol. NMR* **1995**, *5*, 14–24.
- (61) Wishart, D. S.; Bigam, A.; Holm, R. S.; Sykes, B. D. *Biomol. NMR* **1995**, *5*, 67–81.
- (62) Beck, D. A. C.; Alonso, D. O. V.; Inoyama, D.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 12259–12264.
- (63) Ting, D.; Wang, G. L.; Shapovalov, M.; Mitra, R.; Jordan, M. I.; Dunbrack, R. L. *PLOS Comput. Biol.* **2010**, *6*, e1000763.
- (64) Smith, L. J.; Fiebig, K. M.; Schwalbe, H.; Dobson, C. M. *Curr. Biol.* **1996**, *1*, R96–R106.
- (65) Wasylshen, R. E.; Tomlinson, G. *Biochem. J.* **1975**, *147*, 605–607.
- (66) Wang, L. Y.; Eghbalnia, H. R.; Markley, J. L. *J. Biomol. NMR* **2006**, *35*, 155–165.
- (67) Frishman, D.; Argos, P. *Proteins* **1995**, *23*, 566–579.
- (68) We do not include the first and last residues in this calculation because the Ramachandran angles are not defined for these residues.