



Published in final edited form as:

J Chem Theory Comput. 2015 March 10; 11(3): 1315–1329. doi:10.1021/ct5010966.

Molecular Dynamics Simulations of 441 Two-Residue Peptides in Aqueous Solution: Conformational Preferences and Neighboring Residue Effects with the Amber ff99SB-ildn-nmr Force Field

Shuxiang Li, Casey T. Andrews, Tamara Frembgen-Kesner, Mark S. Miller, Stephen L. Siemonsma, Timothy D. Collingsworth, Isaac T. Rockafellow, Nguyet Anh Ngo, Brady A. Campbell, Reid F. Brown, Chengxuan Guo, Michael Schrodt, Yu-Tsan Liu, and Adrian H. Elcock*

Department of Biochemistry, University of Iowa, Iowa City, IA 52242

Abstract

Understanding the intrinsic conformational preferences of amino acids and the extent to which they are modulated by neighboring residues is a key issue for developing predictive models of protein folding and stability. Here we present the results of 441 independent explicit-solvent MD simulations of all possible two-residue peptides that contain the 20 standard amino acids with histidine modeled in both its neutral and protonated states. $^3J_{\text{HNH}\alpha}$ coupling constants and $\delta_{\text{H}\alpha}$ chemical shifts calculated from the MD simulations correlate quite well with recently published experimental measurements for a corresponding set of two-residue peptides. Neighboring residue effects (NREs) on the average $^3J_{\text{HNH}\alpha}$ and $\delta_{\text{H}\alpha}$ values of adjacent residues are also reasonably well reproduced, with the large NREs exerted experimentally by aromatic residues, in particular, being accurately captured. NREs on the secondary structure preferences of adjacent amino acids have been computed and compared with corresponding effects observed in a coil library and the average β -turn preferences of all amino acid types have been determined. Finally, the intrinsic conformational preferences of histidine, and its NREs on the conformational preferences of adjacent residues, are both shown to be strongly affected by the protonation state of the imidazole ring.

Introduction

Molecular dynamics (MD) simulations using pairwise-additive force fields have been used for many years to model the conformational dynamics of biological macromolecules.¹ While current protein force fields generally work quite well, one important recent trend has been

*adrian-elcock@uiowa.edu.

Supporting Information

Plots of standard deviations of Ramachandran maps; comparisons of Ramachandran maps from select repeat simulations; correlation of simulation and experimental $^3J_{\text{HNH}\alpha}$ values computed using residue-specific Karplus parameters; Ramachandran maps from all 441 simulations; fractional secondary structure populations from this work compared with previous work on single residues; neighboring residue effects on fractional populations of other secondary structure types; table of optimized residue-specific Karplus parameters. This material is available free of charge via the Internet at <http://pubs.acs.org>.

the use of NMR data to thoroughly test force fields and to identify areas for improvement.² For example, combined computational and experimental studies of 3J coupling constants in alanine peptides,³ and subsequent MD studies,^{2b, 4} have explored reweighting of conformations sampled during MD in order to improve agreement with experiment. Comparisons with experimental 3J couplings for these and other peptides^{3, 5} have led to the development of improved backbone dihedral energy terms⁶ for the Amber ff99SB force field.⁷ Improved backbone terms for the same force field have also been developed through iterative comparisons with chemical shift data for proteins,⁸ and an extended version of the force field also parameterized against residual dipolar coupling (RDC) data has been reported.⁹ Importantly, in a large-scale comparison of simulated and experimental chemical shifts and 3J couplings for peptide systems, these modified Amber ff99SB force fields produced the best performances of the many force field and water model combinations tested.^{2c}

Combined applications of computational and experimental methods to peptide systems^{3, 10} can be especially valuable for identifying limitations in simulations.¹¹ Eventually, of course, one might hope that MD force fields will develop to the point at which their predictive abilities are beyond question. Until that time, however, there remains a need for good experimental datasets that can be used to test and further refine force fields.¹¹ One such dataset comes from a very recent 2D NMR study¹² that reported $^3J_{H_{NH}\alpha}$ coupling constants and δ_{HN} and $\delta_{H\alpha}$ chemical shifts for each residue in a comprehensive set of 361 blocked two-residue peptides; that work built upon an earlier 1D study performed by the same authors.¹³ While analysis of these data enabled the authors of that work to compile a comprehensive view of neighboring residue effects (NREs) on both 3J couplings and chemical shifts in peptides, the same dataset also clearly provides an excellent new opportunity to test simulation force fields.

In recent work we have described the use of long explicit-solvent MD simulations to model the interaction thermodynamics of all possible pairs of amino acids; the resulting simulation data provided the basis for deriving a simple coarse-grained simulation force field which we named COFFDROP.¹⁴ Here we describe a similar large-scale effort aimed instead at modeling the conformational landscapes of all possible two-residue peptides. In separate work we use these data to derive coarse-grained backbone potential functions for incorporation into COFFDROP, but here we focus on analysis of the conformational properties of the peptides and how the simulation results compare with experimental data. We use a combination of the Amber ff99SB-ildn-nmr force field^{7-8, 15} and the TIP4P-Ew water model¹⁶ to perform all simulations given the success that this combination has achieved in reproducing NMR observables for peptide systems.^{2c} We employ the same procedures to compute $^3J_{H_{NH}\alpha}$ coupling constants¹⁷ and $\delta_{H\alpha}$ chemical shifts¹⁸ as used in that previous study^{2c} and we compare our results to the corresponding experimental data for as many of the peptides as possible.¹² The comparisons indicate a reasonably high level of correspondence between the simulations and experiment and provide the first truly comprehensive description obtained from explicit-solvent MD simulations of the effects of neighboring residues on the conformational preferences of amino acids.

Materials and Methods

Molecular dynamics (MD) simulation protocol

The systems studied here consist of all possible two-residue peptides, each immersed in its own $35 \times 35 \times 35$ Å box containing explicit solvent molecules. Since one of our principal objectives was to sample the conformational behavior that peptides would exhibit when embedded within protein sequences, each two-residue peptide was capped (blocked) with an acetyl group (Ace) at the N-terminus and an N-methyl group (Nme) at the C-terminus in order to mimic the adjacent residues that would be present in proteins. Every two-residue peptide that was simulated was therefore of the form Ace-Xaa-Yaa-Nme, where Xaa and Yaa represent amino acids. We note that throughout this manuscript we use the term “two-residue peptide” to describe these molecules instead of “dipeptide” as the latter is used in multiple ways in the literature (i.e. sometimes it is meant to describe a single amino acid with peptide bonds at both the N- and C-termini – as in the alanine dipeptide¹⁹ – and sometimes it is used to mean two amino acids connected by a single peptide bond¹²). Since we model histidine (His) in both its neutral and protonated states, a total of $21 \times 21 = 441$ possible two-residue peptides have been simulated.

All simulations were performed using the MD software package GROMACS version 4.5.1²⁰ and all systems were modeled using the Amber ff99SB-ildn-nmr^{7-8, 15} force field together with the TIP4P-Ew¹⁶ water model. Systems were first energy minimized using steepest descent minimization for 1000 steps, gradually heated to 298 K over the course of 350 ps, and then equilibrated for a period of 1 ns. Production simulations were then carried out in the NPT ensemble for 300 ns, with the temperature maintained at 298 K using the Nosé-Hoover thermostat,²¹ and the pressure maintained at 1 atm using the Parrinello-Rahman barostat.²² A cutoff of 10 Å was applied to short-range nonbonded interactions and the PME method²³ was used to calculate all long-range electrostatic interactions. Covalent bonds were constrained to their equilibrium lengths using the LINCS algorithm,²⁴ allowing a 2.5 fs time step to be employed. Coordinates of the solutes were collected every 0.1 ps giving a total of 3 million structures for analysis for each two-residue peptide.

Analysis and convergence of molecular dynamics simulations

To assess the completeness of sampling in the MD simulations, we calculated the standard deviations of the population distribution for the backbone dihedral angles ϕ and ψ . For each two-residue peptide, the 300 ns simulation period was divided into three time intervals (0-100, 100-200, 200-300 ns). The GROMACS utility `g_rama` was used to calculate the backbone dihedral angles ϕ and ψ of each residue during the course of each of the three time intervals. 2D histograms of the sampled ϕ and ψ values were then constructed using a bin size of 5° and the standard deviation for each bin was calculated over the three 100 ns intervals. The standard deviations of all bins were summed to obtain total standard deviations for both the N-terminal and C-terminal positions of each two-residue peptide and the two results were averaged to give a single number for each peptide. For the peptides that were found to give the largest average standard deviations (Ile-Pro, Ile-Tyr, Trp-Arg), two independent additional MD simulations, also of 300 ns duration, were performed in order to demonstrate that sampling is likely to be sufficient. A full description of our analysis of

convergence of the simulations is provided in the Supporting Information (Figures S1 & S2).

To aid comparison of the conformational preferences of different peptides, each 2D histogram of ϕ and ψ values was converted into free energy form using the relation $G = -RT \ln (f/f_{\max})$, where f is the frequency of sampling a particular bin in the 2D ϕ, ψ histogram and f_{\max} is the maximum frequency found in any of the bins. Since each of the 441 two-residue peptides has two residues, a total of 882 2D free energy distributions were constructed in this way. The data in these histograms were also expressed in an alternative, more concise form as fractional populations of the four major backbone conformations (right-handed helix α , near right-handed helix α' , β -strand, and PPII) computed by summing the populations in the relevant bins of the histograms; definitions for these secondary structure conformations were taken from the literature:²⁵ α : $-100^\circ < \phi < 0^\circ, -80^\circ < \psi < 40^\circ$; α' : $-160^\circ < \phi < -100^\circ, -80^\circ < \psi < 60^\circ$; β : $-180^\circ < \phi < -100^\circ, 70^\circ < \psi < 180^\circ$; PPII: $-100^\circ < \phi < 0^\circ, 70^\circ < \psi < 180^\circ$.

To obtain average 2D free energy maps and average secondary structure populations that describe the intrinsic conformational preferences of each type of amino acid when at the N- and C-terminal positions, we summed all frequency distributions for peptides that contained the residue of interest, excluding those peptides that had Gly or Pro as neighbors. For example, to calculate the average fractional populations of the different secondary structures for an Ala residue at the N-terminal position, the fractional populations of the N-terminal residue in the peptides Ala-Ala, Ala-Cys, Ala-Gln ... Ala-Val (excluding Ala-Gly and Ala-Pro) were all combined.

To describe the average effects exerted by each type of amino acid on the properties of neighboring amino acids (i.e. neighboring residue effects: NREs), we used the same procedure used by the Cho group to compute NREs on $^3J_{\text{HNH}\alpha}$ and $\delta_{\text{H}\alpha}$ values,¹² which we illustrate as follows. To compute the average NRE exerted by a C-terminal Trp on the α population at the N-terminal position, for example, we consider in turn each of the peptides Ala-Trp, Cys-Trp, Gln-Trp ... Val-Trp (excluding Gly-Trp and Pro-Trp). For Ala-Trp, we calculate the difference between: (a) the α population at the N-terminal Ala in Ala-Trp, and (b) the mean α population at the same position in all peptides Ala-Ala, Ala-Cys, Ala-Gln ... Ala-Val (excluding Ala-Gly and Ala-Pro). We carry out similar calculations for the Cys in Cys-Trp, the Gln in Gln-Trp etc and average the results to obtain the mean change in α population at the N-terminal position induced by the presence of a C-terminal Trp. Similar calculations can be performed for all other types of secondary structure populations and for $^3J_{\text{HNH}\alpha}$ and $\delta_{\text{H}\alpha}$ values (see below).

To compute populations of β -turn conformations, which are defined in terms of the backbone conformations of four consecutive residues $i, i+1, i+2,$ and $i+3$, we identify i with the acetyl capping group (Ace), $i+1$ with the N-terminal amino acid, $i+2$ with the C-terminal amino acid, and $i+3$ with the N-methyl capping group (Nme). We consider all possible β -turns that can be formed with peptide bonds in *trans* configurations using definitions provided in the literature;²⁶ note that these definitions depend only on ϕ and ψ of residues i

+1 and i+2 and so can be meaningfully applied to the capped two-residue peptides studied here even though they do not possess complete amino acids at the i and i+3 positions.

Calculation of $^3J_{\text{HNH}\alpha}$ coupling constants

Since the work of Karplus, 3J coupling constants have been interpreted in terms of the dihedral angle connecting the two coupled nuclei.²⁷ To compute $^3J_{\text{HNH}\alpha}$ coupling constants from the simulations, the ϕ backbone dihedral angles of each amino acid were computed for every simulation snapshot and converted using an empirical Karplus relation of the following form: $^3J(\phi) = A \cos^2(\phi + \theta) + B \cos(\phi + \theta) + C$, where θ denotes a phase shift value, and A, B, C represent parameters. In this study, we used the parameterization of Hu and Bax,¹⁷ for which $A = 7.09$, $B = -1.42$, $C = 1.55$, and $\theta = -60^\circ$. While a large number of alternative parameterizations are available,^{3, 17, 28} including at least one that has been derived from MD simulations of proteins,²⁹ the particular parameterization used here has been selected for consistency with the earlier study of the Pande group^{2c} and with other studies;^{2b, 3} other studies have shown that correlation coefficients between simulation and experimental $^3J_{\text{HNH}\alpha}$ values are insensitive to the specific choice of Karplus parameters.^{2c, 30} As is usual,^{2b, 2c} we assume for most calculations that a single parameterization of the Karplus equation applies equally well to all residue types; we also, however, consider the possibility of deriving residue-specific Karplus parameters in Results.

In comparing with the experimental data reported by the Cho group¹² we omit all peptides containing either Asp or Glu. We have elected to do this because the experiments for these peptides were performed at pH values between 4 and 5 where we expect the sidechains to be partially deprotonated, while our simulations were all performed with the Asp and Glu sidechains in fully deprotonated forms (i.e. more appropriate to a pH of 7). For comparison with experiment, we also omit $^3J_{\text{HNH}\alpha}$ coupling constants of Gly residues since the experimental data do not identify which of the two H α protons is probed: experimentally, it has been shown that the $^3J_{\text{HNH}\alpha}$ values for the two H α atoms can differ enormously (e.g. by 5.1 Hz for Gly79 in staphylococcal nuclease^{28a}).

Calculation of chemical shifts

Calculations of $\delta_{\text{H}\alpha}$ chemical shifts for both residues in each two-residue peptide were performed using the SPARTA+ program;¹⁸ owing to the greater computational expense of these calculations they were carried out only on simulation snapshots sampled at 1 ps intervals (a total of 30,000 snapshots for each peptide). SPARTA+, in common with other chemical shift prediction algorithms,³¹ uses information on the ϕ , ψ angles at neighboring residues i-1 and i+1 when calculating the chemical shifts at position i. This makes use of the program to calculate chemical shifts in our peptides challenging, since the first residue is capped with an acetyl group and so does not have a complete neighbor at the i-1 position, and the second residue is capped with a N-methyl group and so does not have a complete neighbor at the i+1 position. This problem can be mitigated by renaming the two capping groups as Ala residues: this is done by renaming the appropriate atoms and residues in the structure (pdb) file of each snapshot. This simple approach allows us to obtain quite good results for the $\delta_{\text{H}\alpha}$ chemical shifts, although it works less well for δ_{HN} chemical shifts, which are in any case predicted more poorly by SPARTA+;¹⁸ we therefore report only the former

here. We also note that while a chemical shift prediction program has recently been reported that has been specifically parameterized using MD data on proteins,³² the peptides studied here are apparently too short to be used with that program.

Results

Comparison of experimental and simulated $^3J_{\text{HNH}_\alpha}$ coupling constants

Before analyzing the conformational behavior of the peptides in detail it is first important to determine the extent to which they are consistent with experimental data.¹² Figure 1A compares the simulated and experimental $^3J_{\text{HNH}_\alpha}$ coupling constants for all peptides that do not contain an Asp or Glu residue; owing to uncertainties in experimental assignments we also omit coupling constants for Gly residues (see Methods). The overall level of agreement between the 544 computed and experimental $^3J_{\text{HNH}_\alpha}$ coupling constants is quite reasonable: $r^2 = 0.58$. When this figure is replotted including data for Gly residues the r^2 actually improves significantly to 0.68, but as expected, when data for peptides containing Asp or Glu are added, the r^2 worsens again to 0.64 (not shown). The degree of agreement with experiment obtained here is similar to that reported recently³⁰ when the same force field and Karplus parameterization was applied to $^3J_{\text{HNH}_\alpha}$ coupling constants of 19 dipeptides (i.e. capped single amino acids):³³ the Pearson correlation coefficient between simulation and experiment is 0.76 here, while that reported previously for the dipeptides is 0.62.³⁰ In addition, however, the simulations reported here also reproduce the experimental observation¹² that, within any given peptide, the $^3J_{\text{HNH}_\alpha}$ coupling constants of residues at the C-terminal position are generally significantly higher than those at the N-terminal position: in Figure 1A, C-terminal $^3J_{\text{HNH}_\alpha}$ coupling constants (red triangles) are consistently shifted upwards relative to N-terminal $^3J_{\text{HNH}_\alpha}$ coupling constants (blue circles).

With the exception of the Asp/Glu-containing peptides, all of the experiments were performed at pH 2;¹² at this pH value we expect all histidine sidechains to be in a fully protonated form. We should therefore expect to obtain better agreement with experiment from those simulations that modeled the protonated form of the residue (Hip) than from those that modeled the neutral form (His). Figures 1B and 1C show that this expectation is fulfilled. Figure 1B compares the computed and experimental $^3J_{\text{HNH}_\alpha}$ coupling constants for those peptides that contain histidines using data from simulations that employed the neutral His residue type; the agreement is clearly very poor ($r^2 = 0.15$). Figure 1C, on the other hand, shows a corresponding comparison using data obtained from simulations that employed the charged Hip residue type; in this case, the agreement is much better ($r^2 = 0.67$), thereby arguing that the simulations are correctly sensing the overall effects of the sidechain's charge state on the conformational properties of the peptide backbone.

The above results indicate that the simulations perform quite well in reproducing the experimental $^3J_{\text{HNH}_\alpha}$ coupling constants. As might be expected, however, when the data are examined in more detail, significant discrepancies between simulation and experiment emerge. Figure 2A compares the average computed and experimental $^3J_{\text{HNH}_\alpha}$ coupling constants of each type of amino acid when at the N-terminal position of a two-residue peptide. The agreement between the two sets of data is far from perfect ($r^2 = 0.33$): in particular, relative to their experimental values, the simulation-average $^3J_{\text{HNH}_\alpha}$ coupling

constants for Ala, Leu, Trp and Val are too high while those of Asn, Cys, Ser and Thr are too low. Again, these results are in line with those reported recently for simulations of capped single amino acids using the same force field:³⁰ in that work, Ala and Val were again noted as being too high, while Asn, Cys and Ser were too low. Similar discrepancies between the average computed and experimental $^3J_{\text{HNH}_\alpha}$ coupling constants are also apparent at the C-terminal position of the two-residue peptides (Figure 2B), although here the agreement between simulation and experiment is somewhat higher ($r^2 = 0.45$).

As is considered further below, discrepancies between the computed and experimental $^3J_{\text{HNH}_\alpha}$ coupling constants could be caused by a number of different factors, the most likely of which may be the accuracy of the force field. A comparison between simulation and experiment that we anticipate might be less dependent upon potential errors in the force field is to compare the effect of a given residue type on the average $^3J_{\text{HNH}_\alpha}$ coupling constant at a *neighboring* position. Figure 2C shows the effects that each type of amino acid, when present at the N-terminal position, has on the average $^3J_{\text{HNH}_\alpha}$ coupling constant at the C-terminal position of the same peptide: red bars show the experimental values calculated as recently described¹² (see Methods), while blue bars show the corresponding simulation values. As noted by the Cho group,¹² the most noticeable effect in the experimental data is that aromatic residues at the N-terminal position increase the average $^3J_{\text{HNH}_\alpha}$ coupling constants of C-terminal residues by ~ 0.3 Hz. Importantly, this effect is nicely reproduced by the simulations: we find that Phe, Trp and Tyr at the N-terminus cause upward shifts of the average $^3J_{\text{HNH}_\alpha}$ coupling constant at the C-terminus of 0.32, 0.51 and 0.50 Hz, respectively. In fact, the overall level of agreement between the computed and experimental neighboring residue effects (NREs) of N-terminal amino acids is quite high: $r^2 = 0.80$, and for 14 out of the 17 residue types studied, the sign of the NRE is correctly reproduced.

For the NREs caused by amino acids at the C-terminal position (Figure 2D), the agreement is considerably less good but still reasonable ($r^2 = 0.39$). The largest effects seen in the experimental data are provided by the aromatic amino acids Phe and Tyr, which increase the average $^3J_{\text{HNH}_\alpha}$ coupling constants of N-terminal residues by ~ 0.3 Hz, and the positively charged amino acids Arg and Lys, which decrease them by ~ 0.2 Hz. Both of these effects are correctly captured by the simulations. Nevertheless, other significant effects that are predicted by the simulations are not apparent in the experimental data. Gly and Trp, for example, are both predicted to exert positive NREs on the $^3J_{\text{HNH}_\alpha}$ values at the N-terminus, but these are not supported by the experimental data; Hip, on the other hand, is predicted to cause a large negative NRE, and while this is qualitatively supported by the experimental data, the experimental effect is significantly smaller in magnitude.

Origins of discrepancies between experimental and simulated $^3J_{\text{HCH}_\alpha}$ coupling constants

While the overall level of agreement between the computed and experimental $^3J_{\text{HNH}_\alpha}$ coupling constants is quite good, it is clear from the scatter in Figure 1A and from the more detailed comparisons in Figure 2 that there are notable discrepancies. While the obvious and most likely potential source of disagreement is inadequacies of the simulation force field, there are other factors that might play contributing roles. One possible factor is the degree of

convergence of the simulations: it might easily be imagined, for example, that simulations that are less completely sampled might produce worse agreement with experimental $^3J_{\text{HNH}_\alpha}$ coupling constants than simulations that are well sampled. To examine this issue we first arranged the peptides studied experimentally in order of increasing standard deviation of their simulated Ramachandran maps (see Supporting Information). We then computed correlation coefficients between the computed and experimental $^3J_{\text{HNH}_\alpha}$ coupling constants for each of the 10 different groups of peptides with ranks in the range 0-10%, 10-20%, ... 90-100%; we also computed the mean absolute difference between the computed and experimental $^3J_{\text{HNH}_\alpha}$ coupling constants for each ranked group of peptides. The resulting correlation coefficients and average absolute errors are plotted versus the average rank in each group in Figure 3A. There is no compelling relationship between the extent of conformational sampling and the degree of correlation between the computed and experimental $^3J_{\text{HNH}_\alpha}$ coupling constants. In fact, surprisingly, those peptides for which sampling of the Ramachandran maps is probably the best (left-most datapoints in Figure 3A) produce a somewhat worse correlation with experiment (blue circles), and a higher error (red triangles) than those for which sampling is likely to be worst (right-hand side of Figure 3A).

A second possible factor that might contribute to differences between the computed and experimental $^3J_{\text{HNH}_\alpha}$ coupling constants is the identity of the capping group employed at the C-terminus of the peptides. In the experiments of the Cho group¹² a simple amide group (NH_2) was used at the C-terminus, whereas in the simulations an N-methyl group (NHCH_3) was used since a parallel objective of our work was to generate raw data for deriving coarse-grained potential functions for proteins (see Introduction & Methods); at the N-terminus, both the experiments and the simulations used an acetyl group. To test the potential effects of the difference at the C-terminus we randomly selected 20 peptides and performed replicate 300 ns MD simulations using the same amide C-terminal capping group used in the experiments. As shown in Figure 3B, the $^3J_{\text{HNH}_\alpha}$ coupling constants calculated from simulations using the NH_2 capping group (residue-type NHE) are effectively identical with those calculated from simulations using the NHCH_3 capping group (residue-type NME): the difference between the two corresponding $^3J_{\text{HNH}_\alpha}$ coupling constants averages 0.09 ± 0.09 Hz, with the largest discrepancy (0.42 Hz) being for Met-Trp which is likely to be comparatively poorly sampled (see Supporting Information). This suggests that differences in the termini are not likely to be a significant cause of discrepancies with experiment.

With the above two issues dealt with, we can consider whether inaccuracies in the force field are likely to be responsible for differences between the simulated and experimental $^3J_{\text{HNH}_\alpha}$ values. One very real possibility is that differences in the intrinsic conformational preferences of different amino acids might be incorrectly described by the force field.^{30, 34} An indication that this might be the case is provided by Figure 3C, which shows that quite different regression lines are obtained when the computed and experimental $^3J_{\text{HNH}_\alpha}$ values are compared separately for aliphatic (Ile, Leu, Val), aromatic (Phe, Trp, Tyr), 'polar' (Cys, Ser, Thr) and positively charged amino acids (Arg, His, Lys): in particular, the regression line for 'polar' amino acids (yellow) is shifted down by ~ 0.5 Hz relative to that for aliphatic amino acids (blue). We note in passing that an alternative

explanation that might be considered for this result is that different parameterizations of the Karplus equation might be required for different types of amino acids. Figure S3 shows that the agreement between computed and experimental $^3J_{\text{HNH}_\alpha}$ couplings can be significantly improved if Karplus parameters are optimized separately for individual residue types; a full listing of the resulting residue-specific Karplus parameters is provided in Table S1.

Comparison of experimental and simulated chemical shifts

In addition to $^3J_{\text{HNH}_\alpha}$ coupling constants, the Cho group also reported δ_{H_α} and δ_{HN} chemical shifts for both residues in 361 two-residue peptides. To compare with their data, we used the SPARTA+ program¹⁸ (see Methods). Figure 4A compares the computed and experimental δ_{H_α} chemical shifts for all peptides that do not contain an Asp or Glu residue; as before, we also omit values for Gly residues owing to assignment ambiguities (see Methods). The correlation between simulation and experiment is good although the simulated values are clearly shifted upwards by ~ 0.3 ppm relative to experiment: it is possible that this upward shift is a consequence of our artificially redefining the capping groups as alanine residues (see Methods). While the generally high level of agreement is an encouraging result, it is important to determine the extent to which this results from correct sampling of conformations in the simulations: the chemical shift calculations use a neural network that takes as input the identity of the amino acids involved,¹⁸ so it is possible that this information is instead primarily responsible for the good agreement. To explore this issue, we repeated the chemical shift calculations of all peptides with each built in the following idealized backbone conformations: α -helix, β -strand and PPII. If correct sampling of the backbone distributions is responsible for the good agreement shown in Figure 4A then we should expect to obtain poorer agreement with experiment when a single backbone conformation is used in the calculations. Figure 4B compares the computed δ_{H_α} values of all peptides with the corresponding experimental values. All three secondary structures produce good correlations with experiment, with that of the α -helical conformation, surprisingly, being the highest ($r^2 = 0.77$). Importantly, however, the correlation coefficient obtained from the MD simulations is considerably higher ($r^2 = 0.84$), thereby suggesting that MD's reasonably realistic sampling of different backbone conformations plays a significant role in determining the agreement with experiment.

In the same way that NREs on $^3J_{\text{HNH}_\alpha}$ coupling constants can be compared with experiment, it is possible to compare NREs on the average δ_{H_α} chemical shifts. Figure 4C shows the effects that each type of amino acid, when present at the N-terminal position, has on the average δ_{H_α} coupling constant at the C-terminal position of the same peptide: red bars show the experimental values calculated as recently described,¹² while blue bars show the corresponding simulation values. Experimentally, the largest NREs are again due to the aromatic residues, each of which causes a decrease in the average δ_{H_α} chemical shift at a C-terminal position of ~ 0.1 ppm. Qualitatively, these effects are correctly reproduced by the simulations – the three largest NREs are all due to the aromatic residues – but there are other effects predicted by the simulations that are not mirrored in the experiments: e.g. Ile and Val are both predicted to cause large increases in the average δ_{H_α} chemical shifts at the C-terminal position, but much smaller effects are seen experimentally. Despite these discrepancies, the overall level of agreement between the computed and experimental

neighboring residue effects (NREs) of N-terminal amino acids is quite high: $r^2 = 0.70$, with the sign of the NRE being correct for 14 out of the 17 residue types studied. As was the case with the $^3J_{\text{HNH}\alpha}$ coupling constants, the agreement for NREs exerted by C-terminal residues on the average $\delta_{\text{H}\alpha}$ chemical shifts at the N-terminal position is lower (Figure 4D): $r^2 = 0.40$, even though the sign of the NREs is correctly predicted for 13 out of 17 residue types.

Average intrinsic backbone conformational preferences of each amino acid

We proceed now with an analysis of the conformational distributions of all 441 two-residue peptides predicted by the simulations. The individual Ramachandran maps (ϕ/ψ distributions) of all peptides are plotted separately for residues at the N-terminal and C-terminal positions in Figures S4 and S5 respectively. Composite maps that describe the intrinsic backbone conformational preferences of each type of amino acid averaged over all possible types of neighboring residues (see Methods) are shown in Figures 5 and 6 for N-terminal and C-terminal positions respectively. Since in all cases we average only over those neighboring residues that are not Gly or Pro, each of the results shown in Figures 5 and 6 represents an average of $19 \times 300 \text{ ns} = 5.7 \mu\text{s}$ of simulation data; we therefore expect them to be more reliable than Ramachandran maps of any individual system. Figures 7A and 7B show the corresponding average fractional populations of the four major backbone conformations that we consider here (α , α' , β , PPII) for each residue type at the N-terminal and C-terminal positions, respectively. As was the case with the $^3J_{\text{HNH}\alpha}$ coupling constants described earlier, the average fractional populations of ($\alpha + \alpha'$), β and PPII obtained from the present simulations are very similar to those reported recently from replica-exchange simulations of 19 dipeptides (capped single residues) performed using the same force field:³⁴ $r^2 = 0.91$ (Figure S6).

Generally speaking, for all non-Pro and non-Gly residue types, the PPII conformation (red squares) is the most preferred conformation, with β -strand (yellow downward triangles) being the next most preferred conformation (Figures 7A and 7B). Amino acids with less clear preferences for PPII include the negatively charged Asp and Glu, protonated His (Hip) and the small polar amino acids Thr and Ser. Asp and Glu show an increased preference for α -helical conformations (blue circles) at both the N- and C-terminal positions, and when found at the N-terminal position (Figure 7A), their α -helical populations are competitive with those of the PPII conformation. For Hip, on the other hand, the preferred conformation in the simulations is the β -strand, regardless of whether it is at the N- or C-terminal position. Other residues showing comparatively high preferences for the β -strand conformation when at the C-terminal position include the three aromatic amino acids (Phe, Trp and Tyr), Arg and Val. One final notable effect is that while the α conformation is clearly preferred over α' for residues at the N-terminal position (compare blue circles with green upward triangles), this preference largely disappears for residues at the C-terminal position (Figure 7B).

Neighboring residue effects on average backbone conformational distributions

In addition to examining the intrinsic conformational preferences of each type of amino acid, it is also of interest to compute the average effects of each type of amino acid on the conformational preferences of their neighboring residues (see Methods). Figure 7C shows

the effects that each type of amino acid, when present at the N-terminal position, has on the average fractional population of each type of secondary structure at the C-terminal position of the same peptide: positive numbers indicate an increase in the average fractional population, while negative numbers indicate a decrease. Perhaps the most obvious effect is that the presence of an aromatic residue (Phe, Trp, Tyr; also His) at the N-terminal position strongly decreases the chances of finding a PPII conformation at the C-terminal position (red squares) and, with the exception of His, also has the surprising added effect of increasing the preference for the α' conformation (green upward triangles). Importantly, it is this increase in the α' population at the C-terminal position that is most responsible for the increases in the average computed $^3J_{\text{HNH}\alpha}$ coupling constants caused by the aromatic residue types in Figure 2C, not an increase in the β population (see Discussion). A second clear effect is that the presence of the positively charged amino acids Arg and Lys at the N-terminal position increases the average PPII population (red squares) at the C-terminal position while suppressing both the α and α' populations (Figure 7C); the former effect is responsible for the decrease in the average computed $^3J_{\text{HNH}\alpha}$ coupling constants caused by these residue types in Figure 2C. A third significant observation is that very different NREs are exerted by the different protonation states of the histidine sidechain: the neutral form, His, increases the α and α' populations (blue circles and green upward triangles, respectively) at the expense of β and PPII, while the protonated form, Hip, does the opposite.

Figure 7D is a corresponding plot showing the effects that each amino acid type, when present at the C-terminal position, has on secondary structure populations at the N-terminal position. The presence of a Pro at the C-terminal position strongly decreases the chances of finding a N-terminal residue in either the α or α' conformations and accentuates its preference for a PPII conformation; this is in marked contrast to the very minor effects exerted by a N-terminal Pro on the conformational preferences at the C-terminal position (Figure 7C). Again, the presence of a positively charged amino acid (Arg, Hip, Lys) at the C-terminal position increases the population of the PPII conformation at the N-terminal position, and a corresponding but smaller effect results from the presence of a β -branched amino acid at the C-terminal position (Ile, Val). Again, the two protonation states of histidine produce quite different NREs on the conformational properties of N-terminal residues: a C-terminal Hip increases the average PPII population at an N-terminal position by ~8%, while a C-terminal His decreases it by ~5%.

We can compare all of the above changes in populations with corresponding results obtained from a Bayesian statistical analysis of the 'TCB' coil library carried out by the Dunbrack group.³⁵ Figure 7E compares the change in average PPII populations at the C-terminus associated with each type of N-terminal amino acid; Figure 7F shows corresponding effects exerted by C-terminal amino acids on populations at the N-terminus. Analogous plots for NREs exerted on α -helical and β -strand populations are shown in Figure S7. For NREs exerted by N-terminal residues (Figure 7E), correspondence between the simulations and the PDB analysis is quite poor: a scatter plot of the two datasets has $r^2 = 0.28$ and for only 13 out of 20 residue types are the same qualitative effects predicted. For NREs exerted by C-terminal residues (Figure 7F), on the other hand, agreement is much better: $r^2 = 0.78$, and qualitative agreement is obtained for 15 out of 20 residue types. In particular, the very strong

effect exerted by a C-terminal Pro seen in the PDB analysis is reproduced (as expected) by the simulations, as is the tendency for aromatic residues at the C-terminal position to decrease the average populations of PPII for N-terminal residues, and for the β -branched residues Ile and Val to increase the populations of PPII. The tendency for Arg and Lys to modestly increase the PPII conformation at neighboring residues is also a consistent feature of the simulations and the PDB analysis.

β -turn populations

Since our two-residue peptides contain acetyl groups at the N-terminus and N-methyl groups at the C-terminus they can be considered minimal model systems capable of adopting β -turn conformations (see Methods). The average populations of each of the various types of β -turns that can form with all-*trans* peptide bonds are shown in Figure 8. In the simulations, the populations of β -I (blue circles) and β -VIII (red diamonds) turns are, in general, much higher than those of the remaining turn types β -I', β -II and β -II'; exceptions include the finding that β -II' (yellow squares) is common in peptides that contain a Gly at the N-terminal position (i.e. the "i+1" position of a β -turn involving residues i, i+1, i+2 and i+3) and β -II (green downward triangles) is common in peptides containing a Gly at the C-terminal position (i.e. the "i+2" position). For almost all residue types the population of β -I turns is higher than that of β -VIII; exceptions involve peptides containing N-terminal Arg and Lys residues, for which the populations are similar (Figure 8A) and peptides containing C-terminal Ile and Val residues, for which the β -VIII population is clearly higher (Figure 8B).

The preferences of amino acids for forming different β -turn types are expressed in the form of β -turn potentials²⁶ in Figures 8C and 8D for the N-terminal and C-terminal positions, respectively. Each set of symbols refers to a different type of β -turn; the β -turn 'potential' then indicates the fraction of all occurrences of that turn type found in all simulations that involved a residue of a given type. For example, Figure 8C shows that in the combined set of (infrequently sampled) β -I' conformations (cyan upward triangles), the residue most often found at the N-terminal (i+1) position is Hip, with Gly being a close second; Figure 8D, on the other hand, shows that in the combined set of β -II conformations (green downward triangles) – which are also infrequently sampled – the residue at the C-terminal (i+2) position is almost always a Gly.

For the more frequently sampled turn types (β -I, β -II and β -VIII), the simulation data can be compared with statistics compiled from the PDB.²⁶ Given that 60% of all β -turns formed in the simulations are β -I type turns we show results only for this type of turn; for comparison, 57% of the turns identified by Hutchinson and Thornton in their PDB analysis were of this type.²⁶ Figures 8E and 8F compare the β -turn potentials of β -I type turns reported by Hutchinson and Thornton with those calculated here; Figure 8E shows results for residues at the i+1 (N-terminal) position, Figure 8F shows results for the i+2 (C-terminal) position. For the i+1 position, the correspondence is only modest ($r^2 = 0.40$), and the dramatic preference for a Pro residue apparent in the PDB analysis is not mimicked by the simulations. For the i+2 position, on the other hand, the correspondence is more convincing ($r^2 = 0.63$): in particular, the simulations reproduce the finding that polar sidechains have high β -I type turn

potentials while Pro and the β -branched Ile and Val have low β -I type turn potentials. No closer correspondence is obtained when we compare with an updated β -turn database.³⁶

Discussion

The simulations reported here provide a reasonably comprehensive view of the conformational properties of two-residue peptides simulated with a very widely used force field and water model combination. The comparisons with experimental $^3J_{\text{HNH}_\alpha}$ coupling constants and δ_{H_α} chemical shifts show that the simulations achieve an encouraging level of correspondence with experiment. In particular, neighboring residue effects on the $^3J_{\text{HNH}_\alpha}$ and δ_{H_α} values are quite well reproduced (Figures 2C, 2D, 4C and 4D). Discrepancies with experiment are likely to be primarily a function of inadequacies in the force field since we have found that neither the convergence of sampling (Figure 3A) nor the type of capping group (Figure 3B) appear to play significant roles in dictating the level of correspondence with experiment; it appears quite possible, however, that residue-specific Karplus parameterizations might ultimately be required for quantitative reproduction of experiment (Figure S3; see below). Agreement between the simulations and previous analyses of conformational distributions in the PDB is less good (e.g. Figure 7E), but there are obvious structural differences between two-residue peptides and proteins that would place limits on the level of correspondence to be expected.

Owing to the large number of systems simulated here – 441 – the computational demands of the present study were significant, and it was principally for reasons of expense that we chose to sample conformational behavior using equilibrium MD simulations: the combined simulation time of the simulations reported here amounts to 130 μs . In the future, it should be possible to exploit increases in computer power to make use of more effective sampling techniques that have thus far typically only been used on much smaller numbers of systems. Examples include: (a) the metadynamics³⁷ method used to study single amino acids with a variety of force fields³⁸ – and which showed that there was little consensus among the tested force fields regarding backbone conformational preferences – (b) the Hamiltonian replica-exchange MD (REMD) methods that have been used to study neighbor effects on the conformational behavior of a number of small Ala, Val, or Phe-containing peptides,³⁹ and (c) the more conventional replica exchange methods⁴⁰ used in the iterative development of the recent RSFF1³⁰ and RSFF2³⁴ force fields. Improved sampling would have the added benefit of enabling comparisons between simulation and experiment to be performed on an individual peptide-by-peptide basis with confidence. This could be important given that the Cho group has noted¹² that the NREs exerted by a given type of amino acid can vary significantly depending on the identity of the other amino acid being examined: for example, the NRE exerted by a C-terminal Trp on the $^3J_{\text{HNH}_\alpha}$ of an N-terminal Arg is -0.32 Hz while that exerted on an N-terminal Asn is $+0.27$ Hz.

Reproduction of experimental results

Although the force field and water model combination used in this study has already been shown to be one of the best currently available for reproducing a variety of NMR observables in both peptide^{2c} and protein systems,⁴¹ the simulations reported here provide

an opportunity to explore its ability to reproduce a much larger body of NMR data for two-residue peptides. The generally good reproduction of the experimental $^3J_{\text{HNH}\alpha}$ values for two-residue peptides appears notable since the force field was optimized to reproduce NMR data for proteins, not peptide systems,⁸ and since the original parameterization did not consider scalar coupling constants but focused instead on chemical shifts. This, together with the observation that the NREs on $^3J_{\text{HNH}\alpha}$ and $\delta_{\text{H}\alpha}$ values can also be reasonably well captured (see below) reiterates the point made by others that the force field selected here performs quite well;^{2c} it is to be noted, however, that it has recently been shown to be incapable of producing a stable structure for the GB1 hairpin.³⁴

That said, the agreement between the computed and experimental $^3J_{\text{HNH}\alpha}$ values is far from perfect. Sampling issues aside, it appears most likely that this is due to errors in the force field's description of the intrinsic conformational preferences of individual residue types since there are systematic differences between the average $^3J_{\text{HNH}\alpha}$ values measured in the simulations and those measured experimentally (Figures 2C and 2D) that appear to depend on the type of the amino acid. Discrepancies between the computed and experimental $^3J_{\text{HNH}\alpha}$ values for capped single amino acids using the same force field have also been reported recently,³⁰ and the authors of that work have shown that a reparameterization of the Amber ff99SB force field to match conformational distributions observed in a coil library gave $^3J_{\text{HNH}\alpha}$ values for capped single amino acids in much better agreement with experimental values.³⁴ For the future, therefore, it will be of interest to examine whether that new force field, RSFF2, leads to improved agreement with experiment for $^3J_{\text{HNH}\alpha}$ couplings of two-residue peptides.

While we think that the force field is the most likely cause of the discrepancies between computed and experimental $^3J_{\text{HNH}\alpha}$ values, it is intriguing to note that better agreement can also be achieved using Karplus parameters that are optimized independently for each type of amino acid (Figure S3 and Table S1). It has long been recognized that 3J (vicinal) coupling constants can be sensitive to factors in addition to the dihedral angle connecting the nuclei, e.g. substituent effects,⁴² and the possibility of using residue-specific Karplus parameterizations in the analysis of MD simulations has been noted by others.^{2b, 6} A comprehensive determination of all of the factors that determine $^3J_{\text{HNH}\alpha}$ values in peptides and proteins may in the future come from mixed quantum mechanical-molecular mechanical (QM-MM) calculation methods that have recently been applied to proteins.⁴³

Neighboring residue effects

An understanding of the effects that neighboring residues exert on the conformational behavior of other residues is essential for accurate predictions of the conformational ensembles of proteins;¹¹ such effects have been the subject of a very recent comprehensive review by Toal and Schweitzer-Stenner.⁴⁴ NREs can be manifested in a number of ways: the PPII propensities⁴⁵ and chemical shifts⁴⁶ of amino acids in short peptides, for example, can depend significantly upon the sequence in which they are embedded, and $^3J_{\text{HNH}\alpha}$ values in unfolded proteins are typically higher for residues that are preceded by bulky, branched or aromatic residues;⁴⁷ NREs can also be inferred by conformational analysis of protein structure databases.^{35, 48} Here, the focus has been on NREs between amino acids that are

adjacent to each other in sequence, but it should be remembered that statistical analysis of coil conformations indicates that interactions between residues at $i-1$ and $i+1$, for example, can alter the conformational preferences of an intervening residue i ,⁴⁹ and that steric interactions of residues that are not immediate neighbors can also affect the conformational possibilities open to oligopeptides.⁵⁰

To our knowledge, this is the first simulation effort aimed at explicitly reproducing experimental NREs in a comprehensive set of peptide systems. Implicit solvent simulations performed by the Freed and Sosnick groups on a number of two- and three-residue peptides had previously shown that the conformational properties of Ala are sensitive to the identity and conformation of neighboring residues and showed that these NREs were also sensitive to the simulation force field.⁵¹ Monte Carlo simulations performed by the Pappu group using purely steric potential functions explored the influences of neighboring Gly, Ala, Phe, Val and Pro residues on the secondary structure preferences of amino acids in pentameric host-guest peptides;⁵² we find little correspondence with their data (not shown) but this is not surprising given the differences in the peptides and potential functions used in the two studies.

More recently, the Liu group³⁹ has reported using Hamiltonian-REMD simulations together with the GROMOS 53A6 force field⁵³ to explore the conformational free energies of a variety of peptide systems, in particular exploring the NREs exerted by Ala, Phe and Val residues on the conformational behavior of a neighboring Ala. Their study showed that the backbone conformation adopted by the neighboring residue could in some cases play a more significant role in influencing the conformational preferences of a residue of interest than the neighboring residue's identity;³⁹ an important implication of their study, therefore, is that errors in the intrinsic backbone preferences of amino acids – which are apparent for the force field used here (see above) – could lead to errors in their computed NREs. Finally, the Cho group has used equilibrium MD simulations of 30 ns duration each to model the conformational behavior of the two-residue peptides that they studied experimentally;¹³ the computed free energy differences between the PPII and β -strand conformations obtained from their MD simulations, however, yielded no correlation with the corresponding experimentally derived values.⁵⁴ It is not clear to what extent the poor correlation that they obtained might have been due to the comparatively short simulation time, the force field used (Amber ff03⁵⁵), or to the decision to average the two $^3J_{H_{NH}Cl}$ values measured for each two-residue peptide in order to obtain average PPII and β -strand populations.

While the correspondence between the simulation and experimental results reported here is by no means quantitative, one clear success of the simulations is in reproducing the experimental results¹² showing that $\delta_{H_{Cl}}$ chemical shifts and $^3J_{H_{NH}Cl}$ values in two-residue peptides are altered significantly by neighboring aromatic residues. A previous analysis of a protein coil library had suggested that bulky, branched or aromatic residues (analyzed as a single group) tended to increase the β -strand population of following residues,⁴⁷ and this effect had been interpreted in terms of electrostatic solvation free energy differences between bulky and non-bulky residues in a β -strand model of an (Ala)₉ peptide.⁵⁶ It is apparently in light of these earlier studies that the Cho group has interpreted the increased $^3J_{H_{NH}Cl}$ values caused by neighboring aromatic residues as indicating an increase in

the fractional β -strand population.¹² Interestingly, however, while the simulations reported here nicely reproduce the experimental changes in $^3J_{\text{HNH}_\alpha}$ values due to neighboring aromatic residues (Figure 2C), the origins of these changes are quite different: here, they are instead caused primarily by an increase in the α' population. Importantly, both interpretations represent plausible explanations of the experimental results: the ϕ angle ranges of α' and β conformations are similar and their $^3J_{\text{HNH}_\alpha}$ values are also, therefore, expected to be similar. If the increases in α' population predicted by the simulations were to be confirmed by direct experimentation it would provide nice support for the idea of using simulations as a means of rationalizing and interpreting experimental results.

Protonation state effects

A final important result of the work reported here is that the protonation state of the histidine sidechain can significantly affect both its backbone conformational preferences and those of adjacent residues. Although not explicitly commented upon by the authors, the finding that the protonation state of a residue can affect its intrinsic backbone conformational preferences is apparent also in simulation work by the Vondrášek group for histidine and for a number of other amino acids.³⁸ Importantly, the finding that backbone preferences of amino acids can be affected by their protonation state is also supported by experimental work³³ showing pH-dependent changes of ~ 0.5 Hz in the $^3J_{\text{HNH}_\alpha}$ values for Asp and Glu; these have been interpreted in terms of protonation-state dependent alterations in the populations of β -strand conformations.¹⁹ Other experimental work has indicated that the PPII propensity of Asp also depends significantly upon the protonation state of its sidechain.^{45c}

While the effects of sidechain protonation state on the intrinsic conformational preferences of amino acids have been noted elsewhere, the results reported here provide, to our knowledge, the first indication that significant effects can also be exerted on the conformational preferences of neighboring residues. The largest such effect is seen on the average α -helical population of residues C-terminal to histidine: the average equilibrium constant for the α :PPII equilibrium changes from $(0.32/0.25 = 1.3)$ with His at the N-terminus to $(0.04/0.45 = 0.10)$ with Hip at the N-terminus: in free energy terms this amounts to a relative change in the ΔG for the α :PPII interconversion of ~ 1.4 kcal/mol at 298 K. If validated experimentally, such a change would have obvious potential implications for attempts to understand how changes in protonation state contribute to the pH dependence of protein stability.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH R01 GM099865 and R01 GM087290 awarded to A.H.E.

References

1. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nat. Struct. Mol. Biol.* 2002; 9:646–652.
2. (a) Lange OF, van der Spoel D, de Groot BL. Scrutinizing Molecular Mechanics Force Fields on the Submicrosecond Timescale with NMR Data. *Biophys. J.* 99:647–655. [PubMed: 20643085] (b) Best RB, Buchete N-V, Hummer G. Are Current Molecular Dynamics Force Fields too Helical? *Biophys. J.* 2008; 95:L07–L09. [PubMed: 18456823] (c) Beauchamp KA, Lin Y-S, Das R, Pande VS. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput.* 2012; 8:1409–1414. [PubMed: 22754404]
3. Graf J, Nguyen PH, Stock G, Schwalbe H. Structure and Dynamics of the Homologous Series of Alanine Peptides: A Joint Molecular Dynamics/NMR Study. *J. Am. Chem. Soc.* 2007; 129:1179–1189. [PubMed: 17263399]
4. (a) Best RB. Correction. *Biophys. J.* 2008; 95:4494. (b) Wickstrom L, Okur A, Simmerling C. Evaluating the Performance of the ff99SB Force Field Based on NMR Scalar Coupling Data. *Biophys. J.* 2009; 97:853–856. [PubMed: 19651043]
5. Hagarman A, Measey TJ, Mathieu D, Schwalbe H, Schweitzer-Stenner R. Intrinsic Propensities of Amino Acid Residues in GxG Peptides Inferred from Amide I' Band Profiles and NMR Scalar Coupling Constants. *J. Am. Chem. Soc.* 2009; 132:540–551. [PubMed: 20014772]
6. Nerenberg PS, Head-Gordon T. Optimizing Protein–Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides. *J. Chem. Theory Comput.* 2011; 7:1220–1230.
7. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* 2006; 65:712–725.
8. Li D-W, Brüschweiler R. NMR-Based Protein Potentials. *Angew. Chem., Int. Ed.* 2010; 49:6778–6780.
9. Li D-W, Brüschweiler R. Iterative Optimization of Molecular Mechanics Force Fields from NMR Data of Full-Length Proteins. *J. Chem. Theory Comput.* 2011; 7:1773–1782.
10. (a) Toal S, Meral D, Verbaro D, Urbanc B, Schweitzer-Stenner R. pH-Independence of Trilalanine and the Effects of Termini Blocking in Short Peptides: A Combined Vibrational, NMR, UVCD, and Molecular Dynamics Study. *J. Phys. Chem. B.* 2013; 117:3689–3706. [PubMed: 23448349] (b) Xiao X, Kallenbach N, Zhang Y. Peptide Conformation Analysis Using an Integrated Bayesian Approach. *J. Chem. Theory Comput.* 2014; 10:4152–4159. [PubMed: 25221447] (c) Verbaro D, Ghosh I, Nau WM, Schweitzer-Stenner R. Discrepancies between Conformational Distributions of a Polyalanine Peptide in Solution Obtained from Molecular Dynamics Force Fields and Amide I' Band Profiles. *J. Phys. Chem. B.* 2010; 114:17201–17208. [PubMed: 21138254]
11. Schweitzer-Stenner R. Conformational propensities and residual structures in unfolded peptides and proteins. *Molecular BioSystems.* 2012; 8:122–133. [PubMed: 21879108]
12. Jung Y-S, Oh K-I, Hwang G-S, Cho M. Neighboring Residue Effects in Terminally Blocked Dipeptides: Implications for Residual Secondary Structures in Intrinsically Unfolded/Disordered Proteins. *Chirality.* 2014; 26:443–452. [PubMed: 24453185]
13. Oh K-I, Jung Y-S, Hwang G-S, Cho M. Conformational distributions of denatured and unstructured proteins are similar to those of 20 × 20 blocked dipeptides. *J. Biomol. NMR.* 2012; 53:25–41. [PubMed: 22426785]
14. Andrews CT, Elcock AH. COFFDROP: A Coarse-Grained Nonbonded Force Field for Proteins Derived from All-Atom Explicit-Solvent Molecular Dynamics Simulations of Amino Acids. *J. Chem. Theory Comput.* 2014; 10:5178–5194. [PubMed: 25400526]
15. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* 2010; 78:1950–1958.
16. Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* 2004; 120:9665–9678. [PubMed: 15267980]

17. Hu J-S, Bax A. Determination of ϕ and χ_1 Angles in Proteins from ^{13}C - ^{13}C Three-Bond J Couplings Measured by Three-Dimensional Heteronuclear NMR. How Planar Is the Peptide Bond? *J. Am. Chem. Soc.* 1997; 119:6360–6368.
18. Shen Y, Bax A. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR.* 2010; 48:13–22. [PubMed: 20628786]
19. Grdadolnik J, Mohacek-Grosov V, Baldwin RL, Avbelj F. Populations of the three major backbone conformations in 19 amino acid dipeptides. *Proc. Natl. Acad. Sci. USA.* 2011; 108:1794–1798. [PubMed: 21205907]
- 20 (a). Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* 2008; 4:435–447. (b) Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* 2005; 26:1701–1718. [PubMed: 16211538]
- 21 (a). Hoover WG. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A.* 1985; 31:1695–1697. [PubMed: 9895674] (b) Nosé S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* 1984; 81:511–519.
22. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 1981; 52:7182–7190.
23. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J. Chem. Phys.* 1995; 103:8577.
24. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 1997; 18:1463–1472.
25. Cruz VL, Ramos J, Martinez-Salazar J. Assessment of the Intrinsic Conformational Preferences of Dipeptide Amino Acids in Aqueous Solution by Combined Umbrella Sampling/MBAR Statistics. A Comparison with Experimental Results. *J. Phys. Chem. B.* 2011; 116:469–475. [PubMed: 22136632]
26. Hutchinson EG, Thornton JM. A revised set of potentials for β -turn formation in proteins. *Protein Sci.* 1994; 3:2207–2216. [PubMed: 7756980]
27. Karplus M. Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *J. Chem. Phys.* 1959; 30:11–15.
- 28 (a). Vuister GW, Bax A. Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in ^{15}N -enriched proteins. *J. Am. Chem. Soc.* 1993; 115:7772–7777. (b) Vögeli B, Ying J, Grishaev A, Bax A. Limits on Variations in Protein Backbone Dynamics from Precise Measurements of Scalar Couplings. *J. Am. Chem. Soc.* 2007; 129:9377–9385. [PubMed: 17608477] (c) Schmidt J, Blümel M, Löhr F, Rüterjans H. Self-consistent 3J coupling analysis for the joint calibration of Karplus coefficients and evaluation of torsion angles. *J. Biomol. NMR.* 1999; 14:1–12. [PubMed: 21136331] (d) Pardi A, Billeter M, Wüthrich K. Calibration of the angular dependence of the amide proton- α proton coupling constants, $3\text{JHN}\alpha$, in a globular protein: Use of $3\text{JHN}\alpha$ for identification of helical secondary structure. *J. Mol. Biol.* 1984; 180:741–751. [PubMed: 6084720] (e) Wang AC, Bax A. Determination of the Backbone Dihedral Angles ϕ in Human Ubiquitin from Reparameterized Empirical Karplus Equations. *J. Am. Chem. Soc.* 1996; 118:2483–2494. (f) Case DA, Scheurer C, Brüschweiler R. Static and Dynamic Effects on Vicinal Scalar J Couplings in Proteins and Peptides: A MD/DFT Analysis. *J. Am. Chem. Soc.* 2000; 122:10390–10397.
29. Markwick PL, Showalter S, Bouvignies G, Brüschweiler R, Blackledge M. Structural dynamics of protein backbone ϕ angles: extended molecular dynamics simulations versus experimental 3 J scalar couplings. *J. Biomol. NMR.* 2009; 45:17–21. [PubMed: 19629714]
30. Jiang F, Zhou C-Y, Wu Y-D. Residue-Specific Force Field Based on the Protein Coil Library. RSFF1: Modification of OPLS-AA/L. *J. Phys. Chem. B.* 2014; 118:6983–6998. [PubMed: 24815738]
- 31 (a). Han B, Liu Y, Ginzinger S, Wishart D. SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR.* 2011; 50:43–57. [PubMed: 21448735] (b) Kohlhoff KJ, Robustelli P, Cavalli A, Salvatella X, Vendruscolo M. Fast and Accurate Predictions of Protein NMR Chemical Shifts from Interatomic Distances. *J. Am. Chem. Soc.* 2009; 131:13894–13895. [PubMed: 19739624] (c) Xu X-P, Case DA. Probing multiple effects on ^{15}N , $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, and

- $^{13}\text{C}'$ chemical shifts in peptides using density functional theory. *Biopolymers*. 2002; 65:408–423. [PubMed: 12434429]
32. Li D-W, Brüschweiler R. PPM: a side-chain and backbone chemical shift predictor for the assessment of protein conformational ensembles. *J. Biomol. NMR*. 2012; 54:257–265. [PubMed: 22972619]
 33. Avbelj F, Grdadolnik SG, Grdadolnik J, Baldwin RL. Intrinsic backbone preferences are fully present in blocked amino acids. *Proc. Natl. Acad. Sci. USA*. 2006; 103:1272–1277. [PubMed: 16423894]
 34. Zhou C-Y, Jiang F, Wu Y-D. Residue-Specific Force Field Based on Protein Coil Library. RSFF2: Modification of AMBER ff99SB. *J. Phys. Chem. B*. 2014; 119:1035–1047. [PubMed: 25358113]
 35. Ting D, Wang G, Shapovalov M, Mitra R, Jordan MI, Dunbrack RL Jr. Neighbor-Dependent Ramachandran Probability Distributions of Amino Acids Developed from a Hierarchical Dirichlet Process Model. *PLoS Comput Biol*. 2010; 6:e1000763. [PubMed: 20442867]
 36. Guruprasad K, Rajkumar S. Beta- and Gamma-turns in proteins revisited: a new set of amino acid turn-type dependent positional preferences and potentials. *J. Biosci*. 2000; 25:14.
 37. Barducci A, Bussi G, Parrinello M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett*. 2008; 100:020603. [PubMed: 18232845]
 38. Vymtal J, Vondrášek J. Critical Assessment of Current Force Fields. Short Peptide Test Case. *J. Chem. Theory Comput*. 2012; 9:441–451. [PubMed: 26589046]
 39. Xu C, Wang J, Liu H. A Hamiltonian Replica Exchange Approach and Its Application to the Study of Side-Chain Type and Neighbor Effects on Peptide Backbone Conformations. *J. Chem. Theory Comput*. 2008; 4:1348–1359.
 40. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett*. 1999; 314:141–151.
 41. Long D, Li D-W, Walter Korvin F, A. Griesinger C, Brüschweiler R. Toward a Predictive Understanding of Slow Methyl Group Dynamics in Proteins. *Biophys. J*. 2011; 101:910–915. [PubMed: 21843482]
 42. Karplus M. Vicinal Proton Coupling in Nuclear Magnetic Resonance. *J. Am. Chem. Soc*. 1963; 85:2870–2871.
 43. Wang B, He X, Merz KM. Quantum Mechanical Study of Vicinal J Spin–Spin Coupling Constants for the Protein Backbone. *J. Chem. Theory Comput*. 2013; 9:4653–4659. [PubMed: 26589175]
 44. Toal S, Schweitzer-Stenner R. Local Order in the Unfolded State: Conformational Biases and Nearest Neighbor Interactions. *Biomolecules*. 2014; 4:725–773. [PubMed: 25062017]
 - 45 (a). Shi Z, Chen K, Liu Z, Ng A, Bracken WC, Kallenbach NR. Polyproline II propensities from GGXGG peptides reveal an anticorrelation with β -sheet scales. *Proc. Natl. Acad. Sci. USA*. 2005; 102:17964–17968. [PubMed: 16330763] (b) Rucker AL, Payer CT, Campbell MN, Qualls JE, Creamer TP. Host-guest scale of left-handed polyproline II helix formation. *Proteins: Struct., Funct., Bioinf*. 2003; 53:68–75. (c) Brown AM, Zondlo NJ. A Propensity Scale for Type II Polyproline Helices (PPII): Aromatic Amino Acids in Proline-Rich Sequences Strongly Disfavor PPII Due to Proline–Aromatic Interactions. *Biochemistry*. 2012; 51:5041–5051. [PubMed: 22667692]
 - 46 (a). Braun D, Wider G, Wuethrich K. Sequence-Corrected ^{15}N “Random Coil” Chemical Shifts. *J. Am. Chem. Soc*. 1994; 116:8466–8469. (b) Schwarzinger S, Kroon GJA, Foss TR, Chung J, Wright PE, Dyson HJ. Sequence-Dependent Correction of Random Coil NMR Chemical Shifts. *J. Am. Chem. Soc*. 2001; 123:2970–2978. [PubMed: 11457007] (c) Kjaergaard M, Poulsen F. Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J. Biomol. NMR*. 2011; 50:157–165. [PubMed: 21604143]
 47. Penkett CJ, Redfield C, Dodd I, Hubbard J, McBay DL, Mossakowska DE, Smith RAG, Dobson CM, Smith LJ. NMR analysis of main-chain conformational preferences in an unfolded fibronectin-binding protein. *J. Mol. Biol*. 1997; 274:152–159. [PubMed: 9398523]
 - 48 (a). Jha AK, Colubri A, Zaman MH, Koide S, Sosnick TR, Freed KF. Helix, Sheet, and Polyproline II Frequencies and Strong Nearest Neighbor Effects in a Restricted Coil Library. *Biochemistry*. 2005; 44:9691–9702. [PubMed: 16008354] (b) Ormeci L, Gursoy A, Tunca G, Erman B.

- Computational basis of knowledge-based conformational probabilities derived from local- and long-range interactions in proteins. *Proteins: Struct., Funct., Bioinf.* 2007; 66:29–40.(c)
- Betancourt MR, Skolnick J. Local Propensities and Statistical Potentials of Backbone Dihedral Angles in Proteins. *J. Mol. Biol.* 2004; 342:635–649. [PubMed: 15327961]
49. Griffiths-Jones SR, Sharman GJ, Maynard AJ, Searle MS. Modulation of intrinsic ϕ, ψ propensities of amino acids by neighbouring residues in the coil regions of protein structures: NMR analysis and dissection of a β -hairpin peptide. *J. Mol. Biol.* 1998; 284:1597–1609. [PubMed: 9878373]
50. Pappu RV, Srinivasan R, Rose GD. The Flory isolated-pair hypothesis is not valid for polypeptide chains: Implications for protein folding. *Proc. Natl. Acad. Sci. USA.* 2000; 97:12565–12570. [PubMed: 11070081]
51. Zaman MH, Shen M-Y, Berry RS, Freed KF, Sosnick TR. Investigations into Sequence and Conformational Dependence of Backbone Entropy, Inter-basin Dynamics and the Flory Isolated-pair Hypothesis for Peptides. *J. Mol. Biol.* 2003; 331:693–711. [PubMed: 12899838]
52. Tran HT, Wang X, Pappu RV. Reconciling Observations of Sequence-Specific Conformational Propensities with the Generic Polymeric Behavior of Denatured Proteins†. *Biochemistry.* 2005; 44:11369–11380. [PubMed: 16114874]
53. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* 2004; 25:1656–1676. [PubMed: 15264259]
54. Oh K-I, Lee K-K, Park E-K, Jung Y, Hwang G-S, Cho M. A comprehensive library of blocked dipeptides reveals intrinsic backbone conformational propensities of unfolded proteins. *Proteins: Struct., Funct., Bioinf.* 2012; 80:977–990.
55. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 2003; 24:1999–2012. [PubMed: 14531054]
56. Avbelj F, Baldwin RL. Origin of the neighboring residue effect on peptide backbone conformation. *Proc. Natl. Acad. Sci. USA.* 2004; 101:10967–10972. [PubMed: 15254296]

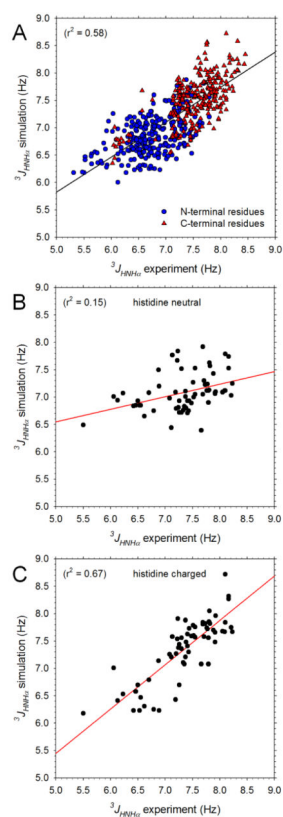


Figure 1. Comparison of computed and experimental $^3J_{HNH\alpha}$ coupling constants

A. Plot comparing simulation and experimental $^3J_{HNH\alpha}$ coupling constants for all non-Gly residues in peptides that do not contain Pro, Asp or Glu; experimental data taken from Jung et al.;¹² black line shows linear regression. **B.** Comparison of simulation and experimental $^3J_{HNH\alpha}$ coupling constants for residues in histidine-containing peptides; simulations modeled histidine using the neutral, His residue type. **C.** Same as B but plotting data obtained from simulations that modeled histidine using the charged, Hip residue type.

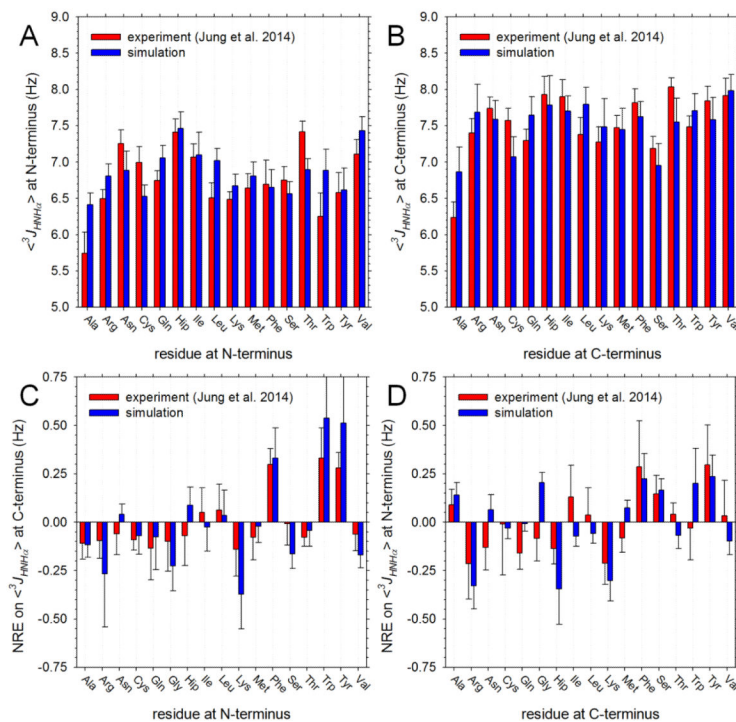


Figure 2. Comparison of computed and experimental ${}^3J_{\text{HNH}\alpha}$ coupling constants averaged by type of amino acid

A. Plot showing average ${}^3J_{\text{HNH}\alpha}$ coupling constant of each type of amino acid when present at the N-terminal position; all averages obtained from data on 17 peptides (all peptides that do not involve Pro, Asp or Glu); error bars indicate standard deviations. **B.** Same as A but showing results for each type of amino acid when present at the C-terminal position. **C.** Plot showing the effect of each type of amino acid, when present at the N-terminal position, on the average ${}^3J_{\text{HNH}\alpha}$ coupling constant of amino acids at the C-terminal position (see text); error bars indicate standard deviations. **D.** Same as C but showing effect of each type of amino acid, when present at the C-terminal position, on the average ${}^3J_{\text{HNH}\alpha}$ coupling constant of amino acids at the N-terminal position.

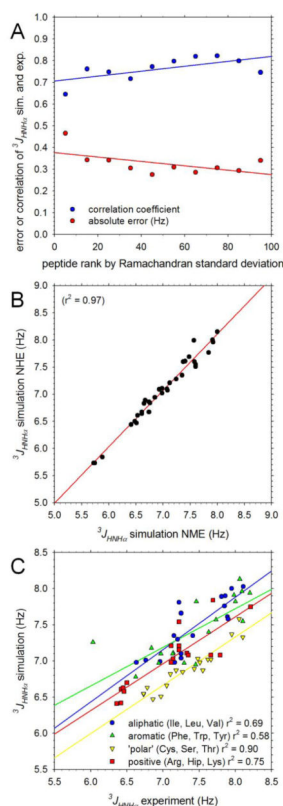


Figure 3. Factors influencing agreement between computed and experimental ${}^3J_{\text{HNH}\alpha}$ coupling constants

A. Plot showing correlation coefficient (blue) and mean unsigned difference (red) between simulation and experimental ${}^3J_{\text{HNH}\alpha}$ coupling constants as a function of the standard deviations of the Ramachandran maps sampled during MD simulations. Peptides are grouped according to their rank in an ordered list of Ramachandran map standard deviations; the datapoints at far-left show the correlation coefficient and error obtained for those peptides with standard deviations in the lowest 10%; the datapoints at far-right show the same for those peptides with standard deviations in the highest 10%. **B.** Plot comparing computed ${}^3J_{\text{HNH}\alpha}$ coupling constants obtained from simulations of identical peptides using NHE (simple amide) and NME (N-methylamide) capping groups. **C.** Plot comparing simulation and experimental ${}^3J_{\text{HNH}\alpha}$ coupling constants for peptides grouped according to their constituent amino acids: datapoints marked aliphatic (blue), for example, are for peptides that contain *only* Ile, Leu or Val residues.

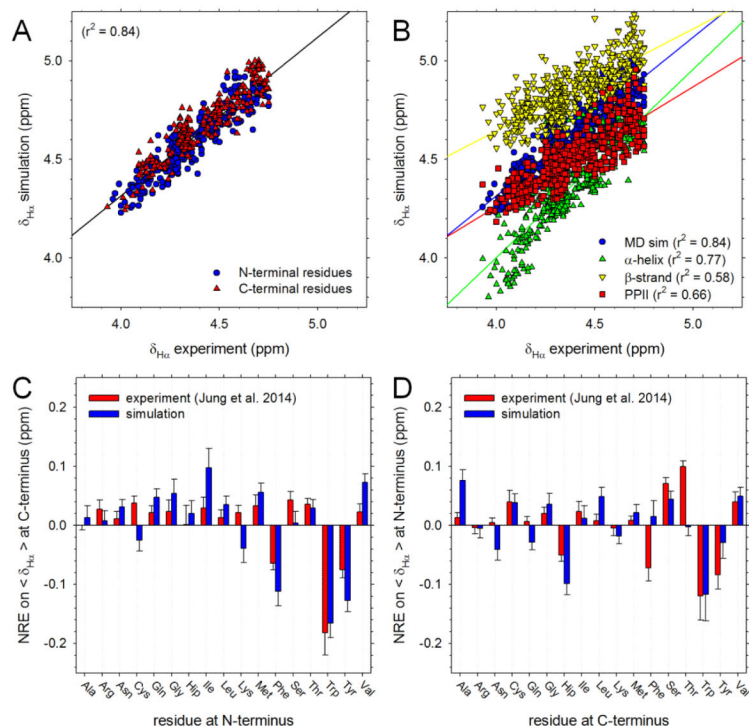


Figure 4. Comparison of computed and experimental δ_{HCl} chemical shifts

A. Plot comparing computed and experimental δ_{HCl} chemical shifts for all non-Gly residues in peptides that do not contain Pro, Asp or Glu. **B.** Same as A, but showing results computed when each peptide is restricted to one of three different backbone conformations. **C.** Plot showing the effect of each type of amino acid, when present at the N-terminal position, on the average δ_{HCl} value of amino acids at the C-terminal position (see text); error bars indicate standard deviations. **D.** Same as C but showing effect of each type of amino acid, when present at the C-terminal position, on the average δ_{HCl} value of amino acids at the N-terminal position.

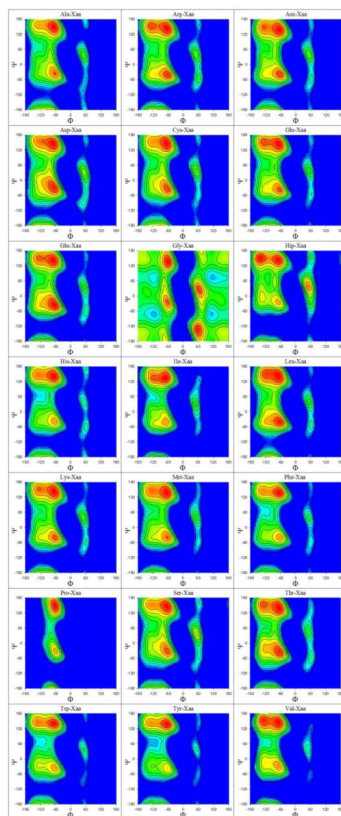


Figure 5. Average Ramachandran maps at the N-terminus

Plots showing simulation Ramachandran maps expressed in free energy form for all 21 types of amino acids averaged over all possible C-terminal residues excluding Pro and Gly. Map at top-left, for example, shows the average Ramachandran map of the N-terminal Ala in all peptides of the form Ala-Ala, Ala-Cys, Ala-Gln, Ala-Glu, etc. Free energies are colored in descending order from blue to red.

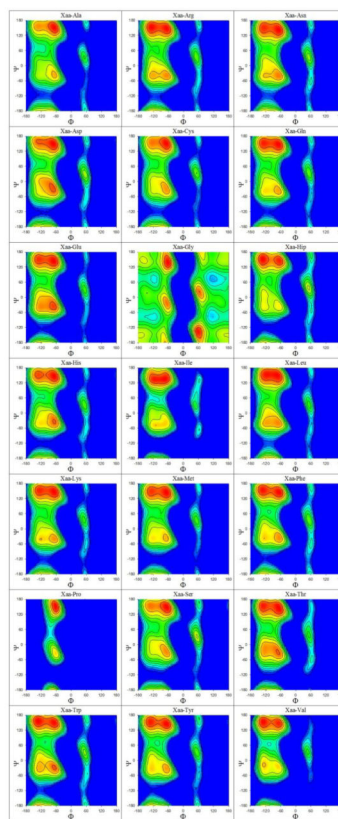


Figure 6. Average Ramachandran maps at the C-terminus

Plots showing simulation Ramachandran maps expressed in free energy form for all 21 types of amino acids averaged over all possible N-terminal residues excluding Pro and Gly. Map at top-left, for example, shows the average Ramachandran map of the C-terminal Ala in all peptides of the form Ala-Ala, Cys-Ala, Gln-Ala, Glu-Ala, etc. Free energies are colored in descending order from blue to red.

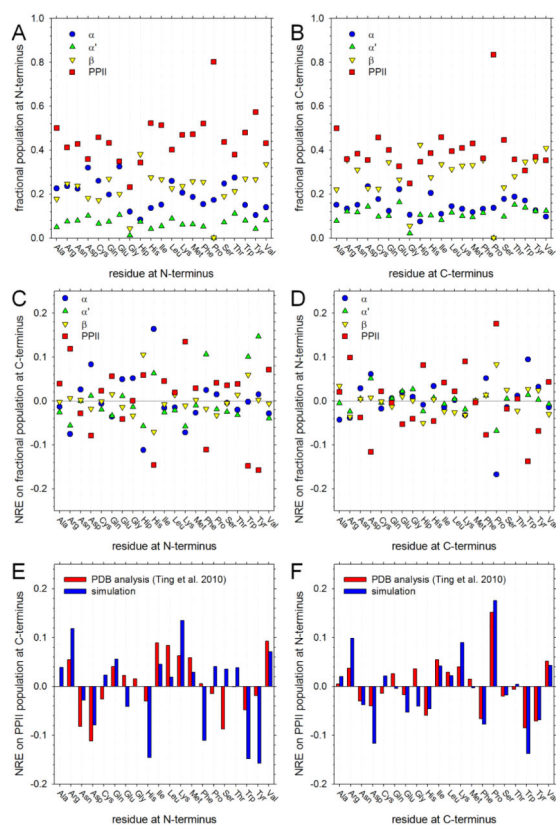


Figure 7. Fractional populations of four major backbone conformations

A. Simulated fractional populations of α , α' , β and PPII conformations for all 21 types of amino acids when present at the N-terminal position; results are averaged over all possible C-terminal residues excluding Pro and Gly. **B.** Same as A but showing results for all 21 types of amino acids when present at the C-terminal position; results are averaged over all possible N-terminal residues excluding Pro and Gly. **C.** Plot showing effect of each type of amino acid, when present at the N-terminal position, on the average fractional populations of α , α' , β and PPII conformations of amino acids at the C-terminal position. **D.** Same as C but showing effect of each type of amino acid, when present at the C-terminal position, on the average fractional populations at the N-terminal position. **E.** Plot comparing the simulated effect of each type of amino acid, when present at the N-terminal position, on the average fractional PPII population at the C-terminal position with that obtained from analysis of a coil library.³⁵ **F.** Same as E but comparing the simulated effect of each type of amino acid, when present at the C-terminal position, on the average fractional PPII population at the N-terminal position.

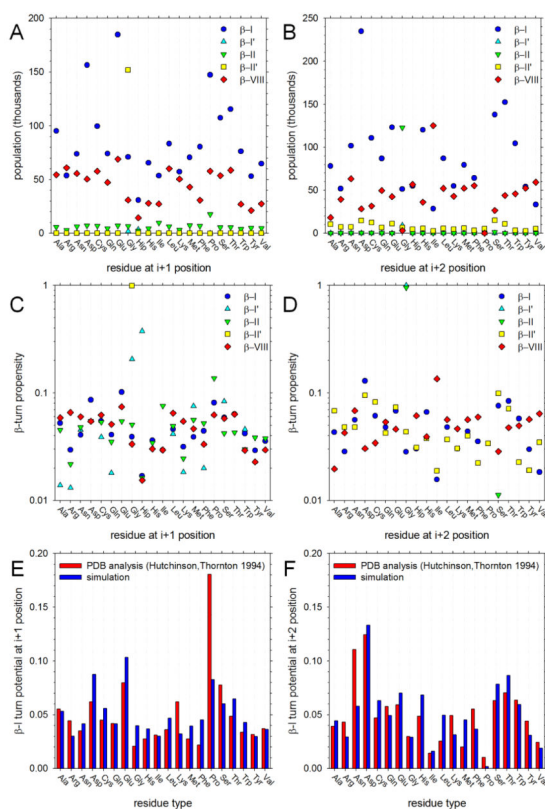


Figure 8. Sampling of β -turn conformations

A. MD-sampled populations of five types of β -turn conformation plotted as a function of the identity of the amino acid at the 'i+1' (i.e. N-terminal) position (see text). **B.** Same as A but plotted as a function of the identity of the amino acid at the 'i+2' (i.e. C-terminal) position. **C.** Plot showing the propensity of each type of amino acid to be found at the 'i+1' position for each type of β -turn (see text). **D.** Same as C but showing the propensity of each type of amino acid to be found at the 'i+2' position. **E.** Plot comparing the MD-computed turn potential of each type of amino acid for the 'i+1' position of type β -I turns with that obtained from analysis of PDB structures.²⁶ **F.** Same as E but showing results for the 'i+2' position.