JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Communication
pubs.acs.org/JACS

Backbone Hydration Determines the Folding Signature of Amino Acid Residues

Olivier Bignucolo,[†,‡] Hoi Tik Alvin Leung,[‡] Stephan Grzesiek,[‡] and Simon Bernèche*[,†,‡]

[†]SIB Swiss Institute of Bioinformatics and [‡]Biozentrum, University of Basel, Klingelbergstrasse 50/70, CH-4056 Basel, Switzerland

Supporting Information

ABSTRACT: The relation between the sequence of a protein and its three-dimensional structure remains largely unknown. A lasting dream is to elucidate the side-chain-dependent driving forces that govern the folding process. Different structural data suggest that aromatic amino acids play a particular role in the stabilization of protein structures. To better understand the underlying mechanism, we studied peptides of the sequence EGAAXAASS (X = Gly, Ile, Tyr, Trp) through comparison of molecular dynamics (MD) trajectories and NMR residual dipolar coupling (RDC) measurements. The RDC data for aromatic substitutions provide evidence for a kink in the peptide backbone. Analysis of the MD simulations shows that the formation of internal hydrogen bonds underlying a helical turn is key to reproduce the experimental RDC values. The simulations further reveal that the driving force leading to such helical-turn conformations arises from the lack of hydration of the peptide chain on either side of the bulky aromatic side chain, which can potentially act as a nucleation point initiating the folding process.

The prediction of the structural properties of a protein from its amino acid sequence remains a major challenge.[1−3] The detailed mechanism driving the protein folding process is unknown, and specifically its dependence on amino acid side chains.[4] The functional importance of intrinsically disordered proteins has stimulated investigation of the relation between their sequence and their conformational tendency.[5,6] In order to improve predictions about the structure of proteins (folded or disordered), a better understanding of the mechanisms by which individual amino acid side chains impact the conformational dynamics of the protein is required.[4,7]

NMR spectroscopy is particularly appropriate to investigate the structural dynamics of peptides in disordered and folded states.[8] Particularly, residual dipolar couplings (RDCs), which arise when molecules are dissolved in anisotropic liquid phases,[9] provide local as well as long-range quantitative structural information on individual chemical bonds.[10−12] The RDC between two nuclei is proportional to the ensemble and time average $\langle(3\cos^2\theta - 1)/2\rangle$, where θ is the instantaneous angle between the internuclear vector and the magnetic field.

In order to investigate the role of individual amino acids on the conformational propensities of a peptide, Dames et al.[13] engineered a series of 14 peptides of sequence EGAAXAASS. The hydrophilic ends ensured solubility, while the nonpolar adjacent residues provided a neutral environment for the systematically single-mutated residue X. They recorded $^1D_{C\alpha H\alpha}$ and $^1D_{NH}$ RDCs of these peptides, performing the alignment measurement with polyacrylamide gels.[14] Most peptides (with X = G, I, V, L, N, Q, T, D, E, or K) produced a relatively flat pattern consistent with a rather extended average conformation with little specific local structure. However, the substitutions with the aromatic amino acids Tyr and Trp resulted in a strong reduction of the RDCs or even changes in their signs at the center of the peptide (black lines in Figure S1 in the Supporting Information (SI)), suggesting the presence of a kink at this position.

All-atom molecular dynamics (MD) simulations provide the most detailed description of peptide structural dynamics with high spatial and temporal resolution. The NMR data can be used to validate the MD simulation trajectories, which can potentially reveal the mechanistic specificity of aromatic amino acids. Here we show through a systematic comparison between simulated and experimental RDCs that the conformations that best reproduce the experimental data correspond to dynamical ensembles of short helices or turns stabilized by backbone hydrogen bonds. We find that one key driving force that increases the folding propensity of peptides containing aromatic residues arises from the lack of hydration of the carbonyl and amide groups on either side of the bulky hydrophobic side chain.

We performed MD simulations in explicit solvent to reproduce previously measured residual dipolar couplings and chemical shifts of peptides of sequence EGAAXAASS.[13] In order to produce adequate sampling, we carried out 7−12 replicated simulations per investigated peptide, each lasting 100 ns. We calculated the $^1D_{C\alpha H\alpha}$ and $^1D_{NH}$ RDCs as well as the ^{1}HN chemical shifts from the coordinates (see Methods in the SI). Figure S1 compares the $^1D_{C\alpha H\alpha}$ and $^1D_{NH}$ RDCs, averaged over all of the replicated simulations, and the experimental values published previously. The predicted RDC patterns of the peptide with X = Gly or Ile are relatively flat, accurately reproducing the experimental values, whereas the profiles obtained for X = Tyr or Trp only partially show the RDC variations along the peptide sequence that are observed experimentally. Both the $^1D_{C\alpha H\alpha}$ and $^1D_{NH}$ RDCs of the peptides with X = Tyr and Trp fluctuate a lot between the replicated simulations as well as within a given simulation, reflecting the fact that the peptides adopt an ensemble of conformations.

Received: January 20, 2015
Published: March 20, 2015

DOI: 10.1021/jacs.5b00660
J. Am. Chem. Soc. 2015, 137, 4300−4303

We took advantage of these conformational variations to investigate the relations between structural order parameters and the RDCs. Time series analysis over 700 ns (seven 100 ns simulations) showed that the conformations of the X = Trp peptide that best reproduced the experimental RDC profile are characterized by a rather compact conformation, as evidenced by structural parameters such as the radius of gyration and the number of intramolecular hydrogen bonds (Figures 1 and S2).

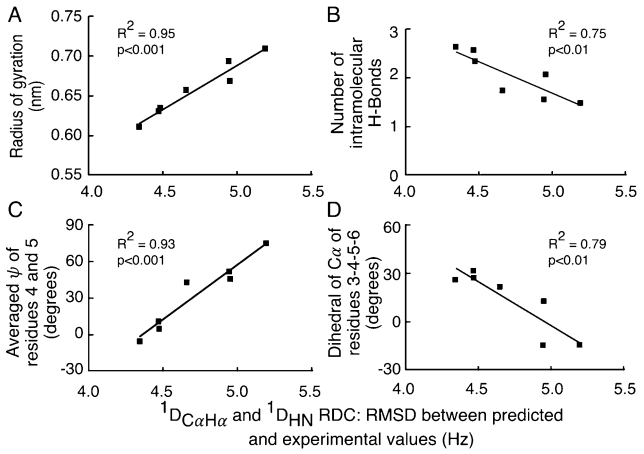

Figure 1. Correlation between different structural parameters and the root-mean-square deviation (RMSD) with respect to experimental RDCs with X = Trp. The RMSD is correlated to (A) the radius of gyration, (B) the formation of intramolecular hydrogen bonds, (C) the ψ angle of residues 4 and 5, and (D) the pseudodihedral angle formed by the Cα atoms of residues 3−6. Each point represents the average over one simulation of 100 ns.

We performed a stepwise regression analysis (see Methods) to identify which hydrogen bonds are statistically relevant. The analysis showed that the conformations that better reproduce the experimental RDC profile involve hydrogen bonds typical of a helix or a turn in the middle of the chain. Precisely, the structurally relevant atomic pairs involve hydrogen bonds between backbone carbonyls of Ala3 or Ala4 and amide groups of residue Ala6, Ala7, or Ser8 (Table S1 in the SI). Analysis of the backbone dihedral angles showed that residues 4 and 5 mainly occupy conformations around $\psi = 150°$ or $-30°$, the latter corresponding to a turn or an α-helical conformation. Clustering of the conformations on the basis of the (ϕ, ψ) dihedral angles revealed that about one-third of all simulation frames contain a short α-helix centered on the X = Trp residue and form the main cluster (Figure S3). The RDC values of this cluster reproduce the characteristics of the experimental profile, as illustrated in Figure S1. The next three clusters account together for about one-third of all conformations and contain β-turns, notably of types I and VIII. Reweighting of the individual conformations to reproduce the RDC and additional J-coupling data[13] while maximizing the entropy (see Methods) confirmed that the first cluster is the most prominent one (Figure S4). Only small readjustments of the cluster population (maximal change per cluster is 4% of the total population) were necessary to reproduce the experimental data within their experimental error, suggesting the overall good accuracy of the MD sampling.

Taken together, these results are consistent with the idea that the conformations of the X = Trp peptide that best reproduce the experimental RDC profile correspond to a turn or a short helix toward the middle of the chain. An analogous analysis of the peptide with X = Tyr showed that this substitution also favors similar conformations, although to a slightly lesser extent than X = Trp (Table S2 and Figures S5 and S6).

We clustered the simulated conformations of the four peptides according to their numbers of intramolecular backbone hydrogen bonds. The left panels of Figure 2 show that the

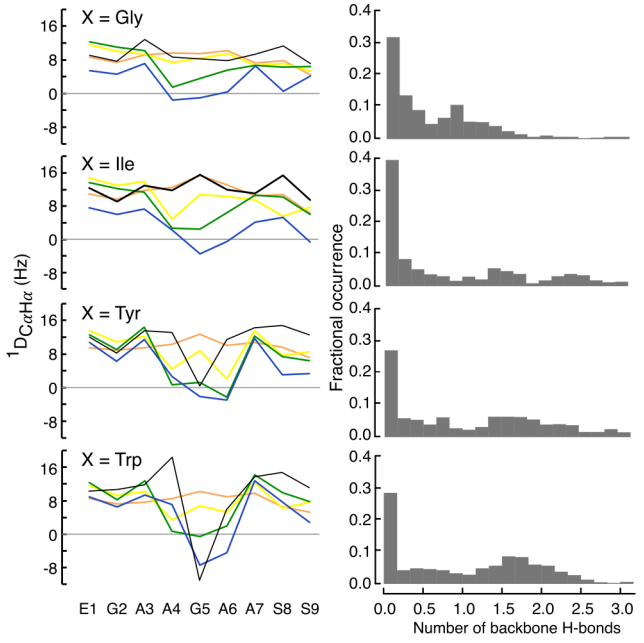

Figure 2. The RDC pattern depends on intramolecular backbone hydrogen bonds. The left panels show for the different peptides the experimental $^1D_{C\alpha H\alpha}$ RDCs (black) and the predicted RDCs for clusters of conformations containing different numbers of backbone hydrogen bonds: 0 (orange), 1 (yellow), 2 (green), 3 (blue). On the right is shown the conformational probability distribution for each peptide as a function of the number of backbone hydrogen bonds (averaged over intervals of 1 ns).

number of hydrogen bonds determines the peptide's RDC profile. Although the averaged values obtained for the peptides with X = Gly or Ile produce rather flat RDC profiles, corresponding to extended peptides with little local structural preference, the conformations that have two or three hydrogen bonds reproduce the distinct dips in the center of the $^1D_{C\alpha H\alpha}$ and $^1D_{NH}$ RDC patterns. On the other hand, the conformations of peptides with X = Tyr or Trp that have only one or no hydrogen bonds produce flat RDC patterns. These different conformations are dynamically visited by the peptides, and thus it is important to consider their probability distribution. On the right side of Figure 2 are shown histograms of the numbers of backbone hydrogen bonds averaged over intervals of 1 ns. The peptides with X = Tyr or Trp both present a maximum around 1.8 bonds (Hartigan's dip test for unimodality, p(Tyr) < 0.01 and p(Trp) < 0.001). The X = Ile peptide also shows some conformations containing more than one hydrogen bond, but with a lower probability than for X = Tyr and Trp. The peptide with X = Gly adopts a non-negligible number of conformations with one hydrogen bond, but the probability of observing more bonds is small. These observations echo the facts that hydrogen bonds become more stable and backbone dihedral fluctuations decrease along the following order of substitutions: Gly → Ile → Tyr → Trp (Figures S7 and S8). The peptides containing

DOI: 10.1021/jacs.5b00660
J. Am. Chem. Soc. 2015, 137, 4300−4303

an aromatic residue adopt significantly fewer conformations with no hydrogen bond in comparison with the Gly and Ile peptides. Similar conformations are visited by all of the peptides, but there is a higher probability of observing more than one hydrogen bond in the peptides containing an aromatic residue and, to some extent, an isoleucine.

The all-atom MD simulations provide a detailed description of the structural dynamics of the peptide and thus allow us to investigate the mechanism by which aromatic residues initiate the folding process. We postulate that their bulky side chains limit the access of water molecules to nearby carbonyl and amide groups. As a consequence, in line with the general understanding of the formation of secondary structure elements,[15] it would be energetically preferable for these backbone functional groups to interact with each other, forming intramolecular hydrogen bonds and favoring peptide folding. To test this hypothesis, we computed the number of water molecules coordinating the carbonyl and amide groups of residues in the middle of the chain. To show that the observed difference in hydration is a potential driving force and not only a consequence of a folded conformation, we compared folded and extended conformations of the Tyr or Trp peptides to conformations of the Gly peptide, which is generally extended. For both the Tyr and Trp peptides, the folded and extended pools respectively contain the conformations corresponding to the 20% lowest and 20% highest RMSDs with respect to the experimental RDC values. Figure 3 shows that the backbone of the X = Trp peptide is significantly less hydrated than that of the peptide with X = Gly, even in its fully extended conformations. The strongest effect is observed for the amide groups of residues 5 and 6. Similar results were obtained for X = Tyr, but the dehydration of the backbone polar groups is significantly less.

Proteins undergoing folding and intrinsically disordered proteins are typically characterized by dynamical ensembles of conformations, with fluctuations involving the formation and release of local secondary structures.[16] Comparison of sequences of intrinsically disordered proteins with natively folded ones showed that disordered regions are generally depleted of specific residues, which were termed "order-promoting amino acids".[6] These include, in decreasing order, Trp, Tyr, and Phe followed by Ile, Leu, and Asn.[17] The bulkiness of the side chains has been proposed to have a direct impact on the local conformation and dynamics of natively unfolded proteins.[18] Our calculations more specifically suggest that bulkier side chains, notably aromatic ones, impede the hydration of neighboring carbonyl and amide groups, favoring the formation of backbone hydrogen bonds and peptide folding. Such elementary folding events are likely to act as nucleation points initiating the folding process and leading to the formation of protein secondary structure elements, without excluding that coalescence of neighboring chains might be essential to stabilize them.[19] The long-standing view that the interaction between backbone functional groups is favored by the formation of hydrophobic pockets[15,20,21] and shielding from solvent[22−25] is thus shown to hold at the scale of a single amino acid side chain. Cooperativity between adjacent side chains is expected to play a key role in defining the level of backbone hydration, suggesting that the processes involved in protein folding are even more local than previously thought.[26] This further reveals that the effective, or biased,[27] conformational search space can involve as little as a few tens of atoms per nucleation point, in line with the mechanism hypothesized by Levinthal in the 1960s as a way to circumvent the protein folding paradox.[28] Here we have provided a mechanistic view at the atomistic scale in which the level of hydration of the main chain is shown to be a determinant of the protein folding process and is defined locally by the characteristics of the lateral chains. These findings contribute to the development of an amino acid-based code to understand the interatomic driving forces defining the tridimensional structure of proteins.

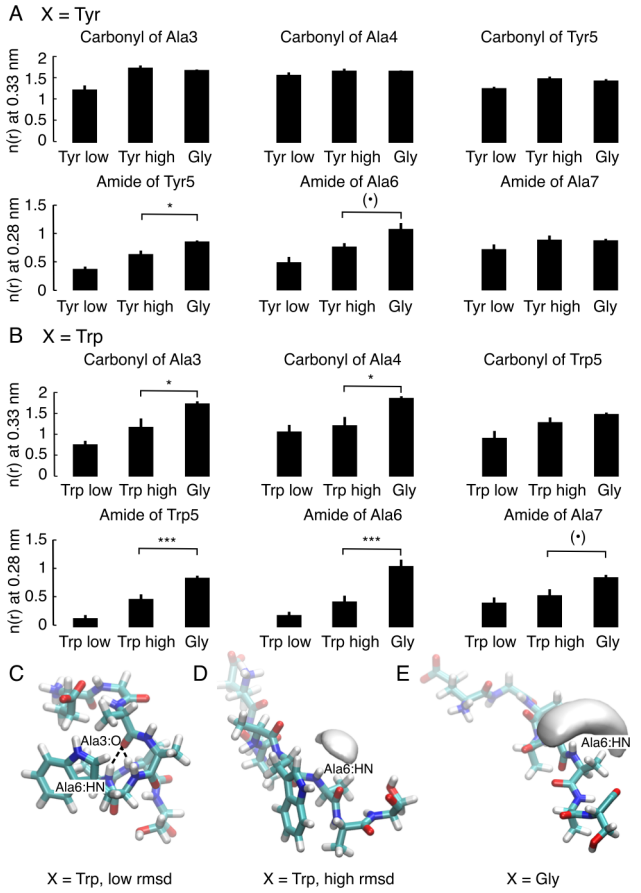

Figure 3. Backbone hydration. (A, B) Histograms showing the numbers of water molecules coordinating the amide hydrogen or carbonyl oxygen atoms of residues 3−7 for the peptides with (A) X = Tyr and (B) X = Trp. Data for X = Gly are shown as a reference. The "low" and "high" labels refer to pools of conformations that have low or high RMSDs with respect to the experimental RDC values. (C−E) Representative molecular structures for X = Trp (low and high RMSD) and X = Gly. The average water density within 3.5 Å of the amide hydrogen of residue Ala6 is shown for X = Trp with high RMSD (D) and X = Gly (E). The water density is isocontoured at 0.016 molecule/Å^3 (i.e., half of the bulk density).

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Methods, tables, and additional figures. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*simon.berneche@isb-sib.ch

Notes

The authors declare no competing financial interest.

DOI: 10.1021/jacs.5b00660
J. Am. Chem. Soc. 2015, 137, 4300−4303

■ ACKNOWLEDGMENTS

We are very grateful to Dr. Navratna Vajpai for help and advice in the initial stages of the project. This work was supported by grants from the Swiss National Science Foundation to S.B. (SNF Professorship 139205) and S.G. (31-149927), and a stipend from the Croucher Foundation to H.T.A.L. Calculations were performed at sciCORE (http://scicore.unibas.ch/) scientific computing core facility at University of Basel.

■ REFERENCES

(1) Anfinsen, C. B. *Science* **1973**, *181*, 223.
(2) Seife, C. *Science* **2005**, *309*, 78.
(3) Dill, K. A.; MacCallum, J. L. *Science* **2012**, *338*, 1042.
(4) Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. *Annu. Rev. Biophys.* **2008**, *37*, 289.
(5) Chouard, T. *Nature* **2011**, *471*, 151−153.
(6) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W. *J. Mol. Graphics Modell.* **2001**, *19*, 26.
(7) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197.
(8) Dyson, H. J.; Wright, P. E. *Methods Enzymol.* **2000**, *339*, 258.
(9) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111.
(10) Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 17908.
(11) Jensen, M. R.; Markwick, P. R. L.; Meier, S.; Griesinger, C.; Zweckstetter, M.; Grzesiek, S.; Bernado, P.; Blackledge, M. *Structure* **2009**, *17*, 1169.
(12) Meier, S.; Blackledge, M.; Grzesiek, S. *J. Chem. Phys.* **2008**, *128*, No. 052204.
(13) Dames, S. A.; Aregger, R.; Vajpai, N.; Bernado, P.; Blackledge, M.; Grzesiek, S. *J. Am. Chem. Soc.* **2006**, *128*, 13508.
(14) Sass, H.; Musco, G.; Stahl, S.; Wingfield, P.; Grzesiek, S. *J. Biomol NMR* **2000**, *18*, 303.
(15) Dyson, H. J.; Wright, P. E.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13057.
(16) Radivojac, P.; Iakoucheva, L. M.; Oldfield, C. J.; Obradovic, Z.; Uversky, V. N.; Dunker, A. K. *Biophys. J.* **2007**, *92*, 1439.
(17) Tompa, P. *Trends Biochem. Sci.* **2002**, *27*, 527.
(18) Cho, M.-K.; Kim, H.-Y.; Bernado, P.; Fernandez, C. O.; Blackledge, M.; Zweckstetter, M. *J. Am. Chem. Soc.* **2007**, *129*, 3032.
(19) Karplus, M.; Weaver, D. L. *Nature* **1976**, *260*, 404.
(20) Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719.
(21) Yang, A.-S.; Honig, B. *J. Mol. Biol.* **1995**, *252*, 351.
(22) Avbelj, F.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10967.
(23) Avbelj, F.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5742.
(24) Drozdov, A. N.; Grossfield, A.; Pappu, R. V. *J. Am. Chem. Soc.* **2004**, *126*, 2574.
(25) Cho, S. S.; Reddy, G.; Straub, J. E.; Thirumalai, D. *J. Phys. Chem. B* **2011**, *115*, 13401.
(26) Dill, K. A.; Fiebig, K. M.; Chan, H. S. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1942.
(27) Zwanzig, R.; Szabo, A.; Bagchi, B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 20.
(28) Levinthal, C. In *Mossbauer Spectroscopy in Biological Systems: Proceedings of a Meeting Held at Allerton House, March 17 and 18, 1969, Monticello, IL*; University of Illinois Press: Urbana, IL, 1969; pp 22−24.

DOI: 10.1021/jacs.5b00660
J. Am. Chem. Soc. 2015, 137, 4300−4303