FOR THE RECORD

Correlation between $^{13}C^{\alpha}$ chemical shifts and helix content of peptide ensembles

DANIEL S. WEINSTOCK, CHITRA NARAYANAN, JEAN BAUM, AND RONALD M. LEVY

BIOMAPS Institute, Rutgers University, Piscataway, New Jersey 08854-8087, USA
Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854-8087, USA

Abstract

Replica exchange molecular dynamics simulations are used to generate three ensembles of an S-peptide analog (AETAAAKFLREHMDS). Percent helicity of the peptide ensembles calculated using STRIDE is compared to percent helicity calculated from $^{13}C^{\alpha}$ chemical shift deviations (CSD) from random coil in order to test the assumption that CSD can be correlated to percent helicity. The two estimates of helicity, one based on structure and the other on CSD, are in close to quantitative agreement, except at the edges of helical stretches where disagreements of as much as 50% can be found. These disagreements can occur by CSDs both as an under- and an overestimate of peptide helicity. We show that underestimation arises due to ensemble averaging of positive CSDs from conformers with torsion angles in the helical region of Ramachandran space with negative CSDs corresponding to conformers of the peptide in the extended region. In contrast, overestimation comes about due to the fact that a large number of conformations with torsion angles in the helical region are not counted as helical by STRIDE due to a lack of correlated helical torsion angles in neighboring residues.

Keywords: chemical shift; peptide ensembles

Supplemental material: see www.proteinscience.org

Chemical shifts have a long history as a tool for determining the secondary structure of proteins (Szilagyi and Jardetzky 1989; Wishart and Sykes 1994a). This methodology was developed to identify secondary structure in native proteins, however, there has increasingly been interest in attempting to find regions of inherent structure within disordered polypeptide systems (Smith et al. 1996; Eliezer et al. 1998; Dyson and Wright 2001; Marsh et al. 2006). In this study we investigate the correlation between ensemble averaged chemical shift deviations (CSD) from random coil and peptide helicity. We demonstrate quantitatively that CSD is a good measure of percent helicity, but that this correspondence can break down at the edges of continuous stretches of helical residues.

A variety of methods have been proposed for characterizing secondary structure using chemical-shift data (Eliezer et al. 1998; Wang and Jardetzky 2002; Marsh et al. 2006). We chose to focus on $^{13}C^{\alpha}$ chemical shifts, which are known to be sensitive mainly to the backbone structure of a polypeptide chain (Spera and Bax 1991). The deviation of the $^{13}C^{\alpha}$ chemical shifts from random coil values can be used as a measure of the population of torsion angles in the α or β regions of Ramachandran space (Eliezer et al. 1998). Small peptides capable of folding into regular secondary structure can serve as useful models for unstructured proteins, since they too exist

in solution as ensembles of inter-converting conformers. Here we use one such peptide, an analog of the S-peptide (Mitchinson and Baldwin 1986) which forms a more stable helix in solution, to illuminate the relationship between CSD and the helicity of the peptide sequence. Using computer simulations, we show that while in general there is a correspondence between helicity and CSD, on the edges of helical segments, CSD can under- and overestimate the percent helicity of the polypeptide ensemble. This is possible both at the ends of the polypeptide sequence and at the edges created when short helical segments are broken by stretches of coil, a motif potentially quite common in natively unstructured proteins.

Materials and Methods

Ensembles of peptide conformations were generated as a function of temperature, ranging from 270 to 690 K, using replica exchange molecular dynamics (REMD) (Sugita and Okamoto 1999) and the OPLS-AA/AGBNP implicit solvent effective potential (Gallicchio and Levy 2004) within the IMPACT molecular simulation package (Banks et al. 2005). The large temperature range used is due to the fact that the helix melts at higher temperatures than observed experimentally because the current generation of implicit solvent models are not parameterized for temperature. This is of little consequence for this study because our goal is to generate ensembles with differing amounts of peptide helicity. Ensemble averaged CSDs were calculated for each ensemble by using SHIFTX (Neal et al. 2003) to calculate chemical shifts for the residues of each peptide conformation, subtracting a reference random coil chemical shift value, and then averaging over all the conformations in the ensemble. For internal consistency in the calculation of CSDs we used simulated random coil values rather than experimental values from the literature. The simulated random coil values were calculated by performing REMD simulations of blocked GGXGG peptides, where X was any of the 20 regular amino acids. These peptides are commonly used experimentally to measure random coil chemical shifts (Wishart et al. 1995; Schwarzinger et al. 2001). For each amino acid, the average chemical shift of a high temperature REMD ensemble was assigned as the simulated random coil chemical shift. Comparisons of the simulation random coil values to two sets of experimental random coil values are included in Supplemental Table S1.

S-peptide, produced by cleaving the first 19 residues off of the N-terminus of ribonuclease A, was the first small peptide shown to form a helix in solution—30% helical at 273 K (Klee 1968). For this study we chose to use a 15-residue analog of S-peptide (AETAAAKFLREHMDS), with increased stability (46% helix at 276 K) (Tirado-Rives and Jorgenson 1991). The analog differs from S-peptide

due to three point mutations—K1A, E9L, Q11E—and the absence of residues 16–19, which do not participate in the helix.

Results and Discussion

Three S-peptide analog ensembles, hereafter referred to as the low (270 K), intermediate (421 K), and high (690 K) temperature ensembles, were selected from the REMD simulations. These three ensembles were chosen because they presented three very different secondary structure profiles. This can be seen by looking at the percent helicity for each ensemble, calculated as the percentage of the ensemble for which each residue was identified as helical by STRIDE (Frishman and Argos 1995). As seen in Figure 1 (red bars), in the low temperature ensemble the peptide is composed of two short helices (residues 2–5 and 10–14) separated by the coil (residues 6–9); the intermediate temperature ensemble has a long helix stretching from residues 3–13; and at high temperature, the helix is fully melted.

$^{13}C^\alpha$ chemical shifts can also be used to determine secondary structure. One method for doing this is the chemical-shift index (CSI) (Wishart and Sykes 1994b), which identifies as a helix any group of three or more consecutive CSDs greater than 0.7 ppm. The secondary structure profile generated in this manner agrees almost entirely with that obtained using STRIDE. Here again we find two short helices separated by residues 6–9 at low temperature, a long helix from residues 3–13 at intermediate temperature, and no helix at high temperature (Fig. 2). There are two discrepancies between STRIDE and the CSI. One is at residue E2 in the low temperature ensemble, on the edge of the first short helix, which is
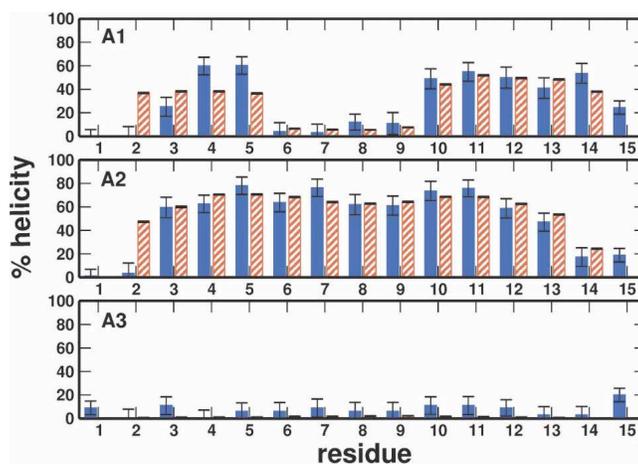

Figure 1. Comparison of percent helicity calculated as the percentage of conformations with a residue assigned as helix by STRIDE (red crosshatches) to percent helicity calculated as % helix (residue i) = $\Sigma_j(CSD_j/CSD_\alpha)$, where $CSD_\alpha = 3.1$ ppm (blue) for the low (*A1*), intermediate (*A2*), and high (*A3*) temperature ensembles.
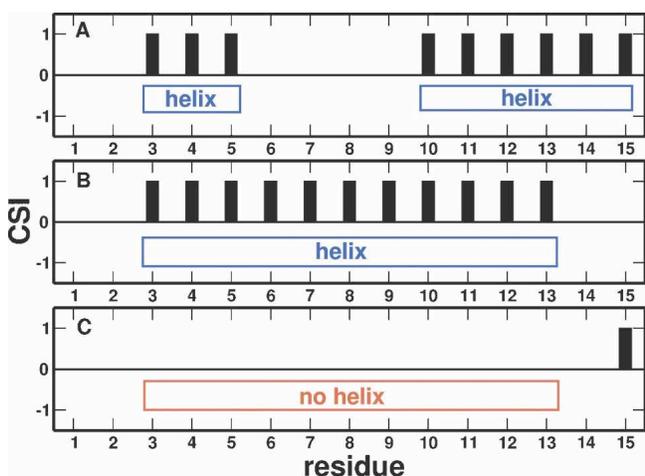
Figure 2. Chemical-shift index (CSI) plotted for each residue of the S-peptide analog for the low (*A*), intermediate (*B*), and high (*C*) temperature ensembles. Segments of the peptide sequence predicted to be helical by CSI are noted.

37% helical according to STRIDE, but not part of the helix according to CSI. The other is at residue S15 in the low temperature ensemble. For this residue, we find the opposite situation—0% helicity according to STRIDE, but part of the helix according to CSI.

An alternative way to generate information about secondary structure from $^{13}C^{\alpha}$ chemical shifts, one which allows for direct comparison with the percent helicity determined by STRIDE, is to assume a linear relationship between CSD and helicity (Eliezer et al. 1998; Dyson and Wright 2001). The percent helicity of each residue can then be calculated using the expression % helix (residue i) $\Sigma_j(CSD_j/CSD_{\alpha})$, where the sum is over all the members of the peptide ensemble. This places each ensemble averaged CSD on a continuum between a maximal CSD value which is equated with 100% helix, and the random coil reference value, which is assumed to be 0% helix. The maximum CSD ($CSD_{\alpha} = 3.1$) is the average chemical shift deviation of a fully helical residue (Spera and Bax 1991).

Overall, there is close to quantitative agreement between the percent helicities calculated from STRIDE and those derived from CSDs (Fig. 1). Again, both methods predict the same average secondary structure profiles for each of the three ensembles: two short helices separated by coil at low temperature, one long helix at intermediate temperature, and no helix at high temperature, respectively. Differences in the helicities occur at the edges of helical stretches, with the percentages calculated from CSDs able to both under- and overestimate the helicity of a particular residue. Underestimation of the helicity by CSD is seen at residue E2 in both the low temperature ensemble (% helix = 0% by CSD compared to 37% by

STRIDE) and in the intermediate temperature ensemble (% helix = 3% by CSD compared to 47% by STRIDE) (Fig. 1A,B). Overestimation is found at residues A4, A5, and S15 in the low temperature ensemble (% helix = 68%, 68%, and 29% by CSD compared to 38%, 37%, and 0% by STRIDE) (Fig. 1A).

$^{13}C^{\alpha}$ chemical shifts depend primarily on backbone torsion angles (de Dios et al. 1993), so Ramachandran plots of the residues for which there is a disagreement between the two helicity calculations can help illustrate why the estimate of helicity based on CSD differs from the helicity calculated using STRIDE. Figure 3 presents Ramachandran plots of two residues. In one, E2 in the low temperature ensemble (Fig. 3A,B), CSD underestimates the helicity: 0% versus 37% from STRIDE. In the other, A5 in the low temperature ensemble (Fig. 3C,D), CSD overestimates the helicity predicting the residue to be 68% helical when the STRIDE helicity value is again 37%. Figure 3, A and C show that 3685 out of 10,000 conformers for E2 and 3653 out of 10,000 conformers for A5, or 37% in each case, are identified by STRIDE as helical and that all of these conformers have backbone ϕ/ψ angles for these residues which are within the α region of Ramachandran space.

Though the Ramachandran plots for the conformers identified by STRIDE as being helical at E2 and A5 are almost identical, the torsion angles of these residues for the other members of the low temperature ensemble have two different distributions in ϕ/ψ space (Fig. 3B,D). For E2, where CSD underestimates the helicity, Figure 3B shows that close to 50% of the conformers in the ensemble have E2 torsion angles outside of the helical
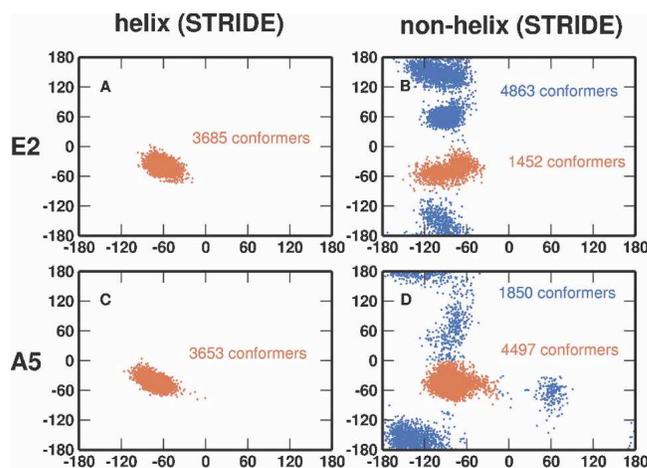

Figure 3. Ramachandran plots of residues E2 and A5 for conformers in the low temperature ensemble which STRIDE identifies as helical (*A,C*) and non-helical (*B,D*). Conformers with torsion angles in the α region of Ramachandran space are shown in red and conformers not in the α region are shown in blue.

section of ϕ/ψ space. Ensemble averaging of the negative CSDs, contributed by the half of the ensemble with E2 torsion angles outside the α region of Ramachandran space, together with the positive chemical shift deviations from the conformations which populate the helical region, yields an average CSD close to 0. Essentially, due to the averaging, the conformations with E2 torsion angles outside of the helical region are masking the presence of a significant fraction (51%) with torsion angles inside the helical region.

While there is still ensemble averaging taking place for residue A5, Figure 3D shows that less than 20% of the conformers in the low temperature ensemble have A5 torsion angles outside of the helical region. This means that the average chemical shift deviation is dominated by contributions from the helical region. In this case, the discrepancy between the percent helicity calculated from STRIDE and that calculated from the CSDs is related to the fact that CSD is a more localized property than helicity. Helicity depends not only on the ϕ/ψ angles of individual residues but also the correlation between consecutive ϕ/ψ pairs and $i,i + 4$ hydrogen bonds, while the chemical shift primarily depends on the ϕ/ψ angles of a particular residue (Eliezer et al. 1998). For residue A5 in the low temperature ensemble, this means that there are a large number of conformers, 4497 out of the 10,000 peptide conformations in the ensemble, with torsion angles in the α region of the Ramachandran plot that are not counted as helical by STRIDE. Even though these conformations are not included in the structural measurement of percent helicity, they all contribute positive CSDs to the ensemble average, thereby inflating the CSD measure of helicity relative to that based on STRIDE.

Chemical shift deviations for each of the amino acid residues were calculated using random coil values obtained from REMD simulations of blocked GGXGG peptides. This is the most consistent way to calculate CSDs for this study; any errors introduced through the simulation protocols and the use of SHIFTX to calculate chemical shifts are equally reflected in both the chemical shifts of the S-peptide and the GGXGG peptides used as references. Using the literature values of the random coil chemical shifts to calculate CSDs does not change the major conclusion of this work—chemical shift deviations are in general a good predictor of peptide helicity for heterogeneous ensembles like the S-peptide. Also, when using random coil chemical shifts from the literature (Wishart et al. 1995; Schwarzinger et al. 2001), we still observe situations, such as residue E2 in the low temperature ensemble, where the CSD underestimates the percent helicity. Overestimation of the percent helicity by CSD at the ends of some helical stretches is, however, diminished when using the literature random coil values, because of a small, systematic increase in the literature

random coil values relative to the simulated random coil shifts (see the Supplemental material).

Conclusion

In conclusion, computer simulations of an S-peptide analog have confirmed the correlation between chemical shift deviations from random coil and the percent helicity of a peptide sequence. We have shown that a given $^{13}C^\alpha$ CSD, taken as a fraction of the average chemical shift deviation of a residue in a fully formed helix, can be used as a quantitative measure of the percent helicity at that residue. There are a number of different chemical shift measures, such as CSI (Wishart and Sykes 1994b), SSP (Marsh et al. 2006), and the probability based methods of Jardetzky (Wang and Jardetzky 2002), that make use of this correlation to predict secondary structure propensities. While this correlation holds in most cases, it is possible for the chemical shift deviations to significantly overestimate or underestimate the helicity at the edges of helical segments. These two types of disagreement between CSD and helicity can occur even at edges caused by short turn or coil segments separating two short helical stretches. This is a pattern that is quite likely to be found in parts of large unstructured proteins, where care must be taken in defining the boundaries of segments with helical propensity.

Acknowledgments

References

Banks, J.L., Beard, H.S., Cao, Y., Cho, A.E., Damm, W., Farid, R., Felts, A.K., Halgren, T.A., Mainz, D.T., Maple, J.R., et al. 2005. Integrated modeling program, applied chemical theory (IMPACT). *J. Comput. Chem.* **26:** 1752–1780.

de Dios, A.C., Pearson, J.G., and Oldfield, E. 1993. Secondary and tertiary structure effects on protein NMR chemical shifts: An ab initio approach. *Science* **260:** 1491–1496.

Dyson, H.J. and Wright, P.E. 2001. Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. *Methods Enzymol.* **339:** 259–270.

Eliezer, D., Yao, J., Dyson, H.J., and Wright, P.E. 1998. Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat. Struct. Biol.* **5:** 148–155.

Frishman, D. and Argos, P. 1995. Knowledge based protein secondary structure assignment. *Proteins* **23:** 566–579.

Gallicchio, E. and Levy, R.M. 2004. AGBNP: An analytic implicit solvent model suitable for molecular dynamics and high-resolution modeling. *J. Comput. Chem.* **25:** 479–499.

Klee, W.A. 1968. Studies of the conformation of ribonuclease S-peptide. *Biochemistry* **7:** 2731–2736.

Marsh, D., Singh, V.K., Zongchao, J., and Forman-Kay, J.D. 2006. Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Sci.* **15:** 2795–2804.

Mitchinson, C. and Baldwin, R.L. 1986. The design and production of semisynthetic ribonucleases with increased thermostability by incorporation

of S-peptide analogues with enhanced helical stability. *Proteins* **1:** 23–33.

Neal, S., Nip, A.M., Zhang, H., and Wishart, D.S. 2003. Rapid and accurate calculation of protein ^{1}H, ^{13}C, and ^{15}N chemical shifts. *J. Biomol. NMR* **26:** 215–240.

Schwarzinger, S., Kroom, G.J., Foss, T.R., Chung, J., Wright, P.E., and Dyson, H.J. 2001. Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.* **123:** 2970–2978.

Smith, L.J., Fiebig, K.M., Schwalbe, H., and Dobson, C.M. 1996. The concept of a random coil: Residual structure in peptides and denatured proteins. *Fold. Des.* **1:** R95–R106. doi: 10.1016/S1359-0278(96)00046-6.

Spera, S. and Bax, A. 1991. Empirical correlation between protein backbone conformation and Cα and Cβ ^{13}C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **113:** 5490–5492.

Sugita, Y. and Okamoto, Y. 1999. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314:** 141–151.

Szilagyi, L. and Jardetzky, O. 1989. α-Proton chemical shifts and protein secondary structure. *J. Magn. Reson.* **83:** 441–449.

Tirado-Rives, J. and Jorgenson, W.L. 1991. Molecular dynamics simulations of the unfolding of an α-helical analogue of ribonuclease A S-peptide in water. *Biochemistry* **30:** 3864–3871.

Wang, Y. and Jardetzky, O. 2002. Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci.* **11:** 852–861.

Wishart, D.S. and Sykes, B.D. 1994a. Chemical shifts as a tool for structure determination. *Methods Enzymol.* **239:** 363–392.

Wishart, D.S. and Sykes, B.D. 1994b. The ^{13}C chemical-shift index: A simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR* **4:** 171–180.

Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., and Sykes, B.D. 1995. ^{1}H, ^{13}C, and ^{15}N random coil NMR chemical shifts of the common amino acids I. Investigations of nearest-neighbor effects. *J. Biomol. NMR* **5:** 67–81.