Position dependence of the ^{13}C chemical shifts of α -helical model peptides. Fingerprint of the 20 naturally occurring amino acids

JORGE A. VILA,^{1,2} HÉCTOR A. BALDONI,³ AND HAROLD A. SCHERAGA¹

¹Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301, USA

²Universidad Nacional de San Luis, Facultad de Ciencias Físico, Matemáticas y Naturales, Instituto de Matemática Aplicada San Luis, CONICET, San Luis, Argentina

³Universidad Nacional de San Luis, Departamento de Química, San Luis, Argentina

(RECEIVED June 12, 2004; FINAL REVISION July 28, 2004; ACCEPTED August 5, 2004)

Abstract

The position dependence of the ^{13}C chemical shifts was investigated at the density functional level for α -helical model peptides represented by the sequence $\text{Ac}-(\text{Ala})_i\text{-X}-(\text{Ala})_j\text{-NH}_2$, where X represents any of the 20 naturally occurring amino acids, with $0 \leq i \leq 8$ and $i + j = 8$. Adoption of the *locally dense* basis approach for the quantum chemical calculations enabled us to reduce the length of the chemical-shift calculations while maintaining good accuracy of the results. For the 20 naturally occurring amino acids in α -helices, there is (1) significant variability of the computed ^{13}C shielding as a function of both the guest residue (X) and the position along the sequence; for example, at the N terminus, the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ shieldings exhibit a uniform pattern of variation with respect to both the central or the C-terminal positions; (2) good agreement between computed and observed $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts in the interior of the helix, with correlation coefficients of 0.98 and 0.99, respectively; for $^{13}\text{C}^\alpha$ chemical shifts, computed in the middle of the helix, only five residues, namely Asn, Asp, Ser, Thr, and Leu, exhibit chemical shifts beyond the observed standard deviation; and (3) better agreement for four of these residues (Asn, Asp, Ser, and Thr) *only* for the computed values of the $^{13}\text{C}^\alpha$ chemical shifts at the N terminus. The results indicate that $^{13}\text{C}^\alpha$, but not $^{13}\text{C}^\beta$, chemical shifts are sensitive enough to reflect the propensities of some amino acids for specific positions within an α -helix, relative to the N and C termini of peptides and proteins.

Keywords: ^{13}C chemical shifts; α -helical peptides; helix breaker; locally dense basis approach

It is well documented that $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts in polypeptides are influenced mainly by backbone geometry and, therefore, are valuable quantities with which to identify secondary structure (Spera and Bax 1991; Kuszewski et al. 1995; Iwadata et al. 1999), with no influence of amino acid sequence (Iwadata et al. 1999; Xu and Case 2002). The influence of the backbone geometry appears as an approximately 4 ppm and 2 ppm chemical-shift separation between

α -helical and β -sheet residues from the corresponding experimental $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts, respectively (Spera and Bax 1991). Despite the enormous effort and progress in this field during the last few years (Iwadata et al. 1999; Wishart and Case 2001; Xu and Case 2001, 2002; Oldfield 2002), a detailed characterization of the factors affecting the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts for *each* of the 20 naturally occurring amino acids remains to be elucidated. As noted by Xu and Case (2002), “deciphering such effects from empirical data alone is a complex undertaking.” For this reason, Iwadata et al. (1999) and Xu and Case (2002), using semiempirical and ab initio calculations, respectively, carried out an analysis of the factors that affect the $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, and $^{13}\text{C}'$ chemical shifts. Among other factors that

Reprint requests to: Harold A. Scheraga, Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301, USA; e-mail: has5@cornell.edu; fax: (607) 254-4700.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.04930804>.

influence the shielding of these nuclei, Iwadata et al. (1999) discussed the effect of a backbone hydrogen bond on the ^{13}C chemical shifts, concluding that hydrogen bonding has a small but nonnegligible influence. They also illustrated the importance of such an effect, noting the dependence of ^{13}C chemical shifts on the position within α -helical structures. Later, Xu and Case (2002) extended this analysis by using quantum mechanical calculations on β -sheet and α -helical model peptides. However the position dependence within the α -helix for each of the 20 naturally occurring amino acids was not computed.

Figure 1 shows that there are clearly differences between the ends and the interior of an α -helical structure, that is, the end residues involve different numbers of intramolecular hydrogen bonds than the interior residues. Thus, certain questions arise and are addressed here: (1) How far into the interior does the end effect penetrate? (2) Are these effects amino-acid specific? and (3) How do these effects influence the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts for each of the 20 naturally occurring amino acids? Furthermore, it is necessary to demonstrate whether or not the position dependence of ^{13}C chemical shifts reflects the propensities observed for some amino acids to occupy the N or the C terminus of α -helices in proteins and peptides. If this correlation exists, its characterization could play a very important role (1) for refinement of NMR-derived 3D models, and (2) for accurate prediction of the secondary structure of a protein when the ^{13}C chemical shift assignments and the amino acid sequence are the only available information.

The existence of a distribution of amino-acid-residue preferences near the N and C termini of a helical segment has been known since the work of Kotelchuck et al. (1969). They carried out a statistical analysis of the distribution of α -helical and non- α -helical regions of the various dipeptide types appearing in a sample of seven proteins of known sequence and structure. They reported "...a sharp change in the nature of the observed dipeptide types when the helix-

Figure 1. Diagram of the backbone hydrogen-bond pattern for the sequence $\text{Ac}-(\text{Ala})_9\text{-NH}_2$ in the α -helical conformation (Poland and Scheraga 1970). Alanine residues are numbered from 1–9. Dotted vertical lines separate each amino acid residue in the sequence. Hydrogen bonds connecting the amide hydrogen and carbonyl groups are represented by horizontal lines. Unsatisfied hydrogen-bonded nitrogen (N) and carbonyl Carbon (C) are in bold. α -Carbons are in italics. If the N and C termini were not blocked, that is, if they were NH_2 and COOH , respectively, there would be four instead of three unsatisfied hydrogen-bonded NH and CO groups.

coil boundary is crossed..." These results were interpreted as evidence for the importance of short-range interactions in determining protein conformation. This concept of the dominance of short-range interactions (Scheraga 1973) was the basis for the host–guest approach to determine the helix-forming propensity of the 20 naturally occurring amino acids (Wojcik et al. 1990) and for the success (Lewis et al. 1970) in predicting the tendency toward helix formation in unfolded proteins (that are poised to fold) that matched the native structure; this concept later received direct experimental confirmation from fluorescence resonance energy transfer experiments (Navon et al. 2001).

Subsequently, Richardson and Richardson (1988) also showed that the observed distribution of amino acids at the N or the C terminus of α -helices is quite different from the interior positions. Almost simultaneously, a possible explanation of why amino acids residues occur with certain frequencies at these positions was suggested by Presta and Rose (1988) based on the depletion of hydrogen bonds at the ends (Fig. 1). Following the suggestions mentioned above, the distribution of amino acids at these sites has been examined in peptides (Nicholson et al. 1988; Bruch et al. 1991; Chakrabartty et al. 1993; Doig and Baldwin 1995) and proteins (Serrano and Fersht 1989; Lecomte and Moore 1991; Aurora et al. 1994; Zhukovsky et al. 1994).

From all this evidence, it is clear that both the amino acid composition and the structural properties at the N and C termini of the α -helical structure, shown in Figure 1, must be taken into account in the computation of the ^{13}C chemical shift. As was already noted, for the terminally blocked $\text{Ac}-(\text{Ala})_9\text{-NH}_2$ peptide, the first and last three amide hydrogens and carbonyl groups at the N and C termini, respectively, are not involved in $i-(i+4)$ backbone hydrogen bonds. The influence of the backbone hydrogen bond pattern, characteristic of the α -helix, on the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ shielding for each of the 20 naturally occurring amino acids will be discussed. It is worth noting that the number of unsatisfied backbone hydrogen bonds is four if the sequence shown in Figure 1 were unblocked.

The Results and Discussion section of this paper is organized as follows: In section I, we use ab initio quantum chemical calculations on α -helical model peptides to compute the position dependence of the ^{13}C chemical shift (here $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$) for each of the 20 naturally occurring amino acids by assuming that the residues are neutral, except for Aspartate at all positions for which we considered both protonation states. In section II, we analyze the effect of the state of ionization on the ^{13}C chemical shift at the specific position in the middle of the helix for *all* the ionizable residues considering both charged and uncharged states. The expected different chemical shifts at the end and in the middle of the α -helix, and their correlation with the values observed by Wang and Jardetzky (2002) will also be discussed in section III. This analysis will focus on residues for

which the differences between computed and observed values are beyond the standard deviation, namely for residues Asn, Asp, Ser, Thr, and Leu. The ^{13}C chemical shifts for Gly and Pro, which represent the most flexible and the most constrained amino acid residue, respectively, will also be analyzed in this section.

The section on Concluding Remarks deals with a discussion of the results and their role in the characterization and refinement of proteins. Finally, in Materials and Methods, the convention and approaches adopted to describe the α -helical model peptides as well as to compute the quantum chemical ^{13}C chemical shifts are described.

Results and Discussion

I. Position dependence effect on the ^{13}C chemical shifts

We have been able to characterize the position dependence of the $^{13}\text{C}^\alpha$ chemical shifts for each of the naturally occurring amino acids. Figure 2A shows, as an example, the variation of the $^{13}\text{C}^\alpha$ chemical shifts obtained for Lys as a function of the position of this residue along the α -helix. This figure shows that the $^{13}\text{C}^\alpha$ chemical shifts of the residues at the N terminus, that is, the first turn of the helix, are more shielded (upfield shift) with respect to the central residues of the molecule, viz., for residues in positions 4–7. This trend of the data is observed for 18 out of 19 residues. The exception is Gly because the most deshielded (downfield shift) $^{13}\text{C}^\alpha$ chemical shifts occur at position 1, as shown in Figure 2B. This particular behavior for Gly will be discussed in section III in connection with the N- and C-terminal propensity of this residue.

When we move from the C terminus to the middle of the helix, the pattern of changes for the $^{13}\text{C}^\alpha$ shielding shows a multifaceted behavior. Some amino acids (Lys, Arg, His,

Gln, Glu, Leu, and Met) are deshielded (downfield shift), in agreement with the empirical observation of Iwadate et al. (1999); for others (Phe, Ile, Asn, Ala, Ser, Asp, Tyr, Thr, Val, and Cys), there are, on average, no significant changes, while for Trp there is a shielding of the $^{13}\text{C}^\alpha$ atom.

For the N terminus, the $^{13}\text{C}^\beta$ shielding shows, as a characteristic pattern, that the first two residues are deshielded (downfield shift) with respect to the central residues of the molecule, viz., those occupying positions 4–7, as shown in Figure 3A for Ala. On the other hand, when we move from the C terminus to the middle of the α -helix these position dependence changes do not exhibit a unique behavior. This is illustrated by comparing Figure 3B for Ile with Figure 3A for Ala. Based on this evidence, when we move from the C terminus to the N terminus, we can categorize the amino acids residues as (1) showing a monotonic shielding (upfield shift for Ala, Asn, Asp, Cys, Phe, and Ser), (2) a monotonic deshielding (downfield shift for Val, Thr, Met, and Ile), or (3) a nonregular behavior (for Arg, Gln, Glu, His, Leu, Lys, Trp, and Tyr); of course, Gly has no C^β , and Pro is treated separately in section III.

This analysis shows that $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ shielding computed at the N terminus shows a uniform pattern of variations with respect to the rest of the molecule, while in the middle of the helix, as well as for the C terminus, the pattern of changes is not uniform. This asymmetry observed with the position may constitute valuable information to assess positional preferences of the amino acids, as we will discuss below. The origin of this effect could be attributed to the already-noted different hydrogen-bonded character between the ends and the interior of the α -helical conformation (see Fig. 1). For this reason, it is useful to provide the range of variations for the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts. Figure 4 shows the range of variation, defined as $\Delta^\xi = [^{13}\text{C}^\xi_{\text{max}} - ^{13}\text{C}^\xi_{\text{min}}]$, with $\xi = \alpha$ or β , for each of the naturally occurring amino acids except for $^{13}\text{C}^\beta$ for Gly and proline. $^{13}\text{C}^\xi_{\text{max}}$ and $^{13}\text{C}^\xi_{\text{min}}$ represent the *maximum* and

Figure 2. (A) Plot of the changes in the computed $^{13}\text{C}^\alpha$ chemical shift of Lys from the peptides $\text{Ac}-(\text{Ala})_i\text{-Lys}-(\text{Ala})_j\text{-NH}_2$, for $0 \leq i \leq 8$ (with $i + j = 8$), in the α -helical conformation. (B) Plot of the changes in the computed $^{13}\text{C}^\alpha$ chemical shift of Gly from the peptides $\text{Ac}-(\text{Ala})_i\text{-Gly}-(\text{Ala})_j\text{-NH}_2$, for $0 \leq i \leq 8$ (with $i + j = 8$), in the α -helical conformation.

Figure 3. (A) Plot of the changes in the computed $^{13}\text{C}^\beta$ chemical shift of Ala from the peptides Ac-(Ala) $_i$ -Ala-(Ala) $_j$ -NH $_2$, for $0 \leq i \leq 8$ (with $i + j = 8$), in the α -helical conformation. (B) Plot of the changes in the computed $^{13}\text{C}^\beta$ chemical shift of Ile from the peptides Ac-(Ala) $_i$ -Ile-(Ala) $_j$ -NH $_2$, for $0 \leq i \leq 8$ (with $i + j = 8$), in the α -helical conformation.

minimum chemical shift values computed for each residue X in the sequence Ac-(Ala) $_i$ -X-(Ala) $_j$ -NH $_2$ (with $i + j = 8$; $0 \leq i \leq 8$). The mean value and standard deviation for Λ^ξ (with $\xi = \alpha$ or β , computed over all the naturally occurring amino acids) are 1.92 ± 0.53 ppm and 2.25 ± 1.26 ppm for $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts, respectively. It is worth noting that the major factor affecting chemical shifts is the backbone geometry, inducing changes of about 4 ppm between α -helix and β -sheet conformations for $^{13}\text{C}^\alpha$ and about 2 ppm for $^{13}\text{C}^\beta$ chemical shifts, respectively (Spera and Bax 1991). By comparison, the site dependence of the ^{13}C chemical shift along the α -helix has a nonnegligible effect of up to about 3 ppm for the $^{13}\text{C}^\alpha$ chemical shift (of Tyr) and up to about 6 ppm for the $^{13}\text{C}^\beta$ chemical shift (of Ile) as shown in Figure 4.

Figure 4. Range of variation of the chemical shifts along the α -helix, defined as $\Lambda^\xi = [^{13}\text{C}^\xi_{\text{max}} - ^{13}\text{C}^\xi_{\text{min}}]$, with $\xi = \alpha$ or β , for each of the naturally occurring amino acids; except for $^{13}\text{C}^\beta$ for Gly and proline. Open and filled bars denote the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical-shift ranges of variations, respectively.

II. Evaluation of the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts in the middle of the α -helix

The difference (Δ) between the ^{13}C chemical shift for both charged and uncharged side chain and the values observed by Wang and Jardetzky (2002) for the ionizable residues Asp, Arg, Glu, His, Tyr, and Lys were computed for the central residue of the α -helix, that is, with $\Delta^\xi = [^{13}\text{C}^\xi_{\text{comp}} - ^{13}\text{C}^\xi_{\text{obs}}]$, with $\xi = \alpha$ or β . Results for Δ^α are shown in Figure 5. From this figure, we observe that, for any ionizable amino acid residue other than Tyr, the computed value of Δ^α value for a charged side chain is *larger* than the observed standard deviation. However, for uncharged side chains, the computed values of Δ^α are *smaller* than the observed standard deviations. This observation is in agreement with the result of Xu and Case (2002), who noted that often “electron distributions of charged molecules in water resemble those of the corresponding neutral species in the gas phase more closely than they resemble the gas-phase charged moiety.” This result indicates that, even though the average values observed by Wang and Jardetzky (2002) contain charged and uncharged side chains, it is better to use uncharged than charged side chains, except for Asp, to compute the $^{13}\text{C}^\alpha$ chemical shifts. Asp is the only ionizable group for which both a charged and an uncharged side chain lead to Δ^α values that are not within the observed standard deviation, as can be seen in Figure 5. Consideration of site position dependence, together with the protonation states of the side chain for this residue, will be discussed in section III.

Figure 6, A and B, shows the correlation between $^{13}\text{C}^\alpha$ (and $^{13}\text{C}^\beta$) chemical shifts computed for residues in the middle of the α -helix, for each of the 20 naturally occurring amino acids (using neutral residues) and the observed mean values of Wang and Jardetzky (2002) from a database containing >6100 amino acids. Similar correlation coefficient

Figure 5. Bars indicate the difference, $\Delta\alpha = [^{13}\text{C}_{\text{comp}}^{\alpha} - ^{13}\text{C}_{\text{obs}}^{\alpha}]$, between computed $^{13}\text{C}^{\alpha}$ chemical shift for the central position in the α -helical model peptide and the mean-value observed by Wang and Jardetzky (2002). Open bars represent values computed with charged side chains. The filled bars represent values computed with uncharged side chains. The observed standard deviation are indicated by the symbol ($\pm$).

results were obtained for both $^{13}\text{C}^{\alpha}$ and $^{13}\text{C}^{\beta}$ chemical shifts, respectively, that is, $R = 0.97$ (slope of 0.91 for the correlation line) and $R = 0.99$ (slope of 0.98 for the correlation line). The vertical lines in Figure 6A (and 6B) represent the standard deviations of the mean values observed by Wang and Jardetzky (2002).

It is known (Wishart et al. 1995), that a large spread of the ^{13}C chemical shift values (from 10 to 80 ppm as observed in Figure 6B for the $^{13}\text{C}^{\beta}$ chemical shifts; cf. 45 to 70 ppm observed in Figure 6A for the $^{13}\text{C}^{\alpha}$ chemical shifts) could mask the significance of the smaller deviations in Figure 6B compared to Figure 6A. For this reason, we carried out a comparison, residue by residue, of the difference (Δ) between the ^{13}C chemical shift computed for the central residue of the α -helix and the corresponding mean value observed by Wang and Jardetzky (2002). Figure 7A shows

the results of this comparison for the $^{13}\text{C}^{\alpha}$ chemical shift (as open bars) and Figure 7B the corresponding results for the $^{13}\text{C}^{\beta}$ chemical shift. In these figures, the vertical lines denote the standard deviation of the mean value observed by Wang and Jardetzky (2002). Five out of 20, and 15 out of 19 residues exceed the observed standard deviation for the $^{13}\text{C}^{\alpha}$ and $^{13}\text{C}^{\beta}$ chemical shift, respectively. This example clearly illustrates that the better agreement observed in Figure 6B, compared to Figure 6A, in terms of the correlation coefficient for the $^{13}\text{C}^{\beta}$ and $^{13}\text{C}^{\alpha}$ chemical shifts, was only apparent.

III. Analysis of the N- and C-terminal effects

Figure 7A (as open bars) shows that five residues possess computed $^{13}\text{C}^{\alpha}$ chemical shifts, which exceed the standard deviation observed by Wang and Jardetzky (2002), namely, Asn, Asp, Ser, Thr, and Leu. These residues, together with Gly and Pro will be analyzed in detail below.

Asn, Asp, Ser, Thr, and Leu

There is evidence (Kotelchuck et al. 1969; Presta and Rose 1988; Richardson and Richardson 1988) indicating the existence of different preferences of the amino acids between the N and C termini and the interior positions in the α -helix. This effect has been noted for peptides (Nicholson et al. 1988; Fairman et al. 1989; Bruch et al. 1991; Lyu et al. 1992; Doig and Baldwin 1995) and proteins (Serrano and Fersht 1989; Serrano et al. 1992; Zhukovsky et al. 1994). In particular, N-terminal preferences for peptides and proteins agree very well among themselves (Chakrabarty et al. 1993; Doig and Baldwin 1995). Among all these studies, Richardson and Richardson (1988) pointed out that Asn, Asp, Ser, Thr, and Gly, which are helix breakers (Wojcik et al. 1990), exhibit stronger preferences than the rest of the

Figure 6. (A) Correlation between the observed $^{13}\text{C}^{\alpha}$ chemical shifts (Wang and Jardetzky 2002) and the computed values at position 5 in the middle of the α -helix for each of the 20 naturally occurring amino acids. $R = 0.97$; slope of 0.91 for the correlation line. Only uncharged residues were used for this comparison. (B) Correlation between the observed $^{13}\text{C}^{\beta}$ chemical shifts (Wang and Jardetzky 2002) and the computed values at position 5 in the middle of the α -helix for each of the 19 naturally occurring amino acids. $R = 0.99$; slope of 0.98 for the correlation line. Only uncharged residues were used for this comparison.

Figure 7. (A) Open bars indicate the difference, $\Delta^\alpha = [^{13}\text{C}^\alpha_{\text{comp}} - ^{13}\text{C}^\alpha_{\text{obs}}]$, between computed $^{13}\text{C}^\alpha$ chemical shift for the *central position* in the α -helical model peptide and the mean value observed by Wang and Jardetzky (2002). The observed standard deviations are indicated by the symbol ($\perp$). All ionizable residues plotted as open bars were considered as uncharged. The black filled bars show the differences obtained for the $^{13}\text{C}^\alpha$ chemical shifts computed for residues Asn, Asp (considered here as Asp^-), Ser, Gly, and Thr at the N-terminal position. (B) Open bars indicate the difference, $\Delta^\beta = [^{13}\text{C}^\beta_{\text{comp}} - ^{13}\text{C}^\beta_{\text{obs}}]$, between computed $^{13}\text{C}^\beta$ chemical shift for the *central position* in the α -helical model peptide and the mean value observed by Wang and Jardetzky (2002). The observed standard deviations are indicated by the symbol ($\perp$). All ionizable residues were considered uncharged in this comparison.

naturally occurring amino acids to be at the N terminus in proteins. Conceivably, the mean values observed by Wang and Jardetzky (2002) should reflect these properties. To examine this possibility, we recomputed the values of Δ^α by using the $^{13}\text{C}^\alpha$ chemical shifts at the N terminus (shown in Fig. 7A as black-filled bars), in place of those computed in the middle of the α -helix (shown in Fig. 7A as open bars), for residues Asn, Ser, and Thr. The best agreement, in terms of Δ^α , for each of these residues was found with the $^{13}\text{C}^\alpha$ chemical shifts value computed at position 2 from the N terminus, except for Leucine.

The value of Δ^α computed for leucine (1.3 ppm), for the central position of the α -helix, shown in Figure 7A, is outside of the observed standard deviation (0.98 ppm). This disagreement cannot be explained by site position preferences. In fact, the value of Δ^α computed from the N terminus gives a higher disagreement (2.9 ppm). To understand whether or not the source of such differences could be due to the side-chain orientation that is adopted by this residue, we carried out additional conformational searches that included five (out of six clustered) conformations (as explained in Materials and Methods). The results indicate that the computed $^{13}\text{C}^\alpha$ chemical shift from the leading-member conformation of family number 3 (57.97 ppm) is closer to the observed one (57.54 ± 0.98 ppm) than the $^{13}\text{C}^\alpha$ chemical shift computed for the lowest energy conformation of family number 1 (56.26 ppm). However, the total energy of the leading member of family number 3 is 2.3 kcal/mole higher in energy than the lowest energy conformation used to compute the value shown in Figure 7A. Conceivably, side-chain interactions (not taken into account in our calculations) could stabilize this high-energy conformation and, hence, provide better agreement with the value observed by Wang and Jardetzky (2002).

The best agreement, in terms of Δ^α , with respect to the observed mean value for Asp, was obtained with the $^{13}\text{C}^\alpha$ chemical shifts computed at position 2 from the N terminus, by using the charged side chain (Asp^-). As already noted at the beginning of this subsection, an analysis for the N-terminal preferences in the α -helical conformation, observed for both peptides (Doig and Baldwin 1995) and proteins (Richardson and Richardson 1988), leads to a good correlation coefficient ($R = 0.86$). From their analysis, Asn, followed by Asp^- , are the amino acids residues with the highest propensity for the N-terminal position. This is a reflection of the tendency for their side chains to form a hydrogen-bonded ring with the backbone in a *nonhelical* backbone conformations (Lewis et al. 1973), and thereby serve as a helix breaker.

After considering the N-terminal preferences for Asn, Asp^- , Ser and Thr (shown as black-filled bars in Fig. 7A), and the remaining amino acids (shown as open bars in the same figure), 19 out of 20 $^{13}\text{C}^\alpha$ chemical shifts are within the observed standard deviation.

Gly

Ananthanarayanan et al. (1971) noted that the low value of the Zimm-Bragg parameter s (Zimm and Bragg 1959) found for glycine, in the temperature range of 0–70°, provides a quantitative basis for classifying this residue as a *helix breaker* in proteins. Kotelchuck et al. (1969) suggested an N- to C-terminal direction of helix propagation, which implies that Gly possesses a stronger preference for the C-terminal position, in line with the prediction made that this residue is a *helix-breaker* (Lewis et al. 1970; Ananthanarayanan et al. 1971). Later, in agreement with this finding, Schellman (1980) noted that about one-third of all

α -helices terminate with a Gly at the C terminus. Richardson and Richardson (1988) found that Gly possesses the sharpest preference among all amino acid residues for the C terminus than for the N terminus in *proteins*. Rules to explain this preference in proteins were proposed by Aurora et al. (1994). However, Doig and Baldwin (1995) found no preference for Gly at the C terminus in their experiments on α -helical *peptides*. Nevertheless, all the above evidence indicates that glycine shows a very low propensity for the middle of the α -helix.

In our theoretical calculations, good agreement in terms of Δ^α (-0.8 ppm) was obtained for the computed $^{13}\text{C}^\alpha$ chemical shifts at position 5, that is, in the middle of the α -helix, as shown in Figure 7A. Despite this good agreement, we checked whether the value of Δ^α could be improved by considering the preference of Gly for the N- or the C-terminal position. Using the $^{13}\text{C}^\alpha$ chemical shifts computed at the 1 to 9 positions, shown in Figure 2B, we found that the best agreement in terms of Δ^α , that is, the one that led to $\Delta^\alpha = 0.1$ ppm in Figure 7A (as a black bar), is found with the value of the $^{13}\text{C}^\alpha$ chemical shifts (47.15 ppm) computed at position 1 (compared to the mean observed value of 47.02 ± 0.90 ppm). By contrast, the computed values of Δ^α for $^{13}\text{C}^\alpha$ chemical shifts at both the 8 and 9 positions gave values of -0.4 and -1.0 ppm, respectively. These results show that better agreement with the observed values, in terms of Δ^α , is found for the values of the $^{13}\text{C}^\alpha$ chemical shifts computed at the N terminus than those in the middle of the helix.

Pro

The computed value plotted in Figure 7A for proline is in good agreement with the observed mean value of Wang and Jardetzky (2002). The lowest energy conformation, that is, with $\phi = -133.18$ and $\psi = 78.85$, from which this value was computed, shows that the alanine residue preceding proline lies in the D region of the map of Zimmermann et al. (1977), while proline itself is in the α -region of the Ramachandran map, with a *trans* peptide bond and the pyrrolidine ring in the up (U) conformation. This particular conformation seen for alanine is in agreement with the observation of MacArthur and Thornton (1991) that, for alanine preceding proline, the entire region $-180 < \psi < 60$ on the Ramachandran map is completely inaccessible, with ϕ showing the characteristic preferences of the Ramachandran map.

For the lowest energy conformation identified in the EDMC conformational search, all alanines following proline displayed backbone dihedral angles in the α -region of the Ramachandran map. Hence, in our calculations, proline appears at the N-terminal position of the α -helix, in agreement with the observation of MacArthur and Thornton (1991), who noted that proline in the first turn is almost exclusively in the first position. Because of this observation,

we initiated the calculations by placing proline *only* in the middle of the helix.

End effects

Our results show that modification of the $^{13}\text{C}^\alpha$ chemical shifts to include the correction for the N-terminal preferences for Asn, Asp⁻, Ser, Thr, and Gly residues, does not improve the correlation coefficient found in Figure 7A significantly, viz., only from $R = 0.97$ to 0.98 . However, the slope of the correlation line in Figure 8 shows a noteworthy change, viz., from 0.91 to 0.99 (which should be compared with the ideal value of 1.0). Comparison of Figure 7A and Figure 7B shows a greater disagreement for $^{13}\text{C}^\beta$ than for $^{13}\text{C}^\alpha$ chemical shifts between computed and mean values observed by Wang and Jardetzky (2002). The reason for such behavior can be found in the analysis of the physical factors that influence the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts (Wishart and Case 2001; Neal et al. 2003), such as ring current effects, torsional (ϕ/ψ) effects, dihedral angles χ 's effects, etc.

Concluding remarks

In this work, we carried out ab initio quantum chemical calculations of the ^{13}C chemical shifts for an α -helical model peptide. In particular, we carried out a detailed comparison of the computed ^{13}C chemical shifts for the α -helical model peptide with observed mean values and standard deviations obtained from an NMR database. From such a comparison, we conclude that (1) the good agreement seen between computed and observed ^{13}C chemical shifts shows that the lowest energy conformation obtained from a gas-phase potential (represented by ECEPP/3) constitutes a good α -helical representation with which to model the observed mean value for the ^{13}C chemical shift; (2) use of

Figure 8. Correlation between the observed $^{13}\text{C}^\alpha$ chemical shifts (Wang and Jardetzky 2002) and the computed values for each of the 20 naturally occurring amino acids, after taking into account the correction for the N-terminal preferences of residues Asn, Asp⁻, Ser, Gly, and Thr. $R = 0.98$; slope of 0.99 for the correlation line.

uncharged side chains, on average, gives better results when compared with observed ^{13}C chemical shifts than charged residues do; (3) consideration of the N-terminal preference for some amino acids improves the agreement between computed and observed $^{13}\text{C}^\alpha$ chemical-shift values; and (4) the position dependence of the computed $^{13}\text{C}^\alpha$ chemical shifts, mainly at the N terminus, is a fingerprint that characterizes the 20 naturally occurring amino acids in the α -helical conformation. This effect appears to be a consequence of the asymmetry between ends and interior of α -helices, which in turn, is a consequence of the uneven distribution of the backbone hydrogen bonds. In particular, this site dependence may constitute valuable information that should be considered, among other important contributions: (1) for refinement of NMR-derived 3D models, and (2) for accurate prediction of secondary structure of proteins when, the ^{13}C chemical-shift assignments and the sequence are the only available information.

Materials and methods

Peptide design

A plot of the frequency distribution of amino acids versus helical length, based on data from the PDB, shows that, on average, a helix length between 5 and 14 residues accounts for more than 75% of the total number of observed helices in proteins (Penel et al. 1999). Based on this observation, an α -helix of nine residues seems to be a good model with which to study the position dependence of the ^{13}C chemical shift because (1) it represents a tradeoff between short and long helices observed in proteins; (2) this length is in close agreement with a previous study of the average length of helices in proteins, which found that the distribution is broad with a mean length of ten residues (Barlow and Thornton 1988); and (3) quantum chemical calculations for a helix of nine residues is feasible in reasonable computational time by using accurate approaches (Vila et al. 2004), as explained later.

It should be noted that, in a terminally blocked α -helix, the initial three NH groups and the final three CO groups differ from the rest by not being able to participate in the complete intramolecular hydrogen bonding that is characteristic of the α -helix, as shown in Figure 1. In this figure, it can also be seen that all the central five residues (numbers 3–7) are bracketed by three backbone hydrogen bonds. They thus differ from the N- and C-terminal residues that are bracketed by fewer hydrogen bonds. Based on these observations, we chose, as an α -helical model, peptides with the sequences $\text{Ac}-(\text{Ala})_i\text{-X}-(\text{Ala})_j\text{-NH}_2$ (with $i+j=8$; $0 \leq i \leq 8$) and X as any of the 20 naturally occurring amino acids. Cysteine was studied *only* in the reduced (–SH) form.

Alanine was selected as the dominant component of the model peptide because (1) among all the amino acids, this residue is one of the most frequently occurring ones in an α -helical conformation (Chou and Fasman 1974; Chen et al. 2003), and (2) its methyl side chain will not exhibit much interference with the side chain of the guest (X) residue.

Modeling the α -helical peptide and clustering analysis

The α -helical conformation for the sequence $\text{Ac}-(\text{Ala})_9\text{-NH}_2$ was generated with the Electrostatically Driven Monte Carlo (EDMC)

method (Ripoll and Scheraga 1988). An all-atom representation of the chain was used with the ECEPP/3 force field (Momany et al. 1975; Némethy et al. 1983, 1992; Sippl et al. 1984). Starting from a canonical α -helix ($\phi = -60.0^\circ$; $\psi = -40.0^\circ$), the lowest energy minimum conformation (-24.69 kcal/mole) identified among 300 accepted conformations (by the Metropolis criterion) was adopted as the α -helix model conformation for this peptide. In this conformation, *all* the alanine residues are in the α -helical region of the Ramachandran map (Ramachandran et al. 1963), with $\phi = -60.0^\circ \pm 10^\circ$ and $\psi = -40.0^\circ \pm 10^\circ$.

When Ala was replaced by any amino acid (X), except for proline (see later section), in the sequence: $\text{Ac}-(\text{Ala})_i\text{-X}-(\text{Ala})_j\text{-NH}_2$ (with $i+j=8$; $0 \leq i \leq 8$), the backbone was kept fixed at the lowest energy minimum identified during the EDMC conformational search for the $\text{Ac}-(\text{Ala})_9\text{-NH}_2$ peptide. Adoption of this conformation will ensure that the site-position preference of X is influenced only by the hydrogen-bonded backbone geometry, the charge characteristic of the α -helix, and the nature of X.

Only for leucine, a classification of the accepted conformations generated with the EDMC procedure was carried out with the clustering procedure used by Vila et al. (2002, 2003) to study statistical-coil peptides in solution, that is, through a minimal tree (the Minimal Spanning Tree [MST]) method (Kruskal 1956). The minimal tree was then partitioned in terms of a specified RMSD cutoff, leading to a defined number of families. The families resulting from the RMSD clustering procedure were ranked in increasing order according to their total free energy. For each family, both the number of conformations and the set of dihedral angles of the lowest energy member were stored. We refer to the lowest energy conformation of a family as the leading member.

Modeling the side-chain orientation

Polar or ionizable groups have a strong tendency to be exposed to the solvent, and hence, great mobility of their side chains is expected to occur. The side-chain position of nonpolar groups in an α -helix will be dominated by side-chain–side-chain interactions, mainly in the interior of a protein. In modeling the α -helical conformation, no attention is paid to side-chain–side-chain interactions or to side-chain mobility. It is not feasible to model both of these effects by quantum chemical calculations with the available computational resources. Therefore, the following procedure was used: For each guest residue X in the peptide $\text{Ac}-(\text{Ala})_4\text{-X}-(\text{Ala})_4\text{-NH}_2$ (with X being any naturally occurring amino acid other than glycine and proline), we carried out an EDMC conformational search in which the *only* allowed variables were the dihedral angles χ of residue X. During this conformational search, *all* the backbone dihedral angles (ϕ, ψ) were fixed at the values obtained for the lowest energy conformation of the α -helix, obtained with the sequence $\text{Ac}-(\text{Ala})_9\text{-NH}_2$, as described in the previous section. The dihedral angles χ , from the lowest energy conformation, identified in this restricted EDMC conformational search for each guest residue X, were adopted as the *representative* side-chain orientation for each respective amino acid, independent of the backbone position that it occupied within the helical structure. The ^{13}C chemical shifts computed in this manner, for each of the 20 naturally occurring amino acids in the middle of the α -helix, were compared with the observed mean values of Wang and Jardetzky (2002), as shown in Figure 6, A and B.

In contrast to the adoption of the low-energy backbone conformation of $\text{Ac}-(\text{Ala})_9\text{-NH}_2$ for 19 residues in the middle position of the α -helix, a different approach was taken for proline. For this residue, only in position X₅ in the sequence $\text{Ac}-(\text{Ala})_4\text{-X}_5-(\text{Ala})_4\text{-NH}_2$, an EDMC conformational search was carried out taking into

account (1) a canonical α -helical structure ($\phi = -60.0^\circ$; $\psi = -40.0^\circ$) as the starting conformation, (2) *all* backbone and side-chain dihedral angles of the whole chain considered as variables during the minimization procedure, (3) *cis* $\leftrightarrow$ *trans* isomerization of the peptide group for the proline residue, and (4) both up (U) and down (D) packing conformations of the pyrrolidine ring, which pertain to the ($\phi = -53.0^\circ$ and $\chi^1 = -28.1^\circ$) and ($\phi = -68.8^\circ$ and $\chi^1 = 27.4^\circ$) positions, respectively, of the C γ atom of this residue. The ^{13}C chemical shift was computed *only* for the lowest energy conformation identified in this conformational search for proline in the middle of the helix, that is, for the peptide Ac-(Ala) $_4$ -Pro-(Ala) $_4$ -NH $_2$.

To understand why there was disagreement between the computed and observed ^{13}C chemical shifts for leucine (as shown in the Results and Discussion section), a more detailed analysis of the side-chain conformational preferences was carried out. The 100 accepted side-chain conformations obtained with the EDMC conformational search in which the dihedral angles χ of the leucine side chain were the only variables, were clustered at an RMSD of 0.2 Å, without a cutoff in energy. The leading member of the first five (out of six) families was used to compute the range of variation for the ^{13}C chemical shifts, as a function of the side-chain orientation. The energy of the sixth family was too high to be of significance.

Modeling the position-dependence of the ^{13}C chemical shifts

The position dependence of the ^{13}C chemical shifts for a given guest residue X, other than proline, was computed by starting with X at the N terminus of the α -helix, and shifting it by one residue at a time up to the C terminus; that is, we computed the ^{13}C chemical shift for the guest residue X in the sequence Ac-(Ala) $_i$ -X-(Ala) $_j$ -NH $_2$ (with $i + j = 8$) for $0 \leq i \leq 8$. The backbone dihedral angles (ϕ, ψ) were always fixed at the values found for the α -helix model conformation of poly (L-alanine) defined earlier, and the dihedral angles (χ) were those found for the lowest energy conformation of the guest residue X in the middle of the helix, as explained in the previous section. Following this procedure, that is, fixing the dihedral angles ϕ , ψ , and χ , we determined the changes that occur in the ^{13}C chemical shift *only* as a consequence of the change in position along the α -helix.

The proline residue was studied *only* in the middle of the helix because (1) there is experimental evidence showing that proline can be classified as a *helix breaker* (Lewis et al. 1970; Altmann et al. 1990) and (2) data from the PDB indicate that proline residues exhibit a preference to be either in the first turn of an α -helix or after the C terminus (MacArthur and Thornton 1991).

Quantum-chemical calculations of the ^{13}C chemical shift

It is possible to obtain theoretical shielding values of good quality by using large basis sets located *only* on the atoms whose shifts are of interest while the rest of the atoms in the molecule are treated with more modest basis sets (Chesnut and Moore 1989). This is called the *locally dense* basis approach, and its use enables us to minimize the length of the chemical-shift calculations while maintaining the accuracy of the results (Laws et al. 1995; Vila et al. 2003, 2004). Using this approach, we were able to show (Vila et al. 2004) that a fairly small increase in the quality of the ^{13}C chemical shift, by treating sets of three, five or seven consecutive residues, instead of one, with a locally dense basis set did not justify the large decrease (by more than 17 times) in speedup.

Based on this observation, in this work we decided to compute the ^{13}C chemical shifts by treating a *single* residue, that is, the guest residue X, with a *locally dense* [6-311+G(2d,p)] basis set, while the rest of the molecule is treated with the simpler 3-21G basis set. This notation refers to the basic basis sets of Pople and colleagues (Hehre et al. 1986) as implemented in Gaussian-98 (Frisch et al. 1998). All the calculated isotropic shielding values (σ) were referenced with respect to a tetramethylsilane (TMS) ^{13}C chemical shift scale (δ), as described previously (Vila et al. 2002).

Acknowledgments

This research was supported by grants from the NIH (GM-14312 and TW006335) and the NSF (MCB00-03722). Support was also received from the National Research Council of Argentina (CONICET) (PIP-02485), and from the Universidad Nacional de San Luis (P-328402), Argentina. Part of this research was conducted using (1) the resources of the Cornell Theory Center, which receives funding from Cornell University, New York State, federal agencies, foundations, and corporate partners; and (2) the National Partnership for Advanced Computational Infrastructure at the Pittsburgh Supercomputing Center, which is supported in part by the NSF (MCA99S007P).

References

- Altmann, K.H., Wojcik, J., Vázquez, M., and Scheraga, H.A. 1990. Helix-coil stability constants for the naturally occurring amino acids in water. XXIII. Proline parameters from random poly(hydroxybutylglutamine-co-L-proline). *Biopolymers* **30**: 107–120.
- Ananthanarayanan, V.S., Andreatta, R.H., Poland, D., and Scheraga, H.A. 1971. Helix-coil stability constants for the naturally occurring amino acids in water. III. Glycine parameters from random poly(hydroxybutylglutamine-co-glycine). *Macromolecules* **4**: 417–424.
- Aurora, R., Srinivasan, R., and Rose, G. 1994. Rules for α -helix termination by glycine. *Science* **264**: 1126–1130.
- Barlow, D.J. and Thornton, J.M. 1988. Helix geometry in proteins. *J. Mol. Biol.* **201**: 601–619.
- Bruch, M.D., Dhingra, M.M., and Gierasch, L.M. 1991. Side chain-backbone hydrogen bonding contributes to helix stability in peptides derived from an α -helical region of carboxypeptidase A. *Proteins* **10**: 130–139.
- Chakrabarty, A., Doig, A.J., and Baldwin, R.L. 1993. Helix capping propensities in peptides parallel those in proteins. *Proc. Natl. Acad. Sci.* **90**: 11332–11336.
- Chen, H., Zhou, X., and Ou-Yang, Z.-C. 2003. Classification of amino acids based on statistical results of known structures and cooperativity of protein folding. *Phys. Rev. E* **65**: 061907.
- Chesnut, D.B. and Moore, K.D. 1989. Locally dense basis-sets for chemical-shift calculations. *J. Comp. Chem.* **10**: 648–659.
- Chou, P.Y. and Fasman, G.D. 1974. Conformational parameters for amino-acids in helical, β -sheet, and random coil regions calculated from proteins. *Biochemistry* **13**: 211–222.
- Doig, A.J. and Baldwin, R.L. 1995. N- and C-capping preferences for all 20 amino acids in α -helical peptides. *Protein Sci.* **4**: 1325–1336.
- Fairman, R., Shoemaker, K.R., Stewart, J.M., and Baldwin, R.L. 1989. Further studies of the helix dipole model: Effects of a free α -NH $^+$ or α -COO $^-$ group on helix stability. *Proteins* **5**: 1–7.
- Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Zakrzewski, V.G., Montgomery Jr., J.A., Stratmann, R.E., Burant, J.C., et al. 1998. Gaussian 98, Revision A.7. Gaussian Inc., Pittsburgh, PA.
- Hehre, W.J., Radom, L., Schleyer, P., and Pople, J.A. 1986. *Ab initio molecular orbital theory*. John Wiley and Sons, New York.
- Iwade, M., Asakura, T. and Williamson, M.P. 1999. C $^\alpha$ and C $^\beta$ carbon-13 chemical shifts in proteins from an empirical database. *J. Biomol. NMR* **13**: 199–211.
- Kotelchuck, D., Dygert, M., and Scheraga, H.A. 1969. The influence of short-range interactions on protein conformation. III. Dipeptide distributions in proteins of known sequence and structure. *Proc. Natl. Acad. Sci.* **63**: 615–622.

- Kruskal Jr., J.B. 1956. On the shortest spanning sub tree of a graph and the traveling salesman problem. *Proc. Am. Math. Soc.* **7**: 48–50.
- Kuszewski, J., Qin, J.A., Gronenborn, A.M., and Clore, G.M. 1995. The impact of direct refinement against C-13(α) and C-13(β) chemical-shifts on protein-structure determination by NMR. *J. Magn. Reson. Ser. B* **106**: 92–96.
- Laws, D.D., Le, H., de Dios, A.C., Havlin, R.H., and Oldfield, E. 1995. A basis size dependence study of Carbon-13 nuclear magnetic resonance spectroscopic shielding in Alanine and Valine fragments: Toward protein shielding hypersurfaces. *J. Am. Chem. Soc.* **117**: 9542–9546.
- Lecomte, J.T.J. and Moore, C.D. 1991. Helix formation in apocytochrome b5: The role of a neutral histidine at the N-cap position. *J. Am. Chem. Soc.* **113**: 9663–9665.
- Lewis, P.N., Gö, N., Gö, M., Kotelchuck, D., and Scheraga, H.A. 1970. Helix probability profiles of denatured proteins and their correlation with native structures. *Proc. Natl. Acad. Sci.* **65**: 810–815.
- Lewis, P.N., Momany, F.A., and Scheraga, H.A. 1973. Energy parameters in polypeptides. VI. Conformational energy analysis of the N-acetyl N'-methyl amides of the twenty naturally occurring amino acids. *Israel J. Chem.* **11**: 121–152.
- Lyu, P.C., Zhou H.X., Jelveh, N., Wemmer, D.E., and Kallenbach, N.R. 1992. Position-dependent stabilizing effects in α -helices-N-terminal capping in synthetic model peptides. *J. Am. Chem. Soc.* **114**: 6560–6562.
- MacArthur, M.W. and Thornton, J.M. 1991. Influence of proline residues on protein conformation. *J. Mol. Biol.* **218**: 397–412.
- Momany, F.A., McGuire, R.F., Burgess, A.W., and Scheraga, H.A. 1975. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. *J. Phys. Chem.* **79**: 2361–2381.
- Navon, A., Ittah, V., Landsman, P., Scheraga, H.A., and Haas, E. 2001. Distributions of intramolecular distances in the reduced and denatured states of bovine pancreatic ribonuclease A. Folding initiation structures in the C-terminal portions of the reduced protein. *Biochemistry* **40**: 105–118.
- Neal, S., Nip, A.M., Zhang, H., and Wishart, D.S. 2003. Rapid and accurate calculation of protein ^1H , ^{13}C and ^{15}N chemical shifts. *J. Biomed. NMR* **26**: 215–240.
- Némethy, G., Pottle, M.S., and Scheraga, H.A. 1983. Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids. *J. Phys. Chem.* **87**: 1883–1887.
- Némethy, G., Gibson, K.D., Palmer, K.A., Yoon, C.N., Paterlini, G., Zagari, A., Rumsey, S., and Scheraga, H.A. 1992. Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. *J. Phys. Chem.* **96**: 6472–6484.
- Nicholson, H., Becktel, W.J., and Matthews, B.W. 1988. Enhanced protein thermo-stability from designed mutations that interact with α -helix dipoles. *Nature* **336**: 651–656.
- Oldfield, E. 2002. Chemical shifts in amino acids, peptides, and proteins: From quantum chemistry to drug design. *Annu. Rev. Phys. Chem.* **53**: 349–378.
- Penel, S., Morrison, R.G., Mortishire-Smith, R.J., and Doig, A.J. 1999. Periodicity in α -helix lengths and C-capping preferences. *J. Mol. Biol.* **293**: 1211–1219.
- Poland, D. and Scheraga, H.A. 1970. *Theory of helix-coil transitions in biopolymers*. Academic Press, New York.
- Presta, L.G. and Rose, G.D. 1988. Helix signal in proteins. *Science* **240**: 1632–1641.
- Ramachandran, G.N., Ramakrishnan, C., and Sasisekharan, V. 1963. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**: 95–99.
- Richardson, J.S. and Richardson, D.C. 1988. Amino acid preferences for specific locations at the ends of α -helices. *Science* **240**: 1648–1652.
- Ripoll, D.R. and Scheraga, H.A. 1988. On the multiple-minima problem in the conformational analysis of polypeptides. II. An electrostatically driven Monte Carlo method-tests on poly(L-alanine). *Biopolymers* **27**: 1283–1303.
- Schellman, C. 1980. The α_L conformation at the ends of helices. In *Protein folding* (ed. R. Jaenicke), p. 53. Elsevier/North-Holland, New York.
- Scheraga, H.A. 1973. On the dominance of short-range interactions in polypeptides and proteins. *Pure Appl. Chem.* **36**: 1–8.
- Serrano, L. and Fersht, A.R. 1989. Capping and α -helix stability. *Nature* **342**: 296–299.
- Serrano, L., Sancho, J., Hirshberg, M., and Fersht, A.R. 1992. α -Helix stability in proteins. I. Empirical correlations concerning substitution of side-chains at the N- and C-caps and replacement of alanine by glycine or serine at solvent-exposed surfaces. *J. Mol. Biol.* **227**: 544–559.
- Sippl, M.J., Némethy, G., and Scheraga, H.A. 1984. Intermolecular potentials from crystal data. 6. Determination of empirical potentials for O–H...O=C hydrogen bonds from packing configurations. *J. Phys. Chem.* **88**: 6231–6233.
- Spera, S. and Bax, A. 1991. Empirical correlation between protein backbone conformation and C^α and C^β and ^{13}C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **113**: 5490–5492.
- Vila, J.A., Ripoll, D.R., Baldoni, H.A., and Scheraga, H.A. 2002. Unblocked statistical-coil tetrapeptides in aqueous solution: A theoretical study. *J. Biomol. NMR* **24**: 245–262.
- Vila, J.A., Baldoni, H.A., Ripoll, D.R., and Scheraga, H.A. 2003. Unblocked statistical-coil tetrapeptides in aqueous solution: Quantum-chemical computation of the Carbon-13 NMR chemical shifts. *J. Biomol. NMR* **26**: 113–130.
- . 2004. Fast and accurate computation of the ^{13}C chemical shifts for an alanine-rich peptide. *Proteins Struct. Funct. Bioinf.* **57**: 87–98.
- 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. and Case, D.A. 2001. Use of chemical shifts in macromolecular structure determination. *Methods Enzymol.* **338**: 3–34.
- Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., and Sykes, B. 1995. H-1, C-13 and N-15 random coil NMR chemical-shifts of the common amino-acids. 1. Investigations of nearest-neighbor effects. *J. Biomol. NMR* **5**: 67–81.
- Wojcik, J., Altmann, K.H., and Scheraga, H.A. 1990. Helix-coil stability constants for the naturally occurring amino acids in water. XXIV. Half-cystine parameters from random Poly(hydroxybutylglutamine-co-S-methylthio-L-cysteine). *Biopolymers* **30**: 121–134.
- Xu, X.-P. and Case, D.A. 2001. Automated prediction of ^{15}N , $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$ and $^{13}\text{C}'$ chemical shifts in proteins using a density functional database. *J. Biomol. NMR* **21**: 321–333.
- . 2002. 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* **65**: 408–423.
- Zhukovsky, E.A., Mulkerrin, M.G., and Presta, L.G. 1994. Contribution to global protein stabilization of the N-capping box in human growth hormone. *Biochemistry* **33**: 9856–9864.
- Zimm, B.H. and Bragg, J.K. 1959. Theory of the phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.* **31**: 526–535.
- Zimmerman, S.S., Pottle, M.S., Némethy, G., and Scheraga, H.A. 1977. Conformational analysis of the 20 naturally occurring amino acid residues using ECEPP. *Macromolecules* **10**: 1–9.