J Biomol NMR (2007) 37:137–146
DOI 10.1007/s10858-006-9118-6

ARTICLE

Effects of side-chain orientation on the ^{13}C chemical shifts of antiparallel β-sheet model peptides

Myriam E. Villegas · Jorge A. Vila · Harold A. Scheraga

Received: 6 January 2006 / Accepted: 3 November 2006 / Published online: 19 December 2006

Abstract The dependence of the ^{13}C chemical shift on side-chain orientation was investigated at the density functional level for a two-strand antiparallel β-sheet model peptide represented by the amino acid sequence Ac-$(Ala)_3$-X-$(Ala)_{12}$-NH_2 where X represents any of the 17 naturally occurring amino acids, i.e., not including alanine, glycine and proline. The dihedral angles adopted for the backbone were taken from, and fixed at, observed experimental values of an antiparallel β-sheet. We carried out a cluster analysis of the ensembles of conformations generated by considering the side-chain dihedral angles for each residue X as variables, and use them to compute the ^{13}C chemical shifts at the density functional theory level. It is shown that the adoption of the *locally-dense* basis set 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 17 naturally occurring amino acids in an antiparallel β-sheet, there is (i) good agreement between computed and observed $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts, with correlation coefficients of 0.95 and 0.99, respectively; (ii) significant variability of the computed $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts as a function of χ^1 for all amino acid residues except Ser; and (iii) a smaller, although significant, dependence of the computed $^{13}C^{\alpha}$ chemical shifts on χ^{ξ} (with $\xi \geq 2$) compared to χ^1 for eleven out of seventeen residues. Our results suggest that predicted $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts, based only on backbone (ϕ,ψ) dihedral angles from high-resolution X-ray structure data or from NMR-derived models, may differ significantly from those observed in solution if the dihedral-angle preferences for the side chains are not taken into account.

Keywords Chemical shift prediction · Torsional side-chain effects · Neighboring residue effects · Structure calculations

Introduction

β-sheet structures, first proposed by Pauling & Corey (1951, 1953), are the second most common secondary-structure element in proteins after α-helices (Creighton 1984). More than 75% of protein domains contain β-sheet elements, and their topology has been studied extensively (Chothia et al. 1977; Zhang and Kim 2000). In globular proteins, α-helices and β-sheets are short, with α-helices ranging from 10 to 15 residues and β-sheets consisting of 2–6 strands with 3–10 residues in each strand (Creighton 1984; Rossmeisl et al. 2004).

Most α-helices are well represented by canonical values of the dihedral angles, i.e., with $\phi = -62° \pm 8°$ and $\psi = -42° \pm 10°$ (Spera and Bax 1991), although some exceptions appear for solvent exposed α-helices which are often bent. This regularity of the backbone

Electronic supplementary material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s10858-006-9118-6 and is accessible for authorized users.

M. E. Villegas · J. A. Vila
Facultad de Ciencias Físico Matemáticas y Naturales, Instituto de Matemática Aplicada San Luis, Universidad Nacional de San Luis, CONICET, Ejército de Los Andes, San Luis 950-5700, Argentina

J. A. Vila · H. A. Scheraga (✉)
Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301, USA
e-mail: has5@cornell.edu

dihedral angles enabled us to study, at the density functional level, the influence of the backbone hydrogen-bond pattern, characteristic of the α-helix, on the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shielding for each of the 20 naturally occurring amino acids (Vila et al. 2004a).

Although it is well documented that $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts depend on secondary structure (Spera and Bax 1991; Kuszewski et al. 1995; Iwadate et al. 1999), with no influence of amino acid sequence (Iwadate et al. 1999; Xu and Case 2002), a detailed characterization of the factors affecting the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$chemical shifts for *each* of the 20 naturally occurring amino acid in β-sheets remains to be elucidated. For example, there are not enough data for some amino acid residues to clearly define the dependence of the chemical shift on the χ^1 dihedral angle (Iwadate et al. 1999). Preliminary attempts to clarify this dependence by ab initio quantum chemical investigation of a selected set of individual amino acids, viz., Gly, Ala, Val, Ile, Ser and Thr, were carried out by the Oldfield group (Havlin et al. 1997; Pearson et al. 1997). Their results rationalize many interesting behaviors of the observed chemical shifts in proteins, such as the increase in the isotropic shielding of sheet over helical geometries as well as the existence of a strong dependence of the ^{13}C chemical shifts on *both* the backbone dihedral angles (ϕ,ψ) and the side-chain geometry, e.g., variations up to 12 ppm were reported for valine residues as a function of the χ^1 dihedral angle. Later, Xu and Case (2002) used density functional calculations on β-sheet model peptides to study conformational effects on ^{15}N, $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ and $^{13}C'$ chemical shifts, associated with hydrogen bonding, backbone conformation, and side-chain orientation. However, characterization of side-chain effects on ^{13}C chemical shifts was studied for a *canonical* planar pleated β-sheet model peptide, for all the naturally occurring amino acids, except Gly and Pro, assuming three possible values for the χ^1dihedral angle (–60°, 60° and 180°), and the χ^2 and χ^3 effect was studied only for Ile and Glu, assuming some particular fixed combinations of the χ^1 values. However, the majority of β-sheets are twisted, rather than planar, with a right-hand twist in the approximately ±30° range for the backbone dihedral angles (Chothia et al. 1977; Chou and Scheraga 1982; Chou et al. 1982; Creighton 1984). Thus, the conformational parameters for β-sheets may deviate from those for planar pleated sheets and, hence, difficult to model by using canonical values. The fact that β-sheets in proteins appear as parallel or antiparallel strands, or a combination of both, only exacerbates the modeling problem.

Clearly, it is not feasible to model both backbone dihedral angles and side-chain effects simultaneously by quantum chemical calculations with the available computational resources. Hence, among all possible sets of backbone dihedral angles to represent the two-strand antiparallel β-sheet model peptide, we decided to adopt those belonging to the 16-residue segment from Gly-41 to Glu-56 of the B3 binding domain of protein G (pdb code 1P7E, shown in Fig. 1) (Ulmer et al. 2003). In particular, in this paper we focus on the effects of side-chain orientation on the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts for this two-stranded antiparallel β-sheet model peptide. The anticipated results will (i) yield a range of variation of the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts as a function of the side-chain orientations (χ's) for all naturally occurring amino acid residues, except for Ala, Gly and Pro; (ii) determine whether or not side chain-side chain interactions may be revealed by the observed ^{13}C chemical shifts, taken from a database containing >6,100 amino acids (Wang and Jardetzky 2002); (iii) indicate the extent to which there is a relationship between the observed ^{13}C chemical shifts (Wang and Jardetzky 2002) and the most frequently observed χ^1 dihedral angles in proteins (McGregor et al. 1987); and (iv) provide an understanding of the physical reasons for the disagreement between the computed and observed ^{13}C chemical shifts for all naturally occurring amino acids.

Methods

a. Peptide design

In this work, no distinction is made between β-hairpins and β-sheet residues for the following reasons. β-hairpins,

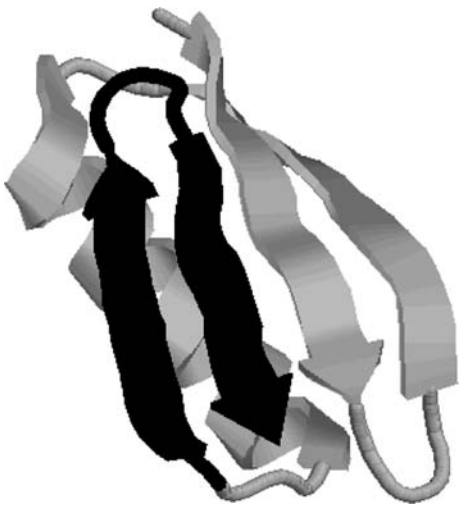

Fig. 1 Ribbon diagram of the protein 1P7E. The anti-parallel β-sheet template selected as a model peptide is highlighted in black, and belongs to the 16-residue C-terminal part (Gly41-Glu56) of the protein

the smallest β-sheet motif, consist of two antiparallel hydrogen-bonded β-strands linked by a loop region and occur frequently in globular proteins as both a separate motif as well as forming part of an extended β-sheet structure (Sibanda and Thornton 1991). Santiveri et al. (2001) reported that the average *conformational shifts*, defined as the deviation of the ^{13}C chemical shifts from their corresponding statistical-coil values, for the residues of β-hairpin strands (from 23 selected proteins), have values of –1.6 ± 1.3 ppm and 1.7 ± 1.8 ppm, for $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts, respectively. These values are in good agreement with those reported by Spera and Bax (1991) based on 126 β-sheet residues, namely, –1.5 ± 1.2 ppm and 2.2 ± 1.9 ppm, for $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts, respectively. In other words, no distinction seems to have been made for *conformational shifts* between residues in β-sheets or β-hairpins. Likewise, the averaged values observed by Wang and Jardetzky (2002) do not discriminate between residues in β-hairpins and β-sheets. Wang and Jardetzky (2002) reported averaged $^{1}H^{N}$, ^{15}N, $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ chemical shifts, together with their standard deviations, categorized according to three types of secondary structure, namely, statistical-coil, β-strand and α-helix. On the other hand, we focus on the computation of the effects of side-chain conformation on the ^{13}C chemical shifts for amino acids in antiparallel β-sheets without consideration of the size and characteristic of the loop region.

Creighton (1984) pointed out, ''β-sheets can consist entirely of parallel or antiparallel strands or can have a mixture of the two. Purely parallel sheets are least frequent; purely antiparallel sheets are most common. Antiparallel sheets often consist of just two or three strands, whereas parallel sheets always have at least four. Mixed sheets usually contain 3–15 strands. Adjacent strands in a sheet tend to be the strands that are also adjacent in the primary structure. This correlation is greatest for antiparallel strands and least for parallel strands.'' We quote this observation to illustrate the complexity of modeling a β-sheet secondary structure element. Computation of the ^{13}C chemical shift at the density functional level for β-sheet model peptides only exacerbates this problem since quantum chemical calculations usually are very demanding in terms of CPU time. Hence, we decided to model the backbone-dihedral angles of the antiparallel β-sheet peptide with the values of the backbone dihedral angles adopted by the 16 amino acid residues of the antiparallel C-terminal portion of the B3 binding domain of protein 1P7E (Ulmer et al. 2003). The protein 1P7E consists of a central α-helix packed against a four-stranded antiparallel-parallel-antiparallel β-sheet (shown in Fig. 1). The reasons for this choice were: (i) the adopted β-sheet model peptide satisfied both the Creighton (1984) and Rossmeisl et al. (2004) criterion for this secondary structural element, i.e., β-sheets typically consist of 2–6 strands with 5–10 amino acid residues in each strand; (ii) the two-stranded antiparallel β-sheet model peptide derived from protein 1P7E, shown in Fig. 2, is not a planar structure held together by hydrogen bonds but is rather a twisted structure of the type that is frequently observed in β-sheets (Chothia et al. 1977; Chothia and Janin 1981; Chou et al. 1982); (iii) the total number of amino acid residues involved, viz., sixteen, makes it feasible to compute the ^{13}C chemical shifts at the density functional level in a reasonable amount of computational time with the available resources, i.e., by using the locally-dense basis set approach (as explained in the Methods section d); and (iv) the protein 1P7E has been solved by X-ray diffraction to a resolution of 1.1 Å (Derrick and Wigley 1994) and by refinement of the X-ray structure with dipolar couplings (Ulmer et al. 2003).

Based on these considerations for the backbone dihedral angles, we model the proposed antiparallel β-sheet peptides with the alanine-based sequences Ac-$(Ala)_3$-X-$(Ala)_{12}$-NH_2 where X is any of the 17 naturally occurring amino acids, and the backbone dihedral angles are those of the corresponding region of protein 1P7E. Cysteine was studied *only* in the reduced (–SH) form.

Alanine exhibits a low frequency of occurrence in β-sheets (Chou and Fasman, 1974; Burgess et al. 1974; McGregor et al. 1987). However, this residue was selected as the dominant component of the model peptide because its methyl side chain will not exhibit much interference with the side chain of the guest (X) residue. Although Gly is even a simpler amino acid residue, with a single hydrogen as a side chain, it possesses an unusual conformational freedom and hence it is not a good model for the rest of the naturally occurring amino acids.

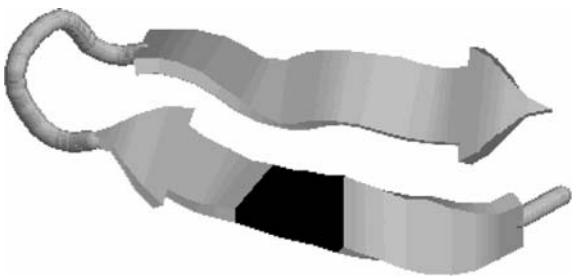

Fig. 2 The corresponding *regularized* alanine-based template of the anti-parallel β-sheet model peptide as explained in Methods section b. The position of the guest amino acid residue, i.e., X, in the sequence Ac-$(Ala)_3$-X-$(Ala)_{12}$-NH_2, is highlighted in black

b. Conversion of the native structure of 1P7E to rigid ECEPP/3 geometry

In order to carry out the present study, the experimentally determined conformation of protein 1P7E was *regularized, i.e.*, all residues were replaced by the standard ECEPP/3 residues (Némethy et al. 1992) in which bond lengths and bond angles are fixed (rigid geometry approximation), and hydrogen atoms were added.

The conversion process was carried out by generating the new rigid-geometry conformation from the N-terminus by adding one residue at a time and minimizing the *rmsd* between all heavy atoms in the generated fragment and the corresponding fragment in the experimental structure (Ripoll et al. 2005). The procedure was iterated until the C-terminal group was added to the chain. The final conformations resulting from this regularization procedure are quite close to the experimental ones within an rmsd of 0.13 Å for all the heavy atoms.

c. Modeling the side-chain orientation and clustering procedure

Spera and Bax (1991) showed that a clear distinction could be made between the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts for α-helices and β-sheet residues in proteins, which suggests a major direct effect of the backbone torsion angles. However, other effects such as hydrogen bonding and side-chain orientation might also be important. For example, the effect of the hydrogen bond on the position dependence of the ^{13}C chemical shifts has recently been investigated for α-helical structures (Vila et al. 2004a). In the latter application, starting from a canonical α-helix ($\phi = -60.0°$; $\psi = -40.0°$), the backbone was always kept fixed at the lowest-energy-minimum identified among 300 accepted conformations (by the Metropolis criterion). Adoption of this conformation ensured that the site-position preference for each amino acid residue is influenced only by the hydrogen-bonded backbone geometry, the charge characteristic of the α-helix, and the nature of the residue (with no consideration of side-chain orientation).

As already mentioned, a similar analysis for β-sheet structures is more difficult because the backbone dihedral angles for this secondary structure element deviate considerably from ideality due to the twist among strands, i.e., residues in β-sheets show substantial variation ($\langle\phi\rangle = -114° \pm 32°$; $\langle\psi\rangle = 142° \pm 20°$) while in α-helices these angles have a much narrower distribution ($\langle\phi\rangle = -62° \pm 8°$; $\langle\phi\rangle = 42° \pm 10°$) [Spera and Bax 1991]. Thus, the question that arises is: to what extent are the chemical shifts of the amino acids residues in a β-sheet affected by the side-chain orientation? The basis for such a query arises from the fact that the three torsion angles ϕ, ψ and χ^1 are not independent of each other over the whole range but they involve a common bond ($N–C^{\alpha}$) [Dumbrack and Karplus 1993, 1994; Chakrabarti and Pal 1998]. However, we did not restrict our analysis only to the χ^1 dihedral angle or to the three staggered positions of the γ-atom ($\chi^1 = 60°$, 180° and –60°). Because the complexity of the side-chain effects grows with the size of the side chain, starting from Ala, for each guest residue X in the peptide Ac-$(Ala)_3$-X-$(Ala)_{12}$-NH_2 (with X being any naturally occurring amino acid other than alanine, proline and glycine because of their low propensities to form a β-sheet structure), we carried out an Electrostatically Driven Monte Carlo (EDMC) [Ripoll and Scheraga 1988] 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 *regularized* structure of the protein. 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, i.e., 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 ECEPP/3 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*.

All accepted side-chain conformations for each amino acid residue, other than Ala, Gly and Pro, obtained with the EDMC conformational search in which the dihedral angles χ of the side chain were the only variables, were clustered at an *rmsd* of 0.1 Å except for Arg, Met, Phe and Thr, for which *rmsd* values of 0.42 Å, 0.28 Å, 0.05 Å, and 0.05 Å, respectively, were adopted, without a cutoff in energy. The adopted *rmsd* is a trade-off between the number of families and the amount of information desired; not more than 8 families were considered for any amino acid residue. The leading member of each family was used *only* to rank the families, and the DFT calculations were carried out for each leading member.

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

There is sound evidence (Laws et al. 1995; Havlin et al. 1997; Pearson et al. 1997) that, 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), it is possible to obtain theoretical shielding values of good quality. Use of this approximation, called the *locally-dense* basis set approach, enabled us to minimize the length of the chemical-shift calculations while maintaining the accuracy of the results (Vila et al. 2003, 2004a, b). In particular, we were able to show (Vila et al. 2004b) that a fairly small increase in the accuracy of the calculated ^{13}C chemical shift, by treating sets of three, five or seven *near-neighbor* residues in a left-handed all-*trans* PP_{II} conformation, instead of one, with a *locally-dense* basis set did not justify the large decrease (by more than 17 times) in speedup. However, near-neighbor residues in β-sheets may include residues *far away* in the sequence, such as those in consecutive strands of the β-sheet. For this reason, the dependence of the computed $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts and the CPU time required to carry out such calculations on the number of near-neighbor residues in both *sequence and space*, treated with a locally-dense basis set, was determined by a simple method that involves a series of quantum chemical calculations indicated as P*n* (with n = 1–5) in Eq. (1). For these calculations, a few consecutive alanine residues (those within the parentheses and in bold face) are treated with a *locally dense* basis set, while the rest of the molecule is treated with the simpler 3-21G basis set. In Eq. (1), the residues involved in the β-turn conformation are denoted in italics.

$$\begin{aligned}
&\text{Ac - AAA (}\underline{\mathbf{A}}\text{) AA} - \textit{AAAA} - \text{AAAAAA - NH}_2 \\
&= \text{(P1)}; \\
&\text{Ac - AAA (}\underline{\mathbf{A}}\text{) AA} - \textit{AAAA} - \text{AA (}\mathbf{A}\text{) AAA - NH}_2 \\
&= \text{(P2)}; \\
&\text{Ac - AA (}\mathbf{A}\underline{\mathbf{A}}\mathbf{A}\text{) A} - \textit{AAAA} - \text{AAAAAA - NH}_2 \\
&= \text{(P3)}; \\
&\text{Ac - AA (}\mathbf{A}\underline{\mathbf{A}}\mathbf{A}\text{) A} - \textit{AAAA} - \text{A (}\mathbf{AAA}\text{) AA - NH}_2 \\
&= \text{(P4)}; \\
&\text{Ac - (}\mathbf{AAA}\underline{\mathbf{A}}\mathbf{AA}\text{)} - \textit{AAAA} - \text{(}\mathbf{AAAAAA}\text{) - NH}_2 \\
&= \text{(P5)};
\end{aligned} \tag{1}$$

The alanine in position 4 in these sequences is underlined because its ^{13}C chemical shifts were used for comparing the results from the different calculations. In other words, for a chosen calculation, the reported chemical-shift values *always* correspond to the Ala residue in position 4. Sequences P*n* were designed to test the influence on the computed values of the ^{13}C chemical shifts for Ala in position 4 when a *supplementary* locally-dense basis is applied to: (i) the nearest-neighbor residues in the sequence (P3); (ii) the nearest-neighbor residues in space (P2), i.e. interactions that occur between consecutives strands; and (iii) the nearest neighbors in sequence *and* space (P4 and P5).

As seen from Fig. 3, the computed $^{13}C^{\alpha}$ chemical shifts for Ala in position 4 from P1 led to similar values as those derived from P2, P3, P4 and P5 calculations. In addition, the results of all the calculations are themselves within the standard deviation observed by Wang and Jardetzky (2002) indicating that there is no significant dependence on the number of residues treated with a locally-dense basis set and no significant influence of residues treated with a locally dense basis set belonging to the nearest strand. Figure 4 illustrates the speedup (S) of the calculation (Sosa et al. 1998). As was noted in previous analyses (Vila et al. 2004b), there is an extraordinary reduction of computational cost without significant loss of accuracy.

Based on this observations, in this work we decided to compute the ^{13}C chemical shifts by treating a *single* residue, i.e., the guest residue X in position 4, with a

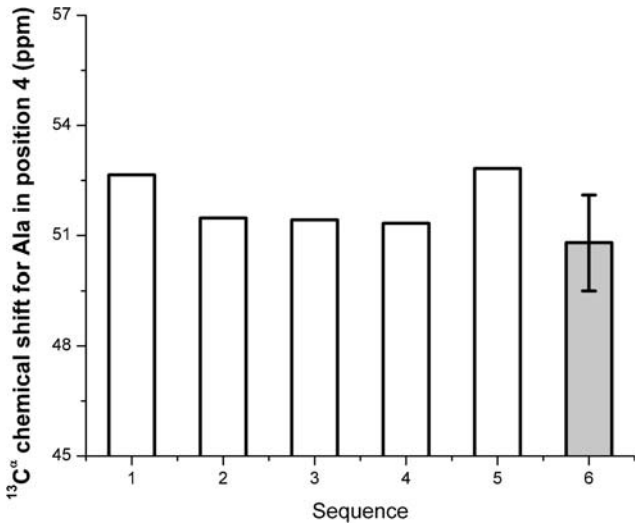

Fig. 3 $^{13}C^{\alpha}$ chemical shifts for Ala in position 4 in the antiparallel β-sheet model peptide [Ac-$(Ala)_{16}$-NH_2] (shown in Fig. 2) versus the number of neighbor residues treated with the locally-dense basis set approach, i.e., by using the sequence P1, P2, P3, P4 and P5, as explained in Methods section d. Grey filled bar, P*6*, represents the observed mean-average value of Wang and Jardetzky (2002) with their corresponding standard deviation ([])

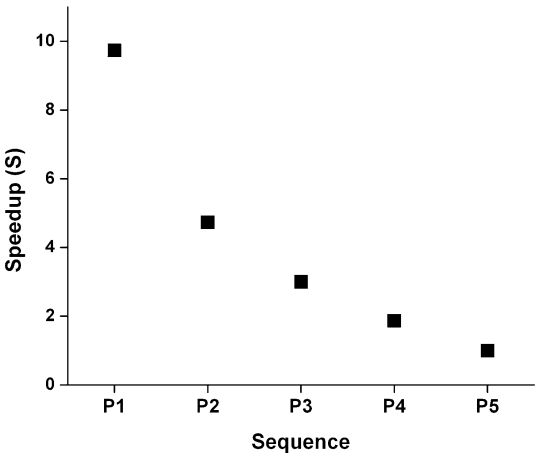

Fig. 4 Speedup (S) of the calculation versus the number of neighbor residues treated with the locally-dense basis set approach, i.e., by using the sequence P1, P2, P3, P4 and P5, as explained in Methods. The Speedup (Sosa et al. 1998) is defined as $S = T_s/T_p$ where T_s represents the serial run time and T_p the time to do the same calculation in P ($P > 1$) processors

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 co-workers (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). 2,2-Dimethyl-2-silapentane-5-sulfonic acid (DSS) was used directly or indirectly as the zero chemical shift reference of ^{1}H, ^{15}N and ^{13}C for all the proteins from which the data were collected (Wang and Jardetzky 2002). To reference the calculated isotropic shielding values (σ) with respect to DSS, a correction factor of 1.7 ppm (Wishart et al. 1995) was added to *all* computed ^{13}C chemical shifts.

Even though the average values observed by Wang and Jardetzky (2002) contain charged and uncharged side chains, uncharged rather than charged side chains were used to compute the ^{13}C chemical shifts for *all* the ionizable groups. The reason to adopt this criterion is based on the observation of Xu and Case (2002) 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...''

e. Comparison between predicted and observed chemical shifts

For each amino acid residue (μ), except Ala, Gly and Pro, we computed the *mean* $^{13}C^{\xi}$ (with $\xi = \alpha$ or β) chemical shift value as: $\langle {}^{13}C\mu^{\xi}\rangle = (1/N)\ \Sigma_{\{\rho\}}{}^{13}C_{\rho}{}^{\xi}$, where $1 \le \rho \le N$, and N represents the total number of clustered families of the ensemble (see Table S1 of supplementary material). The *mean* $^{13}C^{\xi}$ chemical shifts, rather than Boltzman-averaged values, obtained for each amino acid residue are used to compare with the ones observed by Wang and Jardetzky (2002). The reason to adopt this criterion is that the Boltzman-averaged value is dominated by the leading member of family No. 1 among all the clustered families of the ensemble. In addition, in the current procedure, the total energy for each conformation corresponds to that of an alanine-based protein, with X replacing one alanine, and hence it does not contain contributions from side chain-side chain interactions, because these interactions are sequence and conformationally dependent. By contrast, the experimentally observed values for a given residue pertain to an ensemble of conformations influenced by the rapid interplay among *all* these interactions. For a given amino acid residue (μ), the deviations of the chemical shifts of the leading member of each family from the corresponding observed values is computed as $\Lambda^{\xi} = [{}^{13}C^{\xi}_{\mu,\mathrm{comp}} - {}^{13}C^{\xi}_{\mu,\mathrm{obs}}]$.

Results and discussion

I-a. Range of variations between computed and observed ^{13}C chemical shifts

In general, we found that the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts, except for Ser, were *all* quite sensitive to the value of χ^1 (see Table S1 of supplementary material), i.e., variation of χ^1 gave rise to variation in the ^{13}C chemical shifts greater than the standard deviation observed by Wang and Jardetzky (2002). As a consequence, the computed ^{13}C chemical shifts for the leading member of each family show a wide range of variations. In particular, for 14 out of 17 amino acid residues there is at least one conformation, among *all* the leading members of the clustered ensemble, showing good agreement (within the standard deviation) with the $^{13}C^{\alpha}$ averaged-values observed by Wang and Jardetzky (2002) [see Table S1 of supplementary material]. Exceptions are Asn, Lys and Glu.

Spera and Bax (1991) noted that the major factor affecting the chemical shifts is the backbone geometry, inducing changes of about 4 ppm in $^{13}C^{\alpha}$ chemical shifts and about 2 ppm for $^{13}C^{\beta}$ chemical shifts between α-helix and β-sheet conformations, respectively. Likewise, the side chains provide similar $^{13}C^{\alpha}$ chemical shifts variation, Λ^{ξ}, for an antiparallel β-sheet model

peptide, i.e., up to 3.4 ppm for Λ^{α} is seen for Met. However, larger variations, namely 8.2, 5.8 and 5.0 ppm for Cys, Phe and Lys, respectively, are seen from the computed values of Λ^{β} (see Table S1 of supplementary material).

I-b. Relationship between the χ^1dihedral angle of the leading members and the most frequently observed value in proteins

McGregor et al. (1987) analyzed the relationship between side-chain dihedral-angle preference and main-chain conformation, from a set of 61 protein structures solved at 2.0 Å or better resolution. Their study pertained to all naturally occurring amino acids except Ala, Pro and Gly, and showed that the preferences of each amino acid residue for certain side-chain conformations are significantly affected by secondary structure. In their analysis of β-sheets, the two residues at either end of a β-strand were treated separately from those in the interior; thus, the results reported by McGregor et al. (1987) for only the *interior* of the β-strand were used for comparison with our results because our calculation did not include effect of changes in the N- or C-termini of the strands.

Based on the observation of McGregor et al. (1987) and the chemical shifts shown in the supplementary material, we find that, for only 6 out of these 17 amino acid residues, the leading member of the family possesses a χ^1 dihedral angle in agreement with the most frequently observed value in proteins, and simultaneously shows good agreement, within the standard deviation, with the mean averaged value observed by Wang and Jardetzky (2002) for the $^{13}C^{\alpha}$ chemical shifts. The 11 residues which do not satisfy these conditions are: Val, Leu, Cys, His, Trp, Asn, Gln, Thr, Lys, Arg, and Glu.

I-c. Analysis of the variations between computed and observed chemical shifts

Analysis of the variation (Λ^{α}) between the computed mean values and the observed chemical shifts (shown in Fig. 5) indicates that 9 out of 17 of the naturally occurring amino acids, except for Leu, Met, His, Asn, Gln, Thr, Lys, Arg, and Glu, show either good agreement or only a small departure (seen for Leu, Ile and Ser in Fig 5) from the observed standard deviations. In particular, the disagreements of His, Lys and Arg may be due to the use of uncharged side chains. Disagreements with $\Lambda^{\alpha} > 0.5$ ppm are seen for Met, Asn, Gln and Thr (see Fig. 5). In particular, Met and Thr show the most notable deviations, namely $\Lambda^{\alpha} = 3.4$ and 3.5 ppm, respectively. For each of these residues, we can identify a pair of conformations (see Table S1 of supplementary material) with a variation (Λ^{α}) among them of at least ~2.0 ppm. Those selected pairs of conformations, for either Met or Thr, have the following common characteristic: (a) the difference in χ^1 dihedral angles of this pair is less than 2°; and (b) the difference in the χ^{ξ} (with $\xi \geq 2$) of this pair is greater than 60° (see for example families 1 and 4 for Met, and families 1 and 3 for Thr, in Table S1 of supplementary material). This example illustrates not only the sensitivity, and hence the origin of the disagreements of the computed $^{13}C^{\alpha}$ chemical shifts, to variation in the side-chain torsional angles beyond χ^1, but also the difficulty in obtaining an appropriate ensemble of conformations to represent the experimentally observed values in solution. In other words, most of the disagreements observed in Fig. 5 might be a consequence of the limited number of conformational families used. Conceivably, computation of the $^{13}C^{\alpha}$ chemical shifts for *all* conformations in the generated ensemble, i.e., without carry out a clustering procedure, could lead to a significant improvement in the agreement between computed and observed $^{13}C^{\alpha}$ chemical shifts shown in Fig. 5.

We found errors greater than the standard deviations for the differences (Λ^{β}) between the computed mean values and the observed chemical shifts for all the amino acid residues, except for Ser (see Table S1 supplementary material). All the $^{13}C^{\beta}$ chemical shifts

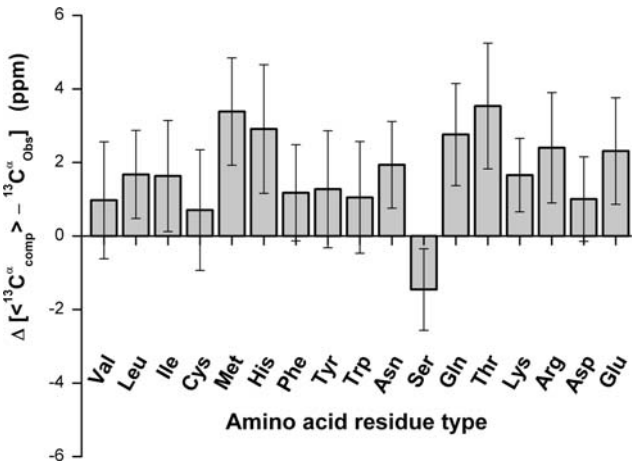

Fig. 5 Bars indicate the difference, $\Delta^{\alpha} = [<^{13}C^{\alpha}_{comp}> - {}^{13}C^{\alpha}_{obs}]$, between the computed *mean* $^{13}C^{\alpha}$ chemical shift value for each of the 17 naturally occurring amino acids in the β-sheet model peptide, as explained in Methods section e and the experimentally observed value by Wang and Jardetzky (2002). Only uncharged residues (for the reasons explained in the Methods section d) were used for this comparison. The observed standard deviations of Wang and Jardetzky (2002) are indicated by the symbol ([])

which had been computed relative to TMS and corrected to a DSS reference in the Supplementary material, show a deviation of ~3 ppm. The following discussion in terms of the correlation coefficient enables us to illustrate the agreement between the calculated and the experimentally observed values without the influence of this scale change.

Figures 6a, b show the correlation between the averaged $^{13}C^{\alpha}$ (and $^{13}C^{\beta}$) chemical shifts for each of the 17 naturally occurring uncharged amino acid residues

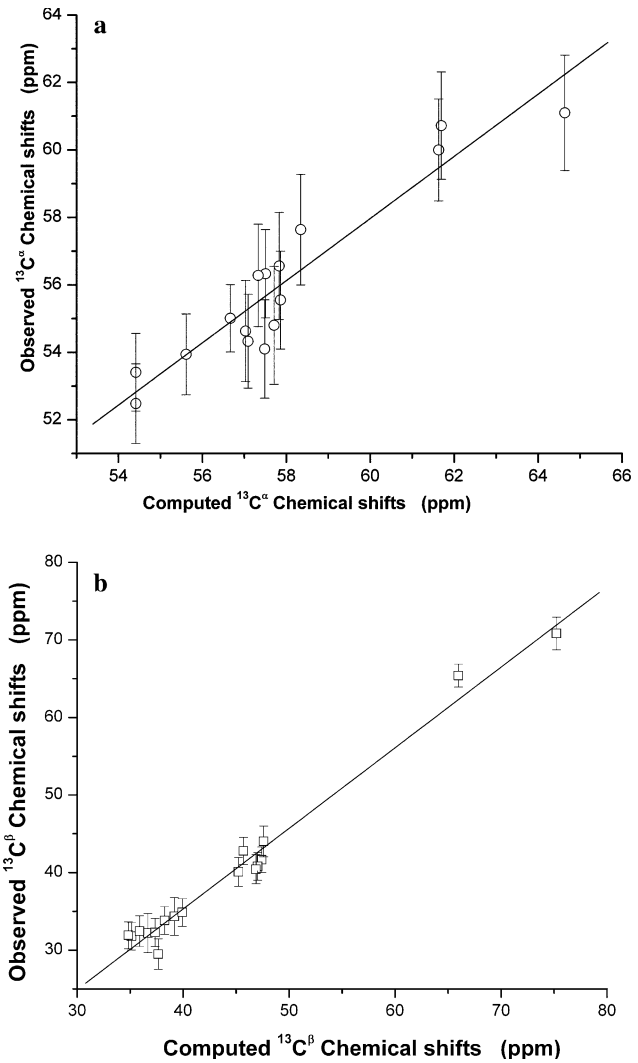

Fig. 6 (**a**) Correlation between the observed $^{13}C^{\alpha}$ chemical shifts of both charged and uncharged residues (Wang and Jardetzky, 2002) and the averaged computed values for each of the 17 naturally occurring uncharged amino acids (except for Gly, Ala and Pro, as explained in Methods section c). $R = 0.95$; slope of 0.92 for the correlation line, after removing Ser from the comparison (as explained in Results and Discussion section I-c). (**b**) Same as Fig. 6a, but for $^{13}C^{\beta}$ chemical shifts, except for the absence of Ala and Pro. $R = 0.99$; slope of 1.06 for the correlation line

and the observed mean-values of Wang and Jardetzky (2002). Similar correlation coefficient results were obtained for both $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts, respectively, viz., $R = 0.91$ (slope of 0.68 for the correlation line) and $R = 0.99$ (slope of 1.06 for the correlation line). However, removing Ser from the computation of the correlation coefficient of $^{13}C^{\alpha}$ improves the agreement significantly, i.e., $R = 0.95$ (slope of 0.92 for the correlation line).

II. Effect of the side-chain χ values on the ^{13}C chemical shifts

Although no systematic search was carried out to identify the effect of the side-chain χ on the ^{13}C chemical shifts, analysis of the χ^1 and χ^{ξ} values (with $\xi \geq 2$) of the leading member reveals some interesting features concerning such influence. Even though the dependences of the computed $^{13}C^{\alpha}$ chemical shifts on the value of the side-chain χ^{ξ} (with $\xi \geq 2$) are smaller than those observed for χ^1, except for Ser, they are significantly large for many side chains such as Leu, Arg, Ile, Thr, His, Phe, Asn, Ser, Gln, Met and Lys. For these 11 residues, changes occurring primarily in the χ^{ξ} (with $\xi \geq 2$) dihedral angle lead to changes in the $^{13}C^{\alpha}$ chemical shifts greater than the standard deviation observed by Wang and Jardetzky (2002) [see Table S1 of the supplementary material].

Concluding remarks

The dependence of the ^{13}C chemical shift on side-chain orientation was investigated here for a two-strand antiparallel β-sheet model peptide represented by the amino acid sequence Ac-$(Ala)_3$-X-$(Ala)_{12}$-NH_2 where X represents any of the 17 naturally occurring amino acids, i.e., not including alanine, glycine and proline. A detailed comparison of the computed ^{13}C chemical shifts with the observed mean-values and standard deviations obtained from an NMR-database by Wang and Jardetzky (2002) for residues in β-strands allows us to conclude that: (i) for 9 of 17 of the naturally occurring amino acid residues (except Ala, Gly and Pro), the computed mean values for the $^{13}C^{\alpha}$ chemical shifts are in agreement with, or show only a small departure from, the observed values of Wang and Jardetzky (2002). However, for 14 out of 17 amino acid residues, there is at least one conformation, among *all* the leading members of the clustered ensemble, showing good agreement with the $^{13}C^{\alpha}$ observed values; (ii) a significant influence, in terms of the observed standard deviations of Wang and Jardetzky (2002), of the χ^1

dihedral angle on the computed values for the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts was found for most of the naturally occurring amino acid residues; (iii) to a less extent, but of considerable importance for accurate chemical shift predictions, we observed that variation of the χ^{ξ} values (with $\xi \geq 2$) induces changes in the $^{13}C^{\alpha}$ chemical shifts greater than the observed standard deviations for 11 out of 17 residues. For these 11 residues, there is at least one pair of conformations for which the χ^1 side-chain torsional angles are similar, i.e., they show differences lower than ~9°, but exhibit a significant difference, namely larger than 60°, in some of the remaining χ^{ξ} torsional values (with $\xi \geq 2$); and the computed $^{13}C^{\alpha}$ chemical shifts difference (Λ^{α}) between these pairs of conformations are larger than the observed standard deviation of Wang and Jardetzky (2002); and (iv) only for 6 out of these 17 amino acid residues, the leading member of the family possessing a χ^1 dihedral angle in agreement with the most frequently observed value in proteins (McGregor et al. 1987) simultaneously shows good agreement for the $^{13}C^{\alpha}$ chemical shifts, within the standard deviation, with the averaged values observed by Wang and Jardetzky (2002). This result indicates that modeling side-chain conformations by using the most frequently observed values in proteins for χ^1 may not be an accurate approximation.

Good agreement between experimental and ab initio ^{13}C theoretical chemical shifts derived from high-resolution X-ray structures, is lost in several cases (Pearson et al. 1997). The reason for such disagreements observed in calculations with the residue valine involves many possible factors: among others, Pearson et al. (1997) conjecture: ''...that the backbone structures of proteins in solution are generally well defined by X-ray ϕ, ψ values from high-resolution structures, but that χ^1 torsions (in valine) are poorly described...'' In addition, after a detailed comparison involving X-ray and NMR-derived protein models of ubiquitin, Xu and Case (2001) suggested that ''...the most representative side-chain conformations may not be presented in any particular PDB entry...'' In line with these observations, in this work we have been able to show that, for *most* of the naturally occurring amino acid residues in β-sheet strands, proper consideration of side-chain preferences, in particular but *not* limited to the χ^1 dihedral angle, may be necessary for good agreement between observed and computed ^{13}C chemical shifts.

Acknowledgments This research was supported by grants from the National Institutes of Health (GM-14312 and TW-6335), and the National Science Foundation (MCB00-03722). Support was also received from the National Research Council of Argentina (CONICET) [PIP-02485] and from the Universidad Nacional de San Luis [UNSL] (P-328402), Argentina. This research was conducted using the resources of two Beowulf-type clusters located at (a) the Instituto de Matemática Aplicada San Luis (CONICET-UNSL) and (b) the Baker Laboratory of Chemistry and Chemical Biology, Cornell University.

References

Burgess AW, Ponnuswamy PK, Scheraga HA (1974) Israel J Chem 12:239–286
Chakrabarti P, Pal D (1998) Protein Eng 11:631–647
Chesnut DB, Moore KD (1989) J Comp Chem 10:648–659
Chothia C, Levitt M, Richardson D (1977) Proc Natl Acad Sci USA 74:4130–4134
Chothia C, Janin J (1981) Proc Natl Acad Sci USA 78:4146–4150
Chou K-C, Scheraga HA (1982) Proc Natl Acad Sci USA 79:7047–7051
Chou K-C, Pottle M, Némethy G, Ueda Y, Scheraga HA (1982) J Mol Biol 162:89–112
Chou PY, Fasman GD (1974) Biochemistry 13:211–222
Creighton TE (1984) Proteins: Structure and Molecular Properties. W.E. Freeman and Company, New York, pp. 186, 223
Derrick JP, Wigley DB (1994) J Mol Biol 243:906–918
Dumbrack RL Jr, Karplus M (1993) J Mol Biol 230:543–574
Dumbrack RL Jr, Karplus M (1994) Nat Struct Biol 1:334–340
Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA Jr, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA (1998) Gaussian 98, Revision A.7, Inc., Pittsburgh, PA
Havlin RH, Le H, Laws DD, de Dios AC, Oldfield E (1997) J Am Chem Soc 119:11951–11958
Hehre WJ, Radom L, Schleyer P, Pople JA (1986) Ab Initio Molecular Orbital Theory. John Wiley and Sons, New York
Iwadate M, Asakura T, Williamson MP (1999) J Biomol NMR 13:199–211
Kruskal JB Jr (1956) Proc American Math Soc 7:48–50
Kuszewski J, Qin JA, Gronenborn AM, Clore GM (1995) J Magn Reson Ser B 106:92–96
Laws DD, Le H, de Dios AC, Havlin RH, Oldfield E (1995) J Am Chem Soc 117:9542–9546
McGregor JM, Islam SA, Sternberg MJE (1987) J Mol Biol 198:295–310
Némethy G, Gibson KD, Palmer KA, Yoon CN, Paterlini G, Zagari A, Rumsey S, Scheraga HA (1992) J Phys Chem 96:6472–6484
Pauling LC, Corey RB (1951) Proc Natl Acad Sci USA 37:729–740
Pauling LC, Corey RB (1953) Proc R Soc London Ser B 141:21–33
Pearson JG, Le H, Sanders LK, Godbout N, Havlin RH, Oldfield E (1997) J Am Chem Soc 119:11941–11950
Ripoll DR, Scheraga HA (1988) Biopolymers 27:1283–1303
Ripoll DR, Vila JA, Scheraga HA (2005) Proc Natl Acad Sci USA 102:7559–7564

Rossmeisl J, Nørskov JK, Jacobsen KW (2004) J Am Chem Soc 126:13140–13143
Santiveri CM, Rico M, Jiménez MA (2001) J Biom NMR 19:331–345
Sibanda BL, Thornton JM (1991) Methods Enzymol 202:59–82
Sosa CP, Ochterski J, Carpenter J, Frisch MJ (1998) J Comp Chem 19:1053–1063
Spera S, Bax A (1991) J Am Chem Soc 113:5490–5492
Ulmer TS, Ramirez BE, Delaglio F, Bax A (2003) J Am ChemSoc 125:9179–9191
Vila JA, Ripoll DR, Baldoni HA, Scheraga HA (2002) J Biomol NMR 24:245–262
Vila JA, Baldoni HA, Ripoll DR, Scheraga HA (2003) J Biomol NMR 26:113–130
Vila JA, Baldoni HA, Scheraga HA (2004a) Protein Sci 13:2939–2948
Vila JA, Baldoni HA, Scheraga HA (2004b) Proteins: Struct Funct Bioinf 57:87–98
Wang Y, Jardetzky O (2002) Protein Sci 11:852–861
Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson JH, Oldfield E, Markley JL, Sykes BD. (1995) J Biomol NMR 6:135–140
Xu X-P, Case DA (2001) J Biomol NMR 21:321–333
Xu X-P, Case DA (2002) Biopolymers 65:408–423
Zhang C, Kim S-H (2000) J Mol Biol 299:1075–1089