Journal of Biomolecular NMR, **26:** 113–130, 2003.
KLUWER/ESCOM

Unblocked statistical-coil tetrapeptides in aqueous solution: Quantum-chemical computation of the carbon-13 NMR chemical shifts

Jorge A. Vila[a,d], Héctor A. Baldoni[b], Daniel R. Ripoll[c] & Harold A. Scheraga[d,*]
[a]*Universidad Nacional de San Luis, Facultad de Ciencias Físico Matemáticas y Naturales, Instituto de Matemática Aplicada San Luis, CONICET, Ejército de Los Andes 950, 5700 San Luis, Argentina;* [b]*Universidad Nacional de San Luis, Departamento de Química, Chacabuco 917, 5700 San Luis, Argentina;* [c]*Cornell Theory Center, Cornell University, Ithaca, NY 14853-3801, U.S.A.;* [d]*Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301, U.S.A.*

Received 22 November 2002; Accepted 21 February 2003

Key words: ab initio (DFT) chemical shifts, clustering, random coils, statistical coils, vicinal coupling constant

Abstract

We recently reported a theoretical characterization of representative ensembles of *statistical-coil* conformations for tetrapeptides with unblocked termini in aqueous solution, at pH 7. The results showed good agreement between the computed Boltzmann-averaged and experimentally-determined values for both the vicinal coupling constants $^3J_{NH\alpha}$ and the α-proton chemical shifts. Here, we carry out a cluster analysis of the ensembles of conformations generated in that study, and use them to compute the Boltzmann-averaged values of the quantum-chemical ^{13}C chemical shifts for different amino acids in the unblocked tetrapeptides GGXA (where X stands for Phe, Arg, His, Glu, Ile, Lys, Gln, Tyr, Leu, Thr, Ala, Gly and Val). The values of the ^{13}C chemical shifts in these thirteen amino acids (for which experimental data are available) were computed by using Density Functional Theory with a 6−311+G(2d,p) basis set. Good agreement is found in terms of both the correlation coefficient (R) and standard deviations of the difference between the computed Bolztmann-averaged and the NMR-determined values for the ^{13}C chemical shifts. These results suggest that it may be possible to build a reliable theoretically-derived database of chemical shifts for *statistical-coil* residues. The results of the current study contribute to our understanding of the relations between chemical shifts, dihedral angles and vicinal coupling constants, $^3J_{NH\alpha}$. In addition, they can shed light as to how the *statistical-coil* conformation is related to the conformational preference of more structured states, such as the α-helical conformation.

Introduction

In recent years, there has been a renewed interest in the relation between characteristic secondary-structure states, such as the α-helix and β-strand, on the one hand, and ^{13}C chemical shifts on the other (Spera and Bax, 1991; Wishart and Sykes, 1994; Havlin *et al.*, 1997; Pearson *et al.*, 1997; Tjandra and Bax, 1997; Cornilescu et al., 1999; Iwadate et al., 1999; Santiveri et al., 2001; Wishart and Case, 2001; Xu and Case, 2001; Sun et al., 2002). However, none of these studies focused on *statistical-coil* peptides, mainly because of the intrinsic difficulties associated with the characterization of unstructured states, i.e., the experimentally-determined (NMR) chemical-shift values for *statistical-coil* peptides are not associated with a unique set of canonical dihedral angles, making a theoretical description of non-structured states difficult to achieve. Studies of the factors that affect the chemical shift are very important because NMR methods that are used to determine secondary structure rely heavily on a comparison with the chemical shifts of the so-called *statistical coil* [frequently, but erroneously, referred to as a *random coil* (Vila et al., 2002)], e.g., through the use of the ^{13}C chemical-shift index (Wishart and Sykes, 1994).

*To whom correspondence should be addressed. E-mail: has5@cornell.edu

^{13}C chemical shifts for the 20 naturally-occurring amino acid residues as models of *statistical-coil* peptides have been determined experimentally by Richarz and Wüthrich (1978) at different pH's, ranging from 4.1 to 12.5, at t = 35 °C. The ^{13}C chemical shifts for thirteen of these amino acids residues were obtained from linear unprotected tetrapeptides GGXA (with X = Phe, Arg, His, Glu, Ile, Tyr, Leu, Thr, Gly, Val, Pro, Met and Trp). Values for the remaining seven amino acids, viz., Lys, Gln, Ala, Ser, Asp, Asn, Cys, were obtained from corresponding terminally-blocked peptides. These two sets of tetrapeptides differ by the addition of protecting groups at (i) the C-terminus, (ii) both termini or (iii) by the addition of one Ala residue at the C-terminus. However, as Richarz and Wüthrich showed, variations of the amino acid sequences of the peptides have, at most, very small effects on the ^{13}C chemical shifts of the amino acid residue in position X.

More recently, Wishart et al. (1995) showed that their experimentally-determined $^{13}C^{\alpha}$ chemical shifts for the terminally-protected linear hexapeptide Gly-Gly-X-Y-Gly-Gly (where Y is Ala or Pro and X represents any of the 20 common amino acids) are in good agreement with previous experimentally-determined values by Richarz and Wüthrich (1978), Spera and Bax (1991) and Thanabal et al. (1994). Best agreement was found with the measurements of Richarz and Wüthrich (1978). Wishart et al. (1995) also found that the $^{13}C^{\beta}$ chemical shift measurements from all four experimental studies are highly correlated (R > 0.99), with a standard deviation of differences among these experimental studies ranging from 0.17 ppm (for the comparison with Richarz and Wüthrich, 1978) to 0.7 ppm (for the comparison with Thanabal et al., 1994). Lately, Schwarzinger et al. (2001) analyzed the NMR data for a series of peptides with the sequence Ac-GGXGG-NH_2 (where X represents any of the 20 common amino acids) to determine the variations in the chemical shifts at position X when four glycine residues flank the central position X. They found that the effect of local sequence variations on $^{13}C^{\alpha}$ is quite small.

It has long been recognized that the $^{13}C^{\alpha}$ and side-chain carbon chemical shifts of a given residue in a *statistical-coil* peptide are not sequence dependent (Howarth and Lilley, 1978), except when an amino acid residue is followed by proline (Iwadate et al., 1999). In agreement with this observation, the experimental evidence mentioned above (Wishart et al., 1995; Schwarzinger et al., 2001) indicates that *statistical-coil* ^{13}C chemical shifts show small, but not zero, variation with changes in sequence. This small sensitivity to the environmental changes and to the nature of the neighboring residues, makes ^{13}C chemical shifts ideally suited for theoretical determination (Schwarzinger et al., 2001). On the other hand, there is experimental evidence suggesting that backbone ^{15}N chemical shifts are influenced by the nature of the preceding amino-acid residue (Braun et al., 1994) while the carbonyl ^{13}C chemical shifts are influenced by the nature of the following residue (Yao et al., 1997).

From a theoretical point of view, the experimental evidence indicating that $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts are not strongly affected by sequence effects is very important information because it would be computationally highly-demanding to reproduce each environmental condition for each different sequence. Based on these facts, we chose the set of values of ^{13}C chemical shifts, reported by Richarz and Wüthrich (1978) as a reference set to which to compare the Boltzmann-averaged values of the ^{13}C chemical shifts determined from quantum-chemical calculations for our ensembles of conformations representing the statistical coil. The statistical-coil ensembles of conformations for thirteen amino acids in the unblocked tetrapeptide GGXA (where X stands for Phe, Arg, His, Glu, Ile, Lys, Gln, Tyr, Leu, Thr, Ala, Gly and Val) were obtained from a conformational analysis at pH 7 following the procedure described by Vila et al., (2002). Bundi and Wüthrich (1979a) also studied these peptides experimentally at pH 7, at t = 35 °C, as a model of a *statistical coil*. These authors reported several parameters of interest such as spin-spin coupling constants, e.g., $^{3}J_{NH\alpha}$, $^{3}J_{H\alpha H\beta}$, as well as ^{1}H-nmr chemical-shift values for fragments such as αNH, αCH, βCH, suggesting that such data are suitable 'statistical-coil' parameters for conformational studies of polypeptides chains.

Even though analysis of molecular conformations of peptides and proteins relies on a comparison with the corresponding spectral parameters for the statistically-coiled polypeptide chain, to our knowledge there has been no theoretical derivation of the Boltzmann-averaged values of ^{13}C chemical shifts for statistically-coiled peptides. The absence of theoretical values of the ^{13}C chemical shifts for non-structured states can be due to (i) the intrinsic difficulties of choosing a good representation of the statistical-coil ensemble of conformations; (ii) lack of an appropriate methodology to treat the complexity inherent in the vast amount of conformations that constitute the sta-

tistical coil ensemble; (iii) difficulties in adopting an appropriate basis set for the quantum-chemical computation of ^{13}C chemical shifts that meets the requirements of both accuracy and reasonable computational time.

In this work, we show that the Boltzmann-averaged values for all ^{13}C chemical shifts, i.e., $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C^{\gamma}$, $^{13}C^{\delta}$, $^{13}C^{\varepsilon}$ and $^{13}C^{\zeta}$, for thirteen amino acids in the sequence GGXA, obtained from quantum chemical calculations of a selected number of conformations derived from a clustering procedure, show close agreement with the experimentally-determined values obtained by Richarz and Wüthrich (1978). The theoretical analysis used to obtain the statistical-coil conformational ensembles is based on the EDMC (Electrostatically Driven Monte Carlo) method (Ripoll and Scheraga, 1988; Ripoll et al., 1996) combined with a fast and reliable MBE (Multigrid Boundary Element) method (Vorobjev et al., 1994, 1995; Vorobjev and Scheraga, 1997).

Conformational analysis

The total number of accepted conformations, computed by Vila et al. (2002) for each amino acid in the statistical-coil tetrapeptides, is shown in Table 1. These values range from 3500 to 5500. The computational time to evaluate the chemical shifts for all the accepted conformations, by an *ab initio* calculation of the ^{13}C chemical shift for each conformation, would be prohibitively large for the required high level of accuracy. Therefore, we adopted a tradeoff between accuracy and computational cost for the calculation of the Boltzmann-averaged values of the ^{13}C chemical shift for each amino acid. Thus, a two-step procedure was followed, namely: (i) Reduction in the number of conformations representing the ensemble of a statistical coil (by using a clustering analysis), and (ii) selection of an appropriate theoretical quantum-chemical approach and a relatively medium-sized basis set (described in Methods).

Clustering analysis

Classification of data into groups (clustering) has been addressed in many contexts and by several approaches. Here, we constructed a minimal tree [by using the Minimal Spanning Tree (MST) method described by Ripoll et al. (1999)], and then partitioned the minimal tree in terms of a specified RMSD cutoff, leading to a given number of families. Fairly accurate Boltzmann averages were computed, as explained in Methods. The degree of approximation to compute a Boltzmann-averaged value depends on the RMSD cutoff used during the clustering procedure. If we were to lower the value of the RMSD cutoff, we would obtain a better approximation to the exact solution but at the cost of increasing both the total number of families and the size of the quantum-chemical computation of the ^{13}C chemical shifts, i.e., the tradeoff between accuracy and computational cost for the calculation of the Boltzmann-averaged values of the ^{13}C chemical shifts for each amino acid residue depends on the RMSD chosen. As a compromise, a cutoff of 1 Å RMSD between all heavy atoms in the tetrapeptide, with no cutoff in the energy, was used during the clustering procedure. Subsequently, an independent test, based on calculations of the Boltzmann-averaged vicinal coupling constant, $<^3J_{NH\alpha}>$, was used to show that the selected RMSD cutoff led to an excellent approximation to the exact solution for $<^3J_{NH\alpha}>$.

The families resulting from the RMSD clustering procedure were ranked in increasing order according to their total free energy, given by Equation (1) in Methods. For each family, we evaluated both the number of conformations belonging to that family and the set of dihedral angles of the lowest-energy conformation. We refer to the lowest-energy conformation of a family as the *leading member*. The Boltzmann averages over all the families for a given tetrapeptide were computed by using only the leading member of each family with the procedure explained in Methods.

As can be seen from Table 1 (column 4), the number of families of conformations for each amino acid ranges from 7 (Isoleucine) to 63 (histidine). The distribution of conformations within each family followed basically four different patterns: (a) Families for which all the members are concentrated around the lowest-energy conformation, e.g., for phenylalanine in Figure 1a; (b) families for which there is a high concentration of those conformations that do not contain the lowest-energy conformation, e.g., for glutamic acid in Figure 1b; (c) families for which we can distinguish two groups that include almost all the conformations, e.g., for leucine in Figure 1c; and (d) families for which the population distribution decreases with increasing total energy of their leading-members, e.g., for histidine in Figure 1d.

To compute averages of any parameter of interest, we considered two possible alternative options:

Table 1. Summary of the EDMC runs

Peptide sequence	Number of energy-minimized conformations[a]	Number of accepted conformations[b]	Number of cluster families[c]	Lowest energy (Kcal/mol)
GGFA	112,688	4513	42 (12)	−236.87
GGRA	112,759	5630	14 (10)	−167.02
GGHA	105,321	4860	63 (21)	−99.83
GGEA	130,905	4784	44 (9)	−171.02
GGIA	112,413	5849	7 (7)	−102.06
GGKA	121,448	3787	47 (12)	−119.83
GGQA	104,278	5401	15 (11)	−133,20
GGYA	135,354	5627	13 (13)	−129,48
GGLA	113,445	5376	14 (13)	−100,67
GGTA	103,391	3972	20 (7)	−128,40
GGAA	120,137	4973	19 (9)	−102,16
GGGA	77,515	4339	17 (9)	−62,99
GGVA[c]	92,132	3667	36 (11)	−84,51

[a]These values correspond to the number of generated conformations (Vila et al., 2002) for the runs using the procedure described in Methods. The calculations were carried out at pH 7 and t = 35 °C, with the solvent free energy and free energy of ionization computed by using the solution of the Poisson–Boltzmann equation as described in Methods (Vila et al., 2002).
[b]According to the Metropolis criterion, and reported by Vila et al. (2002).
[c]Total number of families after a minimal-tree cluster analysis of all accepted conformations, i.e., those listed in column 3 within a cutoff of 1 Å RMSD over all heavy atoms in the tetrapeptide, with no cutoff in energy. The number of clustered families containing at least 95% of the total accepted conformations is given in parentheses.

first, we could use all leading members of all families, and second, we could use the leading member of only the most-populated families. Since the *averages* must provide a good description of the statistical ensemble of conformations, we considered only the most-populated families that included 95% of the total number of accepted conformations. This option is also the most inexpensive for the quantum-chemical calculations. With this procedure, for several cases we found a large reduction of the total number of conformations for which the quantum-chemical ^{13}C chemical shifts had to be computed, e.g., for phenylalanine, histidine, glutamic acid and lysine, the number of *ab initio* calculations was reduced by more than 60%.

To assure that the clustering procedure led to a reasonable representation of the statistical-coil ensemble, we re-evaluated the vicinal coupling constant, $^3J_{NH\alpha}$, by using both the *exact* and the approximate solutions for the Boltzmann-averaged values, as explained in Methods. In Table 2, columns 2 and 3, the corresponding results of this procedure are displayed. The computed Boltzmann-averaged values of the vicinal coupling constant $<^3J_{NH\alpha}>$ using the approximate procedure, i.e., by using the leading members of only those families that represent 95% of the total number of accepted conformations (Table 1, column 4 in parentheses), are identical to those obtained previously (Vila et al., 2002) by using all the accepted conformations (Table 1, column 3).

Methods

Evaluation of the total free energy

The total free energy for the *leading-member* of each family is given by:

$$E(\mathbf{r_p}, pH) = E_{int}(\mathbf{r_p}) + F_{vib}(\mathbf{r_p}) + F_{cav}(\mathbf{r_p}) + F_{solv}(\mathbf{r_p}) + F_{inz}(\mathbf{r_p}, \mathrm{pH}), \quad (1)$$

where $E_{int}(\mathbf{r_p})$ is the internal conformational energy of the molecule in the absence of solvent, assumed to correspond to the ECEPP/3 energy (Némethy et al., 1992) of the neutral molecule; F_{vib} $(\mathbf{r_p})$ is the conformational entropy contribution; $F_{cav}(\mathbf{r_p})$ is the free energy associated with the process of cavity creation when transferring the molecule from the gas phase into the aqueous solution; $F_{solv}(\mathbf{r_p})$ is the free energy associated with the polarization of the aqueous solution, and $F_{inz}(\mathbf{r_p}, \mathrm{pH})$ is the free energy associated with the change in the state of ionization of the ionizable

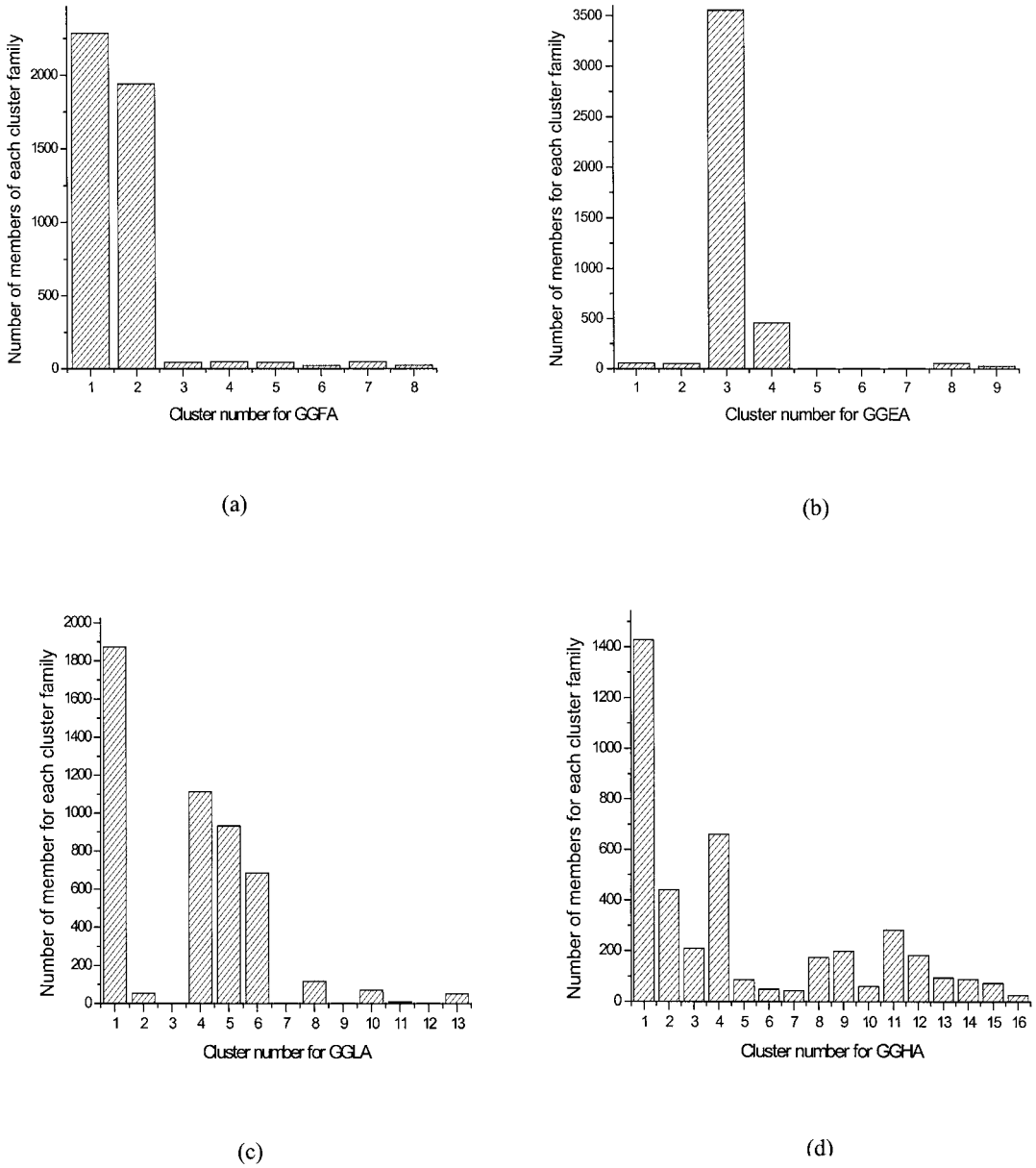


Figure 1. Conformational population distribution versus family number for the unblocked tetrapeptides after clustering all accepted conformations listed in Table 1, column three. Cluster numbers increase with increasing energy of the leading member of each family. Panels (a), (b), (c) and (d) show the data for Gly-Gly-Phe-Ala, Gly-Gly-Glu-Ala, Gly-Gly-Leu-Ala and Gly-Gly-His-Ala, respectively.

groups due to the transfer of the molecule from the gas phase to the solvent, at a fixed pH value.

The contribution to the total free energy from the conformational entropy of the molecule, F_{vib} ($\mathbf{r_p}$), has been approximated by the harmonic vibrational contribution (Gō and Scheraga, 1969; Zimmerman et al., 1977) of each conformation obtained by using the ECEPP/3 potential function. F_{cav} ($\mathbf{r_p}$) describes the free energy of creation of a cavity to accommodate a zero-charge peptide molecule, i.e., with all partial atomic charges set to zero. As shown previously (Sitkoff et al., 1994; Simonson and Brünger, 1994), F_{cav} ($\mathbf{r_p}$) can be considered as the free energy of transfer of a nonpolar molecule from the gas phase to water. This free energy is proportional to the solvent-accessible surface area of the molecule. The term F_{solv}

Table 2. Computed[a] value of the vicinal coupling constant, $^3\mathbf{J}_{NH\alpha}$, for specific amino acids in the tetrapeptides

Sequence	$^3J_{NH\alpha}$ (Hz) Theoretical[b]	$^3J_{NH\alpha}$ (Hz) Theoretical[c]	$^3J_{NH\alpha}$ (Hz) Experimental[d]
GGFA	**9.5**	**9.5**	9.4
GGRA	6.2	6.2	6.9
GGHA	7.0	7.0	8.0
GGEA	**6.6**	**6.6**	7.0
GGIA	**6.8**	**6.8**	7.0
GGKA	**6.8**	**6.8**	6.5
GGQA	**5.8**	**5.8**	6.0
GGYA	**7.3**	**7.3**	6.8
GGLA	7.4	7.4	6.5
GGTA	**6.7**	**6.7**	6.9
GGAA	**6.6**	**6.6**	6.5
GGGA	6.8	6.8	5.6
GGVA	**6.9**	**6.9**	7.0

[a]The theoretical values of the coupling constants were computed from the calculated values of the dihedral angle ϕ by using the Karplus relation (Karplus, 1959, 1963): $^3J_{NH\alpha} = \cos^2 \varphi - B \cos\varphi + C$, with $\varphi = |\phi - 60.0|$ and A = 6.4, B = 1.4 and C = 1.9, as parameterized by Pardi et al., (1984). Calculations were carried out at pH 7 and t = 35 °C, with the solvent free energy and free energy of ionization computed by using the solution of the Poisson–Boltzmann equation as described in Methods. The values within the Bundi and Wüthrich (1979a) experimental errors (± 0.5 Hz) are shown in boldface.
[b]Boltzmann-averaged theoretical values, computed by using the leading member of only those families, given by the minimal-tree cluster analysis procedure for the ensemble of families containing at least 95% of all accepted conformations, as explained in Methods. The values within the Bundi and Wüthrich (1979a) experimental errors (± 0.5 Hz) are shown in boldface.
[c]Boltzmann-averaged theoretical values, computed over *all* accepted conformations. The values within the Bundi and Wüthrich (1979a) experimental errors (± 0.5 Hz) are shown in boldface.
[d]Experimental values from Bundi and Wüthrich (1979a) at pH 7 and t = 35 °C. The experimental error in all cases is ± 0.5 Hz .

($\mathbf{r_p}$) is obtaining by using the fast Multigrid Boundary Element (MBE) method developed by Vorobjev and Scheraga (1997), and F_{inz} ($\mathbf{r_p}$, pH) is calculated by using the general multi-site titration formalism (Bashford and Karplus, 1990; Yang et al., 1993; Vorobjev et al., 1994).

Evaluation of the Boltzmann-averaged value of the vicinal coupling constant $<^3J_{NH\alpha}>$ and $<^{13}C>$ chemical shifts

The exact evaluation of the Boltzmann-averaged value of the vicinal coupling constant $^3J_{NH\alpha}$ for each residue, listed in Table 2 (column 3), follows the procedure described earlier (Vila et al., 2001), i.e.:

$$< {}^3J_{NH\alpha} >_i = \sum_{j=1}^{N} {}^3J_{NH\alpha,i}(j) \exp[-\beta E(\mathbf{r_j}, pH)]/Z \quad (2)$$

with

$$Z = \sum_{j=1}^{N} \exp[-\beta E(\mathbf{r_j}, pH)], \quad (3)$$

where *i* represents any of the thirteen residues listed in Table 1, *j* represents any of the accepted conformations referred to in Table 1 for the corresponding residue *i*, Z is the partition function, $\beta = 1/RT$ with R being the gas constant, T is the temperature in °K, N is the total number of accepted conformations for the peptide *i*, $E(\mathbf{r_j}, pH)$ is the total free energy of conformation *j* as given by Equation 1, and $^3J_{NH\alpha}(j)$ is the vicinal coupling constant for conformation *j* and is computed by using the Karplus relationship (Karplus, 1959, 1963) which connects the size of the spin–spin coupling constant, $^3J_{NH\alpha}(j)$, and the intervening dihedral angle ϕ_j , through the equation:

$$^3J_{NH\alpha}(j) = A \cos^2 \varphi_j - B \cos \varphi_j + C \quad (4)$$

with $\varphi_j = |\phi_j - 60.0°|$ and A = 6.4, B = 1.4 and C = 1.9 as parameterized by Pardi et al. (1984).

A similar process can be used to compute an approximate value of the Boltzmann-averaged vicinal coupling constant $<^3J_{NH\alpha}>_i$ for the clustered conformations (Table 2, column 2) for each of the thirteen residues, *i*:

$$< {}^3J_{NH\alpha} >_i = \sum_{j=1}^{M} \Omega_j \, {}^3J_{NH\alpha,i}(j) \exp[-\beta E(\mathbf{r_j}, pH)]/Z \quad (5)$$

with

$$Z = \sum_{j=1}^{M} \Omega_j \exp[-\beta E(\mathbf{r_j}, pH)] \quad (6)$$

and

$$\sum_{j=1}^{M} \Omega_j = N, \quad (7)$$

where N is the total number of accepted conformations displayed in Table 1 (column 3), *j* represents any of the families referred to in Table 1 for the corresponding residue *i*, Ω_j is the number of conformations of family j, M is the total number of families listed in Table 1 (column 4), $E(\mathbf{r_j}, pH)$ is the total free energy of the

leading-member of family *j* as given by Equation 1, and $^{3}J_{NH\alpha}(j)$ is the vicinal coupling constant and is computed by using Equation 4 with the dihedral angle ϕ_j given by the *leading member* of family *j*. It is important to note that, by allowing the cutoff value of the RMSD to approach zero, $\Omega_j \rightarrow 1$ and $M \rightarrow N$, and, therefore, we will recover the exact solution given by Equation 2.

Calculation of approximate Boltzmann-averaged values of the chemical shifts, $<^{13}C^{P}>_i$ (with P = α, β, γ, δ, ε or ζ) for each of the thirteen residues, *i*, listed in Table 1, can be accomplished by using Equation 5 after replacing $^{3}J_{NH\alpha,i}(j)$ by the corresponding quantum-chemical shift $^{13}C_i(j)$ computed from the *leading-member* of family *j*, as explained in Methods.

As indicated in Figures 1a–d, the cluster number increases with increasing free energy of the leading member of the family. This means that higher-number clusters will not make any significant contribution to the Boltzmann average, as described by Equation 5, because the leading member of such families are much higher in energy than the leading-members of lower-number clustered families. For this reason, we chose a cutoff in the total number of cluster families for each amino acid *i*, M_i, used in both Equations 5 and 6. In particular, we considered only the ensemble of families that represent at least 95% of the total number of accepted conformations. Thus, we had to treat only M_i' leading-member conformations (where $M_i' < M_i$) when computing Boltzmann averaged values of either $<^{3}J_{NH\alpha}>_i$ or $<^{13}C^{P}>_i$ for each amino acid *i*.

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

The chemical shielding tensor (σ) relates the effective magnetic field felt at a probe nucleus to the applied spectrometer field. Methods for *ab initio* calculations of absolute shielding tensors at very accurate levels of approximation are now available, and implementation of such algorithms within *ab initio* and density functional (DFT) methods were reviewed recently (Helgaker et al., 1999).

In the present work, we have computed ^{13}C shielding tensors for all the carbon atoms of the linear unblocked tetrapeptide GGXA (where X stands for the 13 residues in Table 1). We have also computed the corresponding ^{13}C shielding tensors for the side chain of histidine for both deprotonated (pH 7) and protonated (pH 3) states (see footnote a of Table 3). The input geometries were kept frozen as those obtained for the *leading member* of each family resulting from the clustering procedure. The geometries corresponded to those defined in the ECCEP/3 force field (Némethy et al., 1992). No geometry optimization was used because such optimization by *ab-initio* (HF) or DFT methods has only a very small effect on the computed chemical shifts (Pearson et al., 1997; Vila et al., 2002).

The $^{13}C_i^{P}$ shielding tensors (where P refers to α, β, etc.) were obtained by using the method proposed by Cheeseman et al. (1996) as implemented in the Gaussian 98 (Frisch et al., 1998) suite of programs. The effects of electron correlation were addressed by using hybrid density functional theory methods (DFT) (Parr and Yang, 1989). The selected DFT methods employ two different exchange-correlation functionals, (i) Becke's three-parameter functional (Becke, 1993) in combination with (ii) nonlocal correlation provided by the Lee-Yang-Parr expression (Lee et al. 1988; Miehlich et al. 1989), which contains both local and non-local terms. The resulting B3LYP functional makes use of an efficient implementation of the Gauge-Invariant Atomic Orbitals (GIAO) method originally proposed by Ditchfield (1974), and produces very good quality NMR calculations including electron correlation at a low computational cost with medium-size basis sets. Moreover, the B3LYP functional has proven to be a very good choice to predict magnetic shielding tensors for a great variety of compounds containing ^{13}C, ^{15}N, ^{17}O and ^{1}H (Chesnut, 1996; Cheeseman et al., 1996).

A uniform triple split valence basis set 6-311G was used for all the atoms in the molecules. This basis set was supplemented with one extra set of diffuse *s*-functions and two extra sets of polarized *d*-functions on all heavy atoms plus one extra set of polarized *p*-functions on all hydrogen atoms. The resulting 6−311+G(2d,p) basis set predicts chemical shifts which are quantitatively good especially for calculated ^{13}C chemical shifts, as shown by Cheeseman et al. (1996).

Comparisons between theoretical and experimental data are useful to make geometrical and stereochemical assignments, and serve to provide a better knowledge of the relationship between chemical shifts and molecular conformations. To facilitate comparisons between theoretical absolute ^{13}C nuclear shielding values and experimental ^{13}C chemical shift data, all the calculated isotropic shielding values (σ) were converted to a tetramethylsilane (TMS) ^{13}C chemical shift scale (δ) by employing the equa-

Table 3. Boltzmann-averaged[a] values of the ^{13}C chemical shifts for the unblocked tetrapeptide GGXA

Residue name (X)	Chemical shift (ppm)											
	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C^{\gamma}$		$^{13}C^{\delta}$		$^{13}C^{\varepsilon}$		$^{13}C^{\zeta}$	
Phe[b]	54.1	45.5	135.7		132.5	130.1	134.6	134.2	127.6	
	(56.2)	(38.4)	(137.8)		(130.4)		(129.9)		(128.3)	
Arg[c]	{58.2}	{33.1}	{27.1}		{46.2}		{159.1}			
	[60.1]	[35.4]	[28.5]		[47.5]		[154.6]			
	(54.6)	(29.6)	(25.4)		(41.8)		(157.9)			
His[d]	58.7	35.3	145.5		111.5		136.5			
	(54.9)	(29.9)	(133.7)		(119.6)		(137.8)			
His[e]	57.1	32.1	141.7		120.0		136.5			
	(53.6)	(28.1)	(129.5)		(119.0)		(135.2)			
Glu[f]	{58.3}	{33.3}	{42.7}		{175.2}					
	[**55.1**]	[**30.4**]	[**37.1**]		[**182.5**]					
	(55.4)	(28.7)	(34.9)		(182.8)					
Ile[g]	64.1	36.4	29.0	16.1	13.1					
	(59.6)	(37.5)	(25.5)	(15.8)	(11.1)					
Lys[h]	{51.8}	{30.1}	{23.3}		{33.7}		{45.2}			
	[**52.9**]	[**32.0**]	[24.5]		[45.7]		[49.0]			
	(54.6)	(31.9)	(23.2)		(27.6)		(40.4)			
Gln[i]	56.0	31.4	37.8		172.1					
	(54.1)	(28.1)	(32.2)		(179.0)					
Tyr[j]	59.3	46.9	132.7		140.8	138.8	117.5	122.2	155.9	
	(56.3)	(37.5)	(129.5)		(131.8)		(116.7)		(155.5)	
Leu[k]	57.4	44.9	30.3		20.5	24.9				
	(53.8)	(41.2)	(25.7)		(22.1)	(23.6)				
Thr[l]	63.4	71.2	22.3							
	(60.1)	(68.4)	(20.0)							
Ala[m]	49.7	16.2								
	(50.8)	(17.7)								
Gly[n]	47.2									
	(43.9)									
Val[o]	71.1	30.6	20.3	19.5						
	(60.7)	(31.4)	(19.6)	(18.6)						

[a]For each residue in this table, we show both the Boltzmann-averaged ^{13}C chemical shifts computed as described in Methods and, in parentheses, the corresponding experimental value determined by Richarz and Wüthrich (1978). Experimental values for residue X were obtained for the unblocked tetrapeptide GGXA (unless noted). For experimental values of ^{13}C chemical shifts determined at several pH's, we quote only those values that were determined at a pH closest to 7, because the ensemble of conformations for the tetrapeptide GGXA were computed at pH 7. The only exception is for histidine (item *e*) for which we also determined the ensemble of conformations at pH 3. Boltzmann-averaged values for arginine, glutamic acid and lysine where computed by using both charged and uncharged side chains, (the corresponding values being shown in braces and brackets, respectively).

[b]Experimental values were obtained at pH 7.7. There is no distinction between $^{13}C^{\delta 1}$ or $^{13}C^{\delta 2}$ and $^{13}C^{\varepsilon 1}$ or $^{13}C^{\varepsilon 2}$ in the experimental data.

[c]Experimental values were obtained at pH 7.0. The results for the charged and uncharged side chain are shown in braces and brackets, respectively.

[d]Experimental values were obtained at pH 9.5.

[e]Experimental values were obtained at pH 5.0.

[f]Expertimental values were obtained at pH 6.5. The results for the charged and uncharged side chain are shown in braces and brackets, respectively. The value computed with an uncharged side chain, giving better agreement with the experimental data than those obtained with a charged side chain, is denoted in bold-face type.

[g]Experimental values were obtained at pH 4.5.

[h]Experimental values were obtained at 7 in the peptide H-Gly-Gly-Lys-Ala-Ala-OH. The results for the charged and uncharged side chain are shown in braces and brackets, respectively. The value computed with an uncharged side chain, giving better agreement with the experimental data than those obtained with a charged side chain, is denoted in bold-face type.

[i]Experimental values were obtained at pH 7 for the protected peptide CF_3CO-Gly-Gly-Gln-L-Ala-OCH_3.

[j]Experimental values were obtained at pH 6.9. There is no distinction between $^{13}C^{\delta 1}$ or $^{13}C^{\delta 2}$ and $^{13}C^{\varepsilon 1}$ or $^{13}C^{\varepsilon 1}$ in the experimental data.

[k]Experimental values were obtained at pH 4.1.

[l]Experimental values were obtained at pH 5.0.

[m]Experimental values were obtained at pH 7.0 for the protected peptide CF_3CO-Gly-Gly-Ala-L-Ala-OCH_3.

[n]Experimental values were obtained at pH 6.0.

[o]Experimental values determined at pH 4.5.

tion: $\delta_{subst,th} = \sigma_{ref,th} - \sigma_{subst,th}$, where the indices denote a theoretical (th) determination, the substance of interest (subst), and a reference substance (ref). The theoretical value obtained for the ^{13}C shielding of the reference substance, TMS, is $\sigma_{TMS,th} = 182.48$ ppm, and was obtained by using the B3LYP/6-311+G(2d,p)//B3LYP/6-31G(d) procedure. The corresponding experimental value for the ^{13}C shielding of TMS is $\sigma_{TMS,exp.} = 188.1$ ppm (Jameson and Jameson, 1987).

Results and discussion

The computed Boltzmann-averaged values of the vicinal coupling constant, $^3J_{NH\alpha}$, obtained by using only the *leading members* of each family, are identical to the previously-computed Boltzmann-averaged values (Vila et al., 2002) obtained by using all the accepted conformations. The results are shown in Table 2, columns 2 and 3. In addition, as was shown previously (Vila et al., 2002), 9 out of 13 computed Boltzmann-averaged vicinal coupling constant agree, within the experimental error ($\pm$ 0.5 Hz), with the NMR-determined values for the vicinal coupling constant $^3J_{NH\alpha}$ (Bundi and Wüthrich, 1979a).

The good results obtained for the vicinal coupling constant, $^3J_{NH\alpha}$ by computing the Boltzman-averaged values for the leading-member conformations of each family support the approach of using such a restricted set of conformations which provides a good representation of the *statistical-coil* peptide. Hence, the computation of the *ab initio* ^{13}C chemical shift was carried out only for each *leading-member* of the same restricted-number of families (shown in parentheses in Table 1, column 4).

Calculations of chemical shifts always rely on a comparison with a reference state, e.g., TMS. This requires an additional quantum-chemical calculation that introduces another source of error related to the method and basis set used (Baldridge and Siegel, 1999). This caveat should be kept in mind when judging the qualitative comparison between the theoretical- and the experimentally-determined ^{13}C chemical shifts.

Charge assignments for the side chains of the arginine, glutamic acid and lysine residues, and the termini of all the sequences

Based on recent calculations of ^{13}C chemical shifts of residue X in blocked pentapeptides with the sequence Ac-GGXGG-NH_2 (where X is Asp, Arg, Glu and Lys), Xu and Case (2002) discussed the question as to whether it is appropiate to treat the *side chains* as charged or uncharged. Consequently, we also considered this question for Arg, Glu and Lys in our unblocked tetrapeptides. The Boltzmann-averaged ^{13}C chemical shifts for these residues, computed by assuming uncharged or charged side chains, are shown in braces and brackets, respectively, in Table 3, with the experimental values given in parentheses. None of the Boltzmann-averaged chemical-shift values for arginine are improved when the side chain is considered as uncharged. When the lysine side-chain is considered as uncharged, a minor enhancement of the agreement with the experimental values is obtained for $^{13}C^{\alpha}$ and $^{3}C^{\beta}$ but not for the $^{13}C^{\gamma}$, $^{13}C^{\delta}$ and $^{13}C^{\epsilon}$ chemical shifts; a significant improvement in the chemical shift values of $^{13}C^{\alpha}$, $^{13}C^{\beta}$ $^{13}C^{\gamma}$ and $^{13}C^{\delta}$ for glutamic acid is obtained when the side chain is assumed to be uncharged. The significant improvement involving the uncharged glutamic acid side chain could be due, among other possible reasons, to the observation that '*the electron distributions of charged molecules in water resemble those of the corresponding neutral species in the gas phase more closely than they do the gas-phase charged moiety*', as noted by Xu and Case (2002).

Finally, it is important to note that the N- and C-terminal groups were considered uncharged in our calculations of the ^{13}C chemical shifts for the X residue in the unblocked peptides GGXA for the following reasons: (a) In our simulations to obtain the conformational ensemble, several peptide sequences have an average degree of charge at the N- or C-terminal groups that is less than $\pm$ 1.0 (Vila et al., 2002), since it represents an average charge over the 2^N possible charge states of the sequence (where N denotes the number of ionizable groups in the sequence), (b) experiments on unblocked peptides with the sequence GGXGG (with X = Gly, Val, Leu or Ile) show that the $^{13}C^{\alpha}$ chemical shift of the central residue residue X remains unchanged over the pH range of 1.32 to 10.43 (Keim et al., 1973); (c) a comparison of the ^{13}C chemical shifts for the common amino acid residues measured in both blocked and unblocked GGXA peptides shows that the differences between corresponding carbon atoms of X are smaller than 0.3 ppm, as noted by Richarz and Wüthrich (1978), and (d) recent experimental evidence for an unblocked glycyl tripeptide indicates that charges at the termini appear to have only a small effect on the C^{α} chemical shifts

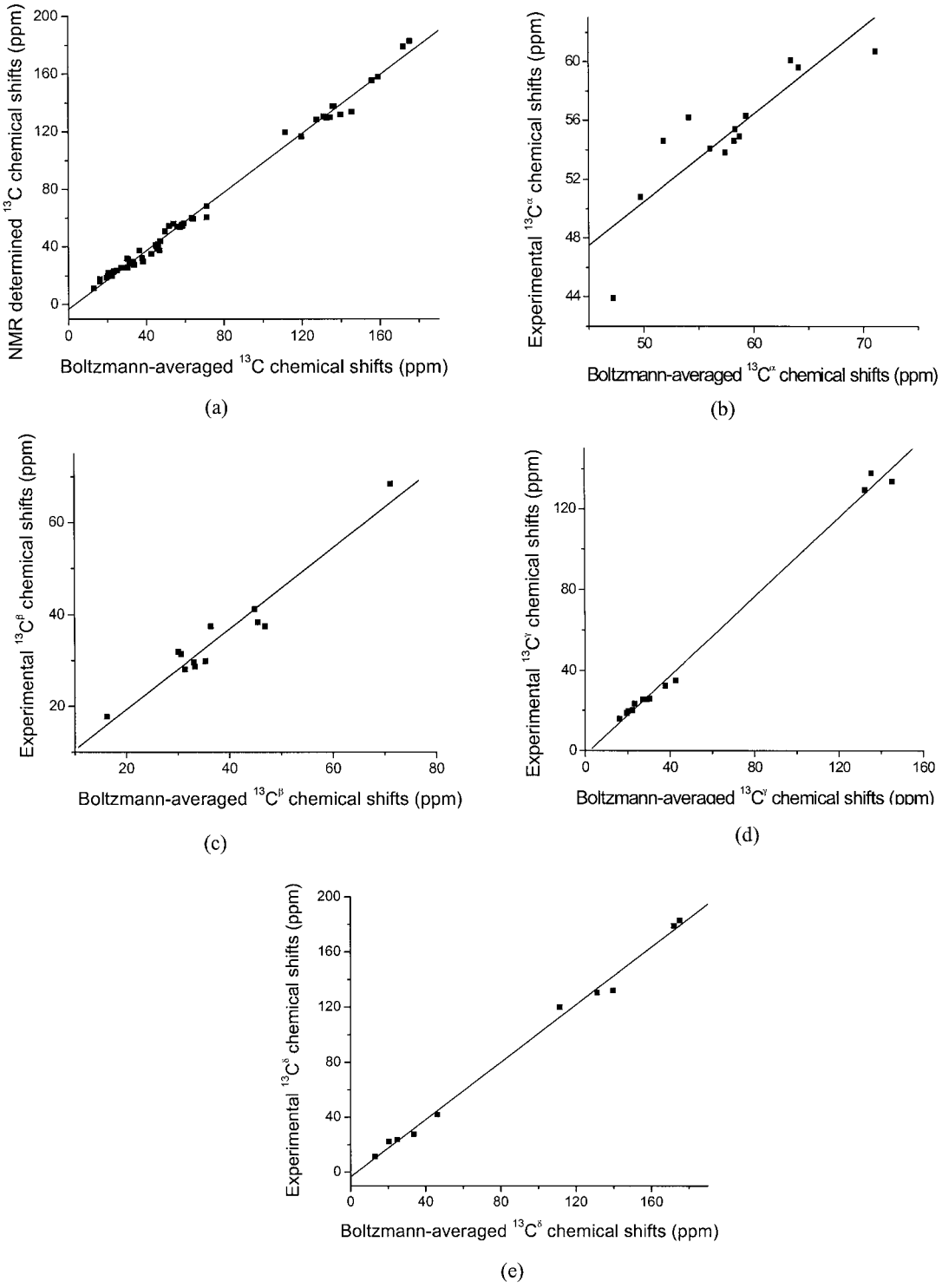


Figure 2. (a) Correlation between the experimentally-determined NMR values of all the ^{13}C chemical shifts of Richarz and Wüthrich (1978) and the corresponding Boltzmann-averaged values determined as explained in Methods and listed in Table 3. R = 0.997; slope of 1.02 for the correlation line; (b) same as (a) for $^{13}C^{\alpha}$ chemical shifts. R = 0.869; slope of 0.60 for the correlation line; (c) same as (a) for the $^{13}C^{\beta}$ chemical shifts. R = 0.965; slope of 0.88 for the correlation line; (d) same as (a) for the $^{13}C^{\gamma}$ chemical shifts. R = 0.997; slope of 0.98 for the correlation line; (e) same as (a) for the $^{13}C^{\delta}$ chemical shifts. R = 0.997; slope of 1.04 for the correlation line.

at adjacent residues (Chekmenev et al., 2002). All the accumulated experimental evidence indicates that the degree of charge at the N- or C- termini is not a source of significant differences for the observed ^{13}C chemical shifts of residue X in the peptide GGXA and, hence, it provides support for our assumption to keep both termini of the peptides uncharged.

General considerations about the Boltzmann-averaged ^{13}C chemical shifts

Figure 2a illustrates the comparison between all the quantum-chemically computed Boltzmann-averaged ^{13}C chemical shifts and the experimentally-determined values by Richarz and Wüthrich (1978). The good correlation coefficient (R = 0.997), with a slope of 1.02 for the correlation line (which should be compared with the ideal value of 1.0), found from this comparison is encouraging, especially after all the considerations that had to be taken into account for the ensemble of statistical-coil peptides in order to make this problem manageable.

It is known that the large dispersion of the ^{13}C chemical shift values (between ∼20 and ∼180 ppm, as shown in Figure 2a) could camouflage disagreements between two sets of data (Wishart et al., 1995). Therefore, caution must be exercised in evaluating the merit of the agreement found between theoretical calculations and experiments, based only on correlation coefficients. In terms of this coefficient, the agreement between the Boltzmann-averaged and the experimentally-determined values for all the ^{13}C chemical shifts, shown in Figure 2a, can be considered as excellent (R = 0.997). In fact, this is comparable to the agreement observed among different experimental studies of the ^{13}C chemical shifts for all 20 naturally occurring amino acids; i.e., Wishart et al. (1995) reported a value of R > 0.995 after comparing their $^{13}C^{\alpha}$ chemical shifts with those obtained by three different experimental studies, viz., those of Richarz and Wüthrich (1978), Spera and Bax (1991) and Thanabal et al. (1994). However, another type of test, viz., the standard deviation of the difference (∼4.0 ppm) between the calculated Boltzmann-averaged and experimentally-determined values of the chemical shifts, displayed in Figure 2a, shows that our agreement is not as good as that found for the corresponding standard deviation of the differences among several experimental studies (∼0.5 ppm) (Wishart et al., 1995). Figure 3 shows significant differences among the standard deviations of the twelve

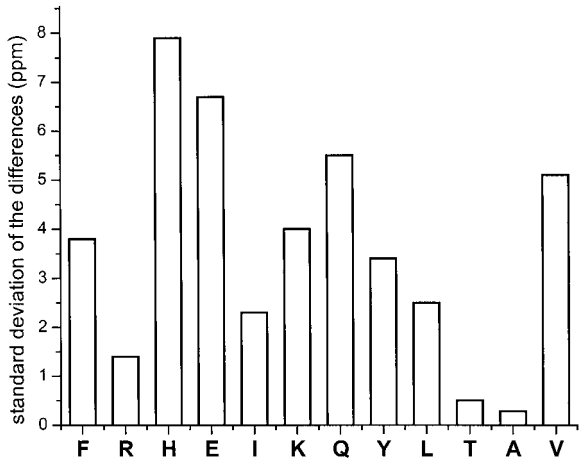


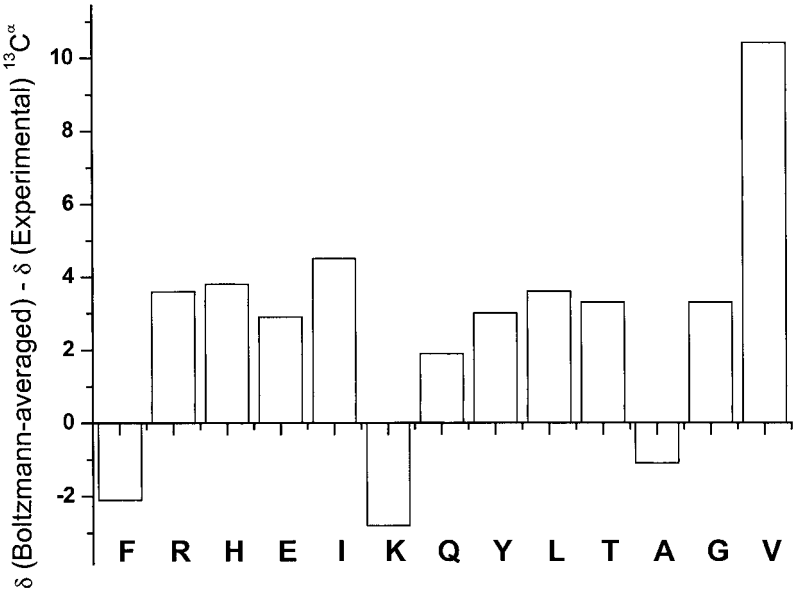
Figure 3. Bar graph showing the standard deviation of the difference between the calculated Boltzamnn-averaged and experimental-determined (Richarz and Wüthrich, 1978) ^{13}C chemical shift values for all the carbon atoms of each of the twelve amino acids (listed in a one-letter code), i.e., after excluding glycine because it contains only a C^{α} atom.

amino acids, i.e., after excluding glycine. It can be seen from Figure 3 that the best agreement in terms of standard deviation is observed for alanine (0.28 ppm) and the worst is for histidine at pH 7 (7.9 ppm).

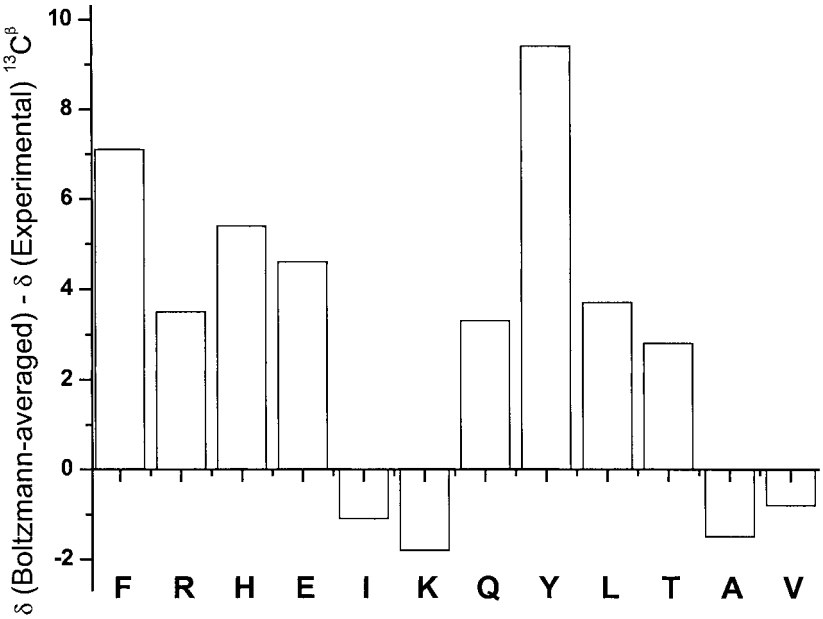
Based on the existing number of experimental assignments for statistical-coil chemical shifts, Wishart and Case (2001) suggested that reasonable deviations from the values of the statistical-coil chemical shifts are ± 5.0 ppm for both $^{13}C^{\alpha}$ and $^{13}C^{\beta}$, respectively. Wishart and Case (2001) also noted that calculated carbon chemical shifts for a wide set of environments are found within ∼3.0–5.0 ppm of experiment in several studies (Cheeseman et al., 1996; Olsson and Cremer, 1996; Wiberg, 1999), with which our calculations are in agreement. In some applications, the standard deviation of the difference between calculated and experimental ^{13}C chemical shifts could be as large as 14.0 ppm, e.g., in methyl bacteriopheophorbide *a* (MBPheo-a), these values range from 4.0 ppm to 14.0 ppm, depending on the optimization of the geometry and the basis set used (Facelli, 1998).

Analysis of the values of the Boltzmann-averaged $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts

e The correlation between theoretical and experimentally-determined $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ values is shown in Figures 2b and 2c, respectively. From these figures, it can be seen that the correlation between the Boltzmann-averaged and the experimentally-

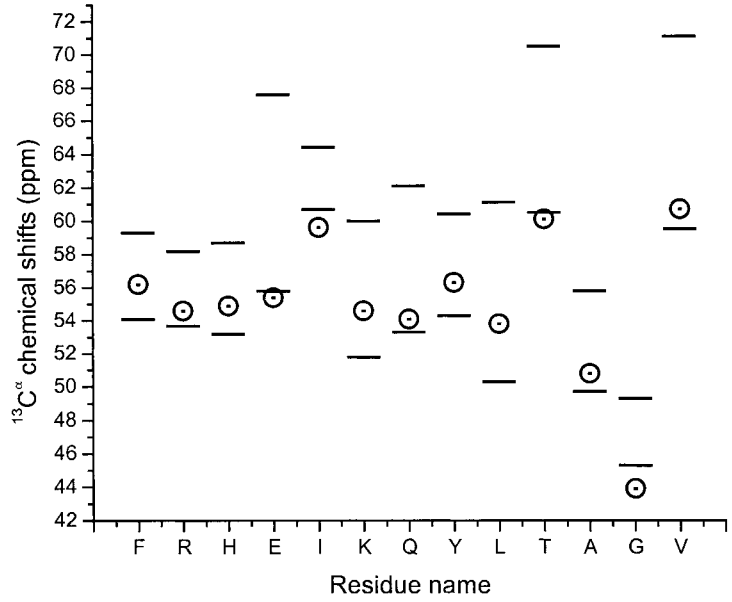


(a)

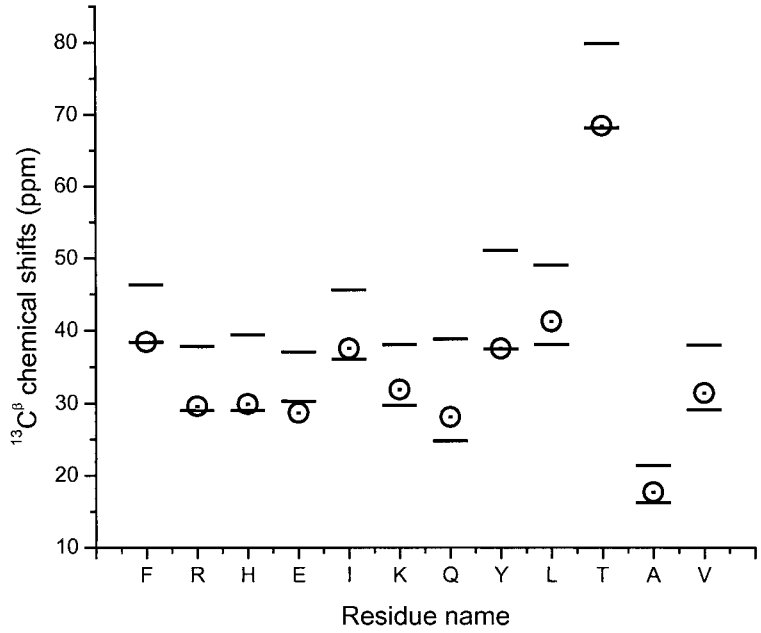


(b)

Figure 4. (a) $^{13}C^{\alpha}$ Chemical shift (δ) difference between experimentally-determined and computed Boltzmann-averaged values for each of the thirteen amino acids (listed in one letter code); (b) same as (a) for the $^{13}C^{\beta}$ Chemical shift (δ) difference for each of the twelve amino acids (listed in a one letter code).



(a)



(b)

Figure 5. (a) Graph showing the maximum and minimum quantum-chemically computed values of the $^{13}C^{\alpha}$ chemical shifts (as horizontal bars) for each of the thirteen amino acids (listed in a one-letter code). The values were computed from the leading member of each family of the corresponding cluster, as explained in Methods. The experimental values (Richarz and Wüthrich, 1978) are also included (as a open circles) for reference; (b) same as (a) for the $^{13}C^{\beta}$ chemical shifts of twelve amino acids (listed in a one-letter code).

determined values is better for the $^{13}C^{\beta}$ chemical shifts (R = 0.965 in Figure 2c) than that for $^{13}C^{\alpha}$ chemical shifts (R = 0.869 in Figure 2b). Figures 4a and 4b present the observed deviations (in ppm) of the computed Boltzmann-averaged chemical shifts (δ) for $^{13}C^{\alpha}$ and $^{13}C^{\beta}$, respectively, from the experimentally determined values.

For ^{13}C atoms of the same type, such us $^{13}C^{\alpha}$ or $^{13}C^{\beta}$, our calculated Boltzmann-averaged chemical shift deviations from the experimentally-determined values (Richarz and Wüthrich, 1978) cover the following range: 10.4 ppm (valine) to −2.8 ppm (lysine) for $^{13}C^{\alpha}$, and 9.4 ppm (tyrosine) to −1.8 ppm (lysine) for $^{13}C^{\beta}$, as shown in Figures 4a and 4b, respectively. It is difficult to provide an explanation for the absolute deviations displayed in Figures 4a and 4b for each ^{13}C chemical shift because these results involve two independent theoretical approaches: (i) The first one is related to the molecular-mechanics calculation that makes use of a mean field potential both to generate the statistical-coil conformational ensemble and to assign the relative free-energies of these conformations (as explained in Methods); and (ii) the second one is related to the quantum-chemical DFT calculations used to compute both the corresponding ^{13}C shielding tensor for each of these conformations as well as the ^{13}C shielding tensor for tetramethylsilane (TMS) that was used to convert isotropic shielding values (σ) of the ^{13}C atoms to the corresponding values of the chemical shifts (δ). Nevertheless our methodology, which involves both theoretical approaches to this problem, seems reasonable, first, because the total free-energy defined by Equation 1 in Methods is a good approach to describe the total free energy of the molecule, even though it could be improved, and second, because it is known that DFT approaches lead to ^{13}C values for which diamagnetic effects are underestimated, and hence, shielding become smaller (Olsson and Cremer, 1996). On the other hand, the observed standard deviations of the differences between the calculated Boltzmann-averaged and the experimentally-determined values of the chemical shifts for the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ data (displayed in Figures 2b and 2c) are ∼3.3 ppm and ∼3.6 ppm, respectively. This range of variation seen for the standard deviations of the differences is within the ∼5.0 ppm chemical-shift separation seen experimentally between helical and sheet residues for both $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ (Spera and Bax, 1991; Oldfield, 2002), indicating that the methodology used is sensitive enough and, hence, it could be used to determine the conformational preferences for residues involved in more structured states, such as α-helix and β-sheet conformations.

Figures 5a and 5b show the highest and lowest values found for the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts, respectively, computed from the *leading-member* conformations of each family. From these figures, it can be seen that there is a broad range of chemical-shift values computed for both $^{13}C^{\alpha}$ and $^{13}C^{\beta}$. In particular, the following minimum and maximum ranges of variation are found: 3.7 ppm (isoleucine) – 11.8 ppm (glutamic acid), and 5.2 ppm (alanine) – 14.0 ppm (glutamine), for $^{13}C^{\alpha}$ and $_{13}C^{\beta}$ chemical shifts, respectively. As an additional illustration, we also included the corresponding experimental NMR-determined values of the chemical shifts obtained by Richarz and Wüthrich (1978) in Figures 5a and 5b.

It is known that $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts are highly sensitive to backbone (ϕ, ψ) dihedral angles and, hence, it is believed that backbone (ϕ, ψ) dihedral angles account for half or more of the observed variation in the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts (Wishart and Case, 2001). Related to these observations, it is worth noting that, as has also been suggested, (a) the $^{13}C^{\alpha}$ chemical shifts are probably comparable to the $^{3}J_{NH\alpha}$ vicinal coupling constants in terms of their ability to identify secondary structure (Luginbühl et al., 1995), and (b) one additional advantage of using chemical-shift-derived dihedral angles instead of vicinal coupling constants is the fact that the former provide data for two backbone dihedral angles (ϕ, ψ) while vicinal coupling constants do so for just one (ϕ) (Wishart and Case, 2001).

In the interest of finding any relationship connecting the experimentally-determined $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts, the (ϕ, ψ dihedral angles, and the vicinal coupling constants $^{3}J_{NH\alpha}$, it is worth noting that 5 residues, out of 9, in Figure 5a, for which we found agreement between theory and experiment in terms of the vicinal coupling constant,$^{3}J_{NH\alpha}$ (displayed in Table 2), exhibit NMR-determined $^{13}C^{\alpha}$ chemical shifts in between the maximum and minimum values computed for the $^{13}C^{\alpha}$ chemical shifts. An equivalent analysis for the $^{13}C^{\beta}$ chemical shifts, from Figure 5b, reveals much better agreement, i.e., 8 residues, out of 9, display such agreement. Despite these variations, our results show clearly that the set of conformations generated by our modified EDMC procedure (Ripoll et al., 1996), as well as the selection of a reduced, but representative set of structures, such as the one obtained by the minimum-tree-spanning clustering method (Ripoll et al., 1999), provides a small,

but high-quality, representation of the statistical-coil ensemble. Such a small set of conformations contains a set of backbone dihedral angles (ϕ, ψ) from which we can reproduce, with acceptable accuracy, both the vicinal coupling constant, $^3J_{NH\alpha}$ and the upper- and lower-limits for the experimentally-observed $^{13}C^{\alpha}$ or $^{13}C^{\beta}$ chemical shifts. In fact, we can see that, for 9 out of 13 (from Figure 5a) and 12 out of 13 (from Figure 5b) cases, the corresponding NMR-determined $^{13}C^{\alpha}$ or $^{13}C^{\beta}$ chemical shifts are within these limits.

The agreement between theory and experiment depends on the scoring function used. For example, using the vicinal coupling constants $^3J_{NH\alpha}$ or the $^{13}C^{\alpha}$ or $^{13}C^{\beta}$ chemical shifts could lead to different results. For these particular scoring functions, the disagreement could be related to the intrinsic restrictions of the vicinal coupling constant, i.e., the vicinal coupling constant is a parameter determined mainly by a single backbone dihedral angle, ϕ, while the corresponding calculations for the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts require the correct assignment of two backbone dihedral angles (ϕ, ψ). The latter observation is particularly relevant for *statistical-coil* peptides because short peptides do not have any chance to develop an ordered structural state, such as an α-helical structure.

All the results obtained in this work are in agreement with the observation that the backbone dihedral angles (ϕ, ψ) account for most of the observed variation in the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts (Wishart and Case, 2001). However, it is plausible, in agreement with the observation of de Dios and Oldfield (1993), that the χ^1 dihedral angles also exert some influence on the $^{13}C^{\beta}$ and $^{13}C^{\alpha}$ chemical shifts. In such a case, their contribution, compared with those resulting from the backbone dihedral angles (ϕ, ψ), should be less important for *statistical-coil* peptides. This statement is based on previous work (Vila et al., 2002), in which we observed that the Boltzmann-averaged vicinal coupling constants, $<^3J_{H\alpha\beta}>$, determined from our statistical-coil ensemble, show a weak correlation with the experimentally-determined values of Bundi and Wüthrich (1979a) ($R \sim 0.63$ and a standard deviation of the difference of $\sim$3.0 Hz); hence, if the dihedral angles χ^1 had greater influence on the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts, this would lead to worse results than those obtained here. On the other hand, the agreement found between the Boltzmann-averaged value for the vicinal coupling constant, $<^3J_{NH\alpha}>$, and the experimentally determined values obtained by Bundi and Wüthrich (1979a) (shown in Table 2) leads to a better correlation ($R \sim 0.77$) and a standard deviation of the difference of $\sim$0.6 Hz, which is notably superior to that obtained for $^3J_{H\alpha H\beta}$ ($\sim$3.0 Hz). The standard deviations (0.6 Hz for $<^3J_{NH\alpha}>$ and 3.0 Hz for $<^3J_{H\alpha H\beta}$) must be interpreted in terms of the experimental error ($\pm$ 0.5 Hz) reported by Bundi and Wüthrich (1979a) for the NMR measurement of the vicinal coupling constants. These results support our statement that the dominant effect on the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts for *statistical-coil* peptides comes mainly from the backbone dihedral angles (ϕ, ψ). This conclusion is in agreement with the work of Spera and Bax (1991), from which we may conjecture that there is a direct effect of the backbone dihedral angles (ϕ, ψ) on electronic structure and hence on shielding, and with recent proposals made by Iwadate et al. (1999) that the influence of the backbone dihedral angles (ϕ, ψ) seems to be the largest single factor controlling $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts.

It is important to note that the above conclusions about the relevance of the backbone dihedral angles (ϕ, ψ) versus the dihedral angle χ^1, in dictating the preference of the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts derived from *statistical-coil* peptides, may not be transferable to the analysis of residues involved in more structured states, such as the α-helix and β-strand, for which the existence of a canonical set of values for the backbone dihedral angles (ϕ, ψ) could make the role of the dihedral angle χ^1 more significant for a correct determination of the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts than in non-structured states.

Analysis of the Boltzmann-averaged $^{13}C^{\gamma}$, $^{13}C^{\delta}$, $^{13}C^{\epsilon}$ and $^{13}C^{\zeta}$ chemical shifts

A linear correlation for the $^{13}C^{\gamma}$ chemical shifts involves an $R = 0.997$ (shown in Figure 2d) and a standard deviation of the difference between the Boltzmann-averaged and experimentally-determined values of Richarz and Wüthrich (1978) of 3.7 ppm. If we exclude His from this comparison, there is a slight improvement of the correlation coefficient ($R = 0.998$) but a more significant lowering of the standard deviation of the difference (2.39 ppm). The computation of the Boltzmann-averaged value of the $^{13}C^{\gamma}$ chemical shifts for both protonated His (charged) at pH 3 (141.7 ppm) or deprotonated His (uncharged) at pH 7 (145.5 ppm), shown in Table 3, does not reveal a significant improvement in the quantitative agreement when compared with the experimentally determined value at pH 5 (129.5 ppm) or at pH 9.5 (133.7 ppm) (Richarz and Wüthrich, 1978). However, from a qual-

itative point of view, it is relevant to show that the experimental values indicate a more up-field value of the chemical shift for the protonated (charged at pH 5) than for the deprotonated (uncharged at pH 9.5) histidine, with which our Boltzmann-averaged values are in agreement. A similar analysis for the $^{13}C^{\delta}$ chemical shift also reveals an R = 0.997 (shown in Figure 2e) but a much higher standard deviation of the differences between the Boltzmann-averaged and the experimentally-determined values of Richarz and Wüthrich (1978) (~5.7 ppm).

We have omitted the corresponding graphical analysis and discussion of the correlation coefficient and standard deviation for the $^{13}C^{\varepsilon}$ and $^{13}C^{\zeta}$ chemical shifts because both the experimental and theoretical data for the thirteen residues are too sparse to provide a reliable analysis.

Concluding remarks

The methodology used in this study involved reasonable computational effort, and several features: (i) Generation of an ensemble of conformations as a representative description of a statistical coil and calculation of the free energies of these conformations with the EDMC and MBE algorithms, (ii) clustering of these conformations into families according to their RMSD, and selection of the lowest-free energy members of each family to reduce the cost of the quantum chemical calculations, (iii) choice of a proper basis set and DFT procedure for quantum chemistry calculations of chemical shifts, i.e., Boltzmann averaging of the computed chemical shifts over the selected conformations, and (iv) comparison with experimental results in terms of correlation coefficients and standard deviations between the computed Boltzmann-averaged and experimentally-determined chemical shifts. Verification of the procedure, including the approximations that were introduced, was provided by the agreement between the $^{3}J_{NH\alpha}$ coupling constants computed by using both the entire ensemble and selected lowest-energy conformations of each family.

By use of the clustering procedure, it was demonstrated that 70% to 90% of the selected reduced ensemble of conformations contained the experimental values of the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts, respectively, as shown in Figures 5a and 5b. These figures exhibited a characteristic pattern, viz., a systematic downfield shift of the quantum-chemically computed ^{13}C chemical shift compared with the corresponding experimentally-determined values of Richarz and Wüthrich (1978). This effect could be due to several possible sources of errors, including: (i) An underestimate of the calculated shielding with pure DFT, or an overestimate if *ab initio* (HF) were used (Bhül et al., 1999), (ii) possible inadequacy of the selected atomic basis set, (iii) DFT treatment of electron correlation effects which can affect the results of NMR calculations significantly for several reasons, including the existence of hydrogen-bond interactions (Malkin et al., 1994), and (iv) error in the calculated value for the ^{13}C shielding of the reference substance, TMS (182.48 ppm), compared to the experimental value of 188.1 ppm; conceivably, this difference could be the main source of error.

It appears that, for statistically-coiled peptides, the main factor determining $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts is the set of backbone dihedral angles (ϕ, ψ) with minor influence from the χ dihedral angles. Thus, for example, for α-helices and β-sheets, database-derived information suggests that there are significant differences between the isotropic chemical shifts of both $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ for those two different conformations (Spera and Bax, 1991; Wishart et al., 1991; Tjandra and Bax, 1997), and that the major factor affecting chemical shifts is the backbone geometry (Iwadate et al., 1999).

Although the conversion of chemical-shift information into quantifiable structural information is a relatively new activity, a promising application of this research would be the use of the information derived from the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts in conjunction with a reliable force field to produce *ab initio* tertiary structure predictions. In other words, the backbone dihedral angles derived from the determined $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts can be used as constraints in a conformational search for the native structure of a protein and, hence, make a significant contribution to the solution of the, as yet unsolved, protein folding problem.

Acknowledgements

This research was supported by grants from the National Institutes of Health (GM-14312 and TW00857), and the National Science Foundation (MCB00-03722). Support was also received from the National Foundation for Cancer Research, the National Research Council of Argentina (CONICET) and Project

No P-328402 of the Universidad Nacional de San Luis, Argentina. Part of this research was conducted using (a) the resources of the Cornell Theory Center, which receives funding from Cornell University, New York State, federal agencies, foundations, and corporate partners and, (b) the National Partnership for Advanced Computational Infrastructure at the Pittsburgh Supercomputing Center which is supported in part by the NSF (MCA99S007P).

References

Baldridge, K.K. and Siegel, J.S. (1999) *J. Phys. Chem*. A, **103**, 4038–4042.

Bashford, D. and Karplus, M. (1990) *Biochemistry*, **29**, 10219–10225.

Becke, A.D. (1993) *J. Chem. Phys*., **98**, 5648–5652.

Bhül, M., Kaupp, M., Malkina, O.L. and Malkin, V.G. (1999) *J. Comp. Chem*., **20**, 91–105.

Braun, D., Wider, G. and Wüthrich, K. (1994) *J. Am. Chem. Soc.,* **116**, 8466–8469.

Bundi, A. and Wüthrich, K. (1979a) *Biopolymers*, **18**, 285–297.

Bundi, A. and Wüthrich, K. (1979b) *Biopolymers*, **18**, 299–311.

Cheeseman, J.R., Trucks, G.W., Keith, T.A. and Frisch, M.J. (1996) *J. Chem. Phys*., **104**, 5497–5509.

Chekmenev, E.Y., Xu, R.Z., Mashuta, M.S. and Wittebort, R.J. (2002) *J. Am. Chem. Soc.*, **124**, 11894–11899.

Chesnut, D.B. (1996) *Rev. Comput. Chem.*, **8**, 245–297.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) *J. Biomol. NMR*, **13**, 289–302.

de Dios, A.C. and Oldfield, E. (1994) *J. Am. Chem. Soc*., **116**, 5307–5314.

Ditchfield, R. (1974) *Mol. Phys*., **27**, 789–807.

Facelli, J.C. (1998) *J. Phys. Chem. B*., **102**, 2111–2116.

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., Dapprich, S., Millam, J.M., Daniels, A.D., Kudin, K. N., Strain, M.C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G.A., Ayala, P.Y., Cui, Q., Morokuma, K., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Cioslowski, J., Ortiz, J. V., Baboul, A.G., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R.L., Fox, D.J., Keith, T., Al-Laham, M.A., Peng, C.Y., Nanayakkara, A., Gonzalez, C., Challacombe, M., Gill, P. M.W., Johnson, B.G., Chen, W., Wong, M.W., Andres, J.L., Gonzalez, C., Head-Gordon, M., Replogle, E.S., and Pople, J.A. (1998) Gaussian 98, Revision A.7, Gaussian Inc., Pittsburgh, PA.

Gō, N. and Scheraga, H.A. (1969) *J. Chem. Phys.*, **51**, 4751–4767.

Havlin, R.H., Le, H., Laws, D.D., deDios, A.C. and Oldfield, E. (1997) *J. Am. Chem. Soc.*, **119**, 11951–11958.

Helgaker, T., Jaszuński, M. and Ruud, K. (1999) *Chem. Rev*., **99**, 293–352.

Howarth, O.W. and Lilley, D.M.J. (1978) *Prog. NMR Spestrosc*., **12**, 1–40.

Iwadate, M., Asakura, T. and Williamson, M.P. (1999) *J. Biomol. NMR*, **13**, 199–211.

Jameson, A.K. and Jameson, C.J. (1987) *Chem. Phys. Lett*., **134**, 461–466.

Karplus, M. (1959) *J. Chem. Phys*., **30**, 11–15.

Karplus, M. (1963) *J. Am. Chem. Soc.*, **85**, 2870–2871.

Keim, P. Vigna, R.A., Marshall, R.C. and Gurd, F.R.N. (1973) *J. Biol. Chem.,* **248**, 6104–6113.

Lee, C, Yang, W. and Parr, R.G. (1988) *Phys. Rev.*, **B37**, 785–789.

Luginbühl, P., Szyperski, T. and Wüthrich, K. (1995) *J. Magn. Reson.*, **B109**, 229–233.

Malkin, V.G., Malkina, O.L., Casida, M.E. and Salahub, D.R. (1994) *J. Am. Chem. Soc.*, **116**, 5898–5908.

Miehlich, B., Savin, A., Stoll, H. and Preuss, H. (1989) *Chem. Phys. Lett.*, **157**, 200–206.

Némethy, G., Gibson, K.D., Palmer, K.A., Yoon C.N., Paterlini G., Zagari, A., Rumsey, S. and Scheraga, H.A. (1992) *J. Phys. Chem.*, **96**, 6472–6484.

Oldfield, E. (2002) *Annu. Rev. Phys. Chem*., **53**, 349–378.

Olsson, L. and Cremer, D. (1996) *J. Chem. Phys.*, **105**, 8995–9006.

Pardi, A., Billeter, M. and Wüthrich, K. (1984) *J. Mol. Biol.*, **180**, 741–751.

Parr, R.G., Yang, W. (1989) *Density Functional Theory of Atoms and Molecules*, Oxford University Press, New York.

Pearson, J.G., Le, H., Sanders, L.K., Godbout, N., Havlin, R.H. and Oldfield, E. (1997) *J. Am. Chem. Soc.*, **119**, 11941–11950.

Richarz, R. and Wüthrich, K. (1978) *Biopolymers,* **17**, 2133–2141.

Ripoll, D.R. and Scheraga, H.A. (1988) *Biopolymers,* **27**, 1283–1303.

Ripoll, D.R., Liwo, A. and Czaplewski, C. (1999) *TASK Quart*., **3**, 313–331.

Ripoll, D.R., Vorobjev, Y.N., Liwo, A., Vila, J.A. and Scheraga H.A. (1996) *J. Mol. Biol.*, **264**, 770–783.

Santiveri, C.M., Rico, M. and Jiménez, M.A. (2001) *J. Biomol. NMR*, **19**, 331–345.

Schwarzinger, S., Kroon, G.J.A., Foss, T.R., Chung, J., Wright, P.E. and Dyson, H.J. (2001) *J. Am. Chem. Soc.*, **123**, 2970–2978.

Schwarzinger, S., Kroon, G.J.A., Foss, T.R., J., Wright, P.E. and Dyson, H.J. (2000) *J. Biomol. NMR*, **18**, 43–48.

Simonson, T. and Brünger, A.T. (1994) *J. Phys. Chem.*, **98**, 4683–4694.

Sitkoff, D., Sharp, K.A. and Honig, B. (1994) *J. Phys. Chem.*, **98**, 1978–1988.

Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc*., **113**, 5490–5492.

Sun, H., Sanders, L.K. and Oldfield, E. (2002) *J. Am. Chem. Soc.*, **124**, 5486–5495.

Thanabal, V., Omecinsky, D.O., Reily, M.D. and Cody, W.L. (1994) *J. Biomol. NMR*, **4**, 47–59.

Tjandra, N. and Bax, A. (1997) *J. Am. Chem. Soc*., **119**, 9576–9577.

Vila, J.A., Ripoll, D.R., Baldoni, H.A. and Scheraga, H.A. (2002) *J. Biomol. NMR*, **24**, 245–262.

Vila, J.A., Ripoll, D.R. and Scheraga, H.A. (2001) *Biopolymers*, **58**, 235–246.

Vorobjev, Y.N. and Scheraga, H.A. (1997) *J. Comput. Chem*., **18**, 569–583.

Vorobjev, Y.N., Scheraga, H.A., Hitz, B. and Honig, B. (1994) *J. Phys. Chem.*, **98**, 10940–10948.

Vorobjev, Y.N., Scheraga, H.A. and Honig, B. (1995) *J. Phys. Chem.*, **99**, 7180–7187.

Wiberg, K.B. (1999) *J. Comput. Chem.*, **20**, 1299–1303.

Wishart, D.S. and Case, D.A. (2001) *Meth. Enzymol*., **338**, 3–34.

Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR*, **4**, 171–180.

Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. and Sykes, B.D. (1995) *J. Biomol. NMR*, **5**, 67–81.

Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.*, **222**, 311–333.
Xu, X.-P. and Case, D.A. (2001) *J. Biomol. NMR*, **21**, 321–333.
Xu, X.-P. and Case, D.A. (2002), *Biopolymers*, **65**, 408–423.
Yang, A-S, Gunner, M.R., Sampogna, R., Sharp, K. and Honig, B. (1993) *Prot. Struct. Funct. Gen.*, **15**, 252–265.
Yao, J., Dyson, H.J. and Wright, P.E. (1997) *FEBS Lett*., **419**, 285–289.
Zimmerman, S.S., Pottle, M.S., Némethy, G. and Scheraga, H.A.(1977) *Macromolecules*, **10**, 1–9.