Journal of Biomolecular NMR, **24:** 245–262, 2002.
KLUWER/ESCOM

Unblocked statistical-coil tetrapeptides and pentapeptides in aqueous solution: A theoretical study

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

Received 5 August 2002; Accepted 30 September 2002

Key words: chemical shifts, conformational search, random coils, solvation free energy, statistical coils, vicinal coupling constant

Abstract

NMR studies of the molecular conformations of peptides and proteins rely on a comparison of the relevant spectral parameters with the corresponding values for so-called *statistical-coil* polypeptides. For this reason, it is necessary to characterize the experimental ensemble of states populated by *statistical-coil* peptides. Such a characterization, however, has proven to be both difficult and sensitive to changes in many environmental parameters such as solvent composition, temperature, pH, as well as the neighboring amino acids in the sequence. As a consequence, a series of significant discrepancies has been reported for some experimentally observed parameters, such as chemical shifts, or vicinal coupling constants, $^3J_{NH\alpha}$, whose values appear to be incompatible with a *statistical-coil* ensemble. In this work, we report the results of a molecular mechanics study of a series of unblocked tetra- and pentapeptides under different pH conditions. These calculations were carried out with explicit consideration of both the coupling between the process of proton binding/release and conformation adopted by the molecule at a given pH and the contribution of the conformational entropy to the total free energy. Good agreement was found between the calculated and experimentally determined values of the vicinal coupling constant, $^3J_{NH\alpha}$, the α-proton chemical shift, and the $^{13}C^{\alpha}$ chemical shift. All the evidence accumulated in these theoretical calculations helps to rationalize some of the unsettled anomalies observed experimentally, and to provide an understanding of the effect of pH and amino acid sequence on the conformational preferences of *statistical-coil* peptides.

Introduction

In the past few years, statistical methods derived from analyses of high-resolution protein crystal structures from the Protein Data Bank have been developed for predicting the values of the vicinal coupling constants, $^3J_{NH\alpha}$ or $^3J_{H\alpha H\beta}$, of amino acids in the *statistical-coil* conformation (Serrano, 1995; Smith et al., 1996; Fiebig et al., 1996; West and Smith, 1998) or for deriving intrinsic (ϕ,ψ) propensities for residues in *coil-regions* (Swindells et al., 1995). These statistical methods are based on the assumption that the effects of specific non-local interactions present in each of the structures included in the database are averaged out. This hypothesis is crucial when attempting to derive the conformational properties of a polypeptide *assumed* to be completely unstructured, i.e., for a peptide in a *statistical-coil state* for which there are no specific non-local interactions between residues.

It is important to remark that a non-structured state, the so-called *statistical-coil* conformation (frequently, but erroneously, referred to as a *random-coil*), corresponds to an energy-weighted ensemble of conforma-

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

tions in which a single residue can occupy any of the regions of the Ramachandran diagram (Ramachandran et al., 1963) with a certain probability specified by the Boltzmann distribution. The probability of occurrence of a particular conformation in the Ramachandran diagram depends on its energy and the temperature of the system, hence the term *statistical coil*. The probabilities of all conformations would be the same (*random coil*) only if their energies were the same, but they are not. Thus, denatured proteins are not random coils. The random coil could exist only at very high temperatures at which the Boltzmann factor would make all conformations in the Ramachandran map, including those involving high-energy steric interactions, equiprobable.

Significant discrepancies (Serrano, 1995) between predicted and experimentally observed $^3J_{NH\alpha}$ vicinal coupling constants for some unstructured peptides have not been addressed by the statistical-based methods. For example, the value of the $^3J_{NH\alpha}$ coupling constant, 9.4 Hz, determined experimentally by Bundi and Wüthrich (1979a) for phenylalanine from a study of the peptide H-Gly-Gly-Phe-Ala-OH at pH 7 appears to be incompatible with that of a *statistical-coil* ensemble when compared with the behavior of the rest of the naturally-occurring amino acids, i.e. with a $^3J_{NH\alpha}$ in the range of 5.6 Hz to 8.0 Hz. In addition, indications of a *non-statistical-coil* structure involving histidine near the terminal COO^- of the peptide H-Gly-Gly-His-Ala-OH have been reported (Jiménez et al., 1986), and some experimental evidence also suggests that the solution conformation of H-Gly-Gly-Glu-Ala-OH includes some species in which the γ carboxyl group of Glu^- could be hydrogen bonded to the backbone amide protons (Bundi and Wüthrich, 1979b), thereby indicating specific departure from a statistical coil conformation.

More recently, O'Connell et al. (1999) made a comparison between two models, one based on a statistical distribution from the structural database and the other based on molecular dynamics simulations. These authors noted that the method based on the structural database should be influenced by both long-range interactions and specific recognizable elements of protein structure. Removal of the regular secondary structures, such as α-helix, β-strands and turns from the database seems to indicate that the calculated (ϕ,ψ) propensities do not change significantly (Swindells et al., 1995). However, as O'Connell et al. (1999) pointed out, some intrinsic difficulties in recognizing *all* the turn and β-strands in the database might indicate that certain *bias* may still be present and will perturb the database-based representation of the statistical-coil state. O'Connell et al. (1999) claimed that their molecular dynamics study of a single-amino-acid model with N-terminal acetyl and C-terminal N′-methylamide blocking groups, should be considered as a better representation of a model system for a *statistical coil* free of long-range interactions. However, as the authors recognized, their theoretical representation for the *statistical-coil* state may not be experimentally realizable. As far as we know, neither treatment of the *statistical coil* described above is able to account for anomalies such as the ones mentioned for phenylalanine, glutamic acid and histidine residues, nor for the pH dependence or the sequence dependence of the experimental results.

A large amount of the experimental work to study *statistical-coil* models that make use of short *unblocked* oligopeptides is based on the assumption that these peptides constitute an unstructured ensemble in solution. Unblocked end groups, on the other hand, are selected to ensure adequate solubility, and are supposed to have no influence on the conformational preference of the residue under investigation (Bundi and Wüthrich, 1979a; Merutka et al., 1995).

We have focused our attention on the short *unblocked* oligopeptides studied by Bundi and Wüthrich (1979a) and Merutka et al. (1995) for the following reasons:

(a) NMR studies of the molecular conformations of peptides and proteins rely on a comparison with the corresponding spectral parameters for the *statistical-coil* polypeptide chain.

(b) Changes in the spectral parameters have been reported to occur with changes in pH, solvent, and amino-acid sequence of the peptide model used to represent the *statistical-coil*, without clear arguments regarding the origin and nature of these changes.

(c) While the properties of the structured state such as the α-helix, β-strand, turn, etc., are well known, those of the unstructured states are not, mainly because of the difficulty in obtaining a good representation of the *statistical-coil* ensemble of conformations.

(d) Wishart and Nip (1998) suggested that changes in the chemical shifts relative to a reference *statistical-coil* state could be particularly helpful in identifying secondary structure, delineating flexible regions, finding hydrogen bonds or detecting aromatic interactions. If this is true, then an analysis of the *statistical-coil* state should contribute to our understanding of how these states are related to the conformational pref-

erences of structured states, such as the α-helical conformation.

Consequently, we have carried out simulations, considering only L-amino acids, in which the equilibrium binding of protons and its dependence on the environmental conditions are considered, for both the unblocked linear tetrapeptides H-Gly-Gly-X-Ala-OH (where X stands for Phe, Glu, His, Ile, Lys, Gln, Arg, Tyr, Leu, Thr, Ala, Val and Gly) studied by Bundi and Wüthrich (1979a) at pH 7 and for the pentapeptide H-Gly-Gly-X-Gly-Gly-OH (where X stands for Phe and His) studied by Merutka et al. (1995) at pH 5. Methionine and tryptophan were omitted from our simulations for the tetrapeptide H-Gly-Gly-X-Ala-OH because Bundi and Wüthrich (1979a) did not report vicinal coupling constants, $^3J_{NH\alpha}$, for these residues. Moreover, we did not carry out calculations on (a) asparagine, aspartic acid and serine because their $^3J_{NH\alpha}$ values were measured on *terminally-protected* peptides, and (b) cysteine because the corresponding experimental value of $^3J_{NH\alpha}$gas obtained for the oxidized (disulfide bonded) state.

These theoretical calculations should enable us to distinguish among the dominant interactions at different pH's to understand the strength and the arrangement of these interactions that influence the conformation of short unblocked oligopeptides and to rationalize some of the anomalies mentioned above, as well as the effect of different sequences on the conformational preference of some amino acids.

The methodology used in this study is based on a Monte Carlo approach, the EDMC (Electrostatically Driven Monte Carlo) method (Ripoll and Scheraga, 1988; Ripoll et al., 1996), combined with a fast and reliable algorithm, the MBE (Multigrid Boundary Element) method (Vorobjev et al., 1994, 1995; Vorobjev and Scheraga, 1997), that provides a solution to the problem of ionization equilibria (Bashford and Karplus, 1990; Yang et al., 1993; Yang and Honig, 1993; Gilson, 1993; Beroza et al., 1995; Vila et al., 1998). The resulting procedure has previously been applied successfully to elucidate the molecular basis of the effect of hydrophobicity on the pK_a of ionizable groups in polypeptides as well as to discuss the nature of the interactions, such as hydrophobicity, charge-charge interactions and solvent polarization effects on the stability of right-handed α-helical conformations (Ripoll et al., 1996; Vila et al., 1998, 2000, 2001).

Methods

Evaluation of the conformational energy

The evaluation of the conformational energy follows the procedure previously published (Ripoll et al., 1996; Vorobjev and Scheraga, 1997; Vila et al., 1998); i.e., the total free energy, $E(\mathbf{r}_p, \text{pH})$, associated with the conformation, $\mathbf{r}_p$, of the molecule in aqueous solution at a given pH, is defined by considering a three-step thermodynamic process (cavity creation, polarization of the solvent, and alteration of the state of proton binding). This free energy involved in transferring the neutral polypeptide from the gas phase to the aqueous solution is given by:

$$\begin{aligned} E(\mathbf{r_p}, \text{pH}) \;=\; & E_{int}(\mathbf{r_p}) + F_{vib}(\mathbf{r_p}) + \\ & F_{cav}(\mathbf{r_p}) + F_{solv}(\mathbf{r_p}) + \\ & F_{inz}(\mathbf{r_p}, \text{pH}), \end{aligned} \tag{1}$$

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 (Momany et al., 1975; Némethy et al., 1983; Sippl et al., 1984; 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} $(\mathbf{r_p}, \text{pH})$ is the free energy associated with the change in the state of ionization of the ionizable 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., all partial atomic charges are 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 non-polar molecule from the gas phase to water. This free energy is proportional to the solvent-accessible surface of the molecule. The term F_{solv} $(\mathbf{r_p})$ is obtaining by using the fast Multigrid Boundary Element (MBE) method developed by Vorobjev and Scheraga (1997), and F_{inz} $(\mathbf{r_p}, \text{pH})$ is calculated by using the general multi-site titration formalism (Bashford and

Karplus, 1990; Yang and Honig, 1993; Vorobjev et al., 1994). The present MBE method provides an accurate and stable calculation of both (a) the potential of mean force between ionized groups of the protein, and (b) the pK shifts of the ionizable groups as a function of the protein environment. The MBE method rapidly and accurately computes the 2^N ionization states, where N is the total number of ionizable groups. Further details of the calculation of the partition function, the pK shift, and the average degree of ionization for each ionizable residue, are provided by Ripoll et al. (1996).

The distinctive features of the present calculations are the following: (i) We take into account the coupling between the process of proton binding/release and the conformation adopted by the molecule at a given pH and (ii) the contribution of the conformational entropy to the total free energy is included.

Generation of the oligopeptides

Polypeptides generally exist as an ensemble of low-energy conformations; consequently, a large part of the computational work was devoted to identifying low-energy structures that constitute a representative set of the conformations in solution at different pH's. The conformations corresponding to the sequence under investigation were constructed by using the ECEPP/3 algorithm (Momany et al., 1975; Némethy et al., 1983, 1992; Sippl et al., 1984). The program considers the complete set of backbone and side-chain dihedral angles as the independent variables while the bond lengths and bond angles of the oligopeptide chain are maintained fixed at their ECEPP/3 values.

The free energy terms, F_{solv} ($\mathbf{r_p}$) and F_{inz} ($\mathbf{r_p}$, pH) associated with electrostatic solvation in Equation 1 are very costly to compute. A full search for the global minimum of the function represented by Equation 1, requiring the energy-minimization of thousands of conformations, is beyond current computational capabilities. For this reason, the following protocol was used to carry out the conformational search to produce a reasonable sampling of the conformational space defined by E($\mathbf{r_p}$, pH) without minimizing this particular function.

The conformational search

At each pH and for each sequence, between 75 000 and 180 000 local minimum conformations were obtained from a series of conformational search runs carried out by using a modified version (Ripoll et al., 1998) of the EDMC method (Ripoll and Scheraga, 1988, 1989; Ripoll et al., 1996, 1998; O'Donnell et al., 1996). After energy-minimization using the Secant Unconstrained Minimization Solver (SUMSL) algorithm (Gay, 1983), in combination with ECEPP/3, their *free energies* were computed by using eq. (1). The objective of these Monte Carlo runs was to sample the low-energy regions of the free energy E($\mathbf{r_p}$, pH). The solvation free energy of the conformations was *always* included in the free energy E($\mathbf{r}_p$, pH). Further details of the procedure can be found in an earlier publication (Ripoll et al., 1996).

Table 1. Summary of the EDMC runs

Peptide sequence	Number of energy-minimized conformations[a]	Number of accepted conformations[b]	Lowest energy (Kcal mol^{-1})
GGFA[c]	112,688	4513	−236.87
GGFGG[d]	179,642	4722	−126.41
GGRA[c]	112,759	5630	−167.02
GGHA[c]	105,321	4860	−99.83
GGHA[e]	81,396	4214	−91.99
GGHGG[d]	139,896	5456	−112.20
GGEA[c]	130,905	4784	−171.02
GGIA[c]	112,413	5849	−102.06
GGKA[c]	121,448	3787	−119.83
GGQA[c]	104,278	5401	−133.20
GGYA[c]	135,354	5627	−129.48
GGLA[c]	113,445	5376	−100.67
GGTA[c]	103,391	3972	−128.40
GGAA[c]	120,137	4973	−102.16
GGGA[c]	77,515	4339	−62.99
GGVA[c]	92,132	3667	−84.51

[a]These values correspond to the number of generated conformations for the runs using the procedure described in Methods.
[b]According to the Metropolis criterion.
[c]Values in this row were computed 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 (Ripoll et al., 1996).
[d]Same as (c), but for pH 5.
[e]Same as (c), but for pH 3.

Evaluation of the Boltzmann-averaged value of the vicinal coupling constants, $\langle {}^3J_{NH\alpha}\rangle$ and $\langle {}^3J_{H\alpha H\beta}\rangle$

The evaluation of the Boltzmann-averaged value of the vicinal coupling constant $^3J_{NH\alpha}$ for each residue follows a published procedure (Vila et al., 2000), i.e.:

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

and

$$Z = \sum_{j=1}^{N} \exp[-\beta E(\mathbf{r_j}, pH)], \quad (2b)$$

where i represents any of the thirteen peptides listed in Table 1, j represents any of the accepted conformations referred to in Table 1 for the corresponding peptide i, Z is the partition function, $\beta = 1/RT$ with R being the gas constant, T is the temperature in K, $E(\mathbf{r_j}, pH)$ is the total free energy of the 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) which connects the size of the spin-spin coupling constant, $^3J_{NH\alpha}$ (j), and the intervening dihedral angle ϕ_j, through the equation:

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

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).

An identical process can be used to compute the vicinal coupling constant $\langle^3J_{H\alpha H\beta}\rangle$ which is related to the torsion angle χ_1 through a Karplus type equation (Karplus, 1959). In this case, the corresponding parameters for Equation 3 are those from De Marco et al. (1978): A = 9.5 ,B = 1.6 and C = 1.8, with φ replaced by χ_1 to compute $^3J_{H\alpha H\beta}{}^2$(j) or by ($\chi_1 - 120$) to compute $^3J_{H\alpha H\beta}{}^3$(j).

Evaluation of the Boltzmann-averaged value of the α-proton chemical shifts

By using the availability of both high resolution X-ray structures of proteins and complete NMR assignments for these structures, Wishart et al. (1991), illustrated the dependence of the α-proton chemical shift on the backbone (ϕ,ψ) dihedral angles. The most important result obtained from this comparison was the absence of any sinusoidal dependence of the chemical shift on the dihedral angle ψ. In its place, the authors found what appears to be a sinusoidal correlation of the dihedral angle ϕ with the chemical shift, and hence, they proposed three empirical functions related to three sinusoidal-dependent phenomena in peptides and proteins shown to be functions of the dihedral angle ϕ; the *first* one is the intranuclear amide proton-to-α-proton distance the *second* one is the dipolar effect arising from the local amide or peptide anisotropy, and the *third* one is the vicinal coupling constant between the amide and α-protons. As the authors noted, the first relationship *does not provide a very good fit* to the experimental data. The second and third relationships reproduce the experimental data with almost similar accuracy; even when it is difficult to say precisely which of the remaining equations is most compatible with the experimental data, the fitting in the positive ϕ region seemed to be slightly worse for the second relationship. For this reason, we adopted the third relationship to evaluate the computed Boltzmann-averaged value of the α-proton chemical shift for each of the thirteen amino acids in the unblocked tetrapeptide GGXA. The correlation with the dihedral angle ϕ, as proposed by Wishart et al. (1991), is obtaining by fitting a Karplus-type equation (Karplus, 1959) to the NMR experimentally-determined values for the chemical shift difference (from the statistical coil), $\Delta\delta_i$:

$$\Delta\delta_i = (\delta_{obs,i} - \delta_{sc,i}) = A\cos^2\varphi_i + B\cos\varphi_i + C, \quad (4)$$

where i indicates the kind of amino acid; $\varphi_i = |\phi_i - 60.0|$; $\delta_{sc,i}$ is the experimentally determined α-proton chemical shift for the *statistical-coil* amino acid of type i, as reported by Wüthrich (1986); $\delta_{obs,i}$ is the observed α-proton chemical shift for the amino acid of type i obtained from four proteins (Wishart et al., 1991), viz., BPTI, thioredoxin, calbindin and ribonuclease A; and A, B and C are the set of parameters to fit the observed data. The result of their procedure led to the following values: A = 1.35, B = 0.25, C = −0.80.

By following a procedure similar to that used for the computation of the vicinal coupling constant (Equation 2a), it is possible to compute the Boltzmann-averaged values of the chemical shift difference $\langle\Delta\delta\rangle_i$ for each of the thirteen amino acids in the unblocked tetrapeptide GGXA by using the expression:

$$\langle\Delta\delta\rangle_i = \sum_{j=1}^{N} \Delta\delta_i(j)\exp[-\beta E(\mathbf{r_j}, pH)]/Z, \quad (5)$$

where $\Delta\delta_i$(j) is computed by using Equation 4; Z represent the partition function given by Equation 2b; N represents the total number of accepted conformations referred to in Table 1 for each amino acid i, and $E(\mathbf{r_j}, pH)$ is the total free energy of conformation j, as given by Equation 1. The values computed by using Equation 5 are shown in column four of Table 3.

The computation of the Boltzmann-averaged value of the α-proton chemical shift $\langle\delta_{theor}\rangle$ for each of the thirteen amino acids in the unblocked tetrapep-

tide GGXA can be computed straightforwardly from Equation 4 as:

$$\langle \delta_{\text{theor}} \rangle_i = \langle \Delta\delta \rangle_i + \delta_{\text{sc,i}}. \quad (6)$$

The corresponding Boltzmann-averaged value of the α-proton chemical shift $\langle \delta_{\text{theor}} \rangle_i$ for each amino acid i, is listed in the second column of Table 3.

It is important to note that the observed empirical correlation of the H^{α} chemical shifts with the dihedral angle ϕ has also been justified in terms of local fields and magnetic susceptibilities (Williamson et al., 1992; Ösapay and Case, 1994).

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

The calculation of the ^{13}C chemical shifts δ requires the calculation of the molecular second-order isotropic nuclear shielding tensor σ. An efficient implementation of such algorithms with *ab initio* and density functional methods was reviewed recently (Helgaker et al., 1999). Such theoretical methods are now widely applied, and it is accepted that the shielding of light nuclei (notably ^{1}H and ^{13}C) can be calculated accurately (Chesnut, 1996; Havlin et al., 1997; Sitkoff and Case, 1997; Xu and Case, 2001; Sun et al., 2002).

A cost-efficient alternative to traditional correlated methods is through density functional theory (DFT), often giving results of quality comparable to that of Møller–Plesset (MP2) for a computational cost of the same order as a Hartree–Fock (HF) calculation (Helgaker et al., 1999). However, numerical evidence indicates that, for accurate calculation of nuclear shielding, a basis set of at least triple split valence quality and with at least one set of polarization functions is needed (Helgaker et al., 1999).

In the present work, we carried out computations of NMR chemical shielding anisotropy tensors with the GIAO (Gauge Invariant Atomic Orbitals) procedure (Wolinski et al., 1990) with both *ab initio* HF (Ditchfield, 1974) and DFT methods as implemented in the *Gaussian 98* suite of computational procedures (Frisch et al., 1998). Density Functional calculations were carried out with the hybrid three parameters B3LYP functional which includes a mixture of the HF exchange and the adiabatic connection methods of Becke (1993) with the correlation functional of Lee et al. (1988). This 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 (Facelli, 1998; Ferraro, 2000; Bagno, 2001).

A triple split valence basis set, such us 6-311G, was used for all the atoms in the molecules. This basis set was added with one extra set of diffuse 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 contains a total of 807 basis functions and 1226 primitive gaussians for the H-Gly-Gly-Phe-Ala-OH peptide.

It was shown by Cheeseman et al. (1996) that the B3LYP/6-311+G(2d,p) level of theory predicts chemical shifts which are quantitatively good especially for ^{13}C. On the other hand, this level of theory includes the effects of electron correlation and is more cost-effective than MP2, representing a reasonable trade-off between accuracy and cost as suggested by Cheeseman et al. (1996).

The calculated isotropic shielding values (σ) of ^{13}C were converted to ^{13}C chemical shifts δ by employing the equation: $\delta_{\text{subst,th}} = \sigma_{\text{ref,th}} - \sigma_{\text{subst,th}}$, 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 tetramethylsilane (TMS) are $\sigma_{\text{TMS,th}} = 182.48$ ppm and 192.58 ppm. These values were obtained by using B3LYP/6-311+G(2d,p)//B3LYP/6-31G(d) and HF/6-311+G(2d,p)//B3LYP/6-31G(d) level of theory, respectively. The corresponding experimental value for the ^{13}C shielding of TMS is $\sigma_{\text{TMS,exp}} = 188.1$ ppm (Jameson and Jameson, 1987).

The theoretical chemical shifts listed in Table 6 were calculated by using the following procedure:
(a) The conformations obtained by the EDMC procedure were not energy minimized at the *ab initio* or at the DFT level of theory; instead, they represent energy-minimized conformations obtained with the geometries defined in the ECCEP/3 force field (Momany et al., 1975; Némethy et al., 1983, 1992; Sippl et al., 1984).
(b) Secondly, we decided to check the hypothesis that *ab initio* geometry optimization produces much better agreement with the experimental values than those not *ab initio* optimized geometries (Facelli, 1998; Ferraro, 2000; Bagno, 2001). To achieve this, we carried out HF/3-21G(d) geometry optimizations in internal coordinates allowing all bond lengths and bond angles to relax but all dihedral angles were kept frozen at the original values.
(c) Once the geometry was obtained, the corresponding isotropic shielding values (σ) for $^{13}C^{\alpha}$ were com-

puted by using the Gaussian98 package (Frisch et al., 1998), and the corresponding chemical shifts were then computed.

Results

Conformational analysis of the oligopeptides H-Gly-Gly-X-Ala-OH in aqueous solution at several pH's

Table 1 shows a summary of a conformational study using the EDMC method for thirteen amino acids in the unblocked tetrapeptide GGXA, i.e., with X= Phe, Arg, His, Glu, Ile, Lys, Gln, Tyr, Leu, Thr, Ala, Gly and Val at pH 7; for GGHA, at pH 3; and for the pentapeptide GGXGG, i.e., with X= Phe or His, at pH 5. In this Table, we list the total number of generated (and energy-minimized) conformations, the total number of accepted conformations, and the lowest-energy conformation found in each search. The number of runs on each oligopeptide, leading to the total number of energy-minimized conformations in column 2 of Table 1, varied from 6 to 11.

Table 2 provides a comparison of the Boltzmann-averaged vicinal coupling constants, $^3J_{NH\alpha}$ computed over all the accepted conformations listed in Table 1. It shows that the theoretically-determined values for $^3J_{NH\alpha}$ agree with the experimental values for the tetrapeptide within experimental error in a reasonable number of cases (9 out of 13). The few exceptions correspond to Arg, His, Leu, and Gly, with the Boltzmann-averaged vicinal coupling constant, $^3J_{NH\alpha}$, for histidine (7 Hz at pH 7) in the tetrapeptide being beyond the range of the reported experimental value (8.0 ± 0.5 Hz at pH 7). As noted in the Introduction, indications of *non-statistical-coil structures* have been reported for oligopeptides involving histidine in the tetrapeptides studied by Bundi and Wüthrich (1979a), and these could be the source of these discrepancies between the observed and computed $^3J_{NH\alpha}$. The particular case of histidine requires more attention, and will be discussed later in the section: *Conformational analysis of the tetrapeptide H-Gly-Gly-His-Ala-OH in aqueous solution*. It is important to note that the differences between the computed Boltzmann-averaged and experimental vicinal coupling constants for glycine of ~ 0.3 Hz in the tetrapeptide, after taking experimental error into account, could be attributed to the small number of data points with positive ϕ angles used by Pardi et al., (1984) during the parameterization of the Karplus (1963) equation.

Table 2. Computed Boltzmann-averaged[a] value of the vicinal coupling constant, $^3J_{NH\alpha}$, for specific amino acids in tetrapeptides and pentapeptides

Sequence	$^3J_{NH\alpha}$ (Hz) Theoretical[b]	$^3J_{NH\alpha}$ (Hz) Experimental[c] ($\pm$0.5 Hz)
GGFA[d]	**9.5**	9.4
GGFGG[e]	6.9	n/a
GGRA[d]	6.2	6.9
GGHA[d]	7.0	8.0
GGHA[f]	7.0	n/a
GGHGG[e]	6.0	n/a
GGEA[d]	**6.6**	7.0
GGIA[d]	**6.8**	7.0
GGKA[d]	**6.8**	6.5
GGQA[d]	**5.8**	6.0
GGYA[d]	**7.3**	6.8
GGLA[d]	7.4	6.5
GGTA[d]	**6.7**	6.9
GGAA[d]	**6.6**	6.5
GGGA[d]	6.4	5.6
GGVA[d]	**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): $^3J_{NH\alpha} = A\cos^2\varphi - B\cos\varphi + C$, with $\varphi = |\phi - 60.0|$ and $A = 6.4$, $B = 1.4$ and $C = 1.9$, as parameterized by Pardi et al. (1984).
[b]Theoretical values within the Bundi and Wüthrich (1979a) experimental errors ($\pm$ 0.5 Hz) are shown in boldface.
[c]Experimental values from Bundi and Wüthrich (1979a) at pH 7 and t = 35 °C. The experimental error in all cases is $\pm$0.5 Hz.
[d]Values in this row were computed 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 (Ripoll et al., 1996).
[e]Same as (d), but for pH 5.
[f]Same as (d), but for pH 3.

The computed Boltzmann-averaged value of the α-proton chemical shift ($\langle\delta_{theor}\rangle$) for each of the 13 amino acids is shown in Table 3 together with the experimentally-determined value (δ_{sc}) obtained from NMR measurements by Wüthrich (1986). From Table 3, it can be seen that the Boltzmann-averaged α-proton chemical shifts, computed as explained in Methods, are in good agreement with the experimental values, i.e., with an average estimated difference of $\sim$0.3 ppm. For nine residues out of thirteen, the computed Boltzmann-averaged values for the α-proton chemical shifts show variations with respect to

Table 3. Computed Boltzmann-averaged[a] α-proton chemical shifts (δ_{theor}) for the residue X in the tetrapeptides H-Gly-Gly-X-Ala-OH

GGXA X=	$\langle\delta_{theor}\rangle$ (ppm) Theoretical[b]	δ_{sc} (ppm) Experimental[c]	$\lvert\Delta\delta\rvert = \lvert\langle\delta_{theor}\rangle - \delta_{sc}\rvert$ (ppm)[d]
Phe[b]	4.92	4.66	**0.26** (0.50/0.42)
Arg	4.10	4.38	0.28 (0.39/0.27)
His[b]	4.48	4.63	**0.15** (0.43/0.34)
Glu	4.07	4.29	**0.22** (0.29/0.56)
Ile[e]	4.83	4.05	0.78 (0.56/0.38)
Lys	4.17	4.36	**0.19** (0.41/0.36)
Gln	4.03	4.37	**0.34** (0.35/0.51)
Tyr	4.50	4.60	**0.10** (0.45/0.43)
Leu[e]	4.11	4.20	**0.09** (0.40/0.31)
Thr	4.14	4.35	**0.21** (0.33/0.50)
Ala	4.14	4.35	**0.21** (0.30/0.32)
Gly	4.61	3.97	0.64 (0.09/0.50)
Val[e]	4.79	4.00	0.79 (0.60/0.38)

[a]The theoretical values of the α-proton chemical shifts were computed from the calculated values of the dihedral angles ϕ by using a Karplus type relation as given by Wishart et al. (1991): $\Delta\delta = A\cos^2\varphi - B\cos\varphi + C$. With A = 1.35, B = 0.25, C = −0.80 $\varphi = |\phi - 60.0|$. δ_{sc} and $\langle\delta_{theor}\rangle$ are the experimentally-measured values of the α-proton chemical shifts for *statistical coil* peptides by NMR, as given by Wüthrich (1986) and the theoretical Boltzmann-averaged values for the α-proton chemical shifts, respectively. The latter were computed as described in Methods.

[b]Values in this row were obtained from the Boltzmann-averaged value of $\langle\Delta\delta\rangle$ by using all the accepted conformation computed at pH 7, as described in Methods.

[c]The experimental values are those reported by Bundi and Wüthrich (1979a) except that the more precise data (Wüthrich, 1986) obtained for Phe, Arg, Lys and Leu are shown.

[d]Theoretical values of $|\Delta\delta|$ that are lower than the absolute values of the observed difference between the average value of the experimental α-proton chemical shift in α-helix or β-strand and the experimental *statistical-coil* value, i.e., $|\delta_{sc} - \langle\delta_\alpha\rangle|$ and $|\delta_{sc} - \langle\delta_\beta\rangle|$, where $\langle\delta_\alpha\rangle$ and $\langle\delta_\beta\rangle$ for each residue were taken from Table 6 of Wishart et al., (1991) are shown in bolface type. The values of $|\delta_{sc} - \langle\delta_\alpha\rangle|$ and $|\delta_{sc} - \langle\delta_\beta\rangle|$ for each residue are shown in parentheses.

[e]The corresponding experimental values reported by Wüthrich (1986) were reduced by 0.18 ppm to account for a previously observed conformational bias (Wishart et al., 1991).

the experimentally-determined values that are *smaller than* the differences between both the experimental α-proton chemical shift and the average α-proton chemical shifts for (a) residues in an α-helix, and (b) for residues in β-strands, respectively. This is denoted in parenthesis in Table 3.

Wishart et al. (1991) indicated that α-proton chemical shifts determined by using an empirical Karplus-type relationship do not provide the same accurate geometrical details as those given by the coupling constants. Regardless of this observation, the close agreement found between the theoretical Boltzmann-averaged and NMR experimentally-determined values for the α-proton chemical shifts, shown in Table 3, seems to indicate that the relationship of Wishart et al. (1991) (Equation 4) is able to distinguish *statistical-coil* α-proton chemical shifts from those that belong to structured states, such as α-helix and β-strand. As noted by Serrano (1995), the dependence of the dihedral angle ϕ on the secondary structure preference, as determined by both the α-proton chemical shifts and the vicinal coupling constant, $^3J_{NH\alpha}$, could provide a basis for recognizing the difference in conformational propensities of the amino acids for unstructured states, i.e., the *statistical-coil.*

Table 4 provides a comparison between the theoretical Boltzmann-averaged values for the vicinal coupling constant, $^3J_{H\alpha H\beta}$, and the experimentally-determined values from NMR by Bundi and Wüthrich (1979a). The values of $^3J_{H\alpha H\beta}$ reflect the Boltzmann-averaged population of the side-chain dihedral angle χ^1. Since NMR resonances of residues with two β protons are not stereo-specifically assigned, it can be assumed (following West and Smith, 1998) that the largest $^3J_{H\alpha H\beta}$ coupling constant belongs to $H^{\beta 2}$; this situation occurs when the population of the least sterically restricted $\chi^1 = -60°$ rotamer is greater than that of the $\chi^1 = -180°$ rotamer. Analysis of the vicinal coupling constants from the experimental data leads to the relative populations of the three staggered χ^1 rotamers; if the vicinal coupling constants $^3J_{H\alpha H\beta^2}$, $^3J_{H\alpha H\beta^3}$ are different, this indicates that the side chain is not populating the three staggered χ^1 rotamers equally.

Table 4 shows that the Boltzmann-averaged values of $^3J_{H\alpha H\beta}$ for Phe in both the tetrapetide studied by Bundi and Wüthrich (1979a) at pH 7 and the pentapeptide studied by Merutka et al. (1995) at pH 5, follow the trend of the experimentally-determined values, i.e., $^3J_{H\alpha H\beta^2} \gg {}^3J_{H\alpha H\beta^3}$ in GGFA, and $^3J_{H\alpha H\beta^2} \gg {}^3J_{H\alpha H\beta^3}$ in GGFGG for both calculated and experimental values. The agreement between experiments and the Boltzmann-averaged populations of the χ^1 rotamers for residues such as phenylalanine and glutamic acid is consistent with a preference of the side chains to be folded against the backbone, making other rotamer states unreachable, as shown below.

The computed Boltzmann-averaged values of $^3J_{H\alpha H\beta}$ for leucine (12.8 Hz and 4.2 Hz) differ from the experimental values (7.2 Hz and 7.2 Hz) determined

Table 4. Computed Boltzmann-averaged[a] vicinal coupling constant $^3J_{H\alpha H\beta}$ for residue X in the peptides H-Gly-Gly-X-Ala-OH and H-Gly-Gly-X-Gly-Gly-OH

Sequence	Theoretical[b] (Hz)		Experimental (± 0.5 Hz)	
	$^3J_{H\alpha H\beta}{}^2$	$^3J_{H\alpha H\beta}{}^3$	$^3J_{H\alpha H\beta}{}^2$	$^3J_{H\alpha H\beta}{}^3$
GGFA[c]	12.7	4.2	10.3[d]	5.6[d]
GGFGG[e]	6.5	5.0	7.7[f]	6.6[f]
GGRA[c]	12.7	2.8	7.6[d]	5.5[d]
GGHA[c]	12.6	4.6	6.9[d]	6.0[d]
GGHGG[e]	4.0	2.8	n/a	n/a
GGEA[c]	12.8	2.8	9.5[d]	4.6[d]
GGIA[c]	–	5.4	–	7.6[d]
GGKA[c]	10.2	2.8	7.8[d]	5.6[d]
GGQA[c]	5.2	2.1	8.8[d]	5.0[d]
GGYA[c]	12.5	4.9	9.0[d]	5.6[d]
GGLA[c]	12.8	4.2	7.2[d]	7.2[d]
GGTA[c]	3.7	–	5.0[d]	–
GGAA[c]	–	12.9	–	7.0[d]
GGVA[c]	–	12.9	–	6.9[d]

[a]The theoretical values of the couplings constants were computed from the calculated values of the dihedral angle ϕ by using the Karplus relation (Karplus, 1959, 1963): $^3J_{H\alpha H\beta} = A\cos^2\varphi - B\cos\varphi + C$, with φ replaced by χ_1 for $^3J_{H\alpha H\beta}{}^2$ or by $(\chi_1 - 120$ for $^3J_{H\alpha H\beta}{}^3$.
[b]These values were computed with the following set of parameters: A = 9.5, B = 1.6 and C = 1.8 as given by De Marco et al. (1978).
[c]Values in this row were computed 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 (Ripoll et al., 1996).
[d]Experimental values from Bundi and Wüthrich (1979a) at pH 7 and t = 35 °C.
[e]Same as (c), but for pH 5.
[f]Experimental values from Merutka et al. (1995) at pH 5 and t = 35 °C.

Table 5. Average[a] degree of charge for the unblocked peptides GGXA and GGXGG

Sequence	Fraction in the NH_3^+ form	Fraction of the charged form of Residue X	Fraction in the COO^- form
GGFA[b]	0.00	n/a	1.00
GGFGG[c]	0.00	n/a	0.36
GGRA[b]	1.00	1.00	0.95
GGHA[b]	0.28	0.10	0.99
GGHA[d]	0.99	1.00	0.83
GGHGG[c]	0.89	1.00	1.00
GGEA[b]	0.02	0.99	1.00
GGIA[b]	0.01	n/a	1.00
GGKA[b]	0.00	1.00	1.00
GGQA[b]	0.00	n/a	1.00
GGYA[b]	0.12	0.00	1.00
GGLA[b]	0.04	n/a	1.00
GGTA[b]	0.01	n/a	0.98
GGAA[b]	0.01	n/a	1.00
GGGA[b]	0.00	n/a	1.00
GGVA[b]	0.01	n/a	1.00

[a]Values of 7.80 and 3.75 were used for the pK_a^o of the ionizable N- and C-terminal α-amino and α-carboxyl groups, respectively (Edsall and Wyman, 1958). The N- and C- terminal groups are indicated as NH_3^+ and COO^-, respectively. The values of 12.50, 6.0, 4.30, 10.50, and 10.10 were adopted as the pK_a^o for the ionizable groups of the residues Arg, His, Glu, Lys, and Tyr, respectively, as an average from the data of Perrin (1972).
[b]Runs in this row were computed at pH 7, with the solvent free energy and free energy of ionization computed by using the solution of the Poisson–Boltzmann equation as described in Methods (Ripoll et al., 1996).
[c]Same as (b), but for pH 5.
[d]Same as (b), but for pH 3.

by Bundi and Wüthrich (1979a). The experimental data of Bundi and Wüthrich (1979a) indicate that the side chain of this residue populates the three staggered χ^1 rotamers equally. However, recent experimental results on side-chain conformations in unfolded hen egg-white lysozyme (Hennig et al., 1999) revealed that leucine displays a strong preference for torsion angles that are clustered around the $\chi^1 = -60°$ rotamer (79% population). The agreement between our computed values for $^3J_{H\alpha H\beta}$ and the experimentally observed χ^1 preference from denatured lysozyme provides an independent test, and shows that the side chain of leucine can preferentially populate one the three staggered rotamers.

In those cases for which there are stereospecifically assigned hydrogens, two cases can be distinguished (see Table 4): (a) Amino acids Phe, Arg, Glu, Gln and Tyr for which the *relative* theoretical values *follow* the trend of the experimental data, and (b) amino acids His, Lys and Leu for which the theoretical values *do not follow* the trend of the experimental data (the results in point (b) may be a consequence of poor sampling of the side-chain conformations of these residues or improper force field parameters).

Conformational analysis of the tetrapeptide H-Gly-Gly-Glu-Ala-OH in aqueous solution

The simulations show that the Boltzmann-averaged values for $^3J_{NH\alpha}$ of GGEA are in good agreement with the experiments, as seen from Table 2.

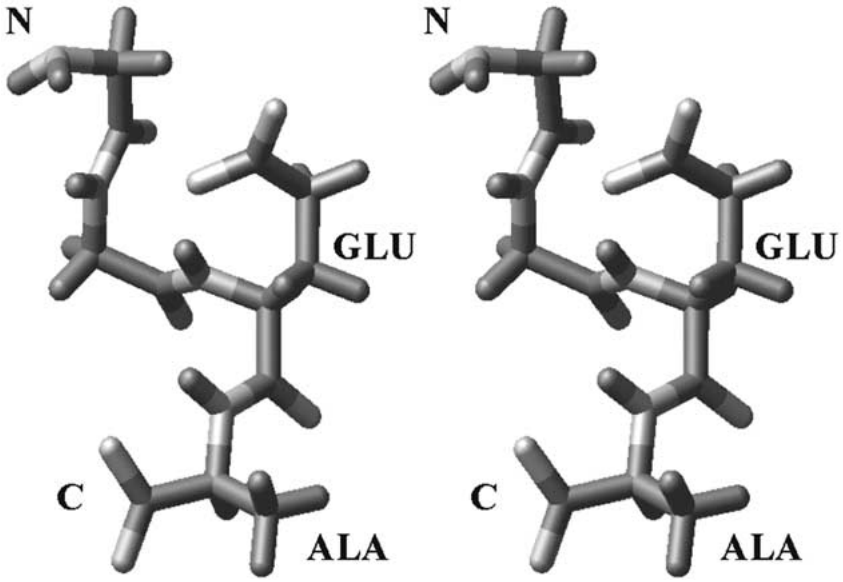

Figure 1. Stereo view of one conformation (the lowest-energy one from family 3) of our *statistical-coil* ensemble found during the simulation carried out for the tetrapeptide H-Gly-Gly-Glu-Ala-OH, at pH 7. This conformation is the most populated in the ensemble, and shows that one of the oxygens of the carboxylate side chain of Glu is in close proximity to the amide proton at positions 2 and 3.

Statistical analysis based on clustering of all the accepted conformations for GGEA with a cutoff RMSD of 1 Å, between all heavy atoms in the tetrapeptide, leads to the partition of the *statistical-coil* ensemble in 44 different families. The lowest-energy conformation for each family is designated as the *leading member.* Moreover, the families are ranked in increasing order of total free energy of their leading member, e.g., the leading member of family 1 belongs to the lowest- energy conformation found for the tetrapeptide GGHA (listed in Table 1). Among all the families of GGEA, the most populated one belongs to the family number 3 in the ranking, and contains $\sim 70\%$ of the total number of conformations in the ensemble. The lowest-energy conformation of family 3 is displayed in Figure 1 (but it is 17.2 Kcal mol^{-1} above the global minimum of Figure 2). The leading member of family 3 displayed in Figure 1 shows that one of the oxygens of the carboxylate side chain of Glu is in close proximity to the amide protons at position 2 and 3. According to Bundi and Wüthrich (1979b), the ensemble of the deprotonated (charged) form of Glu at neutral pH includes species with hydrogen bonds between the Glu side chain and the backbone amide proton of residues at positions 2 or 3, respectively, with which our conformation in Figure 1 is in close agreement. However, it is not clear how such a (computed) high-energy form could have been observed in the experiments of Bundi and Wüthrich(1979b). Besides, the vicinal coupling constant $^3J_{NH\alpha}$ of the conformation shown in Figure 1, is 5.5 Hz, and this value is far from both the experimental value (7.0 Hz) determined by Bundi and Wüthrich (1979a) at pH 7 and from the Boltzmann-averaged computed value of 6.6 Hz, listed in Table 2. Conceivably, this discrepancy may arise from a consideration of only a limited number of conformations instead of the whole ensemble.

The lowest-energy conformation found in our simulations and displayed in Figure 2, on the other hand, shows that the glutamic acid side chain is oriented toward the N-terminus but is not close to the amide protons at positions 2 and 3, the conformation that was seen in Figure 1. Interestingly, this predominant orientation of the glutamic acid side chain in the ensemble is in agreement with predictions made by Bundi and Wüthrich (1979b) that bulky side chains are oriented toward the N-terminus.

Conformational analysis of the tetrapeptide H-Gly-Gly His-Ala-OH in aqueous solution

The above suggestion, that bulky side chains are preferently oriented toward the N-terminus in general, was not found in our simulation for the lowest-energy conformation of histidine where the calculated side-chain position is oriented toward the C-terminus at pH 7 (Figures 3 and 4), in agreement with the experimental

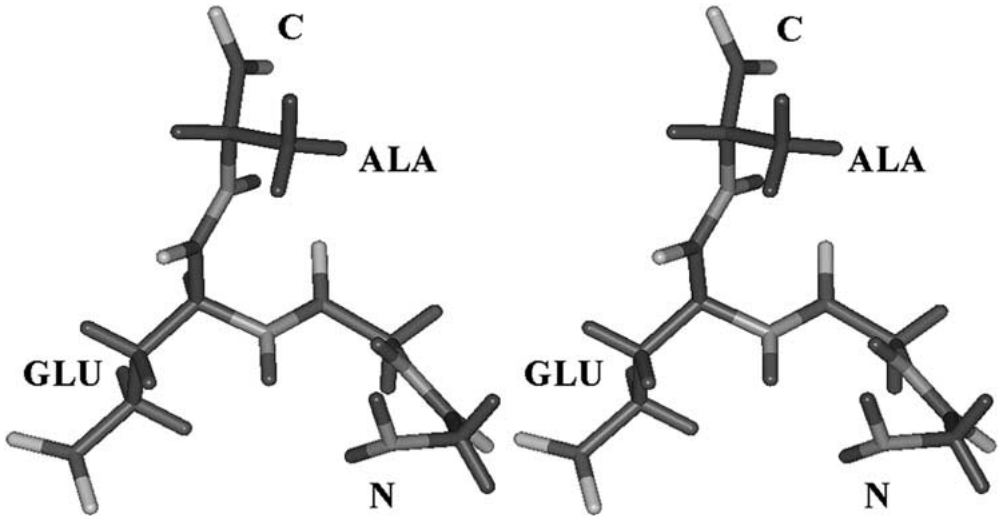

Figure 2. Stereo view of the lowest-energy conformation found during the simulation carried out for the tetrapeptide H-Gly-Gly-Glu-Ala-OH at pH 7.

observation of Jiménez et al. (1986). These authors found signs of non-*statistical-coil* structure near the C-terminus of the GGHA tetrapeptide. According to Jiménez et al. (1986) the presence of a salt bridge between the His side chain and the C-terminal group has been suggested to explain the anomalous titration of the His amide signal found in other peptides having His in the penultimate position. Moreover, histidine is also the residue for which Merutka et al. (1995) found the most significant differences for the side-chain proton chemical shifts when compared with those obtained by Bundi and Wüthrich (1979a).

To investigate the influence of pH and the amino acid sequence on the behavior of histidine, we carried out simulations for the GGHA peptide at pH 7 and 3, and for the GGHGG peptide at pH 5. We found quite a different degree of protonation (average charge distribution) at each pH for the histidine residue in the tetrapeptide and pentapeptide, as shown in Table 5. The average degree of protonation (charge) appears to be similar for both GGHA at pH 3 and GGHGG at pH 5, but quite different for that of GGHA at pH 7 where the histidine is almost fully deprotonated (uncharged). This preferential degree of protonation could explain the differences observed by Merutka et al. (1995) for the chemical shift of the histidine side chain between both GGHA at pH 7 and GGHGG at pH 5, i.e., due essentially to the difference in pH in these two experiments. In addition, the $^3J_{H\alpha H\beta}$ values for the side chain of histidine in the GGHGG peptide (Table 4) seem to be in better agreement with the expected value of a *statistical coil* than the calculated one found for the GGHA peptide. The histidine side chain of the GGHGG peptide populates all three rotamers equally, while the residue seems to have some preferred interaction with the backbone in the GGHA peptide.

There is no structural difference among the lowest-energy conformations found for the GGHA peptide at both pH 7 and pH 3 (the same value of $^3J_{NH\alpha}$ in Table 2), even though the protonation distributions differ at these two pH's (Table 5). However, it is important to note that, at pH 7, there are two low-energy conformations with similar energies, i.e., with a total energy difference of less than 0.7 Kcal mol^{-1}, displaying a preference for the side-chain position of histidine close to the C-terminus, as shown in Figures 3 and 4. One of these conformations, however, brings the uncharged histidine side chain very close to the deprotonated COO^- terminal group, as can be seen in Figure 4. Jiménez et al. (1986) reported that the nonlabile H2 proton of uncharged histidine is able to sense the deprotonation of the COO^- group, suggesting that the histidine side chain and the C-terminus are in close proximity, with which our conformation in Figure 4 is in agreement.

Conformational analysis of the oligopeptides H-Gly-Gly-Phe-Ala-OH and H-Gly-Gly-Phe-Gly-Gly-OH in aqueous solution

The value of the $^3J_{NH\alpha}$ coupling constant determined experimentally for phenylalanine in the tetrapeptide

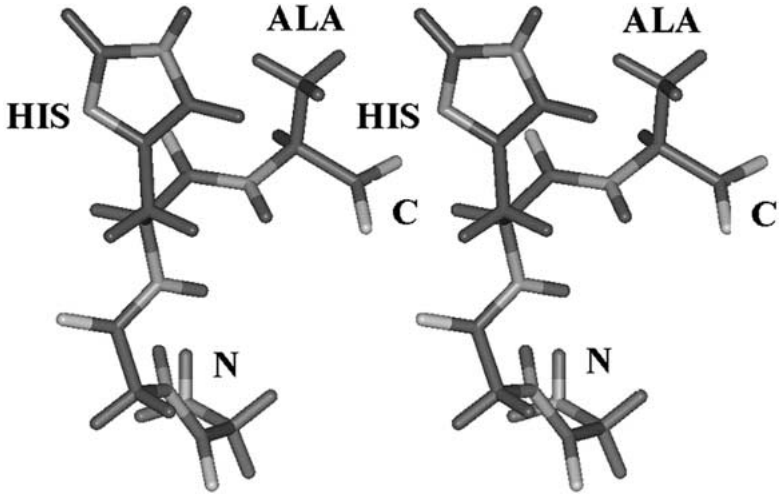

Figure 3. Stereo view of the lowest-energy conformation found during the simulation carried out for the tetrapeptide H-Gly-Gly-His-Ala-OH at pH 7.

GGFA (9.4 Hz) is strange since it is quite different from the values measured for any other amino acids in a *statistical-coil* conformation (Bundi and Wüthrich, 1979a). This behavior has been interpreted by Serrano (1995) as arising from a possible interaction of the phenylalanine side chain with the backbone of the oligopeptide or as a result of an error in the measurement.

Our Boltzmann-averaged value for the $^3J_{NH\alpha}$ coupling constant, computed at pH 7 (9.5 Hz), with the solvent free energy and free energy of ionization properly taken into account by using the solution of the Poisson–Boltzmann equation (Ripoll et al., 1996) as described in Methods, suggests that this anomalous behavior is due to a specific stacking of the phenylalanine side chain with the backbone. This preferential interaction is seen clearly in the lowest-energy conformation found in our simulation (see Figure 5). It can also be seen from this figure that the phenylalanine side chain is orientated towards the N-terminus of the linear tetrapeptide, in agreement with the suggestion made by Bundi and Wüthrich (1979b).

However, some observations have been mentioned by Merutka et al. (1995) regarding both the experimental environmental conditions (pH and temperature) and the existence of sequence dependence effects in the experiments of Bundi and Wüthrich (1979a). The sequence dependence, in particular, refers to the last residue at the C-terminus, i.e., the alanine residue that can influence the conformational preference of the 'X' residue at position 3 of the tetrapeptide, and hence, according to Merutka et al. (1995) induce some kind of systematic error because the alanine residue has a preference (Merutka et al., 1995) 'for backbone' dihedral angles in the α-region of the conformational space. In order to avoid these effects, Merutka et al. (1995) selected the pentapeptide H-Gly-Gly-X-Gly-Gly-OH to minimize the influence of end effects by inserting two glycine residues on each side of the central residue X.

To determine whether the proximity of the phenylalanine side chain to the backbone is sequence dependent, or is the result of an error in the measurement (Serrano, 1995), we carried out a series of simulations on the pentapeptide H-Gly-Gly-Phe-Gly-Gly-OH, at pH 5. The Boltzmann-averaged value for the vicinal coupling constant, $^3J_{NH\alpha}$, obtained for the phenylalanine residue (6.9 Hz) in our simulations is far from both the theoretical and experimentally-determined values for phenylalanine in the tetrapeptide H-Gly-Gly-Phe-Ala-OH, i.e., ∼9.5 Hz. The lowest-energy conformation found in our simulation for the peptide H-Gly-Gly-Phe-Gly-Gly-OH at pH 5 is shown in Figure 6. It can be seen that, in this conformation, there is a preferential orientation of the aromatic side chain of Phe toward the C-terminus, at variance with the observation found for the lowest-energy conformation of the peptide H-Gly-Gly-Phe-Ala-OH, at pH 7 (Figure 5). This preferential ordering is in agreement with the proposal of Merutka et al. (1995) who suggested the existence of some specific orientation of the aromatic side chain toward the C-terminus, or some interaction with the C-terminal glycine. As Merutka et al. (1995) also indicated, this effect has been observed for sequences Ar-Aa-Gly, where Ar is an aromatic

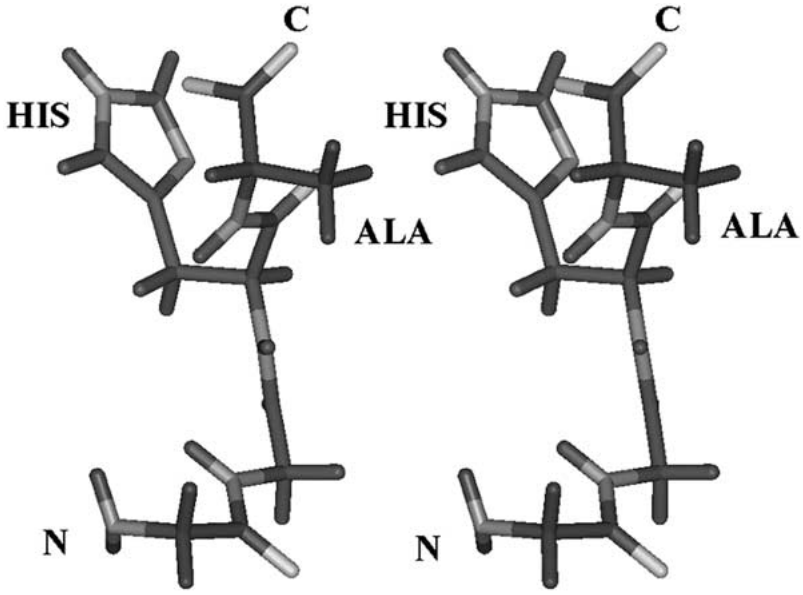

Figure 4. Stereo view of the conformation closest to the lowest-energy one (but 0.65 Kcal mol^{-1} above the one in Figure 3) found during the simulation carried out for the tetrapeptide H-Gly-Gly-His-Ala-OH at pH 7.

residue and Aa is any amino acid (Dyson et al., 1992; Kemmink et al., 1993)

Moderate agreement between the theoretical predictions and the experimental data for $^3J_{H\alpha H\beta}$ in both the tetrapeptide and pentapeptide containing phenylalanine is shown in Table 4. The values obtained for $^3J_{H\alpha H\beta}$ for the pentapeptide indicate an unrestricted side chain, at variance with the values found for the tetrapeptide which seem to indicate some kind of stacking of the side chain against the backbone. This effect can also be inferred from inspection of the lowest-energy conformations shown in Figures 5 and 6.

The largest difference between these Phe-containing tetra- and penta-peptides, beyond the orientation of the Phe side chain, is in the Boltzmann-averaged degree of protonation (charge) at the C-terminus. This end of the chain appears to be fully deprotonated (charged) for GGFA at pH 7 and only partially deprotonated ($\sim$ 36%) for GGFGG at pH 5, as can be seen from Table 5. This could also explain the preferential orientation of the phenylalanine side chain toward the N-terminus in the tetrapeptide. In principle, the phenylalanine side chain in the GGFA tetrapeptide could be oriented to either the N-terminus or C-terminus. However, if it were oriented toward the C-terminus, it would lead to a high energetic cost because the hydrophobic side chain would interfere with the solvation of the fully deprotonated (charged) C-terminus. On the other hand, if the phenylalanine side chain were oriented toward the N-terminus (as in Figure 5), it would not pay an additional energetic cost because the N-terminus is full deprotonated (uncharged). It is worth noting that the phenylalanine side chain of the GGFGG pentapeptide does not have such a restriction because (a) at pH 5 the C-terminus is only partially deprotonated (36%), and (b) the inclusion of one additional glycine in the sequence positions the phenylalanine side chain further away from the C-terminus in the sequence.

It is clear that the hydrophobic side chain of phenylalanine can be partially shielded from solvent by interacting with the backbone. However, the preferential orientation of the side chain will be dominated by both the pH at which the experiments are conducted and the type and amount of residues flanking the phenylalanine. The conformational preference of the aromatic side-chain of phenylalanine is certainly coupled with the degree of ionization (Ripoll et al., 1996; Vila et al., 1998) and *vice versa*; the state of ionization and the conformation are coupled (Laskowski and Scheraga, 1954).

A cluster analysis for the tetrapeptide GGFA, similar to the one already described for the tetrapeptide GGEA, reveals that family 1, out of 42 families, i.e. the one for which the *leading member* of the fam-

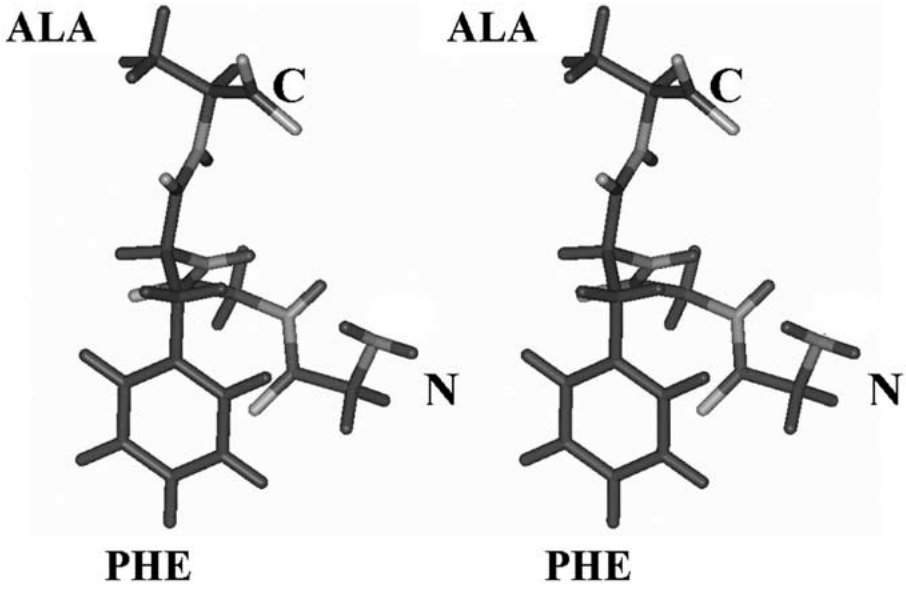

Figure 5. Stereo view of the lowest-energy conformation found during the simulation carried out for the peptide H-Gly-Gly-Phe-Ala-OH at pH 7.

ily is the lowest-energy conformation of the *statistical coil* ensemble, is the most populated family containing more than 49% of the total members of the conformation ensemble. Thus, it seems to be appropriate to compare some experimentally-derived magnitudes, such us the $^{13}C^{\alpha}$ chemical shifts, with the theoretical values derived from the lowest-energy conformation.

Quantum chemical calculation of the $^{13}C^{\alpha}$ chemical shift

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 variations in the backbone dihedral angles (ϕ, ψ) account for half or more of the observed variations in the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts (Wishart and Case, 2001). Calculations of the $^{13}C^{\alpha}$ chemical shifts for the lowest-energy conformation found for the tetrapeptide GGFA provides an independent assessment of the agreement obtained with the vicinal coupling constant which depends only on the dihedral angle ϕ through the Karplus relationship (Equation 3).

Comparison of theoretically-determined values for the H^{α} and $^{13}C^{\alpha}$ chemical shifts with the corresponding experimentally determined value for these nuclei are ideally suited for estimation of secondary structure in proteins based on chemical shifts, with the larger chemical shift dispersion of $^{13}C^{\alpha}$ usually giving better results (Schwarzinger et al., 2001). Furthermore, it is important to note that the statistical-coil chemical shift of these two nuclei seems to depend mainly on the attached side chain, with minor corrections based on the local amino acid sequence (Schwarzinger et al., 2001). The results of H^{α} chemical shifts have already been discussed in connection with Table 3.

Table 6 shows the results of *Quantum Chemical* calculations for the $^{13}C^{\alpha}$ chemical shift and the NMR experimentally determined value by Richarz and Wüthrich (1978). The calculations were carried out for both the lowest-energy conformation found for the tetrapeptide GGFA, displayed in Figure 5, and for one, out of 8, randomly selected initial conformations of the EDMC search. For each case, we carried out calculations of the $^{13}C^{\alpha}$ chemical shift with and without geometry optimization of internal coordinates following the procedure described in Methods.

As seen in Table 6, the agreement of the theoretical values with the experimental value is significant for the calculation carried out with the density functional theory (DFT) approach. On the other hand, the calculations carried out by using the Hartree–Fock (HF) approach led to results that are far from the experimental value. It is known that HF calculations can be improved by using the density functional theory-type calculations (Pearson et al., 1997); i.e., for higher accuracy in chemical shift computations, the inclusion of electron correlation in the theoretical treatment ap-

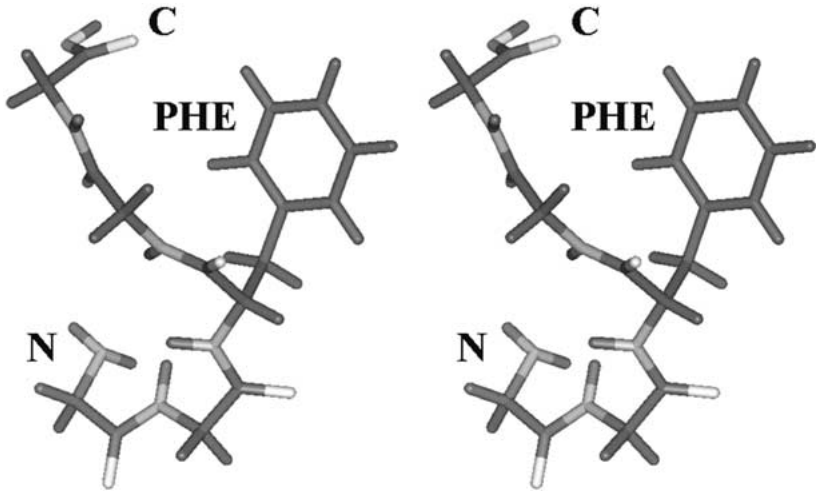

Figure 6. Stereo view of the lowest-energy conformation found during the simulation carried out for the pentapeptide H-Gly-Gly-Phe-Gly-Gly-OH at pH 5.

pears to be important. In addition, numerical evidence indicates that, whenever the HF and DFT results are significantly different, i.e., when the electron correlation effects are large, the DFT values are more reliable (Helgaker et al., 1999). In conformity with this observation, it has been shown (Wang et al., 2001) that the HF method is less appropriate than B3LYP when the system exhibits strong electron correlation effects.

Regardless of the quantitative comparison, it is worth observing that *both* calculations, i.e., those with HF and DFT, *predict* that the chemical shift for the $^{13}C^{\alpha}$ corresponding to phenylalanine in the lowest-energy conformation of the tetrapeptide with or without geometry optimization (47 ppm and 55 ppm, respectively) appears upfield in comparison with the corresponding computed value for one of the random conformations (52 ppm and 59 ppm, respectively). Thus, we can infer that, in both calculations, the $^{13}C^{\alpha}$ atom is predicted to be better shielded in the lowest-energy conformation than in the random one.

After comparing the experimental value (56 ppm) of the $^{13}C^{\alpha}$ chemical shift with those computed for both the lowest-energy (55 ppm) and the randomly-generated *initial* conformations (59 ppm), we can infer that calculations carried out with DFT by using the basis set 6-311+G(2d,p) is fairly sensitive to both backbone and side-chain conformational preference of phenylalanine.

Table 6. Quantum chemical computation of the $^{13}C^{\alpha}$ chemical shift δ for the Residue Phe in the tetrapeptide H-Gly-Gly-Phe-Ala-OH

Theoretical[a] (ppm)		Experimental[b] (ppm)
HF/6-311+G(2d,p)	B3LYP/6-311+G(2d,p)	
47.2[c] (51.7[d])	54.3[c] (59.7[d])	56.2
47.2[e] (52.2[f])	54.6[e] (58.6[f])	56.2

[a] *Ab initio* and DFT values for $^{13}C^{\alpha}$ chemical shifts were calculated as explained in Methods. All the theoretical and experimental values are in ppm relative to TMS (tetramethylsilane).

[b] Experimentally-determined $^{13}C^{\alpha}$ chemical shift from Richarz and Wüthrich (1978) at pH 7.7 and t = 35 °C.

[c] *Ab initio* and DFT calculated values of $^{13}C^{\alpha}$ chemical shifts for the lowest-energy conformation found for the peptide GGFA at pH 7, without geometry optimization, as explained in Methods.

[d] *Ab initio* and DFT calculated values of $^{13}C^{\alpha}$ chemical shifts for one, out of eight, initial random conformations of the EDMC search, without geometry optimization, as explained in Methods.

[e] *Ab initio* and DFT calculated values of $^{13}C^{\alpha}$ chemical shifts for the lowest-energy conformation found for the peptide GGFA at pH 7 and t = 35 °C. This value was obtained after geometry optimization in internal coordinates allowing all bond lengths and bond angles to relax, but all dihedral angles were frozen at the original values, as explained in Methods.

[f] *Ab initio* and DFT calculated value of $^{13}C^{\alpha}$ chemical shifts for one, out of eight, initial random conformations of the EDMC search. This value was obtained after geometry optimization in internal coordinates allowing all bonds lengths and bonds angles to relax, but all dihedral angles were frozen at the original values, as explained in Methods.

Discussion and conclusion

The agreement found between the calculated and the experimentally-derived parameters for *statistically-coiled* amino acids in different sequences and at different pH's reveals that solvent polarization effects and electrostatic interactions play a dominant role in determining the conformational preferences of both backbone dihedral angles ϕ and side-chain dihedral angles χ^1 in *statistically-coiled* peptides.

Analysis of the populations of conformations for phenylalanine in both a tetrapeptide at pH 7 and a pentapeptide at pH 5 reveals that the surprisingly high value for $^3J_{NH\alpha}$, when compared with values for other *statistically-coiled* amino acids, seems to be due to a preferential interaction with the backbone, as suggested by Serrano (1995) and Merutka et al. (1995) and not to an error in the measurement (Serrano, 1995).

The preference of the phenylalanine side chain to orient towards the C-terminus in the pentapetide GGFGG at pH 5 (Figure 6), or towards the N-terminus in the tetrapeptide GGFA at pH 7 (Figure 5), appears to be related to both (a) the differences in the sequences of both peptides, which place two glycine residues between phenylalanine and the C-terminus in the pentapeptide, in place of a single alanine in the tetrapeptide; and (b) the difference in the pH conditions under which the experiments were carried out. This difference seems to lead to different protonation (charge) distributions that, in turn, play a significant role in determining the observed conformation. In fact, the deprotonation (low charge of the amino group) at the N-terminus and high average protonation (low charge of the carboxyl group) at the C-terminus for the pentapeptide at pH 5 allows the side chain to explore the conformational space more freely, without the high energetic cost due to the interference of the side chain with the solvation of the ionized C-terminus. This effect is also reflected in both the theoretical and experimental values of $^3J_{H\alpha H\beta}$ showing sampling of a completely different set of values of the dihedral angle χ^1, as can be seen in Table 4.

The results of our DFT calculations of the $^{13}C^{\alpha}$ chemical shift of phenylalanine for the lowest-energy conformation of the tetrapeptide H-Gly-Gly-Phe-Ala-OH at pH 7 (55 ppm) show good agreement with the NMR experimentally-determined value obtained by Richarz and Wüthrich (1978) (56 ppm). On the other hand, a comparison with a randomly-generated conformation, used as one of the eight starting points of the EDMC searches (59 ppm), reveals that the DFT calculations for the $^{13}C^{\alpha}$ chemical shift are sensitive enough to discriminate between conformations and, in particular, the positions of the side chain, as was already noted by Schwarzinger et al. (2001).

Close agreement is found between the Boltzmann-averaged and the experimentally determined value of the α-proton chemical shifts, i.e., $\langle\delta_{theor}\rangle$ and δ_{sc} respectively, as shown in Table 3. The Boltzmann-averaged value of $\langle\delta_{theor}\rangle$ was determined, as explained in Methods, by using an empirical relationship given by Wishart et al. (1991) which relates the $\Delta\delta$ value$_s$ i.e. $\Delta\delta = (\langle\delta_{theor}\rangle - \delta_{obs})$, with the dihedral angle ϕ through a Karplus-type equation. The results, shown in Table 3, seem to indicate that this relation is useful for discriminating *statistical-coil* preferences from that of a structured state, such as α-helix and β-strand.

The ability to distinguish structured conformations from non-structured ones by analysis of the vicinal coupling constant, $^3J_{NH\alpha}$, was already discussed by Pardi et al. (1984). Here, we were able to show that the empirical relationship between the dihedral angle ϕ and the H^{α} chemical shift, introduced by Wishar et al. (1991), is also sensitive enough to reproduce the Boltzmann-averaged values for the α-proton chemical shift and hence to discriminate the *statistical-coil* from the α-helix and β-strand structured states. In others words, $^3J_{NH\alpha}$ is frequently used to distinguish between residues that belong to structured states such as α-helix (with $^3J_{NH\alpha} < 6.0$ Hz) from those that are not. This is not surprising since both NMR spectral parameters, viz., α-proton chemical shifts and $^3J_{NH\alpha}$, share a common *sinusoidal* dependence with the backbone dihedral angle ϕ (Wishar et al., 1991). The assumption that the α-proton chemical shift is influenced mainly by the preceding residue, and hence by the dihedral angle ϕ has been corroborated by Merutka et al. (1995) after comparing the proton chemical shifts for the pentapeptide H-Gly-Gly-X-Gly-Gly-OH at pH 5 (with X being one of the 20 naturally occurring amino acids) with those obtained by Bundi and Wüthrich (1979a) for the tetrapeptides H-Gly-Gly-X-Ala-OH at pH 7. The most notable exception to the observation that the preceding residue influences ϕ, was histidine, for which Merutka et al. (1995) observed a significant departure from the expected value. From our simulations, however, we have been able to rationalize this difference. We showed that the pattern of the average protonation (charge) distribution for both the tetrapeptide and the pentapeptide differs significantly, and that

this may be the main source of the observed inconsistencies between both experiments. In that sense, we agree with the suggestion of Merutka et al. (1995) who proposed that the difference in chemical shift may be caused by the difference in pH between both studies. The conformation, in turn, is sensitive to the state of ionization of the amino acid residues (Neurath et al., 1944) and, *vice versa*, the state of ionization and the conformation are coupled (Laskowski and Scheraga, 1954). It should be noted that the computed values of the fraction of the α-amino NH_3^+ form (column 2 of Table 5) differ from those that would be expected from the *Null or Zero Interaction* model (Yang and Honig, 1993), i.e., a model in which all ionizable residues are assumed to titrate independent of any interactions with their neighbors; the reported computed values reflect the interactions between the ionizable groups and its neighbors.

The importance of the protonation (charge) distribution and the specific amino acid sequence for the conformational preferences of some amino acids, shown in this study, indicates that (i) hydrophobic, charged or highly polar neutral residues in position X of the tetra- or pentapeptide will compete for solvation with the main-chain backbone CO and NH groups of nearest-neighbor residues, as well as with the unblocked ends groups; therefore, a predisposition to adopt a specific set of conformations, could be determined largely by sequence-dependent effects, as in the analysis of specific-sequence copolymers made by Scheraga et al. (2002) showing that bulky side chains, from both charged or highly polar residues, may sequester water away from the backbone and hence influence the conformational preferences of nearest-neighbor residues; (ii) the tetrapeptides H-Gly-Gly-X-Ala-OH seem to adopt favored conformations for several amino acids, such as histidine, glutamic acid and phenylalanine, and hence, such interactions are reflected in the measured *statistical-coil* state; and (iii) The *bias* introduced by specific interactions between a given residue and its neighbors means that, in such cases, these ensembles do not represent a truly statistical coil, i.e., they correspond to a *biased* energy-weighted sampling of the conformational space of the residue under consideration.

Acknowledgements

This research was supported by grants from the National Institutes of Health (GM-14312 and 5 R03 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 the (a) 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

Bagno, A. (2001) *Chem. Eur. J.*, **7**, 1652–1661.
Bashford, D. and Karplus, M. (1990) *Biochemistry*, **29**, 10219–10225.
Becke, A.D. (1993) *J. Chem. Phys.*, **98**, 5648–5652.
Beroza, P., Fredkin, D.R., Okamura, M.Y. and Feher, G. (1995) *Biophys. J.*, **68**, 2233–2250.
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.
Chesnut, D.B. (1996) *Rev. Comput. Chem.*, **8**, 245–297.
De Marco, A., Llinás, M. and Wüthrich K. (1978) *Biopolymers*, **17**, 617–636.
Ditchfield, R. (1974) *Mol. Phys*., **27**, 789–807.
Dyson, H.J., Sayre, J.R., Merutka, G., Shin, H.-C., Lerner, R.A. and Wright, P.E. (1992) *J. Mol. Biol.* **226**, 819–835.
Edsall, J.T. and Wyman J. (1958) In *Biophysical Chemistry*, Volume I, (Table IX), Academic Press Inc., New York, p. 536.
Facelli, J.C. (1998) *J. Phys. Chem. B*., **102**, 2111–2116.
Ferraro, M.B. (2000) *J. Mol. Struct. (Theochem.)*, **528**, 199–209.
Fiebig, K.M., Schwalbe, H., Buck, M., Smith, L.J. and Dobson, C.M. (1996) *J. Phys. Chem.*, **100**, 2661–2666.
Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Zakrzewski, V.G., Montgomery, J.A., Jr., 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., 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, Inc., Pittsburgh PA.
Gay, D.M., (1983) *ACM Trans. Math. Software*, **9**, 503–524.
Gilson, M.K. (1993) *Prot. Struct. Funct. Genet.*, **15**, 266–282.
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.
Hennig, M., Bermel, W., Spencer, A., Dobson, C.M., Smith, L.J. and Schwalbe, H. (1999) *J. Mol. Biol.*, **288**, 705–723.
Jameson, A.K., and Jameson, C.J. (1987) *Chem. Phys. Lett.*, **134**, 461–466.
Jiménez, M.A., Nieto, J.L., Rico, M., Santoro, J., Herranz, J. and Bermejo, F.J. (1986) *J. Mol. Struct.*, **143**, 435–438.
Karplus, M. (1959) *J. Chem. Phys.,* **30**, 11–15.
Karplus, M. (1963) *J. Am. Chem. Soc*., **85**, 2870–2871.
Kemmink, J., Van Mierlo, C.P.M., Scheek, R.M. and Creighton, T.E. (1993) *J. Mol. Biol.*, **220**, 312–322.
Laskowski, Jr. M. and Scheraga, H.A. (1954) *J. Am. Chem. Soc*., **76**, 6305–6319.
Lee, C, Yang, W. and Parr, R.G. (1988) *Phys. Rev. B.*, **37**, 785–789.
Merutka, G., Dyson, H.J. and Wright, P.E. (1995) *J. Biom. NMR*, **5,** 14–24.
Momany, F. A., McGuire, R.F., Burgess, A.W. and Scheraga H.A. (1975) *J. Phys. Chem.*, **79**, 2361–2381.
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.
Némethy, G., Pottle, M.S. and Scheraga H.A. (1983) *J. Phys. Chem.*, **87**, 1883–1887.
Neurath, H. Greenstein, J.P., Putnam, F.W. and Erickson, J.O. (1944) *Chem. Revs*., **34**, 157–265.
O'Connell, T.M., Wang, L., Tropsha, A. and Hermans, J. (1999) *Prot. Struct. Funct. Genet.*, **36**, 407–418.
O'Donnell, T.J., Hotovy, S.G., Pottle, M.S., Ripoll, D.R. and Scheraga, H.A. (1996) In *Lecture Notes in Computer Science*, Vol. 1067: *High-Performance Computing and Networking*, Liddel, H. Colbrook, A., Hertzberger, B. and Sloot, P., Eds., Springer-Verlag, Berlin, Heidelberg, New York, pp. 365–373.
Ösapay, K. and Case, D.A. (1994) *J. Biomol. NMR*, **4**, 215–230.
Pardi, A., Billeter, M. and Wüthrich, K. (1984) *J. Mol. Biol.*, **180**, 741–751.
Pearson, J.G., Le, H., Sanders, L.K., Godbout, N., Havlin, R.H., and Oldfield, E. (1997) *J. Am. Chem. Soc.*, **119**, 11941–11950.
Perrin, D.D. (1972) *Dissociation Constants of Organic Bases in Aqueous Solution*, Butterworths, London.
Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V. (1963) *J. Mol. Biol.*, **7**, 95–99.
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. and Scheraga, H.A. (1989) *J. Protein Chem.,* **8**, 263–287.
Ripoll, D.R., Liwo, A. and Scheraga, H.A. (1998) *Biopolymers*, **46**, 117–126.
Ripoll, D.R., Vorobjev, Y.N., Liwo, A., Vila, J.A. and Scheraga H.A. (1996) *J. Mol. Biol*., **264**, 770–783.
Scheraga, H.A., Vila, J.A. and Ripoll, D.R. (2002) *Biophys. Chem.*, in press.
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.
Serrano, L. (1995) *J. Mol. Biol.*, **254**, 322–333.
Simonson, T. and Brünger, A.T. (1994) *J. Phys. Chem.*, **98**, 4683–4694.
Sippl, M.J., Némethy, G. and Scheraga, H.A. (1984) *J. Phys. Chem.*, **88**, 6231–6233.
Sitkoff, D. and Case, D.A. (1997) *J. Am. Chem. Soc.*, **119**, 12262–12273.
Sitkoff, D., Sharp, K.A. and Honig, B. (1994) *J. Phys. Chem.*, **98**, 1978–1988.
Smith, L.J., Bolin, K.A., Schwalbe, H., MacArthur, M.W., Thornton, J.M. and Dobson, C.M. (1996) *J. Mol. Biol.*, **255**, 494–506.
Sun, H., Sanders, L.K. and Oldfield, E. (2002) *J. Am. Chem. Soc.*, **124**, 5486–5495.
Swindells, M.B., MacArthur, M.W. and Thornton, J.M. (1995) *Nat. Struct. Biol*., **2**, 596–603.
Vila, J.A., Ripoll, D.R. and Scheraga, H.A. (2000) *Proc. Natl. Acad. Sci. USA*, **97**, 13075–13079.
Vila, J.A., Ripoll, D.R. and Scheraga, H.A. (2001) *Biopolymers*, **58**, 235–246.
Vila, J.A., Ripoll, D.R., Villegas, M.E., Vorobjev, Y.N. and Scheraga, H.A. (1998) *Biophys. J.*, **75**, 2637–2646.
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.
Wang, B., Fleischer, U., Hinton, J.F. and Pulay, P. (2001) *J. Comp. Chem.*, **22**, 1887–1895.
West, N.J. and Smith, L.J. (1998) *J. Mol. Biol.*, **280**, 867–877.
Williamson, P.M., Asakura, T., Nakamura, E. and Demura, M. (1992) *J. Biomol. NMR*, **2**, 93–98.
Wishart, D.S. and Case, D.A. (2001) *Meth. Enzymol.*, **338**, 3–34.
Wishart, D.S. and Nip, A.M. (1998) *Biochem. Cell Biol.*, **76** 153–163.
Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.*, **222**, 311–333.
Wolinski, K., Hinton, J.F. and Pulay, P. (1990) *J. Am. Chem. Soc.*, **112**, 8251–8260.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York, NY, p. 17.
Xu, X.-P. and Case, D.A. (2001) *J. Biomol. NMR*, **21**, 321–333.
Yang, A.-S. and Honig, B. (1993) *J. Mol. Biol.*, **231**, 459–474.
Yang, A.-S., Gunner, M.R., Sampogna, R., Sharp, K. and Honig, B. (1993) *Prot. Struct. Funct. Genet.*, **15**, 252–265.
Zimmerman, S.S., Pottle, M.S., Némethy, G. and Scheraga, H.A. (1977) *Macromolecules*, **10**, 1–9.