

# $^1\text{H}$ , $^{13}\text{C}$ and $^{15}\text{N}$ random coil NMR chemical shifts of the common amino acids.

## I. Investigations of nearest-neighbor effects

David S. Wishart, Colin G. Bigam, Arne Holm\*, Robert S. Hodges and Brian D. Sykes\*\*

*Protein Engineering Network of Centres of Excellence, Department of Biochemistry,  
University of Alberta, Edmonton, AB, Canada T6G 2S2*

Received 25 July 1994  
Accepted 26 August 1994

*Keywords:*  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  NMR chemical shifts; Random coil; Amino acid; Peptide

### Summary

In this study we report on the  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  NMR chemical shifts for the random coil state and nearest-neighbor sequence effects measured from the protected linear hexapeptide Gly-Gly-X-Y-Gly-Gly (where X and Y are any of the 20 common amino acids). We present data for a set of 40 peptides (of the possible 400) including Gly-Gly-X-Ala-Gly-Gly and Gly-Gly-X-Pro-Gly-Gly, measured under identical aqueous conditions. Because all spectra were collected under identical experimental conditions, the data from the Gly-Gly-X-Ala-Gly-Gly series provide a complete and internally consistent set of  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  random coil chemical shifts for all 20 common amino acids. In addition, studies were also conducted into nearest-neighbor effects on the random coil shift arising from a variety of X and Y positional substitutions. Comparisons between the chemical shift measurements obtained from Gly-Gly-X-Ala-Gly-Gly and Gly-Gly-X-Pro-Gly-Gly reveal significant systematic shift differences arising from the presence of proline in the peptide sequence. Similarly, measurements of the chemical shift changes occurring for both alanine and proline (i.e., the residues in the Y position) are found to depend strongly on the type of amino acid substituted into the X position. These data lend support to the hypothesis that sequence effects play a significant role in determining peptide and protein chemical shifts.

### Introduction

The interpretation of chemical shifts plays an increasingly important role in peptide and protein NMR studies. Over the past few years a number of methods have been developed which employ chemical shifts in identifying secondary structure (Pastore and Saudek, 1990; Wishart et al., 1991a; Wishart and Sykes, 1994a), in monitoring folding transitions (Reily et al., 1992), in quantifying main-chain flexibility (Wishart et al., 1991b) and in refining tertiary structures (Gippert et al., 1990). Many of these applications evolved out of a series of statistical and experimental studies which systematically analyzed the effects of protein conformation and flexibility on  $^1\text{H}$ ,  $^{13}\text{C}$

and  $^{15}\text{N}$  chemical shifts (Saito, 1986; Williamson, 1990; Spera and Bax, 1991; Wishart et al., 1991b). This renewed interest in chemical shifts among experimentalists has also led to a renewed interest among theoreticians. In particular, a number of very powerful and remarkably accurate semiempirical (Ösapay and Case, 1991; Herranz et al., 1992; Williamson et al., 1992) and ab initio quantum-mechanical approaches (De Dios et al., 1993a,b) have recently emerged which permit the direct calculation of  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  and  $^{19}\text{F}$  chemical shifts. The rapid development and surprising success of these approaches suggests that protein structures may eventually be determined from chemical shift information alone.

Critical to both the experimental and theoretical suc-

\*Permanent address: Chemistry Department, Royal Veterinary and Agriculture University, Frederiksberg, Denmark.

\*\*To whom correspondence should be addressed.

*Abbreviations:* DIEA, diisopropylethylamine; DMF, dimethylformamide; DMSO, dimethyl sulfoxide; DSS, 2,2-dimethyl-2-silapentane-5-sulfonic acid; HMQC, heteronuclear multiple-quantum coherence; HOBt, *N*-hydroxybenzotriazole; MBHA, 4-methylbenzylhydramine; NOE, nuclear Overhauser effect; TBTU, 2-(benzotriazole-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate; TFE, trifluoroethanol; TMS, tetramethylsilane; TOCSY, total correlation spectroscopy; TSP, 3-(trimethylsilyl)propionate, sodium salt.

cesses of the past five years has been the availability of tables listing the so-called 'random coil' chemical shifts of the 20 common amino acids. For much of the last two decades, these random coil values have served both as the bench-marks for experimentalists and the correction factors for theoreticians. The random coil chemical shift is typically defined as the experimentally measured chemical shift of an amino acid residue within a peptide, which is free to access all sterically allowed regions of its conformational space. Historically, random coil chemical shifts have been measured using short, substituted linear peptides with such sequences as Gly-Gly-X-Ala (Richarz and Wüthrich, 1978; Bundi and Wüthrich, 1979a) or Gly-Gly-X-Gly-Gly (Keim et al., 1973a,b,1974; Merutka et al., 1995) (where X is the amino acid of interest) under a variety of solvent conditions. The assumption has always been that these peptides were sufficiently short so as to remain completely unstructured (see, however, Bundi and Wüthrich (1979b)) and that the presence of neighboring glycine (or alanine) residues prevented any steric perturbations of the residue being measured. Over the past 20 years random coil  $^1\text{H}$  chemical shifts have been determined for amino acids in water (Bundi and Wüthrich, 1979a), DMSO (Masson and Wüthrich, 1973) and aqueous TFE (Jimenez et al., 1992; Merutka et al., 1995) while random coil  $^{13}\text{C}$  chemical shifts have been measured for amino acids in water (Keim et al., 1973a,b,1974; Richarz and Wüthrich, 1978), DMSO (Grathwohl and Wüthrich, 1974), aqueous TFE (Thanabal et al., 1994), aqueous acetonitrile (Thanabal et al., 1994) and the solid state (reviewed by Saito in 1986). To date there has been only one published study of random coil  $^{15}\text{N}$  chemical shifts of amino acids in aqueous solution (Glushka et al., 1989,1990 (correction)). However, this work was based on extrapolated values from N-acetylated amino acids (measured in DMSO).

Random coil chemical shifts can also be inferred from statistical data. Using chemical shift information collected from previously assigned protein resonances, several groups have produced estimates of 'average' or 'coil' chemical shifts (Gross and Kalbitzer, 1988; Szilagy and Jardetzky, 1989; Wishart et al., 1991b) which seem to serve as reasonably good proxies for experimentally measured 'random coil' chemical shifts (Wishart and Sykes, 1994b). The fact that there is no perfect agreement between the two data sets (particularly for amide chemical shifts and for certain hydrophobic amino acids (Wishart et al., 1991b)) could be due to the inherent shortcomings of either the experimental measurements, the statistical sampling methods or both. The difficulties inherent to statistical approaches include the broad variation in sample conditions, the exclusion of ring current corrections, the limited sample size (for histidine, methionine and tryptophan), the probable conformational bias, and the indeterminacy of whether 'averaged' or 'coil'

chemical shifts would best represent the true value for 'random coil' chemical shifts.

Experimental measurements of random coil shifts also have their share of problems. For instance, solvent, salt, pH and temperature variations can lead to significant differences between any two sets of 'random coil' chemical shifts (Howarth, 1978; Richarz and Wüthrich, 1978). Furthermore, the choice of different reference compounds (such as TMS, TSP, DSS or dioxane) can cause considerable variation in the quoted values of random coil chemical shifts (Thanabal et al., 1994; Wishart and Sykes, 1994a). Similarly, the use of unprotected or protected tetrapeptides (Bundi and Wüthrich, 1979a), pentapeptides (Keim et al., 1973a,b,1974; Spera and Bax, 1991; Thanabal et al., 1994) or hexapeptides (vide infra) containing either glycine or alanine can introduce unspecified 'end-group effects', nearest-neighbor perturbations or steric interactions that will differ for each class of peptide or for each class of amino acid (Bundi and Wüthrich, 1979b). The fact that even pentapeptides have been shown to adopt relatively stable, sequence-dependent structures in aqueous solution (Dyson et al., 1988; Dyson and Wright, 1991) also calls into question the validity of the assumption that short linear peptides can serve as good models for random coils. Finally, as a consequence of the fact that no single set of  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  random coil shifts has been collected under identical conditions, using the same set of peptides and the same set of chemical shift references, random coil chemical shifts have had to be tabulated largely on an ad hoc basis (Wüthrich, 1986; Wishart and Sykes, 1994b).

The lack of internal consistency between various data sets, combined with the dearth of reliable  $^{15}\text{N}$  data and the persistent discrepancies between experimental and statistically derived random coil values, clearly suggests that better and more complete random coil chemical shift measurements must be performed. Our lack of understanding concerning peptide sequence effects or nearest-neighbor interactions and their influence on the intrinsic chemical shifts of amino acids implies that this, too, is an issue which must be further explored. Indeed, there is already considerable evidence for the influence of at least one amino acid (proline) on the  $^{13}\text{C}$  chemical shifts of a large number of residues (Torchia et al., 1975; Howarth, 1978). The question, therefore, arises as to whether this kind of sequence effect may be observed for other amino acids, for other positions or for other nuclei.

In this report we seek to resolve the problems of inconsistency with random coil measurements by providing complete  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  random coil shifts for the protected linear hexapeptide Ac-Gly-Gly-X-Ala-Gly-Gly-NH<sub>2</sub> (where X is any of the 20 common amino acids). By using the same set of protected, fully denatured (in 1 M urea) linear peptides with the same consistent set of references at relatively low peptide concentrations, we believe we

have been able to avoid the possible limitations of previous studies. Furthermore, by systematically varying the residues substituted into the third and fourth positions of this peptide model, we have been able to explore the effects of nearest-neighbor interactions on intrinsic amino acid chemical shifts. It is expected that these results could have important implications for the analysis of chemical shift variations in peptides (Williamson, 1990; Zhou et al., 1992), in the monitoring of secondary shift changes during folding (Reily et al., 1992), in the characterization of denatured proteins (Evans et al., 1991; Neri et al., 1992) and in the development of semiempirical and quantum-mechanical theories of chemical shifts (De Dios et al., 1993a,b; Ösapay and Case, 1994).

## Materials and Methods

To acquire a complete set of  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  random coil chemical shifts while at the same time studying nearest-neighbor effects, it was necessary to develop a peptide-solvent system that avoided many of the possible shortcomings of earlier models. Based on previously published studies (Keim et al., 1973a,b,1974; Richarz and Wüthrich, 1978), it was decided that a linear hexapeptide with the sequence Gly-Gly-X-Y-Gly-Gly could serve as the best (i.e. simplest) model for studying all 400 ( $20 \times 20$ ) amino acid pairs. The presence of two glycine residues on the N-terminus and two glycine residues on the C-terminus was expected to allow maximal flexibility of the two central residues, while at the same time preventing many of the problems associated with peptide 'end-group' and terminal charge effects. However, subsequent tests comparing unprotected hexapeptides with their protected (N-acetylated, C-amidated) counterparts revealed that protected peptides generally yielded more stable chemical shift measurements (particularly for  $^{13}\text{C}$  carbonyl and  $^1\text{H}$  amide resonances). Because of this, it was decided that the protected hexapeptide model (Ac-Gly-Gly-X-Y-Gly-Gly-NH<sub>2</sub>) would be adopted for all subsequent chemical shift measurements\*.

In order to ensure that these peptides could be fully solubilized and that no residual secondary structures would remain in solution, all peptides were dissolved in a buffer containing 1 M urea and 50 mM phosphate (pH ~5). The chaotropic properties of urea are well known and its presence in limited concentration was found to be sufficient to solubilize all but one peptide (Gly-Gly-Trp-Ala-Gly-Gly). Furthermore, the presence of urea serves to eliminate any detectable trace of residual structure or peptide aggregation.

\*Thanabal and co-workers (Thanabal et al., 1994), in their recent study of  $^{13}\text{C}$  chemical shifts, independently selected a similar kind of protected pentapeptide model.

## Peptide synthesis

All peptides with the sequence Gly-Gly-X-Ala-Gly-Gly were synthesized on an Applied Biosystems 430A solid-phase peptide synthesizer using a Boc-Gly-MBHA resin. After completion of the chain synthesis, the peptides were N-acetylated using a mixture of 10% acetic anhydride and 5% DIEA in dichloromethane. The blocked peptides were cleaved from the resin using anhydrous hydrogen fluoride in the presence of anisole to produce the free peptide. Purification (if necessary) was accomplished by reversed-phase HPLC using a C<sub>8</sub> semipreparative column. The peptide composition was verified by plasma-desorption mass spectrometry. Because of the poor solubility characteristics of the peptide Gly-Gly-Trp-Ala-Gly-Gly, this compound had to be resynthesized with free amino- and carboxy-termini.

Peptides with the sequence Gly-Gly-X-Pro-Gly-Gly were manually synthesized on a multi-well peptide synthesizer (A. Holm, personal communication) using flow resin NovaSyn PR 500 coupled to Fmoc-Gly-OH. Coupling of all remaining amino acids was achieved using one equivalent of TBTU, one equivalent of HOBt and two equivalents of DIEA for each equivalent of Fmoc amino acid. Peptides were cleaved and deprotected with a mixture of 95% TFA/5% H<sub>2</sub>O. After deprotection, the peptides were acetylated with 10% acetic anhydride in DMF and washed with DMF (2 $\times$ ) and MeOH (4 $\times$ ) before drying overnight. Purity and composition were verified by reversed-phase HPLC and plasma-desorption mass spectrometry. Limitations with the technology required the *t*-Boc chemistry described above for the syntheses of the peptides with the sequence pairs Arg-Pro, Cys-Pro, His-Pro, Met-Pro and Trp-Pro.

## NMR methods

All spectra were collected on a Varian Unity 300 MHz spectrometer ( $^1\text{H}$  frequency = 299.92 MHz,  $^{13}\text{C}$  frequency = 74.42 MHz,  $^{15}\text{N}$  frequency = 30.39 MHz), equipped with either a 5 mm inverse detection probe (for  $^1\text{H}$  and  $^{15}\text{N}$  NMR) or a 10 mm broadband direct detection probe (for  $^{13}\text{C}$  NMR). For all  $^1\text{H}$  and  $^{13}\text{C}$  studies, peptide samples with a dry weight of 5–7 mg were dissolved in 500  $\mu\text{l}$  of a deuterated buffer containing 99.9% D<sub>2</sub>O (Sigma), 1.0 M deuterated urea and 50 mM phosphate (pH 5.1, uncorrected meter reading). Sample concentrations were typically ~20 mM. If necessary, further pH adjustments were made using small aliquots of dilute DCl or NaOD (MSD isotopes) to give a sample pH of  $5.0 \pm 0.3$ . Heteronuclear  $^{15}\text{N}$  studies were conducted on the same samples after they had been lyophilized and redissolved in 500  $\mu\text{l}$  of 95% H<sub>2</sub>O/5% D<sub>2</sub>O. For all studies reported here, the sample temperature was maintained at 25 °C.

One-dimensional  $^1\text{H}$  data were acquired with a  $^1\text{H}$  sweepwidth of 4000 Hz and an acquisition time of 3.98 s.

The residual HDO signal was suppressed by presaturation. When necessary, TOCSY (Braunschweiler and Ernst, 1983) spectra were collected using 128  $t_1$  increments and spectral widths of 3000 Hz in both dimensions. Acquisition times were set to 0.341 s, relaxation delays were 2.0 s and spin-lock (MLEV-17) mixing times were 50 ms. Data were zero-filled to produce a matrix of  $2K \times 2K$  complex points and processed using a shifted sine-bell weighting function. Quadrature detection was achieved using the method of States et al. (1982). All  $^1\text{H}$  spectra were referenced to internal DSS at 0.0 ppm.

$^1\text{H}$ - $^{15}\text{N}$  HMQC experiments (Bax and Subramanian, 1986) were collected with a  $^{15}\text{N}$  sweepwidth of 2000 Hz and a  $^1\text{H}$  sweepwidth of 4000 Hz. A total of 1024 complex points were collected along the  $t_2$  dimension ( $^1\text{H}$ ) and 64 increments along the  $t_1$  dimension ( $^{15}\text{N}$ ). Data in both dimensions were zero-filled to create a  $2K \times 1K$  data set. The relaxation delay for this experiment was typically 1.2 s and the refocusing delay was set at 5.3 ms. Typical collection times were about 6 h. All  $^1\text{H}$ - $^{15}\text{N}$  HMQC spectra were referenced to internal DSS (for the  $^1\text{H}$  dimension) and external liquid ammonia (for the  $^{15}\text{N}$  dimension).

All 1D  $^{13}\text{C}$  spectra were acquired with a sweepwidth of 18 000 Hz. Protons were decoupled throughout the acquisition, pulse and delay periods using broadband Waltz-16 decoupling. Acquisition times for 1D spectra were typically 2.0 s. Total acquisition times were typically 20 h. Where necessary,  $^1\text{H}$ - $^{13}\text{C}$  HMQC experiments were collected with a  $^{13}\text{C}$  sweepwidth of 10 000 Hz and a  $^1\text{H}$  sweepwidth of 4000 Hz. A total of 1024 complex points was collected along the  $t_2$  dimension ( $^1\text{H}$ ) and 128 increments along the  $t_1$  dimension ( $^{13}\text{C}$ ). Data in both dimensions were zero-filled to create a  $2K \times 1K$  data set. The relaxation delay for this experiment was typically 1.5 s and the refocusing delay was set at 3.6 ms. All  $^{13}\text{C}$  spectra were referenced to internal DSS at 0.0 ppm.

#### *Chemical shift referencing*

One of the most critical features in measuring random coil chemical shifts is the issue of chemical shift referencing. It is essential that a stable, well-defined referencing system be chosen in order to ensure experimental consistency and reproducibility. In the past, a wide variety of internal and external reference compounds has been used for both  $^1\text{H}$  and  $^{13}\text{C}$  random coil chemical shift measurements including TMS (Howarth, 1978), TSP (Bundi and Wüthrich, 1979a; Spera and Bax, 1991),  $\text{CS}_2$  (Keim et al., 1974) and dioxane (Richarz and Wüthrich, 1978). A similar variety of reference compounds has also been employed in many  $^{15}\text{N}$  studies, including liquid  $\text{NH}_3$ , nitromethane, urea,  $\text{NH}_4\text{Cl}$  and  $\text{NH}_4\text{NO}_3$ . While the differences in most common  $^1\text{H}$  standards (DSS, TSP and TMS) are trivially small, the variations in  $^{13}\text{C}$  standards (dioxane, TMS, TSP and DSS) and  $^{15}\text{N}$  standards ( $\text{NH}_3$ ,

$\text{NH}_4\text{Cl}$ ,  $\text{NH}_4\text{NO}_3$  and nitromethane) can be quite substantial (Thanabal et al., 1994; Wishart and Sykes, 1994b). A case in point concerns the use of internal dioxane for  $^{13}\text{C}$  chemical shift referencing. Depending on how it is measured, the dioxane signal can be found at 66.6 ppm (when measured with respect to external TMS), at 67.5 ppm (when measured against internal TMS) or at 69.4 ppm (when measured against internal TSP). In fact, it is this variability among common standards that accounts for the most significant proportion of observed discrepancies among published random coil chemical shift values (compare, for instance, the  $^{13}\text{C}$  values quoted by Richarz and Wüthrich (1978) and those quoted by Thanabal et al. (1994)).

A broad literature survey, in conjunction with a series of detailed chemical shift referencing experiments conducted in our laboratory, indicated that the ideal  $^1\text{H}$  and  $^{13}\text{C}$  reference compound would have to be: (i) water soluble (i.e., a viable internal standard); (ii) insensitive to solvent, temperature and pH variations; (iii) a commonly used primary (zero-point) standard; (iv) unlikely to interact with a peptide or protein; and (v) directly or indirectly detectable in low concentrations. Among the possible  $^1\text{H}$  and  $^{13}\text{C}$  reference compounds tested, only 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) met all five specifications. All other tested compounds failed in at least one of the selection criteria. For instance, TMS was found to be insoluble in water and likely to interact with hydrophobic peptides and proteins, TSP was found to be quite sensitive to pH variations both as a  $^1\text{H}$  standard (DeMarco, 1977) and as a  $^{13}\text{C}$  standard (Wishart et al., 1995), while dioxane was found to resonate too far downfield to serve as a good primary (zero-point) standard for either  $^1\text{H}$  or  $^{13}\text{C}$  spectroscopy. Consequently, internal DSS was adopted as the zero-point standard for all  $^1\text{H}$  and  $^{13}\text{C}$  chemical shift measurements reported in this study. Typically a 10  $\mu\text{M}$  concentration of DSS was found sufficient for detection by  $^1\text{H}$  NMR, while a 1 mM concentration was sufficient for direct detection by  $^{13}\text{C}$  NMR. The presence of 1 M urea was found not to affect the chemical shift of DSS in any significant manner ( $< 0.01$  ppm). While we advocate adopting DSS as a universal  $^1\text{H}$  and  $^{13}\text{C}$  standard, there are now several published tables which provide the necessary conversion factors to switch to TMS or TSP as alternative reference standards (Wishart and Sykes, 1994a,b).

The situation for  $^{15}\text{N}$  referencing is somewhat more clear-cut. There is only one primary zero-point standard which has been universally adopted by the NMR community, i.e., liquid  $\text{NH}_3$ . However, because the apparent difficulty of working with liquid  $\text{NH}_3$ , many spectroscopists have chosen to use a variety of indirect referencing methods (Live et al., 1984) or secondary reference compounds (Witanowski et al., 1993). As a result, a good deal of confusion has arisen over the use of these compounds

and the proper approach to employing them as references. This may be, in part, the result of the fact that many workers adopt secondary or tertiary reference values measured on iron core (transverse field) magnets (Srinivasan and Lichter, 1977) without making the necessary bulk susceptibility corrections to apply these numbers to superconducting (parallel field) magnets. In order to avoid the problems of secondary references and bulk susceptibility corrections, we decided to measure all  $^{15}\text{N}$  chemical shifts with respect to a single sample of external liquid ammonia. This sample was prepared and calibrated as follows.

Ammonia gas was condensed into a chilled ( $-195\text{ }^\circ\text{C}$ ) thick-walled 5 mm (o.d.) NMR tube and sealed with an oxy-gas torch, yielding a liquid ammonia sample with a column height of  $\sim 5$  cm. This sample was placed coaxially within a 10 mm thin-walled NMR tube containing 2 ml of 99.9%  $\text{D}_2\text{O}$  and a  $^{15}\text{N}$  spectrum was recorded (at 30.32 MHz) with the lock signal tuned to  $\text{D}_2\text{O}$  and the temperature maintained at  $25\text{ }^\circ\text{C}$ . The exact frequency offset (from the carrier) of the  $^{15}\text{N}$  signal arising from the liquid ammonia was then determined and this value was used to calculate the zero-point (0 ppm) for all subsequent  $^{15}\text{N}$  HMQC spectra collected on the spectrometer. The  $\Xi$  ratio (the ratio of the  $^{15}\text{N}$  and  $^1\text{H}$  zero-point frequencies) was determined to be 0.101329118.

This approach offers several advantages. First, it is direct in that it does not depend on the measurement of any secondary or tertiary reference compounds (such as nitromethane or  $\text{NH}_4\text{Cl}$ ), nor does it depend on any extrapolations to 0 ppm. Second, it is relatively simple and

reproducible. Finally, the reference value we obtain is found to yield  $^{15}\text{N}$  chemical shifts which are in good agreement (vide infra) with the 'average' or 'coil'  $^{15}\text{N}$  chemical shifts reported by Wishart et al. (1991b). This provides a good indication that our referencing procedure is not only reasonable, but that it is consistent with the bulk of data already published on peptides and proteins.

## Results and Discussion

Tables 1–3 provide complete  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  random coil chemical shift assignments for the hexapeptide Gly-Gly-X-Ala-Gly-Gly (where X is any of the 20 common amino acids). The assignments presented here were made using a variety of standard approaches. Many resonances were assigned by comparison of two or more 1D  $^1\text{H}$  (or  $^{13}\text{C}$ ) spectra of slightly different analogs. This approach permitted the rapid identification of most resonances belonging to the variable (i.e. substituted) amino acids. These 1D  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra were further complemented by a number of 2D correlated experiments (TOCSY,  $^{15}\text{N}$  HMQC and  $^{13}\text{C}$  HMQC) which were used to resolve ambiguous assignments or to separate degenerate resonances. The assignment task was made much easier by virtue of the fact that a considerable body of  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shift data had already been accumulated for the 20 amino acids of interest (Richarz and Wüthrich, 1978; Wüthrich, 1986; Wishart et al., 1991b). This information was used in confirming and clarifying many of our initial assignments.

Because of the conditions under which these peptides

TABLE 1  
RANDOM COIL  $^1\text{H}$  CHEMICAL SHIFTS FOR THE 20 COMMON AMINO ACIDS WHEN FOLLOWED BY ALANINE

Residue	NH	H $^\alpha$	H $^\beta$	Others
Ala	8.24	4.32	1.39	
Cys (reduced)	8.32	4.55	2.93, 2.93	
Cys (oxidized)	8.43	4.71	3.25, 2.99	
Asp	8.34	4.64	2.72, 2.65	
Glu	8.42	4.35	2.06, 1.96	$\gamma\text{CH}_2$ 2.31, 2.31
Phe	8.30	4.62	3.14, 3.04	2,6H 7.28; 3,5H 7.38; 4H 7.32
Gly	8.33	3.96		
His	8.42	4.73	3.29, 3.16	2H 8.58; 4H 7.29
Ile	8.00	4.17	1.87	$\gamma\text{CH}_2$ 1.45, 1.16; $\gamma\text{CH}_3$ 0.91; $\delta\text{CH}_3$ 0.86
Lys	8.29	4.32	1.84, 1.75	$\gamma\text{CH}_2$ 1.44, 1.44; $\delta\text{CH}_2$ 1.68, 1.68; $\epsilon\text{CH}_2$ 2.99, 2.99; $\epsilon\text{NH}_2$ 7.81
Leu	8.16	4.34	1.62, 1.62	$\gamma\text{CH}_2$ 1.59; $\delta\text{CH}_3$ 0.92, 0.87
Met	8.28	4.48	2.11, 2.01	$\gamma\text{CH}_2$ 2.60, 2.54; $\epsilon\text{CH}_3$ 2.10
Asn	8.40	4.74	2.83, 2.75	$\gamma\text{NH}_2$ 7.59, 6.91
Pro	—	4.42	2.29, 1.94	$\gamma\text{CH}_2$ 2.02, 2.02; $\delta\text{CH}_2$ 3.63, 3.63
Gln	8.32	4.34	2.12, 1.99	$\gamma\text{CH}_2$ 2.36, 2.36; $\delta\text{NH}_2$ 7.52, 6.85
Arg	8.23	4.34	1.86, 1.76	$\gamma\text{CH}_2$ 1.63, 1.63; $\delta\text{CH}_2$ 3.20, 3.20; $\epsilon\text{NH}$ 8.07
Ser	8.31	4.47	3.89, 3.87	
Thr	8.15	4.35	4.24	$\gamma\text{CH}_3$ 1.21
Val	8.03	4.12	2.08	$\gamma\text{CH}_3$ 0.94, 0.93
Trp <sup>a</sup>	8.25	4.66	3.29, 3.27	2H 7.27; 4H 7.65; 5H 7.18; 6H 7.25; 7H 7.50
Tyr	8.12	4.55	3.03, 2.98	2,6H 7.14; 3,5H 6.84

Chemical shifts are referenced to internal DSS at  $25\text{ }^\circ\text{C}$ , pH  $\sim 5.0$ .

<sup>a</sup> Measured using a peptide with free N- and C-termini.

were measured (1 M urea, low pH, low peptide concentration) and because of the completeness, consistency and accuracy of the data, we believe that these values may be regarded as the current 'best estimates' of the random coil chemical shifts of the 20 common amino acids. However, it is important to note that the values in Tables 1–3 are specific to a standard set of chemical shift references (DSS for  $^1\text{H}$  and  $^{13}\text{C}$ ;  $\text{NH}_3$  for  $^{15}\text{N}$ ) and that conversion to other standards such as TMS, TSP or  $\text{NH}_4\text{Cl}$  requires the addition or subtraction of certain correction factors (see Wishart and Sykes (1994a) for details). For the most part, conversions from DSS to TSP are trivially small (for  $^1\text{H}$ ) and are only significant at the 0.1 ppm level for  $^{13}\text{C}$ . However, conversions to other standards are generally much larger and will significantly alter the chemical shifts presented in these tables.

#### *Comparison with other data ( $^1\text{H}$ chemical shifts)*

At present there are two other published collections of random coil  $^1\text{H}$  chemical shifts. The first was compiled by Bundi and Wüthrich (1979a) and revised slightly in a later publication (Wüthrich, 1986). The second was recently completed by Merutka et al. (1995). The latter survey also addresses the issues of solvent effects and temperature dependencies – two key issues which are particularly important for understanding amide chemical shifts. For the first part of this discussion we shall limit our comparison to Wüthrich's data (1986).

Detailed comparisons between the chemical shifts presented in Table 1 and the data provided by Wüthrich (1986) and Bundi and Wüthrich (1979a) indicate an excellent level of agreement, particularly among the  $\alpha$ -protons and the side-chain aliphatic and aromatic protons ( $R^2 = 0.95\text{--}0.99$ ). However, there are some notable exceptions, the most obvious of which can be found among the amide  $^1\text{H}$  chemical shifts. Careful comparison indicates that Wüthrich's values are consistently higher (average 8.35 ppm versus 8.26 ppm) and substantially more dispersive (8.09–8.75 ppm versus 8.00–8.43 ppm) than the present data. Some amide chemical shifts are significantly different, particularly those for asparagine (8.75 ppm versus 8.40 ppm), valine (8.44 ppm versus 8.03 ppm) and leucine (8.42 ppm versus 8.16 ppm). There could be any number of reasons for these discrepancies, but we believe they are primarily the result of 'end-group' and 'residual structure' effects arising from Bundi and Wüthrich's use of unprotected tetrapeptides in high (50 mM) concentration without the presence of a denaturant. The fact that Bundi and Wüthrich collected their data at 35 °C (pH 7.0), whereas the present data were collected at 25 °C (pH 5.0), could partly account for the difference. However, correcting for the 10 °C temperature discrepancy would only exaggerate the overall chemical shift difference (by  $\sim 0.08$  ppm) – not diminish it. Similarly, systematic temperature and pH differences are not likely to be able to explain the signifi-

cant individual discrepancies seen for the amide  $^1\text{H}$  resonances of asparagine, valine and leucine.

Comparison of the amide  $^1\text{H}$  data in Table 1 with those of Merutka et al. (1995) also reveals some interesting differences. Perhaps the most obvious feature is the extent to which the amide  $^1\text{H}$  shifts of Merutka et al. are shifted downfield relative to the data in Table 1. Even after the necessary temperature corrections (to 25 °C) have been made (using the temperature coefficients supplied by Merutka et al.), their mean amide  $^1\text{H}$  shift is still 0.23 ppm further downfield than the mean amide  $^1\text{H}$  shift reported in Table 1 (8.49 ppm versus 8.26 ppm). Because of the similarity in experimental conditions, it seems likely that the cause of the downfield displacement of Merutka's  $^1\text{H}$  amide chemical shifts lies in the peptide sequence itself. In particular, the peptide model employed by Merutka et al. has the sequence Gly-Gly-X-Gly-Gly, while the model used in the present study has the sequence Gly-Gly-X-Ala-Gly-Gly. It may be that the presence of a glycine (instead of alanine) on the C-terminal side of the X residue is sufficient to cause a  $\sim 0.2$  ppm downfield displacement in amide  $^1\text{H}$  chemical shifts. If this is the case, it would be a striking example of how nearest-neighbor interactions can significantly affect the chemical shifts of peptide backbone nuclei.

Given the obvious differences between the amide  $^1\text{H}$  chemical shifts reported by Wüthrich (1986), Merutka et al. (1995) and in this study, it is necessary to ask the question: Which set of chemical shifts is most representative of 'true random coil' values? Statistical studies of amide chemical shifts conducted by Gross and Kalbitzer (1988), Wishart et al. (1991b) and Wishart and Sykes (1994b) provide support for the validity of the measurements presented in Table 1. In all three of these studies, amide chemical shifts were found to be somewhat further upfield and significantly less dispersive than suggested by the data of Wüthrich or Merutka et al. In particular, average chemical shifts taken from Table 7 in Wishart et al. (1991b) give a 19 amino acid mean of 8.19 ppm, while extrapolated random coil values from Table 7 in Wishart and Sykes (1994b) give a mean of 8.27 ppm. Both of these values are closer to the average of 8.26 ppm (calculated from the present data in Table 1) than Wüthrich's mean of 8.35 ppm or Merutka's mean of 8.49 ppm. Similarly, the dispersion among statistically derived amide chemical shifts matched much more closely with the present data than with those of Wüthrich. In particular, the range among 'coil' values (from a minimum of 7.93 ppm to a maximum of 8.40 ppm) is nearly identical to the values found in Table 1. Overall, the correlation coefficient between the statistically derived amide 'coil' data (Wishart et al., 1991b) and the experimental data presented in Table 1 is  $R^2 = 0.43$ . This is substantially better than the correlation between Wüthrich's data (1986) and the present data ( $R^2 = 0.16$ ) and between Wüthrich's

experimental data and Wishart's (1991b) statistical 'coil' data ( $R^2 = 0.28$ ). Interestingly, the correlation coefficient between the amide  $^1\text{H}$  data in Table 1 and the amide  $^1\text{H}$  shifts reported by Merutka et al. is somewhat higher ( $R^2 = 0.61$ ). This indicates that, despite the systematic chemical shift difference of 0.23 ppm, these two data sets are actually quite similar.

While the most obvious differences in  $^1\text{H}$  chemical shifts tend to be found among the amide protons, it is also worth noting that small but significant differences exist between certain  $^1\text{H}^\alpha$  chemical shifts. This is particularly true with respect to the data reported by Wüthrich (1986). The most obvious  $^1\text{H}^\alpha$  chemical shift differences are seen for histidine (4.63 ppm versus 4.73 ppm), aspartic acid (4.76 ppm versus 4.64 ppm), glutamic acid (4.29 ppm versus 4.35 ppm), isoleucine (4.23 ppm versus 4.17 ppm) and valine (4.18 ppm versus 4.12 ppm). The 0.06–0.12 ppm discrepancy in the histidine, aspartate and glutamate chemical shifts is most likely attributable to the pH differences between the two data sets (7.0 versus 5.0). The slight differences in isoleucine and valine  $^1\text{H}^\alpha$  chemical shifts are somewhat more difficult to explain. Given the tendency of peptides containing hydrophobic amino acids like isoleucine and valine to aggregate, combined with their high propensity to form  $\beta$ -strands, it may be that at the concentrations used by Bundi and Wüthrich, these peptides were forming some kind of aggregated  $\beta$ -sheet structure. Should this have occurred, even for a small portion of the molecules, a slight downfield shift (Wishart et al., 1991b) would have been expected. We believe that our use of 1 M urea (along with lower peptide concentrations) eliminates the possibility of forming these peptide aggregates. Consequently, the  $^1\text{H}^\alpha$  chemical shifts we obtain for isoleucine and valine (and leucine) are seen to be shifted slightly upfield relative to Wüthrich's data. This same upfield trend is also found in the data reported by Merutka et al. (1995).

Apart from these anomalous residues, the agreement between the remaining  $^1\text{H}^\alpha$  chemical shifts for all three data sets is seen to be quite good. Differences are typically no more than  $\pm 0.03$  ppm and these may be due, in part, to the different choice of temperatures (35 °C versus 25 °C versus 5 °C) or the different choice of  $^1\text{H}$  standards (TSP versus DSS versus dioxane).

#### *Comparison with other data ( $^{13}\text{C}$ chemical shifts)*

Since 1978 at least three studies have been published providing experimentally measured random coil  $^{13}\text{C}$  chemical shift information for peptides in aqueous solution. The most complete of these was presented by Richarz and Wüthrich in 1978. This very comprehensive study provided complete  $^{13}\text{C}$  chemical shift information for all observable main-chain and side-chain carbons for each of the 20 common amino acids. All  $^{13}\text{C}$  chemical shifts were reported relative to internal dioxane (set at 67.8 ppm

relative to external TMS). A much more limited study undertaken by Spera and Bax (1991) reported the  $^{13}\text{C}$  chemical shifts for  $\alpha$ - and  $\beta$ -carbons of 19 (of the possible 20) amino acids. In this case, the chemical shifts were given relative to internal TSP. Most recently, Thanabal and co-workers (1994) published a comprehensive compilation of  $^{13}\text{C}$  random chemical shifts obtained from a series of protected linear pentapeptides. However, only the chemical shifts for carbons with attached protons were reported. All of Thanabal's  $^{13}\text{C}$  chemical shifts were referenced relative to internal dioxane (set at 69.4 ppm relative to external TSP). Two other statistical studies have provided estimates of random coil chemical shifts based on data collected from assigned peptides and proteins (Wishart et al., 1991b; Wishart and Sykes, 1994b). The latter study, which is based on a much larger data set, provides random coil chemical shifts for all 20 amino acids, referenced to internal DSS.

#### *$^{13}\text{C}^\alpha$ data*

In Table 2 we present complete  $^{13}\text{C}$  data (including the carbonyl and carboxyl carbon assignments) for all 20 common amino acids, referenced to internal DSS. Because there is considerable variation in the  $^{13}\text{C}$  chemical shifts of reference compounds such as TMS, TSP and DSS (Wishart and Sykes, 1994a; Wishart et al., 1995) it is essential to recalibrate data from the earlier studies in order to make direct comparisons. We find that, on average, the  $^{13}\text{C}^\alpha$  data from Richarz and Wüthrich lie 1.51 ppm upfield, while the values from Thanabal et al. lie 0.30 ppm upfield and those of Spera and Bax lie 0.03 ppm downfield from the present data. Based on previous calibrations of  $^{13}\text{C}$  chemical shift standards, we would have expected the data from Richarz and Wüthrich to lie 1.50 ppm upfield and those of Spera and Bax, and Thanabal et al. 0.17 ppm downfield from the present  $^{13}\text{C}$  chemical shift measurements. This indicates that the agreement between the observed and expected values is actually quite good.

Once the necessary corrections have been made, we find that all four studies provide  $^{13}\text{C}^\alpha$  chemical shifts which are highly correlated ( $R^2 > 0.99$ ). The results from a 'difference deviation' test indicate that the standard deviations of the chemical shift differences are 0.17 ppm between the data in Table 2 and Richarz and Wüthrich's values, 0.36 ppm for Wishart and Sykes's statistical data, 0.37 ppm for Spera and Bax's data, and 0.45 ppm for Thanabal's data. Evidently the lowest standard deviation, and hence the best agreement, is found with the (reference-corrected) measurements of Richarz and Wüthrich.

Among the different data sets, at least four residues exhibit significant variability in their  $^{13}\text{C}^\alpha$  chemical shifts. These are cysteine (cystine), aspartic acid, glutamic acid and proline. The differences are most pronounced when comparisons are made to the data of Thanabal et al.;

however, they are also evident comparing any of the other data sets. The variability observed for aspartic and glutamic acid is likely the result of pH effects, which is to be expected. The situation for proline and cysteine (cystine) may reflect the influence of as yet uncharacterized steric or electronic interactions.

### $^{13}\text{C}^\beta$ data

As with the  $^{13}\text{C}^\alpha$  data, we find that  $^{13}\text{C}^\beta$  chemical shift measurements for all four experimental studies are highly correlated ( $R^2 > 0.99$ ). However, this value must be interpreted cautiously. The inherent dispersion in  $^{13}\text{C}^\beta$  chemical shifts is very large (20–70 ppm) and consequently a general measure of colinearity, such as the Pearson correlation coefficient, will tend to diminish local disagreements between any two data sets. By applying the difference deviation test we used earlier, it is possible to more accurately identify the data set which best agrees with the data in Table 2. The results from such a test indicate the following: the standard deviation of differences (in ppm) between the data in Table 2 and Richarz and Wüthrich's values is 0.16, for Spera and Bax's data it is 0.51 and for Thanabal's data it is 0.71. Once again, the lowest standard deviation (and hence the best agreement) is found with the reference-corrected measurements of Richarz and Wüthrich. It is also worth noting that, while the standard deviation of the differences did not change significantly for the data from Richarz and Wüthrich (0.17 versus 0.16) or from Spera and Bax (0.37 versus 0.51), the difference deviation did change significantly for the data from

Thanabal et al. (0.45 versus 0.71). This appears to be largely a result of the significant deviations seen for the  $^{13}\text{C}^\beta$  chemical shifts of aspartic acid and glutamic acid within the Thanabal data set. If these two outliers are excluded, the measurements of Thanabal et al. actually agree quite well with the data presented in Table 2.

Another interesting feature of other published  $^{13}\text{C}^\beta$  chemical shifts is their systematic displacement relative to their  $^{13}\text{C}^\alpha$  shift values. Because of the different referencing systems employed by these workers, we would have expected the  $^{13}\text{C}^\beta$  chemical shifts from Richarz and Wüthrich to be shifted upfield by  $\sim 1.5$  ppm and those of Thanabal et al., and Spera and Bax downfield by  $\sim 0.2$  ppm relative to the data in Table 2. In fact, we find that the  $^{13}\text{C}^\beta$  shifts of Richarz and Wüthrich are, on average, 1.29 ppm (instead of 1.5 ppm) upfield while the  $^{13}\text{C}^\beta$  shifts of Spera and Bax, and Thanabal et al. are actually 0.36 and 0.13 ppm upfield, respectively, instead of 0.2 ppm downfield. This translates to a net downfield shift of  $\sim 0.2$  ppm for the Richarz and Wüthrich data and a net upfield shift of  $\sim 0.4$  ppm for the data of both Spera and Bax, and Thanabal et al. The reason for these systematic displacements is not clear. However, they may be the result of differences in the peptide sequences used and the influence of nearest-neighbor interactions (i.e., the effects of residues on the C-terminal side of the residue being measured).

As with  $^{13}\text{C}^\alpha$  chemical shift measurements, we find that a number of residues exhibit a high degree of  $^{13}\text{C}^\beta$  chemical shift variability from one study to the next. Among

TABLE 2  
RANDOM COIL  $^{13}\text{C}$  CHEMICAL SHIFTS FOR THE 20 COMMON AMINO ACIDS WHEN FOLLOWED BY ALANINE

Residue	C=O	C $^\alpha$	C $^\beta$	Others
Ala	177.8	52.5	19.1	
Cys (reduced)	174.6	58.2	28.0	
Cys (oxidized)	174.6	55.4	41.1	
Asp	176.3	54.2	41.1	$\gamma\text{CO}$ 180.0
Glu	176.6	56.6	29.9	$\gamma\text{CH}_2$ 35.6; $\delta\text{CO}$ 183.4
Phe	175.8	57.7	39.6	1C 138.9; 2,6CH 131.9; 3,5CH 131.5; 4CH 129.9
Gly	174.9	45.1		
His	174.1	55.0	29.0	2CH 136.2; 4CH 120.1; 5C 131.1
Ile	176.4	61.1	38.8	$\gamma\text{CH}_2$ 27.2; $\gamma\text{CH}_3$ 17.4; $\delta\text{CH}_3$ 12.9
Lys	176.6	56.2	33.1	$\gamma\text{CH}_2$ 24.7; $\delta\text{CH}_2$ 29.0; $\epsilon\text{CH}_3$ 41.9
Leu	177.6	55.1	42.4	$\gamma\text{CH}$ 26.9; $\delta\text{CH}_3$ 24.9, 23.3
Met	176.3	55.4	32.9	$\gamma\text{CH}_2$ 32.0; $\epsilon\text{CH}_3$ 16.9
Asn	175.2	53.1	38.9	$\gamma\text{CO}$ 177.2
Pro	177.3	63.3	32.1	$\gamma\text{CH}_2$ 27.2; $\delta\text{CH}_2$ 49.8
Gln	176.0	55.7	29.4	$\gamma\text{CH}_2$ 33.7; $\delta\text{CO}$ 180.5
Arg	176.3	56.0	30.9	$\gamma\text{CH}_2$ 27.1; $\delta\text{CH}_2$ 43.3; $\epsilon\text{C}$ 159.5
Ser	174.6	58.3	63.8	
Thr	174.7	61.8	69.8	$\gamma\text{CH}_3$ 21.5
Val	176.3	62.2	32.9	$\gamma\text{CH}_3$ 21.1, 20.3
Trp <sup>a</sup>	176.1	57.5	29.6	2CH 127.4; 3C 111.2; 4CH 122.2; 5CH 124.8; 6CH 121.0; 7CH 114.7; 8C 138.7; 9C 129.5
Tyr	175.9	57.9	38.8	1C 130.6; 2,6CH 133.3; 3,5CH 118.2; 4C 157.3

Chemical shifts are referenced to internal DSS at 25 °C, pH  $\sim 5.0$ .

<sup>a</sup> Measured using a peptide with free N- and C-termini.



the most dispersive residues are aspartic acid, glutamic acid, cysteine (cystine) and histidine. As might be expected, the differences in aspartic acid, glutamic acid and histidine are likely pH dependent, since the  $^{13}\text{C}^{\beta}$  resonance is known to be quite sensitive to the charge of an amino acid side chain. Had conditions been more similar between the five studies, it would have been likely that these differences would be much reduced. While the variability in these charged residues is significant, perhaps the most striking difference lies in the chemical shift of the  $\beta$ -carbon of cysteine. Not only is this resonance extremely sensitive to its oxidation state (oxidized: 41.1 ppm; reduced: 28.0 ppm), it is also quite dependent on local steric interactions or end-group effects. Consequently, one must be quite careful in distinguishing between oxidized and reduced species when defining random coil chemical shifts for cysteine (cystine).

#### *$^{13}\text{C}$ carbonyl data*

The  $^{13}\text{C}$  data of Richarz and Wüthrich (1978) provide the only set of experimental carbonyl  $^{13}\text{C}$  chemical shift measurements with which direct comparisons to Table 2 can be made. However, there are two statistical studies (Wishart et al., 1991b; Wishart and Sykes, 1994b) which provide estimates of random coil  $^{13}\text{C}$  carbonyl chemical shifts and these may also be used to assess the measurements given in Table 2. On average, we find that the data of Richarz and Wüthrich are 2.44 ppm upfield compared to the values in Table 2, while the statistical data of Wishart and Sykes are essentially unshifted (only 0.12 ppm downfield) – as expected. Based on their choice of reference compound, Richarz and Wüthrich's values would be predicted to be  $\sim 1.5$  ppm upfield of the values in Table 2, so the observed difference of 2.44 ppm actually represents a systematic upfield shift of  $\sim 0.9$  ppm over the expected shift. Given that we had previously identified a systematic downfield shift ( $\sim 0.4$  ppm) in the  $\beta$ -carbons of the Richarz and Wüthrich data, it seems likely that the  $^{13}\text{C}$  frequency scale used by Richarz and Wüthrich must be somewhat different from the one adopted for our measurements.

Overall, the agreement between the three data sets is not as high as it was for the aliphatic ( $\alpha$  and  $\beta$ ) carbons. Indeed, the correlation coefficient between the carbonyl  $^{13}\text{C}$  data of Richarz and Wüthrich and those of Table 2 is only  $R^2 = 0.74$ , with a difference deviation of 0.53 ppm. Similarly, a comparison of Table 2 to the statistical data of Wishart and Sykes yields a correlation coefficient of 0.75 with a difference deviation of 0.52 ppm. The lower level of correlation and the higher standard deviations for carbonyl carbons are likely the result of end-group and local charge effects, since these carbons seem to be much more sensitive to environmental effects (solvent, H-bonds, local charge density, etc.) than  $\alpha$ - or  $\beta$ -carbons. Because we have undertaken measures to eliminate end-group and

charge effects as much as possible, we believe that the data in Table 2 probably represent a more realistic tabulation of random coil  $^{13}\text{C}$  carbonyl chemical shifts than those of Richarz and Wüthrich.

Just as with the  $^{13}\text{C}^{\alpha}$  and  $^{13}\text{C}^{\beta}$  random coil chemical shift values, there are a number of residues which display considerable variability in their  $^{13}\text{C}$  carbonyl chemical shifts between one data set and the next. In particular, we find large chemical shift deviations for glycine, proline, aspartic acid and cystine. Once again it is reasonable to invoke such explanations as pH effects (for aspartic acid) and nearest-neighbor influences (for glycine, cystine and proline). The situation for glycine, however, is somewhat puzzling because this residue has not previously been found to be prone to a high level of chemical shift variability. The reason for its exaggerated downfield shift (approximately 1.2 ppm relative to both the Richarz and Wüthrich, and Wishart and Sykes data) is not entirely clear and an answer may have to await further experimental work with other substituted peptides.

#### *Comparison with other data ( $^{15}\text{N}$ chemical shifts)*

Only one published set of experimental random coil  $^{15}\text{N}$  chemical shifts is available for comparison with the data presented in Table 3. However, the backbone  $^{15}\text{N}$  measurements reported by Glushka et al. (1989,1990) actually represent extrapolated random coil shifts taken from measurements of individual N-acetylated amino acids in DMSO. Consequently, they might best be regarded as theoretical random coil  $^{15}\text{N}$  chemical shifts. Two statistical studies (Wishart et al., 1991b; Wishart and Sykes, 1994b) also have been published, in which backbone  $^{15}\text{N}$  chemical shifts from a number of previously assigned proteins were tabulated and random coil chemical shifts extracted from these data. In order to compare these two previously published data sets with the data in Table 3, we chose to use the corrected  $^{15}\text{N}$  chemical shifts of Glushka et al. (1990) from their 'Gly, Ala' column, the  $^{15}\text{N}$  chemical shifts of Wishart et al. (1991b) from their 'coil' column, and the extrapolated  $^{15}\text{N}$  random coil estimates of Wishart and Sykes (1994b).

Overall, the agreement between the three data sets is reasonably good. For instance, the correlation coefficient between the  $^{15}\text{N}$  data of Glushka et al. and those of Table 3 is  $R^2 = 0.84$ , with a difference deviation of 1.72 ppm. Similarly, a comparison of Table 3 to the statistical 'coil' data of Wishart et al. yields a correlation coefficient of 0.85 and a difference deviation of 1.36 ppm. Likewise, the random coil estimates of Wishart and Sykes (1994b) yield a correlation coefficient of 0.91 and a difference deviation of 1.04 ppm. On average, the shifts of Glushka et al. are found to be 1.12 ppm downfield of those in Table 3, while the shifts of Wishart et al. are only 0.31 ppm downfield. Based on the smaller difference deviation and the smaller downfield displacement, it appears that the statis-

tical data of Wishart et al. are most similar to the backbone  $^{15}\text{N}$  chemical shifts presented in Table 3. It is also worth noting that, even though the correlation coefficients between the four data sets are generally high ( $> 0.85$ ), there is much less chemical shift precision (i.e., the standard deviations are much higher) among  $^{15}\text{N}$  chemical shifts than among  $^{13}\text{C}$  or  $^1\text{H}$  chemical shifts. This is likely due to the influence of external variables such as solvent, H-bond effects and nearest-neighbor interactions, which seem to play a more significant role in determining  $^{15}\text{N}$  chemical shifts than  $^{13}\text{C}$  and  $^1\text{H}$  chemical shifts.

Because of the many problems of properly referencing  $^{15}\text{N}$  chemical shifts in the past, we were quite surprised to find that the global average of  $^{15}\text{N}$  chemical shifts from assigned proteins would be so close to the random coil values measured in this study (within 0.3 ppm). This unexpectedly high concordance may be the result of the averaging process which, evidently, cancelled out the effects of chemical shifts which were referenced either too far upfield or too far downfield. Regardless of the specific reasons, this result clearly shows that our referencing procedure for random coil  $^{15}\text{N}$  measurements is in good agreement with the bulk of published  $^{15}\text{N}$  chemical shift assignments.

It is also worth noting that the random coil values provided by Glushka et al. are approximately 1.1 ppm downfield of the values in Table 3 and approximately 0.8

TABLE 3  
RANDOM COIL  $^{15}\text{N}$  AND  $^1\text{H}$  AMIDE CHEMICAL SHIFTS  
FOR THE 20 COMMON AMINO ACIDS WHEN FOLLOWED  
BY ALANINE

Residue	$^{15}\text{N}$ chemical shift		$^1\text{H}$ chemical shift	
	NH	Others	HN	Others
Ala	123.8		8.24	
Cys (reduced)	118.8		8.32	
Cys (oxidized)	118.6		8.43	
Asp	120.4		8.34	
Glu	120.2		8.42	
Phe	120.3		8.30	
Gly	108.8		8.33	
His	118.2		8.42	
Ile	119.9		8.00	
Lys	120.4	$\epsilon\text{NH}_3$ 125.9	8.29	$\epsilon\text{NH}_3$ 7.81
Leu	121.8		8.16	
Met	119.6		8.28	
Asn	118.7	$\gamma\text{NH}_2$ 112.7	8.40	$\gamma\text{NH}_2$ 6.91, 7.59
Pro	—		—	
Gln	119.8	$\delta\text{NH}_2$ 112.1	8.32	$\delta\text{NH}_2$ 6.85, 7.52
Arg	120.5		8.23	
Ser	115.7		8.31	
Thr	113.6		8.15	
Val	119.2		8.03	
Trp <sup>a</sup>	121.3		8.25	
Tyr	120.3		8.12	

<sup>1</sup>H chemical shifts are referenced to internal DSS at 25 °C, pH ~5.0.  $^{15}\text{N}$  chemical shifts are indirectly referenced to external  $\text{NH}_3$  at 25 °C.

<sup>a</sup> Measured using a peptide with free N- and C-termini.

ppm downfield of the statistical data of Wishart et al. Evidently, the indirect referencing procedure adopted by Glushka et al. can lead to  $^{15}\text{N}$  chemical shift estimates which are significantly further downfield than those determined by the direct referencing procedure used in this study.

Several residues display considerable variability in their backbone  $^{15}\text{N}$  chemical shifts between each of the three data sets. Depending on which data set is compared, significant differences can be found for histidine, lysine, arginine, asparagine, glutamine and tyrosine (Glushka et al., 1990), as well as for alanine, isoleucine, methionine, valine and tyrosine (Wishart et al., 1991b; Wishart and Sykes, 1994b). This chemical shift variability could arise from any number of experimental or theoretical sources. For instance, the correction factors employed by Glushka et al. seem to be insufficient for amino acids containing amines, amides or hydroxyl groups in their side chain. Evidently, the influence of DMSO on the charge or conformation of these side chains must have had some effect on the measurements of their chemical shifts. Similarly, the relatively small sample size used in the statistical study of Wishart et al. could lead to some spurious values for certain rare residues, such as methionine and tyrosine. The possible influence of nearest-neighbor interactions, which was not accounted for in the statistical study of Wishart et al., could contribute to additional discrepancies, particularly for bulky residues such as valine, isoleucine and tyrosine. It is unfortunate that no other experimental studies are available to confirm the veracity of the  $^{15}\text{N}$  chemical shifts presented in Table 3. However, the generally good agreement with the  $^{15}\text{N}$  statistical data of Wishart et al. gives us every reason to believe that these experimentally measured chemical shift values are very close to the 'true random coil'  $^{15}\text{N}$  values.

#### Results from Gly-Gly-X-Pro-Gly-Gly

We believe that the data presented for the Gly-Gly-X-Ala-Gly-Gly series of peptides (Tables 1–3) provide a very complete summary of 'random coil' chemical shifts of the 20 common amino acids. While this kind of data is generally sufficient for most applications in biomolecular NMR, there is an increasing desire among theoreticians to understand chemical shift displacements at a finer level of detail. In particular, with the advent of chemical shift/structural refinement, there is now an urgent need to understand the influence of sequence effects or nearest-neighbor interactions on chemical shifts. To gain a better insight into these effects on amino acid chemical shifts, we chose to study a series of 20 hexapeptides where alanine (in the fourth position) was replaced by proline. The  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts for the 20 common amino acids in this series of peptides are shown in Tables 4–6. A quick comparison with Tables 1–3 reveals the tremendous effect that proline can have on amino acid

TABLE 4  
RANDOM COIL.  $^1\text{H}$  CHEMICAL SHIFTS FOR THE 20 COMMON AMINO ACIDS WHEN FOLLOWED BY PROLINE

Residue	NH	H <sup>a</sup>	H <sup>b</sup>	Others
Ala	8.19	4.62	1.35	
Cys (reduced)	8.30	4.81	2.93, 2.85	
Asp	8.31	4.90	2.72, 2.57	
Glu	8.34	4.64	2.06, 1.90	$\gamma\text{CH}_2$ 2.31, 2.31
Phe	8.13	4.90	3.14, 2.97	2,6H 7.29; 3,5H 7.38; 4H 7.32
Gly	8.21	4.13		
His	8.37	5.00	3.23, 3.12	2H 8.57; 4H 7.29
Ile	8.06	4.47	1.85	$\gamma\text{CH}_2$ 1.48, 1.16; $\gamma\text{CH}_3$ 0.94; $\delta\text{CH}_3$ 0.86
Lys	8.18	4.60	1.80, 1.74	$\gamma\text{CH}_2$ 1.45, 1.45; $\delta\text{CH}_2$ 1.69, 1.69; $\epsilon\text{CH}_2$ 2.97, 2.97; $\epsilon\text{NH}_2$ 7.81
Leu	8.14	4.63	1.64, 1.64	$\gamma\text{CH}$ 1.57; $\delta\text{CH}_3$ 0.93, 0.93
Met	8.25	4.82	2.04, 1.97	$\gamma\text{CH}_2$ 2.63, 2.56; $\epsilon\text{CH}_3$ 2.12
Asn	8.37	5.00	2.82, 2.67	$\gamma\text{NH}_2$ 7.60, 6.92
Pro	—	4.73	2.31, 1.91	$\gamma\text{CH}_2$ 2.01, 2.01; $\delta\text{CH}_2$ 3.60, 3.60
Gln	8.29	4.65	2.09, 1.93	$\gamma\text{CH}_2$ 2.38, 2.38; $\delta\text{NH}_2$ 7.53, 6.88
Arg	8.20	4.65	1.81, 1.81	$\gamma\text{CH}_2$ 1.67, 1.67; $\delta\text{CH}_2$ 3.21, 3.21; $\epsilon\text{NH}$ 8.07
Ser	8.26	4.78	3.85, 3.85	
Thr	8.15	4.61	4.11	$\gamma\text{CH}_2$ 1.23
Val	8.02	4.44	2.06	$\gamma\text{CH}_3$ 0.97, 0.92
Trp	8.09	4.99	3.32, 3.16	2H 7.25; 4H 7.69; 5H 7.19; 6H 7.27; 7H 7.51
Tyr	8.10	4.84	3.05, 2.89	2,6H 7.15; 3,5H 6.86

Chemical shifts are referenced to internal DSS at 25 °C, pH ~5.0.

chemical shifts when placed on the C-terminal side of a given residue. This was not entirely unexpected, since previous workers had noticed similar effects, albeit on a more limited scale (Torchia et al., 1975; Howarth, 1978). What is significant about these results is that they help to confirm the previous assumption that proline influences not only the main-chain  $^{13}\text{C}$  but also the  $^{15}\text{N}$  and  $^1\text{H}$  chemical shifts of all 20 common amino acids. This result

is important, because it finally places the importance of nearest-neighbor interactions on a solid experimental footing.

In Table 7 we summarize the influence of proline on the main-chain chemical shifts ( $^{15}\text{N}$ ,  $^{13}\text{C}$  and  $^1\text{H}$ ) for each of the common amino acids. To prepare this summary, we subtracted the chemical shift of each residue in the X-Pro series from the same residue in the X-Ala series. On

TABLE 5  
RANDOM COIL.  $^{13}\text{C}$  CHEMICAL SHIFTS FOR THE 20 COMMON AMINO ACIDS WHEN FOLLOWED BY PROLINE

Residue	C=O	C <sup>a</sup>	C <sup>b</sup>	Others
Ala	175.9	50.5	18.1	
Cys (reduced)	173.0	56.4	27.1	
Asp	175.0	52.2	40.9	$\gamma\text{CO}$ 179.9
Glu	174.9	54.2	29.2	$\gamma\text{CH}_2$ 35.0; $\delta\text{CO}$ 183.0
Phe	174.4	55.6	39.1	1C 138.9; 2,6CH 132.1; 3,5CH 131.4; 4CH 129.9
Gly	174.5	44.5		
His	172.6	53.3	29.0	2CH 136.3; 4CH 120.3; 5C 131.2
Ile	175.0	58.7	38.7	$\gamma\text{CH}_2$ 26.9; $\gamma\text{CH}_3$ 17.1; $\delta\text{CH}_3$ 12.7
Lys	174.8	54.2	32.6	$\gamma\text{CH}_2$ 24.6; $\delta\text{CH}_2$ 29.1; $\epsilon\text{CH}_2$ 41.9
Leu	175.7	53.1	41.7	$\gamma\text{CH}$ 27.1; $\delta\text{CH}_3$ 25.1, 23.3
Met	174.6	53.3	32.4	$\gamma\text{CH}_2$ 32.0; $\epsilon\text{CH}_3$ 17.0
Asn	173.6	51.3	38.7	$\gamma\text{CO}$ 177.1
Pro	171.4	61.5	30.9	$\gamma\text{CH}_2$ 27.2; $\delta\text{CH}_2$ 49.7
Gln	174.4	53.7	28.8	$\gamma\text{CH}_2$ 33.4; $\delta\text{CO}$ 180.5
Arg	174.5	54.0	30.2	$\gamma\text{CH}_2$ 26.8; $\delta\text{CH}_2$ 43.4; $\epsilon\text{C}$ 159.6
Ser	173.1	56.4	63.3	
Thr	173.2	59.8	69.8	$\gamma\text{CH}_2$ 21.4
Val	174.9	59.8	32.6	$\gamma\text{CH}_3$ 20.9, 20.1
Trp	174.8	55.7	28.9	2CH 127.3; 3C 111.3; 4CH 122.2; 5CH 124.8; 6CH 120.9; 7CH 114.7; 8C 138.8; 9C 129.6
Tyr	174.8	55.8	38.3	1C 130.7; 2,6CH 133.5; 3,5CH 118.2; 4C 157.3

Chemical shifts are referenced to internal DSS at 25 °C, pH ~5.0.

average, we find that when an amino acid precedes a proline, its  $^1\text{H}^\alpha$  resonance is shifted downfield by 0.29 ppm, its amide  $^1\text{H}$  resonance is shifted upfield by 0.05 ppm, its  $^{13}\text{C}^\alpha$  resonance is shifted upfield by 1.95 ppm, its carbonyl  $^{13}\text{C}$  resonance is shifted upfield by 1.72 ppm and its  $^{15}\text{N}$  resonance is shifted downfield by 0.97 ppm. These averaged proline-induced chemical shifts are plotted in Fig. 1, along with their associated standard deviations. It is interesting to note that nearly all amino acids are shifted approximately equally upfield or downfield (depending on the nucleus), except for glycine. On average, we find that when glycine is adjacent to proline it is only shifted by approximately half as much as the other residues. This is likely due to its unique structure (i.e., the lack of a side chain) and its concomitant lack of steric inhibition by proline. While glycine tends to be under-affected, proline is significantly over-affected. Inspection of Table 7 reveals that its  $^{13}\text{C}$  carbonyl chemical shift is displaced upfield by nearly four times (5.9 ppm) the average value of the other residues. This may reflect the very tight steric limitations and unique peptide geometry which develops when two prolines are placed side by side.

The chemical shift trends which are observed for residues placed next to proline are similar to those observed for residues in  $\beta$ -sheet conformations (Wishart et al., 1991b; Wishart and Sykes, 1994b). In fact, when residues are in a  $\beta$ -sheet, the  $^1\text{H}^\alpha$  resonances are typically shifted downfield by 0.38 ppm, the  $^{13}\text{C}^\alpha$  and carbonyl  $^{13}\text{C}$  resonances are shifted upfield by 1.4 ppm and the  $^{15}\text{N}$  reson-

TABLE 6  
RANDOM COIL  $^{15}\text{N}$  AND  $^1\text{H}$  AMIDE CHEMICAL SHIFTS FOR THE 20 COMMON AMINO ACIDS WHEN FOLLOWED BY PROLINE

Residue	$^{15}\text{N}$ chemical shift		$^1\text{H}$ chemical shift	
	NH	Others	HN	Others
Ala	125.0		8.19	
Cys (reduced)	119.9		8.30	
Asp	121.4		8.31	
Glu	121.7		8.34	
Phe	120.9		8.13	
Gly	109.1		8.21	
His	118.2		8.37	
Ile	121.7		8.06	
Lys	121.6	$\epsilon\text{NH}_2$ 125.9	8.18	$\epsilon\text{NH}_2$ 7.81
Leu	122.6		8.14	
Met	120.7		8.25	
Asn	119.0	$\gamma\text{NH}_2$ 112.8	8.37	$\gamma\text{NH}_2$ 6.92, 7.60
Pro	—		—	
Gln	120.6	$\delta\text{NH}_2$ 112.1	8.29	$\delta\text{NH}_2$ 6.88, 7.53
Arg	121.3		8.20	
Ser	116.6		8.26	
Thr	116.0		8.15	
Val	120.5		8.02	
Trp	122.2		8.09	
Tyr	120.8		8.10	

$^1\text{H}$  chemical shifts are referenced to internal DSS at 25 °C, pH ~5.0.

$^{15}\text{N}$  chemical shifts are indirectly referenced to external  $\text{NH}_3$  at 25 °C.

TABLE 7  
EFFECTS OF PROLINE ON THE CHEMICAL SHIFTS OF PEPTIDE BACKBONE NUCLEI

Residue	$\text{H}^\alpha$	HN	$\text{C}^\alpha$	C=O	NH
Ala	-0.30	0.05	2.0	1.9	-1.2
Cys (reduced)	-0.26	0.02	1.8	1.6	-1.1
Asp	-0.26	0.03	2.0	1.3	-1.0
Glu	-0.29	0.08	2.4	1.7	-1.5
Phe	-0.28	0.17	2.1	1.4	-0.6
Gly	-0.17	0.12	0.6	0.4	-0.3
His	-0.27	0.05	1.7	1.5	0.0
Ile	-0.30	-0.06	2.4	1.4	-1.8
Lys	-0.28	0.11	2.0	1.8	-1.2
Leu	-0.29	0.02	2.0	1.9	-0.8
Met	-0.34	0.03	2.1	1.7	-1.1
Asn	-0.26	0.03	1.8	1.6	-0.3
Pro	-0.31	—	1.8	5.9	—
Gln	-0.31	0.03	2.0	1.6	-0.8
Arg	-0.31	0.03	2.0	1.8	-0.8
Ser	-0.31	0.05	1.9	1.5	-0.9
Thr	-0.26	0.00	2.0	1.5	-2.4
Val	-0.32	0.01	2.4	1.4	-1.3
Trp	-0.33	0.16	1.8	1.3	-0.9
Tyr	-0.29	0.02	2.1	1.1	-0.5
Average	-0.287	0.050	1.95	1.72	-0.97
Standard deviation	0.036	0.056	0.37	1.04	0.56

Chemical shift differences are calculated as  $\delta X_A - \delta X_P$ , where  $\delta X_A$  is the chemical shift of residue X in Gly-Gly-X-Ala-Gly-Gly and  $\delta X_P$  is the chemical shift of residue X in Gly-Gly-X-Pro-Gly-Gly.

ances are shifted downfield by 1.2 ppm. These shifts are quite close to the chemical shift changes observed for our proline-substituted peptides (see above), suggesting that residues which immediately precede a proline will likely have a time-averaged conformation similar to a  $\beta$ -sheet. This may explain why proline residues have a statistically high propensity to be at the end of  $\beta$ -sheets (Wishart, 1991) or to be in the  $i + 1$  position of type I and type II  $\beta$ -turns (Wilmot and Thornton, 1988). In essence, proline may act as both a  $\beta$ -sheet initiator (in the C- to N-direction) and a  $\beta$ -sheet terminator (in the N- to C-direction). While this is an intriguing possibility, it is important to remember that the influence of proline can be manifested not only through steric factors, but through chemical factors as well. In particular, an imide bond formed by an X-Pro pairing is generally thought to be much less electron-withdrawing than an amide bond. Consequently, a carbonyl carbon attached to an imide nitrogen would be expected to be somewhat more shielded than one attached to an amide nitrogen and thus subject to a noticeable upfield shift. Hence, a chemical reason such as imide shielding could be invoked to explain at least some portion of the observed chemical shift trends among proline-containing peptides. Clearly, more experimental information will have to be collected and, perhaps, additional calculations will need to be performed to distinguish which process (chemical or steric) is more important to these nearest-neighbor interactions.

### Nearest-neighbor effects (preceding residues)

While the data presented so far provide important information on how the chemical shift of a residue in the  $i$  position can be affected by the type of residue in the  $i + 1$  position, we have not yet looked at the reverse effect (i.e., how the type of residue in the  $i$  position can affect the chemical shift of the residue in the  $i - 1$  position). Fortunately, data collected from the X-Ala and X-Pro series of peptides actually allow us to determine this reverse effect for at least two residues: alanine and proline. Table 8 lists the main-chain  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts of alanine when preceded by each of the 20 naturally occurring amino acids. Table 9 lists the main-chain  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts of proline when preceded by each of the 20 naturally occurring amino acids.

Inspection of Table 8 reveals the presence of several interesting nearest-neighbor effects. One particularly obvious trend is seen when aromatic amino acids (tryptophan, phenylalanine and tyrosine) are located immediately before alanine (on the N-terminal side). Typically, the alanine  $^1\text{H}^\alpha$ , amide  $^1\text{H}$  and  $^{13}\text{C}^\alpha$  chemical shifts of these Trp-Ala, Phe-Ala and Tyr-Ala pairs are displaced slightly upfield relative to their usual random coil values (Tables 1-3). This chemical shift displacement is most evident for tryptophan, while it is a little less obvious for tyrosine and phenylalanine. The upfield shift is most likely the result of time-averaged ring current effects. However, contributions from steric effects may also be significant. On average,  $^{13}\text{C}$  carbonyl resonances appear to be largely unaffected by the identity of the preceding residue, although some interesting shifts are brought about by the presence of tyrosine or tryptophan. On the other hand, amide  $^1\text{H}$  and amide  $^{15}\text{N}$  chemical shifts seem to display considerable susceptibility to nearest-neighbor effects. In particular, amide  $^1\text{H}$  chemical shifts are seen to range between 8.26-8.51 ppm, while  $^{15}\text{N}$  chemical shifts extend from 123.2 to 128.2 ppm. These ranges are almost an

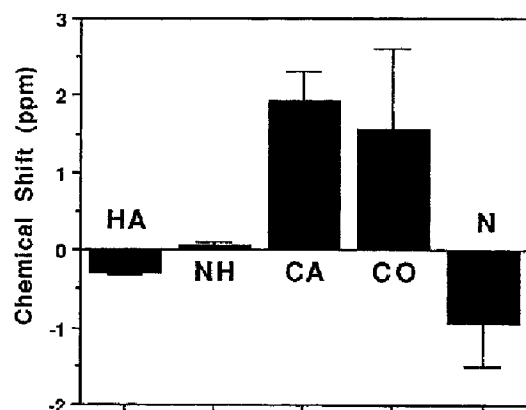


Fig. 1. Graph indicating the proline-induced chemical shift for  $^1\text{H}^\alpha$ , amide  $^1\text{H}$ ,  $^{13}\text{C}^\alpha$ , carbonyl  $^{13}\text{C}$  and  $^{15}\text{N}$  amide nuclei as averaged over all 20 naturally occurring amino acids. The standard deviation is indicated by an error bar.

TABLE 8  
RANDOM COIL CHEMICAL SHIFTS FOR ALANINE WHEN  
PRECEDED BY ONE OF THE 20 COMMON AMINO ACIDS

Preceding residue	$\text{H}^\alpha$	HN	$\text{C}^\alpha$	C=O	NH
Ala	4.33	8.34	52.6	178.5	123.2
Cys (reduced)	4.35	8.51	52.8	178.3	126.7
Cys (oxidized)	4.35	8.50	52.9	178.2	126.1
Asp	4.33	8.38	52.9	178.5	124.8
Glu	4.34	8.44	52.7	178.5	125.2
Phe	4.31	8.38	52.5	178.0	126.4
Gly	4.35	8.30	52.7	178.5	124.0
His	4.37	8.47	52.7	178.2	125.8
Ile	4.34	8.47	52.7	178.3	128.2
Lys	4.34	8.42	52.6	178.3	125.6
Leu	4.32	8.36	52.7	178.4	125.0
Met	4.34	8.40	52.7	178.3	125.1
Asn	4.34	8.38	52.8	178.3	124.7
Pro	4.33	8.50	52.7	178.6	124.4
Gln	4.34	8.44	52.7	178.4	125.3
Arg	4.35	8.44	52.6	178.3	125.4
Ser	4.37	8.42	52.7	178.3	125.9
Thr	4.36	8.43	52.7	178.3	126.4
Val	4.34	8.48	52.7	178.4	127.9
Trp	4.23	8.26	52.3	179.0	126.8
Tyr	4.31	8.35	52.4	177.9	126.8

$^1\text{H}$  and  $^{13}\text{C}$  chemical shifts are directly referenced to internal DSS at 25 °C, pH ~5.0.  $^{15}\text{N}$  chemical shifts are indirectly referenced to external  $\text{NH}_3\text{D}_3$  at 25 °C.

order of magnitude larger than those seen for  $^1\text{H}^\alpha$  and  $^{13}\text{C}$  chemical shifts.

The wide variability in amide  $^{15}\text{N}$  chemical shifts brought on by residues in the  $i - 1$  position had been observed earlier by Kricheldorf (1981) and led others (Glushka et al., 1989) to attempt to calculate neighboring residue 'correction factors'. Based on the present data and those of Kricheldorf (1981), it appears that these correction factors must depend on both the bulkiness and the hydrophobicity of the preceding residue.  $\beta$ -Branched hydrophobic amino acids such as isoleucine and valine in the  $i - 1$  position tend to produce the most significant downfield  $^{15}\text{N}$  shift (~4 ppm), followed by aromatic amino acids such as phenylalanine, tryptophan and tyrosine (~3 ppm). Serine, threonine, cysteine and histidine form a third group of bulky, slightly polar amino acids which generally produce a slightly smaller downfield shift (~2 ppm). The remaining unbranched or long-chain amino acids (leucine, methionine, lysine, etc.) comprise a fourth group, leading to an even smaller downfield shift (~1 ppm). Proline, glycine and alanine represent a fifth group which seems to be mostly neutral (~0 ppm). Such a ranking by amino acid type does not seem possible with amide  $^1\text{H}$  chemical shifts; however, additional data from other peptide constructs may eventually reveal an identifiable trend.

The fact that  $^{15}\text{N}$  amide chemical shifts are so sensitive to nearest-neighbor effects may have been one of the main reasons why very little structural or sequential information has been derived from them in the past. Evident-

TABLE 9  
RANDOM COIL CHEMICAL SHIFTS FOR PROLINE WHEN  
PRECEDED BY ONE OF THE 20 COMMON AMINO ACIDS

Preceding residue	H <sup>α</sup>	C <sup>α</sup>	Preceding residue	H <sup>α</sup>	C <sup>α</sup>
Ala	4.44	63.5	Met	4.44	63.6
Cys (reduced)	4.45	63.8	Asn	4.44	63.8
Asp	4.43	63.9	Pro	4.44	63.5
Glu	4.44	63.7	Gln	4.44	63.6
Phe	4.44	63.7	Arg	4.45	63.5
Gly	4.45	63.7	Ser	4.46	63.8
Ile	4.47	63.7	Thr	4.41	63.8
Ile	4.41	63.8	Val	4.42	63.7
Lys	4.41	63.6	Trp	4.42	63.8
Leu	4.43	63.6	Tyr	4.43	63.8

<sup>1</sup>H and <sup>13</sup>C chemical shifts are directly referenced to internal DSS at 25 °C, pH ~5.0.

ly, these sequence-specific interactions are so large that they tend to hide most of the secondary chemical shifts that might be induced by important structural changes. This suggests that, if one were to attempt to develop a method to derive structural information from <sup>15</sup>N chemical shifts (as has been done for <sup>1</sup>H<sup>α</sup> and <sup>13</sup>C chemical shifts), it would be critical to include neighboring residue corrections. These nearest-neighbor effects on <sup>15</sup>N chemical shifts could be exploited in other ways as well. The fact that the residue type in the *i* - 1 position influences the <sup>15</sup>N chemical shift in the *i* position implies that one could identify and classify unique pairs of amino acids on the basis of their <sup>15</sup>N chemical shifts alone. This could potentially allow one to perform a sequential assignment of short (mostly unstructured) peptides, or even denatured proteins, without the need of NOE information. In addition to the possibility of NOE-independent sequential assignments, evidence has been presented by Glushka et al. (1989) that <sup>15</sup>N chemical shifts could be used to provide detailed information about backbone ψ angles. According to these workers, the <sup>15</sup>N chemical shift of residue *i* correlates quite well with the ψ angle of residue *i* - 1. Evidently, by incorporating more detailed residue-specific corrections, this correlation could be much improved. The possibility that <sup>15</sup>N chemical shifts could be used to provide ψ angle information, combined with the demonstrated potential for <sup>1</sup>H<sup>α</sup> and <sup>13</sup>C<sup>α</sup> data to provide φ angle information (Spera and Bax, 1991; Wishart et al., 1991b) suggests that <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N chemical shifts could be used to determine peptide and protein backbone conformations.

While the results summarized in Table 8 have provided a wealth of information on nearest-neighbor interactions, the same cannot be said for the data in Table 9. Evidently, proline is generally less affected than alanine by residues which precede it. Whether this is a unique property of proline or a unique property of alanine is unclear. An answer will likely have to await further experiments with other residue substitutions.

## Conclusions

In this report, we have presented <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N chemical shift assignments for amino acids in two hexapeptide constructs: Gly-Gly-X-Ala-Gly-Gly and Gly-Gly-X-Pro-Gly-Gly (where X is any of the 20 naturally occurring amino acids). We believe that the chemical shifts reported for Gly-Gly-X-Ala-Gly-Gly (presented in Tables 1-3) represent the most complete set of random coil chemical shift measurements published to date. Comparisons of these experimentally measured random coil chemical shifts with other reported values (Richarz and Wüthrich, 1978; Wüthrich, 1986; Glushka et al., 1990; Spera and Bax, 1991; Merutka et al., 1995; Thanabal et al., 1994) indicate a very high level of agreement for <sup>1</sup>H<sup>α</sup>, <sup>13</sup>C<sup>α</sup> and <sup>13</sup>C<sup>β</sup> chemical shifts and an acceptably high agreement for <sup>15</sup>N and carbonyl <sup>13</sup>C chemical shifts. Almost no statistically significant agreement is found between the amide <sup>1</sup>H chemical shifts reported here and the amide <sup>1</sup>H chemical shifts reported by Wüthrich (1986). Comparisons with statistically derived 'coil' chemical shifts, as determined from previously assigned proteins (Wishart et al., 1991b; Wishart and Sykes, 1994b), indicate a very good level of agreement for all of the random coil <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N chemical shifts reported here. In general, the chemical shifts reported in Tables 1-3 agree better with the statistically derived chemical shifts than any other single experimental data set. As a result, we believe that the random coil chemical shifts reported here are very good approximations to 'ideal' or 'true' random coil values.

In addition to the work on random coil measurements, we have also demonstrated the importance of nearest-neighbor effects in determining amino acid chemical shifts. Comparisons between the chemical shift measurements obtained from Gly-Gly-X-Ala-Gly-Gly and Gly-Gly-X-Pro-Gly-Gly revealed significant systematic chemical shift differences arising from the presence of proline in the peptide sequence. The influence of proline on the chemical shifts of residues which precede it was found to occur for all 20 amino acids and for all three types of nuclei (<sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N). Similar measurements on the influence of residues in the *i* - 1 position on the chemical shift of residues in the *i* position revealed important, sequence-dependent effects on amide <sup>15</sup>N and amide <sup>1</sup>H chemical shifts. Together, these data lend substantial support to the hypothesis that sequence effects play a significant role in determining peptide and protein chemical shifts.

In the future we plan to investigate nearest-neighbor interactions for all 400 possible amino acid pairs. We believe such a data set could provide important information for both theoreticians and experimentalists alike. Experimentalists could use these data to assist in their assignment of native proteins or in their assessment of denatured (or weakly structured) peptides and proteins.

Theoreticians could use them to further refine their theories on  $^1\text{H}^\alpha$  and  $^{13}\text{C}$  chemical shifts and to develop new theories on amide  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts. In addition to these nearest-neighbor studies, we plan to conduct a series of studies on solvent effects, pH effects and thermal variations on amino acid chemical shifts. We expect that these data will further add to our understanding of chemical shifts and could help in future theoretical and experimental developments in biomolecular NMR.

## Acknowledgements

We wish to thank Paul Semchuk for his help in preparing and purifying many of the peptides used in this study. Special thanks are also extended to Charlotte Holm for her assistance in preparing the X-Pro series of peptides. The work of Don Fvjen during the early phases of this project is also gratefully appreciated. Financial support by the MRC Group in Protein Structure and Function (Canada), the Protein Engineering Networks of Centres of Excellence (Canada) and the Alberta Heritage Fund for Medical Research is gratefully acknowledged.

## References

- Bax, A. and Subramanian, J. (1986) *J. Magn. Reson.*, **67**, 565–570.
- Braunschweiler, L. and Ernst, R.R. (1983) *J. Magn. Reson.*, **53**, 521–528.
- Bundi, A. and Wüthrich, K. (1979a) *Biopolymers*, **18**, 285–297.
- Bundi, A. and Wüthrich, K. (1979b) *Biopolymers*, **18**, 299–311.
- De Dios, A.C., Pearson, J.G. and Oldfield, E. (1993a) *Science*, **260**, 1491–1495.
- De Dios, A.C., Pearson, J.R. and Oldfield, E. (1993b) *J. Am. Chem. Soc.*, **115**, 9768–9773.
- Delgarno, D.C., Levine, B.A. and Williams, R.J.P. (1983) *Biosci. Rep.*, **3**, 443–452.
- DeMarco, A. (1977) *J. Magn. Reson.*, **26**, 527–528.
- Dyson, H.J., Rance, M., Houghten, R.A., Lerner, R.A. and Wright, P.E. (1988) *J. Mol. Biol.*, **201**, 161–200.
- Dyson, H.J. and Wright, P.E. (1991) *Annu. Rev. Biophys. Biophys. Chem.*, **20**, 519–538.
- Evans, P.A., Topping, K.D., Woolfson, D.N. and Dobson, C.M. (1991) *Protein Struct. Funct. Genet.*, **9**, 248–266.
- Gippert, G.P., Yip, P.F., Wright, P.E. and Case, D.A. (1990) *Biochem. Pharmacol.*, **40**, 15–22.
- Glushka, J., Lee, M., Coffin, S. and Cowburn, D. (1989) *J. Am. Chem. Soc.*, **111**, 7716–7722.
- Glushka, J., Lee, M., Coffin, S. and Cowburn, D. (1990) *J. Am. Chem. Soc.*, **112**, 2843.
- Grathwohl, C. and Wüthrich, K. (1974) *J. Magn. Reson.*, **13**, 217–225.
- Gross, K.-H. and Kalbitzer, H.R. (1988) *J. Magn. Reson.*, **76**, 87–99.
- Herranz, J., Gonzalez, C., Rico, M., Nieto, J.L., Santoro, J., Jimenez, M.A., Bruix, M., Neira, J.L. and Blanco, F.J. (1992) *Magn. Reson. Chem.*, **30**, 1012–1018.
- Howarth, O.W. (1978) *Prog. NMR Spectrosc.*, **12**, 1–40.
- Jimenez, M.A., Blanco, F.J., Rico, M., Herranz, J., Santoro, J. and Nieto, J.L. (1992) *Eur. J. Biochem.*, **207**, 39–49.
- Keim, P., Vigna, R.A., Marshall, R.C. and Gurd, F.R.N. (1973a) *J. Biol. Chem.*, **248**, 6104–6113.
- Keim, P., Vigna, R.A., Morrow, J.S., Marshall, R.C. and Gurd, F.R.N. (1973b) *J. Biol. Chem.*, **248**, 7811–7818.
- Keim, P., Vigna, R.A., Morrow, A.M., Marshall, R.C. and Gurd, F.R.N. (1974) *J. Biol. Chem.*, **249**, 4149–4156.
- Kricheldorf, H.R. (1981) *Org. Magn. Reson.*, **15**, 162–177.
- Live, D.H., Davis, D.G., Agosta, W.C. and Cowburn, D. (1984) *J. Am. Chem. Soc.*, **106**, 1939–1943.
- Masson, A. and Wüthrich, K. (1973) *FEBS Lett.*, **31**, 114–118.
- Merutka, G., Dyson, H.J. and Wright, P.E. (1995) *J. Biomol. NMR*, **5**, 14–24.
- Neri, D., Wider, G. and Wüthrich, K. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 4397–4401.
- Ösabay, K. and Case, D.A. (1991) *J. Am. Chem. Soc.*, **113**, 9436–9444.
- Ösabay, K. and Case, D.A. (1994) *J. Biomol. NMR*, **4**, 215–230.
- Pastore, A. and Saudek, V. (1990) *J. Magn. Reson.*, **90**, 165–176.
- Reily, M.D., Thanabal, V. and Omecinsky, D.O. (1992) *J. Am. Chem. Soc.*, **114**, 6251–6252.
- Richarz, R. and Wüthrich, K. (1978) *Biopolymers*, **17**, 2133–2141.
- Saito, H. (1986) *Magn. Reson. Chem.*, **24**, 835–852.
- Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.*, **113**, 5490–5492.
- Srinivasan, P.R. and Lichter, R.L. (1977) *J. Magn. Reson.*, **28**, 227–234.
- States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) *J. Magn. Reson.*, **48**, 286–292.
- Szilagyi, L. and Jardetzky, O. (1989) *J. Magn. Reson.*, **83**, 441–449.
- Thanabal, V., Omecinsky, D.O., Reily, M.D. and Cody, W.L. (1994) *J. Biomol. NMR*, **4**, 47–59.
- Torchia, D.A., Lyckla, J.R. and Quattrone, A.J. (1975) *Biochemistry*, **14**, 887–892.
- Williamson, M.P. (1990) *Biopolymers*, **29**, 1423–1431.
- Williamson, M.P., Asakura, T., Nakamura, E. and Demura, M. (1992) *J. Biomol. NMR*, **2**, 83–98.
- Wilmot, C.M. and Thornton, J.M. (1988) *J. Mol. Biol.*, **203**, 221–232.
- Wishart, D.S. (1991) Ph.D. Thesis, Yale University, New Haven, CT.
- Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991a) *FEBS Lett.*, **193**, 72–80.
- Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991b) *J. Mol. Biol.*, **222**, 311–333.
- Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) *Biochemistry*, **31**, 1647–1651.
- Wishart, D.S. and Sykes, B.D. (1994a) *J. Biomol. NMR*, **4**, 171–180.
- Wishart, D.S. and Sykes, B.D. (1994b) *Methods Enzymol.*, **239**, 363–392.
- Wishart, D.S., Bigam, C.G., Yao, J., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995) *J. Biomol. NMR*, submitted for publication.
- Witanowski, M., Stefaniuk, L. and Webb, G.A. (1993) *Annu. Rep. NMR Spectrosc.*, **25**, 1–480.
- Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.
- Zhou, N.E., Zhu, B.-Y., Sykes, B.D. and Hodges, R.S. (1992) *J. Am. Chem. Soc.*, **114**, 4320–4326.