A new model for mapping the peptide backbone: predicting proton chemical shifts in proteins†

José Luis Barneto,* Martín Avalos, Reyes Babiano, Pedro Cintas, José Luis Jiménez and Juan Carlos Palacios

Received 9th October 2009, Accepted 20th November 2009
First published as an Advance Article on the web 23rd December 2009
DOI: 10.1039/b921121g

This paper describes a methodology that correlates experimental chemical shifts (at the alpha proton) of proteins with their geometrical data (both dihedral angles and distances) obtained from 13 representative proteins, which are taken from the Protein Data Bank (PDB) and the BioMagRes Data Bank (BMRB). To this end, the experimentally measured proton chemical shifts of simple amides have been correlated with DFT-based calculated structures, at the B3PW91/6-31G* level. This results in a series of mathematical relationships, which are extrapolated to the above-mentioned proteins giving rise to a modified equation for such skeleta. It is relevant to note that the equation is also supported by a clear comparison with NMR data of a protein beyond the chosen set, such as insulin, even with lower errors. The model also relates the dependence of chemical shifts on hydrophobic and anisotropic effects at the amino acid residues.

Introduction

Peptides and proteins constitute ubiquitous chemicals, which represent equally the clue for life's origin, both catalyzing and promoting a wide range of biological processes that include cell signaling, adhesion, or differentiation to name a few. Certainly the current challenge in the post-genomic age is to identify the structural aspects and functions of all the encoded proteins and this hard work benefits from collaborative high-throughput efforts between X-ray crystallography and NMR techniques for small protein structure determination.[1] Protein data banks such as PDB[2] and BMRB[3] collect data from both diffraction-quality crystals and NMR screening, although a significant number of proteins are amenable only to one technique. Traditionally, methods aimed at determining protein structures use NOE-derived distances with observed and computed chemical shifts.[4] As recognized by Oldfield in an authoritative revision, simply being able to predict chemical shifts and chemical shift tensors in amino acids, peptides, and proteins is, in and of itself, of some interest.[5] And it might be more useful to use experimental chemical shifts to determine or refine further aspects of peptide structure and dynamics.

With impressive progress on high-resolution multidimensional NMR spectroscopy at a breathtaking pace, the earliest protein investigations using the ^{1}H nucleus have largely been complemented by the routine measurement of other nuclei (mostly ^{13}C and ^{15}N) to obtain reliable information for all backbone and side-chain torsional angles.[6] Since the Achilles' heel of protein NMR is associated with both size and time barriers, complementary techniques that include, among others, isotopic labelling of proteins,[7] recursive multidimensional decomposition (R-MDD) to speed recording,[8] measurements on the pico- to nano-second time scale, and residual dipolar couplings,[9] are being employed in conjunction with chemical shift data to predict the most likely local structures as well as conformational preferences of amino acid and peptide residues, which appear to play key roles in biological function[10] and degenerative diseases.[11]

In this investigation we have carried out a systematic study that correlates geometrical data from DFT calculations and experimental δ_H values in protein structures. To address this issue we have constructed the simplest structural model of the peptide bond showing the dependence of the magnetic anisotropy caused by the amide function on distances and torsional angles that can be easily assessed. Overall, this leads to a series of computed chemical shifts values for the C(=O)–N–CH protons that consistently reproduce experimental values obtained for proteins. Based on this analysis, this work proposes a relatively simple empirical formula that accurately reproduces proton chemical shifts using optimized geometries. The predictive value has been tested with a representative number of naturally-occurring proteins and, in addition, such shifts also evidence key relationships with other features, such as the electronic and hydrophobic character of the backbone amino acid.

The paper is organized in different sections, first describing the computational methodology, the approach employed to construct the model based on previous studies with secondary amides, and finally application to protein structures chosen from protein data banks.

Results and discussion

Structure model and general considerations

For proteins the position of the RCH framework is flanked by two neighboring amido groups having a distinctive chemical

Departamento de Química Orgánica e Inorgánica, QUOREX Research Group, Facultad de Ciencias, Universidad de Extremadura, E-06071, Badajoz, (Spain). E-mail: jbarneto@unex.es; Fax: +34 924 271149; Tel: +34 924 289382

† Electronic supplementary information (ESI) available: Tables S1 and S2; Fig. S1 and S2; minimum energy structures for *Z*-amides **1–15** obtained by B3PW91/6-31G* calculations; and NMR spectra from amides **1–15** in $CDCl_3$, DMSO-d_6 and D_2O. See DOI: 10.1039/b921121g

environment. The aforementioned CH protons should thus undergo magnetic deshielding caused by the two amide functional groups. As depicted in Fig. 1 the conformational arrangement of such a peptide bond can be described by a series of torsional angles and nonbonded distances, which in turn define the rotamer states about each of the two amide groups involved. It should be firstly noted that the notation employed for such structural parameters has been arbitrarily chosen for the purpose of this work and should not be confused with the standard nomenclature coined for the torsion angles in protein skeleta.[12]

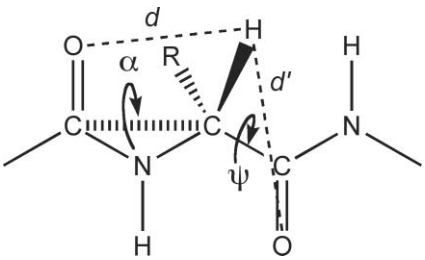

Fig. 1 Distances (d, d') and dihedral angles α (virtual, O=C···C–H) and ψ taken in this work.

In previous works,[13,14] we have evaluated the chemical deshielding caused by the amide function (σ_{amide}) on C(=O)–N–CH protons in 15 secondary amides, which can accurately be correlated with DFT-based structures. Thus, we found not only that the *anti* disposition between the CH proton and the N–H bond appears to be the more stable conformation of simple amides, thereby clarifying previous contradictory spectroscopic and theoretical data, but also that there is a systematic dependence of the deshielding and therefore of the experimental chemical shift (*vide infra*) on the torsion angle and the distance between the shifted proton and the oxygen atom, through the optimized expression (1), which is suitable for estimations in $CDCl_3$, DMSO-d_6 and D_2O solvent systems:[14]

$$\sigma_H = a + 2.16\cos^2\left(\frac{\alpha - 35}{d}\right) \quad (1)$$

where a represents a solvent-dependent parameter, α is the dihedral angle O=C···C–H and d the distance between the shifted proton and the carbonyl oxygen atom (Fig. 1). One should note that in the above expression chemical deshieldings can be fit to a $\cos^2$ function, in a similar way to Karplus-like equations for spin–spin coupling constants, whose structural dependencies have been recently investigated by DFT methods too.[15] In addition, a similar $\cos^2$ function appears on the classical McConnell equation,[16] one of the first attempts to calculate proton shifts based on magnetic anisotropies, although their core parameters (the magnetic anisotropy as a component of the magnetic susceptibility tensor and the distance between the shifted atom and the center of the anisotropic group) are not easily evaluated.

In following a similar methodology,[13,14] other additional secondary amides, which include acetamides, propionamides, isobutyramides, and 2-phenylpropionamides (Fig. 2) have also been investigated by means of B3PW91/6-31G* calculations.[17] The preferential conformational arrangement of such structures is invariably a $Z_{anti,anti'}$ disposition as shown in Fig. 3.

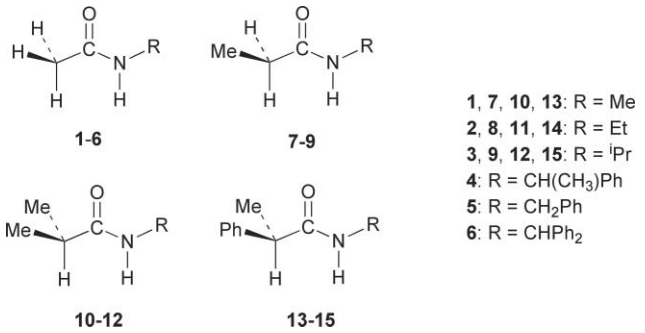

Fig. 2 Set of structures **1–15**.

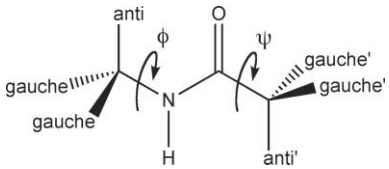

Fig. 3 Arrangement of *anti* and *gauche* protons in a $Z_{anti,anti'}$ structure.

Likewise, the experimental chemical deshieldings caused by the amide function on the protons linked to the carbon atom vicinal to the carbonyl group, N–C(=O)–C*H*, can be easily determined. Such deshieldings are denoted as σ'_{amide} to distinguish them from those on the C(=O)–N–C*H* protons (σ_{amide}).[13] Again, the well-known Shoolery relationship (2)[18] provides an additive dependence of chemical shifts on the different anisotropies of vicinal groups:

$$\delta_{CHXYZ} = 0.23 + \sigma_X + \sigma_Y + \sigma_Z \quad (2)$$

where one of the adjacent atoms or groups is –C(=O)–N, the amide function:

$$\sigma'_{amide} = \delta_{obs} - (0.23 + \sigma_X + \sigma_Y) \quad (3)$$

This equation enables the calculation of σ'_{amide} anisotropies based on published data for σ_X and σ_Y, provided that the corresponding proton chemical shift is known.

Furthermore, it is convenient to consider that the observed peaks in proton NMR spectra would correspond to an average of *anti* and *gauche* protons at the CH carbon atom, and accordingly:

$$\sigma'_{amide} = \frac{\sigma'_{anti}}{n} + (n-1)\frac{\sigma'_{gauche}}{n} \quad (4)$$

where n is the number of protons on the CH carbon atom.

Since the preferential conformation found is $Z_{anti,anti'}$, amides with only one proton on CH (**10–15**) would have no *gauche* protons ($n = 1$) and therefore $\sigma'_{anti} = \sigma'_{amide}$ in that particular case. Afterwards, we decided to take the average of these σ'_{anti} values as σ'_{anti} for amides where $n > 1$, allowing us to calculate σ'_{gauche} by eqn (4) (see Table S1 in the ESI†).

Having determined chemical deshieldings (generically σ'_H), it is possible to find a relationship with geometrical data obtained previously by DFT-methods as shown in Fig. 4, which depicts a good correlation with the ψ dihedral angle (O=C–C–H):

Our data fitted to a third-order polynomial which resembles the $\cos^2/2$ function. Thus, a further linear relationship [eqn (5)] with a regression coefficient of $r^2 = 0.97$ (Fig. 5) could equally be established:

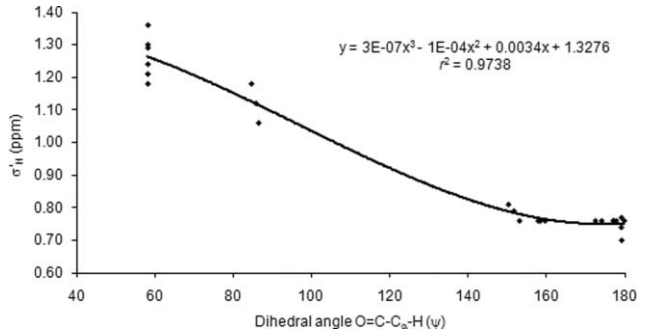

Fig. 4 Plot of chemical deshieldings (σ'_H, ppm) in $CDCl_3$ *versus* ψ dihedral angles for amides **1–15**.

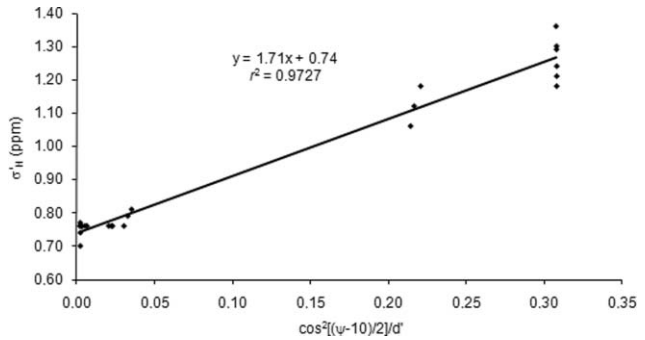

Fig. 5 Linear relationship between σ'_H and the function $\cos^2[(\psi - 10)/2]/d'$ in $CDCl_3$ for amides **1–15**.

$$\sigma'_H = 0.74 + 1.71\frac{\cos^2(\psi - 10)/2}{d'} \quad (5)$$

where d' is the distance between the shifted proton and the carbonyl oxygen atom.

Similar plots were obtained for such amides in DMSO-d_6 and D_2O (see Fig. S1 and S2 in the ESI†), although their regression coefficients were lower (Table 1) because σ values estimated by Shoolery correlations have been parametrized for $CDCl_3$ only.[18] This effect can be expressed quantitatively by means of eqn (6), where c and e are solvent-dependent parameters, in a similar way as a in eqn (1):

$$\sigma'_H = c + e\frac{\cos^2(\psi - 10)/2}{d'} \quad (6)$$

One should, in addition, bear in mind that solvation (not considered in the present study) can also stabilize conformers which are not stable in the gas phase. In this context, a series of recent studies have been focused on the alanine dipeptide using different solvent models, which are able to reproduce and interpret vibrational and Raman spectra in aqueous solution.[19] Moreover, structural modifications may also lead to inconsistencies in both gas-phase calculations and simple continuum models, such as in the zwitterionic species of the L-alanine amino acid.[19,20]

Table 1 Parameters for eqn (6)

Solvent	c	e	r^2
$CDCl_3$	0.74	1.71	0.97
DMSO-d_6	0.74	0.70	0.65
D_2O	0.89	0.75	0.63

Peptide bond correlations

We have obtained the necessary equations to calculate deshieldings on both sides of the amide function. One can assume a Shoolery relationship for –C(=O)–N–CHR–C(=O)–N–protons located in the main chain of a protein skeleton, [eqn (7)]:

$$\delta = 0.23 + \sigma_R + \sigma_H + \sigma'_H \quad (7)$$

where σ_R represents the chemical deshielding influenced by the amino acid residue. Substituting σ_H and σ'_H by means of eqn (1) and (6), eqn (7) now becomes:

$$\delta = 0.23 + a + c + \sigma_R + b\cos^2\frac{(\alpha - 25)}{d} + e\cos^2\frac{(\psi - 10)/2}{d'} \quad (8)$$

Finally, and assuming that peptides are usually studied in an aqueous environment, the appropriate values for D_2O give rise to eqn (9) that correlates the chemical shift of a given CH proton with geometrical data:

$$\delta = 2.87 + \sigma_R + 1.32\cos^2\frac{(\alpha - 25)}{d} + 0.75\cos^2\frac{(\psi - 10)/2}{d'} \quad (9)$$

Applications to amino acid residues

In order to obtain σ_R values, conformational data have been extracted from the protein data bank (PDB) along with observed chemical shift data from the BiomagResDataBank (BMRB) as mentioned below in the Materials and methods section. Replacement of such data in eqn (9) allowed us to obtain 2153 σ_R values and Table 2 lists the average σ_R value for every amino acid.[21] These σ_R data re-introduced in eqn (9) provide calculated chemical shifts (δ_{calc} in contrast with δ_{obs}), whose averages ($\bar{\delta}_{calc}$ and $\bar{\delta}_{obs}$) were obviously identical. Standard deviations (SD) were higher for δ_{obs} than for δ_{calc} data. Table 2 also includes the average error (AE), the average of the difference between δ_{obs} and δ_{calc}, as well as their standard deviations; the global error being 0.41 ± 0.35 ppm.

Data collected in Table 2 are representative enough of the peptide set chosen in this study; the magnitude of δ_{obs} for a given amino acid is quite similar to that reported recently by Zhang and coworkers,[22] with deviations less than 0.14 ppm as evidenced in Table 3 for comparative purposes. The sole exception is found for tryptophan (0.30 ppm), although this fact could be attributed to the small proportion of such a residue in the global sample (19 out of 2035 amino acids).

To validate eqn (9) on a protein different from those of the protein banks employed to estimate σ_R deshielding, we have also calculated alpha proton chemical shifts in insulin which have been compared with experimental values. This results in an average error of 0.27 ppm and 0.21 as the standard deviation (see Table S2 in the ESI†).

An important structure factor that can be correlated with chemical shifts is the hydrophobic (or hydrophilic) character of some amino acids or even protein regions. Hydrophobicity often represents a crucial factor that determines, for instance, antigen activity. We have therefore compared our results with predictions of hydrophobicity based on frequently used scales, such as the Kyte–Doolittle (KD)[23] and Hopp–Woods (HW),[24] in which the

Table 2 Average error and statistical data (in ppm) for every amino acid in proteins

	Amino acid	$\bar{\sigma}_R$	$\bar{\delta}_{calc} = \bar{\delta}_{obs}$	SD_{obs}	SD_{calc}	AE[a]	SD_{AE}	Measured residues
Alanine	ALA	0.75	4.24	0.46	0.14	0.36	0.34	151
Arginine	ARG	0.80	4.29	0.44	0.11	0.39	0.28	113
Asparagine	ASN	1.18	4.64	0.39	0.14	0.32	0.27	77
Aspartate	ASP	1.11	4.59	0.33	0.14	0.28	0.22	119
Cysteine	CYS	1.22	4.70	0.60	0.14	0.53	0.35	35
Glutamine	GLN	0.70	4.19	0.37	0.15	0.32	0.28	111
Glutamate	GLU	0.77	4.26	0.45	0.13	0.37	0.32	149
Glycine	GLY	0.59	3.94	0.52	0.22	0.34	0.35	118
Histidine	HIS	1.09	4.54	0.52	0.16	0.42	0.35	39
Isoleucine	ILE	0.65	4.13	0.63	0.12	0.57	0.37	124
Leucine	LEU	0.90	4.39	0.52	0.12	0.46	0.34	187
Lysine	LYS	0.81	4.30	0.45	0.13	0.39	0.28	138
Methionine	MET	0.76	4.26	0.64	0.11	0.51	0.43	40
Phenylalanine	PHE	1.25	4.73	0.67	0.13	0.61	0.38	74
Proline	PRO	1.01	4.42	0.36	0.14	0.30	0.30	97
Serine	SER	1.05	4.50	0.46	0.13	0.39	0.32	128
Threonine	THR	0.93	4.42	0.54	0.13	0.47	0.38	113
Tryptophan	TRP	0.99	4.49	0.47	0.17	0.42	0.32	19
Tyrosine	TYR	1.21	4.67	0.58	0.12	0.50	0.37	63
Valine	VAL	0.72	4.20	0.68	0.12	0.61	0.39	140
GLOBAL				0.55	0.25	0.41	0.35	2035

[a] $\sum(|\delta_{obs} - \delta_{calc}|)/n$.

Table 3 Average chemical shifts (in ppm) from Zhang and this work

	Zhang *et al.*22		This work	
Amino acid	δ_{obs}	SD	δ_{obs}	SD
ALA	4.29	0.49	4.24	0.46
ARG	4.29	0.50	4.29	0.44
ASN	4.71	0.42	4.64	0.39
ASP	4.62	0.34	4.59	0.33
CYS	4.79	0.64	4.70	0.60
GLN	4.30	0.48	4.19	0.37
GLU	4.28	0.45	4.26	0.45
GLY	3.98	0.41	3.94	0.52
HIS	4.64	0.52	4.54	0.52
ILE	4.23	0.60	4.13	0.63
LEU	4.37	0.52	4.39	0.52
LYS	4.29	0.49	4.30	0.45
MET	4.40	0.53	4.26	0.64
PHE	4.67	0.61	4.73	0.67
PRO	4.40	0.39	4.42	0.36
SER	4.56	0.45	4.50	0.46
THR	4.53	0.51	4.42	0.54
TRP	4.79	0.58	4.49	0.47
TYR	4.70	0.63	4.67	0.58
VAL	4.20	0.61	4.20	0.68
Amino acids in protein structures	26294		2035	

higher errors and standard deviations are found for the most hydrophobic amino acids (Table 4). Moreover, the observed shifts in hydrophobic amino acids show higher standard deviations than the corresponding calculated shifts. This fact suggests that the high AE is linked to hydrophobicity effects, which in aqueous solution produce structures that avoid the interaction with the solvent in protein regions where these amino acids are located, instead of the geometrical factors used by our method. In contrast, hydrophilic amino acids show lower values of AE and standard deviations; and in addition, standard deviations are quite similar to both observed and calculated chemical shifts.

Fig. 6 shows the error percentages (separated by 0.1 ppm intervals) of all computed amino acids. The individual percentages relative to the whole sample are indicated by means of blue bars, while cumulative errors are represented by red bars. Thus, one can observe that for approximately 55% of the cases studied, the mean error is less than 0.4 ppm. Further inspection of errors above the averaging for some amino acid residues reveals that they are often located next to anisotropic groups, such as a phenyl ring or amido linkage. Such groups may be far away in sequence but spatially close due to backbone folding. In stark contrast, amino acid residues with little AE values have no anisotropy sources in their vicinity. Fig. 7 shows an example from a residue with high error (1.42 ppm): the distance between the CH proton in the residue GLN 15 from the 1XNA structure and a carbonylic oxygen is just 2.545 Å.[25] In this case there is another anisotropic group, –N(–C=C), within a short distance: 3.099 Å. The opposite situation is represented by Fig. 8, which depicts a fragment of

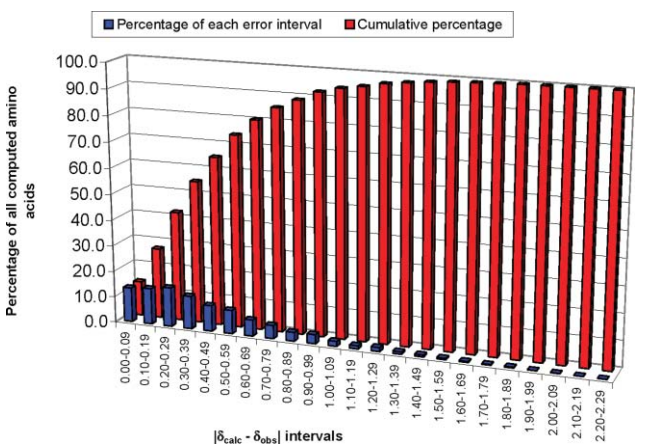

Fig. 6 Errors calculating proton chemical shifts for amino acid residues. Each entry stands for a 0.1 ppm interval.

Table 4 Average errors and standard deviations related to hydrophobicity and hydrophilicity scales

Aminoacid	K–D[a]23	H–W[b]24	AE	SD_{AE}	SD_{obs}	SD_{calc}	Measured residues
ALA	1.8	–0.5	0.36	0.34	0.46	0.14	151
ARG	–4.5	3.0	0.39	0.28	0.44	0.11	113
ASN	–3.5	0.2	0.32	0.27	0.39	0.14	77
ASP	–3.5	3.0	0.28	0.22	0.33	0.14	119
CYS	2.5	–1.0	0.53	0.35	0.60	0.14	35
GLN	–3.5	0.2	0.32	0.28	0.37	0.15	111
GLU	–3.5	3.0	0.37	0.32	0.45	0.13	149
GLY	–0.4	0.0	0.34	0.35	0.52	0.22	118
HIS	–3.2	–0.5	0.42	0.35	0.52	0.16	39
ILE	4.5	–1.8	0.57	0.37	0.63	0.12	124
LEU	3.8	–1.8	0.46	0.34	0.52	0.12	187
LYS	–3.9	3.0	0.39	0.28	0.45	0.13	138
MET	1.9	–1.3	0.51	0.43	0.64	0.11	40
PHE	2.8	–2.5	0.61	0.38	0.67	0.13	74
PRO	–1.6	0.0	0.30	0.30	0.36	0.14	97
SER	–0.8	0.3	0.39	0.32	0.46	0.13	128
THR	–0.7	–0.4	0.47	0.38	0.54	0.13	113
TRP	–0.9	–3.4	0.42	0.32	0.47	0.17	19
TYR	–1.3	–2.3	0.50	0.37	0.58	0.12	63
VAL	4.2	–1.5	0.61	0.39	0.68	0.12	140
GLOBAL			0.41	0.35	0.55	0.25	2035

[a] Coefficient derived both from water-vapor transfer free energies and the interior–exterior distribution of amino acid side-chains. [b] Coefficient based On Levitt's hydrophobicity parameters,[27] taken as the measured free energy of transfer from water to ethanol, or estimated from the relationship between accessible surface area and hydrophobicity when experimental values are not available.

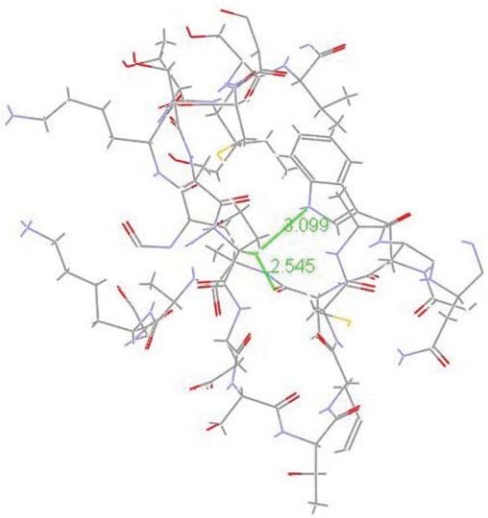

Fig. 7 Example of amino acid residue surrounded by anisotropic groups and having a high calculated error.

1T0W structure;[26] the nearest anisotropic group to the CH proton of the GLN 2 residue, whose error is just 0.09 ppm, is a phenyl ring.

Conclusions

We have developed a methodology that ultimately releases an equation for the prediction of α-proton chemical shifts, in the amide moiety (δ_H for–CO–NH–C*H*(R)–CO–NH–) in different solvents, including an aqueous environment. The algorithm takes into account only geometrical parameters that reflect the anisotropy of the amide groups in the peptide bond. The calculated shifts correlate well with experimental values found for 2035 amino

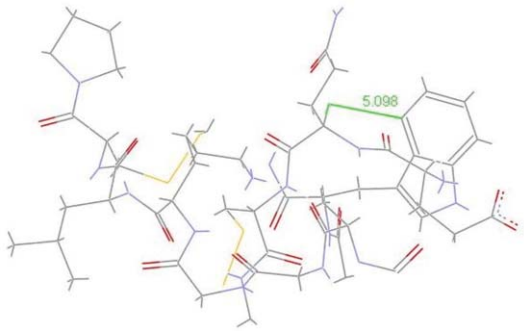

Fig. 8 Example of amino acid residue having no anisotropic groups in its vicinity and therefore with a low calculated error.

acid residues present in 13 different proteins, whose structures have been previously determined by NMR analyses. The averaging error was 0.41 ppm and the standard deviation was 0.35 ppm. To validate the model beyond the above-mentioned proteins, the mathematical relationship has been applied to insulin, finding an error of 0.27 ppm with SD = 0.21 ppm.

Given the popularity of NMR chemical shifts for structure elucidation and assuming that there are numerous software packages containing databases of millions experimental proton shifts, the method provides a fast and reliable alternative to existing prediction programs. One should further expect improvements by means of higher level of theories and beyond standard basis sets. However, the proof of concept has been demonstrated here with the usual B3PW91/6-31G* level, which can be regarded a suitable tradeoff between speed and accuracy for large molecular systems. The correlations are also sensitive to hydrophobic and conformational effects that enhance the anisotropy within the peptide backbone. Such shifts can thus complement or give structural information in cases for which NOE and/or coupling

data are not readily available. Finally, we plan to extrapolate the present methodology to other nuclei of interest and its application to particular protein conformations.

Experimental

Materials and methods

Acetamides **1** and **2**, and propionamide **7** were obtained from commercial suppliers and used without further purification. Acetamides **3–6** were prepared as described in a previous work.[13] Compounds **8–15** were also synthesized by reaction of the corresponding amine and acyl chloride.[28]

DFT calculations were carried out with the Gaussian 03 program package.[29] The stationary points were characterized by frequency calculations to verify that minima have no imaginary frequencies. All NMR data were collected at 400 MHz in perdeuterated solvents ($CDCl_3$, DMSO-d_6, and D_2O, 99.9% D) with chemical shifts referred to tetramethylsilane (TMS) as the internal standard ($\delta = 0.00$ ppm). Structural data for proteins have been taken from the PDB database[2] as well as chemical shifts from BMRB,[3] namely the N-terminal domain of enzyme I from *Escherichia coli* (2EZB);[30] human neutrophil gelatinase-associated lipocalin (1NGL);[31] N-terminal domain of single-strand DNA-repair protein XRCC1 (1XNA);[25] human T-cell linfotropic virus type I capsid protein (1QRJ);[32] N-terminal domain of VCP-like ATPase of thermoplasma (1CZ4);[33] capsid protein from Rous sarcoma virus (1D1D);[34] human prion protein hPrP (1QLX);[35] inserted domain of human leukocyte function associated antigen-1 (1DGQ);[36] CDC42 from human (1AJE);[37] 3-methyladenine DNA glycosylase I (1LMZ);[38] acidic fibroblast growth factor (1RML);[39] truncated hevein of 32 amino acids (1T0W);[26] and homeobox gene Hex-1;[40] dimeric insulin (3INS).[41] They have been chosen without any further consideration of a preferred conformational organization (either α-helix or β-sheet).

Acknowledgements

The authors gratefully acknowledge Dr Jesús Jiménez Barbero for his comments and encouragement, as well as for his hospitality during a short stay of JLB at CIB-CSIC, Madrid. We also thank the Ministry of Education and Science (Grant No. CTQ2007-66641) and Junta de Extremadura (Grants No. PRI07A016 and PRI08A032) for financial support. JLB thanks the Junta de Extremadura for a scholarship.

Notes and references

1 D. A. Snyder, Y. Chen, N. G. Denissova, T. Acton, J. M. Aramini, M. Ciano, R. Karlin, J. Liu, P. Manor, P. A. Rajan, P. Rossi, G. V. T. Swapna, R. Xiao, B. Rost, J. Hunt and G. T. Montelione, *J. Am. Chem. Soc.*, 2005, **127**, 16505–16511; A. A. Yee, A. Savchenko, A. Ignachenko, J. Lukin, X. Xu, T. Skarina, E. Evdokimova, C. S. Liu, A. Semesi, V. Guido, A. M. Edwards and C. H. Arrowsmith, *J. Am. Chem. Soc.*, 2005, **127**, 16512–16517.
2 H. B. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–242.
3 B. R. Seavey, E. A. Farr, W. M. Westler and J. L. Markley, *J. Biomol. Nucl. Magn. Reson.*, 1991, **1**, 217–236.
4 K. Wüthrich, in *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986; S. W. Fesik and E. R. P. Zuiderweg, *Q. Rev. Biophys.*, 1990, **23**, 97–131; G. M. Clore and A. M. Gronenborn, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1991, **23**, 43–92; A. Bax and S. Grzesiek, *Acc. Chem. Res.*, 1993, **26**, 131–138; M. Nilges, *J. Mol. Biol.*, 1995, **245**, 645–660; J. Cavanagh, W. J. Fairbrother, A. G. Palmer, III and N. J. Skelton, in *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, San Diego, 1996.
5 E. Oldfield, in *Annual Review of Physical Chemistry*, ed. S. R. Leone, P. Alivisatos and A. E. McDermott, Annual Reviews, Palo Alto, CA, 2002, vol. 53, pp 349-378.
6 K. D. Park, K. Guo, F. Adebodun, M. L. Chiu, S. G. Sligar and E. Oldfield, *Biochemistry*, 1991, **30**, 2333–2347; S. Spera and A. Bax, *J. Am. Chem. Soc.*, 1991, **113**, 5490–5492; K. Osapay and D. A. Case, *J. Am. Chem. Soc.*, 1991, **113**, 9436–9444; A. C. de Dios, J. G. Pearson and E. Oldfield, *Science*, 1993, **260**, 1491–1496; A. C. de Dios and E. Oldfield, *J. Am. Chem. Soc.*, 1994, **116**, 7453–7454; N. Tjandra and A. Bax, *J. Am. Chem. Soc.*, 1997, **119**, 9576–9577; A. C. de Dios and E. M. Earle, *J. Phys. Chem. A*, 1997, **101**, 8132–8134; D. Sitkoff and D. A. Case, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1998, **32**, 165–190; N. Godbout, R. Havlin, R. Salzmann, P. G. Debrunner and E. Oldfield, *J. Phys. Chem. A*, 1998, **102**, 2342–2350; A. Perczel and A. G. Császár, *Chem.–Eur. J.*, 2001, **7**, 1069–1083; J. Meiler and D. Baker, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 15404–15409; J. Birn, A. Poon, Y. Mao and A. Ramamoorthy, *J. Am. Chem. Soc.*, 2004, **126**, 8529–8534; P. R. L. Markwick and M. Sattler, *J. Am. Chem. Soc.*, 2004, **126**, 11424–11425; M. V. Berjanskii and D. S. Wishart, *J. Am. Chem. Soc.*, 2005, **127**, 14970–14971; B. López-Mendez and P. Güntert, *J. Am. Chem. Soc.*, 2006, **128**, 13112–13122; L. L. Parker, A. R. Houk and J. H. Jensen, *J. Am. Chem. Soc.*, 2006, **128**, 9863–9872; J. A. Vila, D. R. Ripoll and H. A. Scheraga, *J. Phys. Chem. B*, 2007, **111**, 6577–6585; B. J. Wylie, L. J. Sperling, H. L. Frericks, G. J. Shah, W. T. Franks and C. M. Rienstra, *J. Am. Chem. Soc.*, 2007, **129**, 5318–5319; D. F. Hansen, P. Vallurupalli, P. Lundström, P. Neudecker and L. E. Kay, *J. Am. Chem. Soc.*, 2008, **130**, 2667–2675; R. E. London, B. D. Wingad and G. A. Mueller, *J. Am. Chem. Soc.*, 2008, **130**, 11097–11105.
7 M. Kainosho, T. Torizawa, Y. Iwashita, T. Terauchi, A. Mei Ono and P. Güntert, *Nature*, 2006, **440**, 52–57; S. J. Opella, *Nature*, 2006, **440**, 40.
8 V. Y. Orekhov, I. Ibraghimov and M. Billeter, *J. Biomol. Nucl. Magn. Reson.*, 2003, **27**, 165–173; V. Tugarinov, L. E. Kay, I. Ibraghimov and V. Y. Orekhov, *J. Am. Chem. Soc.*, 2005, **127**, 2767–2775; V. Jaravine, I. Ibraghimov and V. Y. Orekhov, *Nat. Methods*, 2006, **3**, 605–607.
9 N. Tjandra and A. Bax, *Science*, 1997, **278**, 1111–1114; J. J. Chou, S. Li and A. Bax, *J. Biomol. Nucl. Magn. Reson.*, 2000, **18**, 217–227; A. Mittermaier and L. E. Kay, *Science*, 2006, **312**, 224–228; J. Gsponer, H. Hopearuoho, A. Cavalli, C. M. Dobson and M. Vendruscolo, *J. Am. Chem. Soc.*, 2006, **128**, 15127–15135; S. A. Dames, R. Aregger, N. Vajpai, P. Bernado, M. Blackledge and S. Grzesiek, *J. Am. Chem. Soc.*, 2006, **128**, 13508–13514; J. Korukottu, M. Bayrhuber, P. Montaville, V. Vijayan, Y.-S. Jung, S. Becker and M. Zweckstetter, *Angew. Chem., Int. Ed.*, 2007, **46**, 1176–1179.
10 C. Branden and J. Tooze, in *Introduction to Protein Structure*, Taylor & Francis, London, U.K., 2nd edn, 1999; J. Kyte, in *Structure in Protein Chemistry*, Garland Science, London, U.K., 2006.
11 For an excellent special issue on amyloid assembly, a topic related to neurodegenerative diseases: *Acc. Chem. Res.*, 2006, **39**, 567–680, ed. R. Wetzel.
12 (*a*) *IUPAC-IUB Recommendations 1971*; IUPAC Information Bulletin No. 10, Pergamon, Oxford, U.K., 1971; (*b*) *Amino Acids, Peptides and Proteins*, ed. G. T. Young, Royal Chemical Society, London, U.K., 1972, Vol. 4, p 455.
13 M. Ávalos, R. Babiano, J. L. Barneto, J. L. Bravo, P. Cintas, J. L. Jiménez and J. C. Palacios, *J. Org. Chem.*, 2001, **66**, 7275–7282.
14 M. Ávalos, R. Babiano, J. L. Barneto, P. Cintas, F. R. Clemente, J. L. Jiménez and J. C. Palacios, *J. Org. Chem.*, 2003, **68**, 1834–1842.
15 H. Zhao, Q. Pan, W. Zhang, I. Carmichael and A. S. Serianni, *J. Org. Chem.*, 2007, **72**, 7071–7082.
16 H. M. McConnell, *J. Chem. Phys.*, 1957, **27**, 226–229.
17 One of the referees has judiciously suggested that improved results could be attained with expanded basis sets. We have performed structure calculations for *E*-*N*-ethyl formamide using aug-cc-pVTZ and 6-311++G**. Relative to the 6-31G* level, very small variations in dihedral angles and distances were observed, 0.6° and 0.031 Å, respectively.
18 J. N. Shoolery, in *Varian Associates Technical Information Bulletin*, Palo Alto, CA, 1959, Vol. 2, No. 3; E. C. Friedrich and K. Gates Runkle,

J. Chem. Educ., 1986, **63**, 127–129; P. S. Beauchamp and R. Marquez, *J. Chem. Educ.*, 1997, **74**, 1483–1485.
19 See for instance: K. J. Jalkanen, M. Elstner and S. Suhai, *J. Mol. Struct. (THEOCHEM)*, 2004, **675**, 61–77; K. J. Jalkanen, I. M. Degtyarenko, R. M. Nieminen, X. Cao, L. A. Nafie, F. Zhu and L. D. Barron, *Theor. Chem. Acc.*, 2008, **119**, 191–210.
20 I. M. Degtyarenko, K. J. Jalkanen, A. A. Gurtovenko and R. M. Nieminen, *J. Phys. Chem. B*, 2007, **111**, 4227–4234; I. M. Degtyarenko, K. J. Jalkanen, A. A. Gurtovenko and R. M. Nieminen, *J. Comput. Theor. Nanosci.*, 2008, **5**, 277–285.
21 Since glycine (118 residues) is the only amino acid with two alpha protons, this fact accounts for the difference between the number of observed chemical shifts—2153—and the number of measured residues—2035.
22 H. Zhang, S. Neal and D. S. Wishart, *J. Biomol. Nucl. Magn. Reson.*, 2003, **25**, 173–195.
23 J. Kyte and R. F. Doolittle, *J. Mol. Biol.*, 1982, **157**, 105–132. The Kyte–Doolittle scale ranges from 4.5 (isoleucine -ILE-, the most hydrophobic amino acid) to –4.5 (arginine -ARG-, the less hydrophobic).
24 T. P. Hopp and K. R. Woods, *Proc. Natl. Acad. Sci. U. S. A.*, 1981, **78**, 3824–3828. The Hopp–Woods scale shows a maximum hydrophilicity value of 3.0 for 4 different amino acids (arginine -ARG-, aspartate -ASP-, glutamate -GLU-, and lysine -LYS-), and a minimum value of –3.4 for tryptophan (TRP).
25 A. G. Marintchev, M. W. Maciejewski and G. P. Mullen, *J. Biomol. Nucl. Magn. Reson.*, 1999, **13**, 393–394.
26 N. H. Andersen, B. Cao, A. Rodríguez-Romero and B. Arreguin, *Biochemistry*, 1993, **32**, 1407–1422; J. L. Asensio, F. J. Cañada, M. Bruix, A. Rodríguez-Romero and J. Jiménez-Barbero, *Eur. J. Biochem.*, 1995, **230**, 621–633; J. L. Asensio, F. J. Cañada, M. Bruix, C. González, N. Khiar, A. Rodríguez-Romero and J. Jiménez-Barbero, *Glycobiology*, 1998, **8**, 569–577; J. L. Asensio, F. J. Cañada, H. C. Siebert, J. Laynez, A. Poveda, P. M. Nieto, U. M. Soedjanaamadja, H. J. Gabius and J. Jiménez-Barbero, *Chem. Biol.*, 2000, **7**, 529–543; N. Aboitiz, M. Vila-Perelló, P. Groves, J. L. Asensio, D. Andreu, F. J. Cañada and J. Jiménez-Barbero, *ChemBioChem*, 2004, **5**, 1245.
27 M. Levitt, *J. Mol. Biol.*, 1976, **104**, 59–107.
28 L. A. LaPlanche and M. T. Rogers, *J. Am. Chem. Soc.*, 1964, **86**, 337–341.
29 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *GAUSSIAN 03*, Gaussian, Inc., Wallingford, CT, 2004.
30 D. S. Garrett, Y.-J. Seok, D.-I. Liao, A. Peterkofsky, A. M. Gronenborn and M. G. Clore, *Biochemistry*, 1997, **36**, 2517–2530.
31 M. Coles, T. Diercks, B. Muehlenweg, S. Bartsch, V. Zoelzer, H. Tschesche and H. Kessler, *J. Mol. Biol.*, 1999, **289**, 139–157.
32 S. Khorasanizadeh, R. Campos-Olivas, C. A. Clark and M. F. Summers, *J. Biomol. Nucl. Magn. Reson.*, 1999, **14**, 199–200.
33 M. Coles, T. Diercks, J. Liermann, A. Groeger, B. Rockel, W. Baumeister, K. K. Koretke, A. Lupas, J. Peters and H. Kessler, *Curr. Biol.*, 1999, **9**, 1158–1168.
34 R. Campos-Olivas, J. L. Newman, Y. Ndassa and M. F. Summers, *J. Biomol. Nucl. Magn. Reson.*, 1999, **15**, 267–268.
35 R. Zahn, A. Liu, T. Luhrs, R. Riek, C. v. Schroetter, F. Lopez Garcia, M. Billeter, L. Calzolai, G. Wider and K. Wüthrich, K., *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 145–150.
36 R. W. Kriwacki, G. B. Legge, U. Hommel, P. Ramage, J. Chung, L. L. Tennant, P. E. Wright and H. J. Dyson, *J. Biomol. Nucl. Magn. Reson.*, 2000, **16**, 271–272.
37 A. Morreale, M. Venkatesan, H. R. Mott, D. Owen, D. Nietlispach, P. N. Lowe and E. D. Laue, *Nat. Struct. Biol.*, 2000, **7**, 384–388.
38 A. C. Drohat, K. Kwon, D. J. Krosky and J. T. Stivers, *Nat. Struct. Biol.*, 2002, **9**, 659–664.
39 A. Pineda-Lucena, M. A. Jiménez, J. L. Nieto, J. Santoro, M. Rico and G. Giménez-Gallego, *J. Mol. Biol.*, 1994, **242**, 81–98; R. M. Lozano, M. Jiménez, J. Santoro, M. Rico and G. Giménez-Gallego, *J. Mol. Biol.*, 1998, **281**, 899–915.
40 I. de la Mata, J. L. Garcia, C. González, M. Menéndez, F. J. Cañada, J. Jiménez-Barbero and J. L. Asensio, *ChemBioChem*, 2002, **3**, 726–740.
41 A. Wlodawer, H. Savage and G. Dodson, *Acta Crystallogr., Sect. B: Struct. Sci.*, 1989, **45**, 99–107; A. D. Kline and R. M. Justice, *Biochemistry*, 1990, **29**, 2906–2913.