Eur. Phys. J. D **20**, 513–530 (2002)
DOI: 10.1140/epjd/e2002-00157-4

THE EUROPEAN PHYSICAL JOURNAL D

Toward direct determination of conformations of protein building units from multidimensional NMR experiments III

A theoretical case study of For–L-Phe–NH_2

A. Perczel[1,a] and A.G. Császár[2,b]

[1] Department of Organic Chemistry, Eötvös University, P.O. Box 32, 1518 Budapest 112, Hungary
[2] Department of Theoretical Chemistry, Eötvös University, P.O. Box 32, 1518 Budapest 112, Hungary

Received 10 January 2002
Published online 13 September 2002 –

Abstract. Chemical shielding anisotropy tensors have been determined, within the GIAO-RHF formalism using a smaller [6-31+G(d)] and two medium-size basis sets [6-311++G(d,p) and TZ2P], for all elements of the conformational library (altogether 27 structures) of the hydrophobic model peptide For–L-Phe–NH_2. The individual chemical shifts and their conformational averages have been compared to their experimental counterparts taken from the BioMagnetic Resonance Bank (BMRB). At the highest level of theory applied, for all nuclei but the amide proton, deviations between statistically averaged theoretical and experimental chemical shifts are as low as a few percent. One-dimensional (1D) chemical shift – structure plots do not allow unambiguous identification of backbone conformations. On the other hand, on chemical shift – chemical shift plots of selected nuclei, *e.g.*, $^{1}H^{N}$ with ^{15}N or ^{15}N with $^{13}C^{\alpha}$, regions corresponding to major conformational motifs have been found, providing basis for the identification of peptide conformers solely from NMR shift data. The 2D $^{1}H^{\alpha}$–$^{13}C^{\alpha}$ as well as the 3D $^{1}H^{\alpha}$–$^{13}C^{\alpha}$–$^{13}C^{\beta}$ chemical shift – chemical shift plots appear to be of special importance for direct determination of conformations of protein building units from multidimensional NMR experiments. 48 pairs of $^{1}H^{\alpha}$–$^{13}C^{\alpha}$ data for phenylalanine residues have been extracted from 18 selected proteins and compared to relevant *ab initio* results, supporting the calculated results. Thus, the appealing idea of establishing backbone folding information of peptides and proteins from chemical shift information alone, obtained from selected multiple-pulse NMR experiments (*e.g.*, 2D-HSQC, 2D-HMQC, and 3D-HNCA), has received further support.

PACS. 31.15.Ar Ab initio calculations – 33.25.+k Nuclear resonance and relaxation

1 Introduction

In the age of proteomics, when the full genes of over 800 species are known [1], among which the humane genome is undoubtedly the most important one, new approaches are required for the structure elucidation of biomolecules. For rational drug design structural biology requires more and more three-dimensional (3D) structures of proteins. In the near future an increased number of proteins will be expressed, produced in considerable quantities, and studied using different fast throughput screening techniques [2]. Thus, an important aim is the development of "fast" 3D structure determination methods for proteins at "low resolution". A quick and direct way to determine the backbone fold of proteins in solution from multidimensional nuclear magnetic resonance (NMR) spectroscopy would be extremely rewarding [3].

Today the full resonance assignment, a prerequisite for the commonly used structure determination procedures, of small or medium-size proteins (up to 15–30 kDa) with ^{13}C, ^{15}N (and perhaps ^{2}H) enriched nuclei can be achieved without any reliance on nuclear Overhauser effect (NOE) information. Resonance assignments are done through specific sets of standardized 3D and 4D NMR experiments [*e.g.*, HNCA, HN(CO)CA, CBCA(CO)NH, HBHA(CO)NH, CC(CO)NH, HCC(CO)NH, and HCCH-TOCSY] [4–6], all based on alternative coherence transfer strategies through homo- and heteronuclear couplings. Under optimal conditions the time required for the execution of a carefully selected set of experiments can be decreased to a few weeks. The most straightforward part of the subsequent automated resonance assignment and validation procedure is the sequential backbone assignment. Nevertheless, even if full resonance assignment is obtained without the use of NOE-type data the only routine technique available at present for structure determination of biomolecules relies on acquiring a large number of constraints, obtained primarily from NOE-type data converted to proton-proton distances [7–16]. These often redundant geometrical data can be augmented with

[a] e-mail: perczel@para.chem.elte.hu
[b] e-mail: csaszar@chem.elte.hu

selected dihedral angle, H-bond, and residual dipolar coupling information [17–21], providing an even more complete set of constraints for determination of the solution structure. This tedious part of the resonance assignment procedure is not yet automated and increases the time required for obtaining even a low-resolution 3D structure of a medium-size protein to months if not to years.

There are alternative routes to the use of NOEs for the quick determination of the mainchain fold. A promising approach is based on residual dipolar couplings [17–21]. The direct use of chemical shift (CS) information for structure determination offers another approach [22–39], investigated in detail below.

It is well established experimentally that the CS of a nucleus of an amino-acid residue is often different at different sites of the protein. (This is in fact what enables resonance assignments.) NMR signal dispersion is due to several factors, among which local conformation can be an important one. If the conformational effect dominates, the information on dihedral angles could be retrieved from CSs. An increasing number of chemical shifts of proteins are now available in databases (*e.g.*, the Biomagnetic Resonance Bank (BMRB) [40] and Ref. [41]), open for statistical analysis. Despite all the effort devoted to this topic [22–39], it is not yet clear whether CSs are good enough markers for all characteristic conformational elements that make up proteins and can thus circumvent the use of NOEs for structure determination. This possibility is further explored in this communication using For–L-Phe–NH_2 as our model.

Although longer polypeptides and especially proteins can exhibit a unique and often well-defined fold, their fragments and building units are inherently flexible. Few details are known about how torsional angles of amino acid residues take their "stable" values when incorporated in proteins. Unlike experiments, computations can determine, at least in principle, the library of all accessible conformers of any amino acid residue. We believe that considering each and every possible conformer is the only way to prove that CS information is indeed useful in itself for determination of the dihedral angle space of proteins. Therefore, theoretical conformational analyses of computationally accessible model systems, like For–L-Phe–NH_2, are of considerable utility and relevance [42,43] and must antedate related theoretical NMR investigations.

Experimental and computational studies have clearly shown a few structure-induced CS changes in proteins (*e.g.*, for ^{1}H with a range of ≈2 ppm, for $^{13}C^{\alpha}$ with a range of ≈8 ppm, for ^{15}N with a range of ≈25 ppm, and for ^{17}O with a range of ≈10 ppm). Among these relationships correlation of the α-helix and the β-sheet secondary structures with $^{13}C^{\alpha}$ CS alterations was established first [35,38,39]. Initiated by the study of Jiao and co-workers [44], it became clear that in glycine residues the $^{13}C^{\alpha}$ CS values shift with respect to their random coil value, δ_{rc}, both in the alpha helical conformation (≈ 2.3 ppm downfield shift) and in the extended conformation (≈ 2.9 ppm upfield shift). Alanine diamide models (*e.g.*, For–L-Ala–NH_2) [24,45] were the subject of further computational studies with similar results. Analysis of experimental chemical shift data during an unfolding study of myoglobin [46] has shown that for all Ala residues within the protein $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{1}H^{\alpha}$, and $^{1}H^{\beta}$ CS values tend to reach, under denaturing conditions (low pH and high urea concentration), their respective random coil values. Thus, when proteins are folded the individual CS values should reflect the secondary and tertiary structure of the macromolecule.

Amino acid residues larger than Ala need to be investigated in order to determine the effect the side chains have on the shielding properties of the peptide backbone. The hydrophobic amino acid residue next in size to Ala is valine (Val). We have recently published [47] an *ab initio* conformational library for valine diamide confirming that Val is also a good conformational model of amino acid residues Ile and Leu. Furthermore, we have found that gas-phase conformational preferences of these models show strong correlation with structures derived from the hydrophobic cluster(s) of proteins. In other words, conformers which are most frequently found in proteins are the ones with the lowest *ab initio* relative energies. Thus, we believe that NMR shielding properties of hydrophobic residues (*e.g.*, those of For–L-Val–NH_2) [25] make NMR CS computations even more closely related to experimental results than otherwise expected. De Dios and Oldfield [37c] concluded that substitution on C^{α} induces a larger shielding shift on C^{β} than on C^{α}. In another related study Pearson and co-workers [50] showed that in a protein like nuclease the $[\phi, \psi]$ values of Val residues can be estimated from C^{α} chemical shifts.

Although the number of experimental NMR chemical shifts for proteins increases rapidly, at present the most viable strategy for establishing and probing structure – chemical shift relationships is offered by the use of *ab initio* electronic structure theory [24–27,48–52]. Most *ab initio* computations related to the NMR spectroscopy of peptides and proteins have focused on the determination of chemical shielding tensors for selected nuclei and the calculation of chemical shielding isotropy (CSI), chemical shielding anisotropy (CSA), and CS values.

Nitrogen-, proton-, and C' (the carbon of the carbonyl group) CSs of the amide groups show a noticeable resonance shift due to sequential neighboring effect. (This holds both for folded and for unfolded states of proteins.) However, the C^{α} and H^{α} nuclei are likely to be the least dependent on sequential information, and the most dependent on local fold. This offers the possibility of determining the 3D structure of peptides and proteins.

The effect of intramolecular H-bonds on shieldings was established by the gauge-including atomic orbital restricted Hartree–Fock (GIAO–RHF) method [53, 54] on molecules containing only a few atoms (*e.g.*, N-methylacetamide interacting with formamide [55,56]) and on systems of considerably larger size (*e.g.*, For-$(Ala)_5$-NH_2) [56]. Not surprisingly, the most significant perturbation caused by H-bonding was observed on the carbon atom of the carbonyl groups directly involved in the formation of the hydrogen bond [57].

In summary, most conformationally dependent NMR calculations on peptide models suggest that it is possible to derive certain information on backbone (and possibly on side-chain) fold from C^{α} CS values. Having presented data on a typical hydrophobic but non-aromatic residue, valine [25], in this paper we provide *ab initio* computed CS information for phenylalanine (Phe), a hydrophobic and aromatic residue.

To the best of our knowledge no detailed NMR chemical shielding studies were published on phenylalanine containing peptide model systems. However, limited information on NMR properties of aromatic amino acid residues is available from the literature. One of the pioneering studies was performed by Toma *et al.* [58], who analysed the preferred solution structure and calculated conformers of dermorphin, a molecule containing three aromatic amino acid residues, two tyrosines and a phenylalanine. Recently, using two cyclic heptapeptides, both containing a phenylalanine residue, an *ab initio* study was published on the transferability of atomic solvation parameters [59]. Molecules as large as bradykinin, which contains two phenylalanine residues, were investigated using molecular mechanics to identify candidates of low-energy structures to assist NMR investigations [60]. The conformational properties of a cyclic retro-dipeptide along with its parent cyclic dipeptide, cyclo(L-Tyr-L-Phe), were studied by ^{1}H NMR and by semiempirical energy calculations [61].

In proteins most phenylalanines are buried, often contributing to the hydrophobic core region of the globular system, where water is mostly excluded. Thus, *ab initio* results, obtained for hydrophobic amino acid residues, are representative of "real" hydrophobic diamide units folded in aqueous media [25]. Our systematic effort already covers model systems that incorporate glycine, alanine, and valine amino acid residues [24,25]. Here we report a library consisting structural and CS information of phenylalanine using its diamide model, For–L-Phe–NH_2. It is our belief that by performing even a simple statistical analysis of the available data we have our best chance to filter out structure–structure, structure–chemical shift, and chemical shift–chemical shift correlations. We note in passing that in this paper we have not studied in an explicit manner the well-known aromatic ring effect on shieldings. Most of our effort was concentrated on finding and correlating NMR CSs of selected nuclei with the main-chain fold.

2 Computational details

Schematic representation of the constituent atoms of our model system, For–L-Phe–NH_2, is shown in Figure 1. Average chemical shifts, usually called random coil shift values, δ_{rc}, of all relevant nuclei are reported therein, as taken from the BioMagnetic Resonance Bank (BMRB) [40].

In search of the lowest appropriate theoretical level and basis set to calculate NMR chemical shielding anisotropy (CSA) tensors Laws and co-workers [62] have tested substandard (STO-3G) and split-valence basis sets [3-21G, 4-31G, and 6-311+G(2d)] for the Ala and Val amino acid

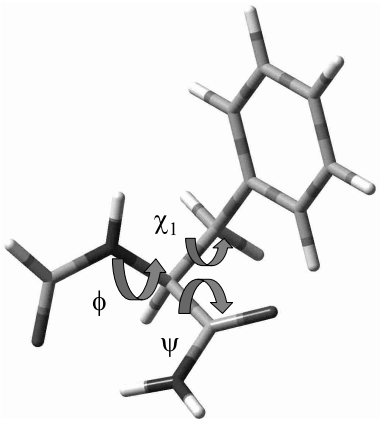


Fig. 1. The For–L-Phe–NH_2 peptide model.

residues. They found that chemical shifts determined by using smaller basis sets correlate well with data obtained with larger basis sets. This suggests that scaling CS data computed under different conditions is possible. Although Pearson *et al.* [63] have recently shown that DFT-type ^{13}C CS calculations are of considerable interest, we have restricted ourselves in this study to fully geometry optimized reference structures and to magnetic shielding tensor calculations with no electron correlation included.

Computation of NMR chemical shielding tensors was performed at the GIAO-RHF (gauge including atomic orbitals restricted Hartree–Fock) [53,54] level employing the following basis sets: 6-31+G(d) [64], 6-311++G(d,p) [64,65], and TZ2P [66]. The computations utilized the Gaussian94 [67] and Gaussian98 [68] electronic structure packages.

Geometries employed in this study for CSA computations have been optimized at the 3-21G RHF and 6-31+G(d) RHF levels of theory. We have established a library that covers all twenty-seven characteristic conformers of For–L-Phe–NH_2. Since some of these structures are clearly not minima on the PES of For–L-Phe–NH_2, for these conformers *constrained* geometry optimizations were performed keeping the ϕ and ψ torsional angles constant at characteristic values.

Three different sets of CSA calculations were generated in this study, labeled A, B and C, as detailed in Scheme 1. Tables 1 to 3 report ϕ, ψ, χ^1, and χ^2 parameters [69], energies, and isotropic chemical shielding values (σ-scale) for these model conformers. For relative CSs (δ-scale) the isotropic chemical shielding values of ^{1}H, ^{13}C, and ^{15}N were referenced to the ^{1}H and ^{13}C values of tetramethylsilane (TMS) and to the ^{15}N value of NH_3. (The geometry of NH_3 was computed at the all-electron aug-cc-pVTZ CCSD(T) level [70], while that of TMS at the 3-21G RHF level.)

Chemical shift values of Phe associated with a random conformation, δ_{rc}, were taken from BMRB [40] (*cf.* Fig. 1 and Tab. 3). Chemical shift re-referencing, when needed, was done in the group of Wishart and was kindly provided for us for the following 18 proteins: 1BPI, 3LZM, 1LZ1, 2RN2, 2RNT, 1SNC, 1HCB, 1UBQ, 1CEX, 1GZI, 5P21,

Sheme 1. Designation of the levels of theory employed in this work for the For–L-Phe–NH_2 model system and the resulting number of structures.

Level	NMR computation	Geometry optimization	No. of structures	Comment
A	6-31+G(d) GIAO-RHF	3-21G RHF	19	full opt. $[\phi, \psi]$ only
B	6-311++G(d,p) GIAO-RHF	6-31+G(d) RHF	16	full opt. $[\phi, \psi]$ only
C1	TZ2P GIAO-RHF	6-31+G(d) RHF	16	full opt. $[\phi, \psi]$ only
C2	TZ2P GIAO-RHF	6-31+G(d) RHF	11	constr. opt. $[\phi, \psi]$ only
C3	TZ2P GIAO-RHF	6-31+G(d) RHF	27	full opt. plus $[\phi, \psi]$ constr.

1ROP, 1ICM, 192L, 1IGD, 3RN3, 2TRX, and 1A2P. Simultaneously, the 3D structures of the same proteins were retrieved from the Protein Data Bank (PDB) [71]. These CS and structural data were used for testing theoretically computed values. (Note that C^{α} CSs were not available for proteins 1LZ1, 2RNT, 1GZI, 1A2P, and 192L. Furthermore, for 2TRX and 1A2P there are two and three sets of PDB coordinates, respectively, with slightly different values.) After omitting all questionable entries, a total of 48 Phe and 48 Tyr residues remained in our database, all with unambiguous ^{1}H and ^{13}C CSs as well as dihedral angle values. We believe that these 48 Phe and 48 Tyr residues form the largest database used for a comprehensive analysis; nevertheless, as turns out from the present study, an even larger experimental database would be desirable to reach definitive conclusions about structure – chemical shift correlations.

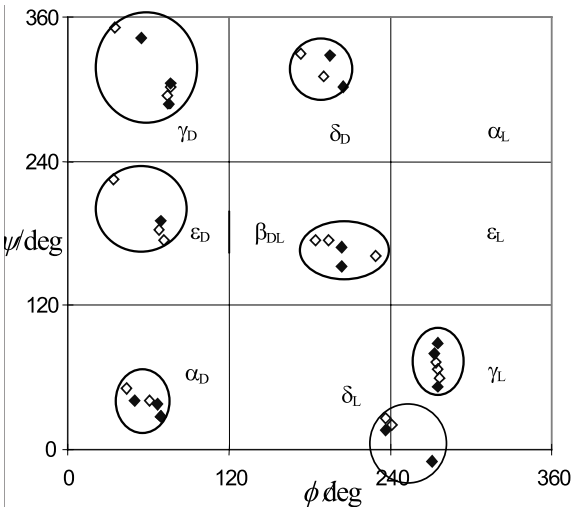


Fig. 2. Locations of different minima of For–L-Phe–NH_2 on a Ramachandran surface determined at the 3-21G RHF [open symbols] and 6-31+G(d) RHF [filled symbols] levels of theory.

3 Structures and energetics

The values of $[\phi, \psi]_i$ torsional angle pairs pinpoint the positions of residue i [72,73] on a Ramachandran surface and determine the main-chain fold of a peptide or protein. Consideration of torsions around the N–C^{α} and the C^{α}–C' bonds within an alpha amino acid residue results in three different orientations of ϕ and ψ. Therefore, nine backbone conformers are expected for each peptide unit –NH–C^{α}HR–C'O– [42,74–76]. These nine easily distinguishable backbone orientations are named [42,74] as follows: $\alpha_{\rm L}$, $\alpha_{\rm D}$, $\beta_{\rm L}$, $\gamma_{\rm L}$, $\gamma_{\rm D}$, $\delta_{\rm L}$, $\delta_{\rm D}$, $\varepsilon_{\rm L}$, and $\varepsilon_{\rm D}$. Location of these 9 conformations of Phe on the Ramachandran surface, determined at two different levels of theory, are given in Figure 2. X-ray diffraction data of proteins confirm the existence of all nine backbone conformations but with very different natural abundance [77]. The $\beta_{\rm L}$- (extended), $\gamma_{\rm L}$- (inverse γ-turn), $\delta_{\rm L}$-, and $\varepsilon_{\rm L}$-type (polyproline II) conformers commonly form the well-known and rather broad β-region, while the $\alpha_{\rm L}$ part, which comprises both 3_{10}- and 4_{13}-helices form the right-handed helical region on the Ramachandran surface. Conformational building units such as $\alpha_{\rm D}$, $\gamma_{\rm D}$, $\delta_{\rm D}$, and $\varepsilon_{\rm D}$ are less frequently assigned in proteins, since they are built from amino acid residues of L relative chirality.

Ab initio grid searches carried out for amino acid diamide models such as For–Xxx–NH_2 and Ac–Xxx–NHMe (where Xxx = Gly [24], Ala [24,78], Ser [79], and Phe [80]) resulted in slightly fewer minima than expected by the above arguments. For example, at the 3-21G RHF level two, while at the 6-311++G(d,p) RHF level three out of the nine backbone conformers of For–L-Ala–NH_2 have not been found [42]. The *ab initio* calculations revealed that the $\alpha_{\rm L}$, $\varepsilon_{\rm L}$, and $\varepsilon_{\rm D}$ conformers are absent most often in the diamide models. Nevertheless, these "missing" backbone orientations do show up when polar side-chains, *e.g.*, in Ser or His [81], are present.

Phe is the simplest representative of the hydrophobic and aromatic amino acid residue group containing Phe, Tyr, and Trp. Since the side chain of Phe can formally be derived from the methyl side-chain of Ala by replacing one H^{β} with a phenyl group, the topological features of For–L-Phe–NH_2 are somewhat similar to the parent amino acid residue. Phenylalanine adopts only a single and slightly shifted *gauche*+-like χ^2 orientation ($\approx 90 \pm 15°$). Thus, the systematic conformational search of phenylalanine dipeptide models (For–L-Phe–NH_2 or Ac–L-Phe–$NHCH_3$) can be performed in three torsional variables: $E = E(\phi, \psi, \chi^1)$. Similarly to valine [82], the 4D torsional potential energy surface of For–L-Phe–NH_2 is expected to have as many as 27 characteristic conformers (three χ^1 orientations for each of the nine typical backbone conformers) [80]. Certain aspects of structural and energetic features of the For–L-Phe–NH_2 model system have already been discussed based on 3-21G RHF [80a] and 6-31+G(d) RHF data [80b]. Here we wish to recapitulate the following: (a) out of the 27 expected structures, 19 conformers at the 3-21G RHF level (most of the $\alpha_{\rm D}$, $\beta_{\rm L}$, $\gamma_{\rm L}$, $\gamma_{\rm D}$, $\delta_{\rm L}$, $\delta_{\rm D}$, and $\varepsilon_{\rm D}$ backbone

Table 1. Selected conformational parameters and relative[a] isotropic chemical shieldings (σ) of For–L-Phe–NH_2 [b].

conf.[c]	Level[d]	ϕ	ψ	χ^1	χ^2	^{15}N	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C'$	$^{1}H^{N}$	$^{1}H^{\alpha}$	$^{1}H^{\beta}_{1}$	$^{1}H^{\beta}_{2}$
$\alpha_D(g+)$	C3	50.1	41.3	52.1	82.1	10.35	–6.76	–3.33	1.34	0.52	0.93	–0.16	0.27
	B					10.26	–6.89	–3.33	1.52	0.39	0.88	–0.18	0.24
	A	43.1	50.3	44.6	78.1	12.30	–4.13	–3.20	4.39	0.26	0.90	–0.12	0.25
$\alpha_D(a)$	C3	66.6	37.9	–139.6	93.5	–1.43	–8.25	–6.47	1.29	–0.36	1.41	1.66	–2.02
	B					–2.14	–8.11	–6.58	1.67	–0.45	1.32	1.69	–2.03
	A	60.6	41.1	–144.6	101.8	0.49	–7.42	–6.81	3.04	–0.65	1.32	1.76	–2.34
$\alpha_D(g-)$	C3	69.6	27.0	–57.0	105.1	11.21	–9.07	1.18	2.10	0.59	1.23	0.54	–1.20
	B					10.85	–8.96	0.89	2.24	0.51	1.14	0.54	–1.21
	A	68.6	28.2	–56.7	98.9	11.74	–8.23	0.64	4.21	0.36	1.17	0.15	–1.17
$\alpha_L(g+)$[f]	C3	–54.0	–45.0	49.9	85.1	7.15	–4.24	–1.81	1.42	0.01	0.21	0.60	–0.38
$\alpha_L(a)$	C3	–54.0	–45.0	–173.4	80.1	1.19	–8.11	–5.40	3.71	0.26	0.54	1.26	–0.21
$\alpha_L(g-)$	C3	–54.0	–45.0	–60.0	111.0	6.53	–7.73	–4.40	1.06	0.35	0.48	1.41	–0.85
$\beta_L(g+)$	C3	–156.1	168.1	57.3	86.6	10.03	–2.05	–2.62	1.56	–1.03	0.14	1.18	–1.04
	B					9.64	–2.31	–2.93	1.76	–1.05	0.08	1.15	–1.05
	A	–175.9	174.6	53.4	84.4	10.52	–2.36	–0.97	2.03	–1.35	0.27	1.49	–1.86
$\beta_L(a)$	C3	–156.5	152.2	–167.9	71.2	3.61	–4.19	–5.98	0.53	–1.57	0.24	0.39	–0.01
	B					3.07	–4.45	–6.21	0.64	–1.56	0.19	0.39	–0.04
	A	–166.6	173.9	–146.6	68.5	2.83	–5.72	–6.39	1.47	–2.18	0.39	0.41	–0.03
$\beta_L(g-)$	C3	–156.0	160.0	–59.6	101.6	10.99	–2.15	–9.93	1.31	–1.05	–0.02	1.43	–0.46
	A	–130.4	160.9	–61.9	91.5	9.95	–1.37	–8.60	0.79	–1.20	–0.24	1.45	–0.34
$\gamma_D(g+)$	C3	55.2	–18.1	68.7	81.3	3.86	–11.48	–7.59	–0.82	0.28	0.78	0.23	–0.06
	B					4.11	–11.74	–7.70	–0.83	0.16	0.70	0.18	–0.09
	A	34.8	8.6	68.6	84.8	4.50	–12.25	–8.43	2.81	0.34	0.44	0.07	0.04
$\gamma_D(a)$	C3	74.7	–72.2	–167.7	80.7	1.30	–13.39	–3.92	3.02	–0.47	1.08	1.17	–1.52
	B					1.17	–13.19	–4.15	2.64	–0.49	1.03	1.14	–1.57
	A	74.5	–65.2	172.6	88.8	2.74	–12.64	–3.47	2.92	–0.62	1.00	1.31	–1.82
$\gamma_D(g-)$	C3	75.9	–54.5	–59.9	105.5	7.71	–14.55	–0.40	1.66	0.25	1.02	0.17	–0.44
	B					7.95	–14.33	–0.62	1.41	0.22	0.97	0.17	–0.45
	A	76.2	–57.6	–57.3	100.3	7.16	–13.44	–0.87	1.21	0.00	0.94	0.01	–0.42
$\gamma_L(g+)$	C3	–85.4	52.5	44.9	80.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	B					0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	A	–83.0	59.8	40.6	76.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\gamma_L(a)$	C3	–85.1	89.2	–168.4	84.7	–3.78	–2.83	–2.05	2.03	–0.19	0.21	1.42	–0.98
	B					–3.98	–2.81	–2.18	2.16	–0.19	0.22	1.40	–1.01
	A	–85.6	72.5	–162.0	94.4	–4.75	–4.02	–1.72	1.78	–0.38	0.20	1.73	–1.41
$\gamma_L(g-)$	C3	–86.9	79.4	–59.1	114.3	–0.24	–1.25	–1.37	1.98	0.19	0.13	0.90	–0.61
	B					–0.24	–1.20	–1.55	1.90	0.14	0.12	0.88	–0.61
	A	–84.4	66.8	–55.1	111.5	–0.70	–1.76	–1.62	1.92	0.02	0.10	0.93	–0.84
$\delta_D(g+)$	C3	–164.4	–32.0	62.2	95.2	9.53	–8.09	–1.46	–1.04	0.14	0.00	1.06	–1.53
	B					9.27	–8.23	–1.60	–1.09	0.14	–0.02	1.04	–1.57
	A	173.8	–31.1	58.0	89.7	7.19	–7.52	–1.41	0.51	0.26	0.13	1.04	–2.21
$\delta_D(a)$	C3	–154.8	–57.4	–174.5	74.8	4.85	–9.23	–5.69	3.37	–0.28	–0.02	0.09	0.33
	B					4.40	–9.17	–5.91	3.25	–0.23	–0.03	0.06	0.29
	A	–169.9	–49.6	–149.4	76.2	4.03	–10.41	–1.41	3.50	–0.22	0.29	0.20	0.00
$\delta_D(g-)$	C3	–159.0	–45.0	–69.8	108.7	9.61	–7.94	–9.94	0.99	–0.05	–0.27	1.74	–0.61
$\delta_L(g+)$	C3	–123.3	15.7	54.4	82.1	5.81	–0.56	–1.65	0.64	0.36	–0.53	–0.14	0.02
	B					5.89	–0.69	–1.68	0.85	0.27	–0.53	–0.14	0.02
	A	–123.1	25.6	52.4	83.1	6.19	0.18	–1.21	1.44	0.29	–0.76	–0.13	0.00
$\delta_L(a)$	C3	–103.0	0.0	–153.4	75.2	–3.74	–2.60	–7.80	0.65	0.19	–0.27	1.30	–0.36
$\delta_L(g-)$	C3	–83.8	–15.7	–59.2	113.7	4.15	–5.67	–3.72	0.19	0.43	0.23	1.59	–1.23
	B					3.49	–5.80	–3.91	0.56	0.33	0.21	1.55	–1.23
	A	–119.6	23.2	–60.0	102.4	4.15	–2.61	–4.63	2.02	0.45	–0.56	1.84	–1.63
$\varepsilon_D(g+)$	C3	60.0	–120.0	59.1	103.7	6.53	–6.20	–4.65	–0.96	–0.62	1.03	0.87	–0.49
	A	34.0	–134.3	71.5	175.7	7.21	–3.17	–5.19	6.94	–0.15	0.66	0.79	–0.85
$\varepsilon_D(a)$	C3	68.5	–169.2	–153.8	59.1	5.13	–8.06	–4.65	2.26	0.00	1.14	0.61	–1.53
	B					4.96	–8.20	–4.59	2.65	–0.13	1.07	0.59	–1.57
	A	67.6	–177.6	–158.3	64.5	7.26	–7.86	–5.72	4.58	–0.15	0.97	0.66	–1.94
$\varepsilon_D(g-)$	C3	60.0	–120.0	–62.6	101.3	7.83	–9.06	–0.46	–0.61	–0.17	1.27	0.50	–0.43
	A	71.5	174.4	–58.4	97.7	14.05	–8.98	–2.86	5.39	0.54	1.14	–0.74	–0.06
$\varepsilon_L(g+)$[g]	C3	–60.0	120.0	41.5	83.1	3.86	0.76	–1.39	–3.66	0.15	0.57	0.80	–0.31
$\varepsilon_L(a)$	C3	–60.0	120.0	–172.8	83.5	–0.56	–5.39	–3.28	–1.77	–0.02	0.73	1.21	–0.54
$\varepsilon_L(g-)$	C3	–60.0	120.0	–59.1	112.7	1.71	–4.00	–2.54	–2.64	0.05	0.56	1.04	–0.54

[a] All relative isotropic shielding values are referenced, as clear from the table, to the absolute isotropic shieldings of the $\gamma_L(g+)$ conformer. Isotropic shielding values (in ppm) for the $\gamma_L(g+)$ conformer are as follows: level A: $^{15}N = 157.43$, $^{13}C^{\alpha} = 154.44$, $^{13}C^{\beta} = 170.92$, $^{13}C' = 27.23$, $^{1}H^{N} = 28.62$, $^{1}H^{\alpha} = 28.64$, $^{1}H^{\beta}_{1} = 28.96$, and $^{1}H^{\beta}_{2} = 30.49$; level B: $^{15}N = 150.26$, $^{13}C^{\alpha} = 136.15$, $^{13}C^{\beta} = 155.65$, $^{13}C' = 13.98$, $^{1}H^{N} = 28.35$, $^{1}H^{\alpha} = 28.99$, $^{1}H^{\beta}_{1} = 28.85$, and $^{1}H^{\beta}_{2} = 30.00$; and level C3: $^{15}N = 151.58$, $^{13}C^{\alpha} = 136.03$, $^{13}C^{\beta} = 155.51$, $^{13}C' = 13.48$; $^{1}H^{N} = 28.16$, $^{1}H^{\alpha} = 28.84$, $^{1}H^{\beta}_{1} = 28.69$, and $^{1}H^{\beta}_{2} = 29.82$. [b] Symbols for relevant conformational types: α_L, β_L, ε_L etc. for backbone, and $g+$, a or $g-$ for the side-chain (χ^1) orientation. [c] See Scheme 1 for details concerning the computational levels. [d] All torsional values in degrees, all chemical shifts in ppm. [f] $\phi = -54°$, $\psi = -45°$ are typical values for helical secondary structural elements in globular proteins. [g] $\phi = -60°$, $\psi = 120°$ are typical values for polyproline II secondary structural element in globular proteins.

Table 2. Selected structural and relative chemical shift (δ) values of For–L-Phe–NH_2 obtained by conformational averaging. (A) Averaged over the individual backbone conformers. (B) Averaged over all backbone conformers using arithmetical averaging. (C) Averaged over all backbone conformers using Boltzman averaging.

bb	sc[a]	Level[b]	^{15}N[c]	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C'$	$^{1}H^{N}$	$^{1}H^{\alpha}$	$^{1}H^{\beta}{}_{1}$	$^{1}H^{\beta}{}_{2}$	ϕ	ψ	ΔE[d]	$p(i)/\Sigma p(i)$[e]
α_L	3	C3	116.02	53.87	35.46	178.32	4.14	3.76	2.68	2.83	–54.00	–45.00	6.07	0.00
α_D	3	A	101.33	53.39	33.44	170.13	4.19	3.03	3.24	3.40	57.42	39.87	6.74	0.00
	3	B	110.31	55.66	35.22	178.94	4.08	3.01	3.06	3.33	62.12	35.39	6.82	0.00
	3	C3	114.27	55.20	34.46	178.81	4.10	2.98	3.09	3.33	62.12	35.39	6.68	0.00
β_L	3	A	101.74	49.95	35.64	172.58	5.76	4.02	2.72	3.05	–157.62	169.81	1.26	0.35
	2	B	110.28	51.05	36.78	179.55	5.53	3.99	2.98	2.87	–156.28	160.12	1.02	0.65
	3	C3	112.77	49.97	37.76	179.25	5.57	4.05	2.77	2.85	–156.19	160.08	2.09	0.64
γ_L	3	A	111.32	48.73	31.43	172.78	4.30	4.06	2.95	3.06	–84.34	66.36	0.52	0.56
	3	B	118.04	49.01	33.45	179.40	4.24	4.01	2.99	2.87	–85.78	73.70	1.14	0.30
	3	C3	122.32	48.54	32.73	179.05	4.35	4.06	3.00	2.88	–85.78	73.70	1.12	0.30
γ_D	3	A	104.71	59.58	34.58	171.70	4.27	3.37	3.38	3.04	61.87	–43.79	5.02	0.02
	3	B	112.22	60.76	36.37	179.68	4.26	3.23	3.25	3.03	68.57	–48.25	5.27	0.01
	3	C3	116.69	60.32	35.56	179.10	4.33	3.21	3.25	3.02	68.57	–48.25	5.24	0.01
δ_L	2	A	104.34	48.01	33.24	172.28	3.81	4.82	2.99	3.13	–121.36	24.41	1.52	0.08
	2	B	111.94	50.91	35.00	180.04	3.93	4.29	3.04	2.93	–103.57	–0.02	2.38	0.03
	3	C3	118.90	50.12	35.98	179.89	4.02	4.36	2.85	2.87	–103.38	–0.01	3.92	0.03
δ_D	2	A	103.90	55.76	33.66	172.00	4.16	3.95	3.22	3.42	–178.05	–40.36	7.11	0.00
	2	B	109.80	56.37	35.96	179.67	4.27	4.15	3.20	2.97	–159.55	–44.66	6.53	0.00
	3	C3	112.98	55.60	37.28	179.28	4.41	4.27	2.81	2.95	–159.37	–44.77	7.95	0.00
ε_L	3	C3	119.31	50.05	33.99	183.08	4.29	3.55	2.75	2.81	–60.00	120.00	3.20	0.02
ε_D	3	A	100.00	53.47	34.91	168.37	4.10	3.24	3.60	3.26	57.69	–165.82	9.25	0.00
	1	B	111.67	55.87	36.80	178.10	4.36	3.06	3.16	3.90	68.53	–169.18	6.95	0.00
	3	C3	114.48	54.95	34.84	180.16	4.61	3.02	3.11	3.17	62.84	–136.39	8.28	0.00

(A)

bb. averages	Level[b]	^{15}N[c]	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C'$	$^{1}H^{N}$	$^{1}H^{\alpha}$	$^{1}H^{\beta}{}_{1}$	$^{1}H^{\beta}{}_{2}$
Average[f] (in ppm)	A	103.88	52.78	33.88	171.33	4.41	3.72	3.16	3.19
	B	112.34	54.30	35.46	179.42	4.35	3.67	3.09	3.07
	C1	116.47	53.77	34.70	179.13	4.42	3.67	3.10	3.07
	C2	116.33	52.33	36.28	180.43	4.43	3.73	2.67	2.82
	C3	116.42	53.18	35.34	179.66	4.43	3.70	2.92	2.97
Standard deviation[g] (in ppm)	A	4.91	4.30	2.80	1.80	0.71	0.59	0.77	0.89
	B	4.54	4.43	2.50	1.24	0.55	0.55	0.62	0.75
	C1	4.24	4.50	2.51	1.26	0.56	0.56	0.58	0.70
	C2	4.55	3.04	3.30	2.12	0.41	0.49	0.38	0.17
	C3	4.44	3.97	2.91	1.75	0.51	0.53	0.57	0.59
Difference[h] (in ppm)	A	16.70	5.41	5.95	4.25	4.00	0.89	–0.21	-0.20
	B	8.24	3.89	4.37	–3.84	4.06	0.94	–0.14	-0.08
	C1	4.11	4.42	5.13	–3.55	3.99	0.94	–0.15	-0.08
	C2	4.25	5.86	3.55	–4.85	3.98	0.88	0.28	0.17
	C3	4.16	5.01	4.49	-4.08	3.98	0.91	0.03	0.02
Accuracy[i] (%)	A	14	9	15	2	48	19	7	7
	B	7	7	11	2	48	20	5	3
	C1	3	8	13	2	47	20	5	3
	C2	4	10	9	3	47	19	10	6
	C3	3	9	11	2	47	20	1	1
Experimental average[j]		120.58	58.19	39.83	175.58	8.41	4.61	2.95	2.99
Experimental standard deviation[k]		4.29	2.66	1.88	2.17	0.73	0.57	0.41	0.37

(B)

bb. averages	Level[b]	$^{15}N^{NH}$[c]	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C'$	$^{1}H^{N}$	$^{1}H^{\alpha}$	$^{1}H^{\beta}{}_{1}$	$^{1}H^{\beta}{}_{2}$
Average [f] (in ppm)	A	108.63	50.12	33.40	172.45	4.94	4.00	3.14	2.83
	B	114.75	51.26	36.75	179.76	5.23	3.96	3.18	2.59
	C1	118.78	50.59	35.90	179.47	5.35	3.97	3.19	2.60
	C2	117.47	51.39	37.47	179.94	5.87	3.92	3.35	2.38
	C3	118.81	50.62	35.87	179.53	5.33	3.96	3.17	2.61
Difference [h] (in ppm)	A	11.95	8.07	6.43	3.13	3.47	0.61	-0.19	0.16
	B	5.83	6.93	3.08	-4.18	3.18	0.65	-0.23	0.40
	C1	1.80	7.60	3.93	-3.89	3.06	0.64	-0.24	0.39
	C2	3.11	6.80	2.36	-4.36	2.54	0.69	-0.40	0.61
	C3	1.77	7.57	3.96	-3.95	3.08	0.65	-0.22	0.38
Accuracy [i] (%)	A	10	14	16	2	41	13	6	5
	B	5	12	8	2	38	14	8	14
	C1	1	13	10	2	36	14	8	13
	C2	3	12	6	2	30	15	14	20
	C3	1	13	10	2	37	14	8	13
Experimental average[j]		120.58	58.19	39.83	175.58	8.41	4.61	2.95	2.99
Experimental standard deviation[k]		4.29	2.66	1.88	2.17	0.73	0.57	0.41	0.37

(C)

[a] sc. stands for the number of side-chain conformers used when chemical shift isotropy (CSI) values (relative to TMS) were arithmetically averaged (for CSI data see Tab. 1). [b] See Scheme 1 for description of theoretical levels. [c] Absolute chemical shielding values are relative to TMS or NH_3. The ^{1}H(TMS), ^{13}C(TMS), and $^{15}N(NH_3)$ values for level A (GIAO-RHF/6-31+G*) are 32.800, 201.239, and 266.937, respectively. The ^{1}H(TMS), ^{13}C(TMS), and $^{15}N(NH_3)$ values for level B (GIAO-RHF/6-311++G**) are 32.415, 195.559, and 262.784, respectively. The ^{1}H(TMS), ^{13}C(TMS), and ^{15}N(NH3) values for level C (GIAO-RHF/TZ2P) are 32.231, 194.688, and 268.698, respectively. [d] Averages of relative energies determined at 6-31+G* RHF (level A), 6-311++G** RHF (level B) or TZ2P RHF (level C) levels of theory. [e] Relative populations are calculated as $\exp(-\Delta E/kT)/\Sigma \exp(-\Delta E/kT)$, where $kT = 0.595371$ kcal mol^{-1} at $T = 300$ K. [f] In level A 19 fully optimized, in level B 16 fully optimized, in level C1 16 fully optimized, in level C2 11 partially optimised plus the $\gamma_L(g+)$, while in level C3 all 27 conformers were used. [g] Standard deviation of the calculated CSI values. [h] Difference between experimentally and *ab initio* determined CSI averages. [i] Accuracy (in %) = $(CSI^{\text{exp.}} - CSI^{\text{calc.}})/CSI^{\text{exp.}}$. [j] Experimental average shift, data taken from BMRB 02-02-1999. [k] Standard deviation of experimental average shifts, data taken from BMRB 02-02-1999.

Table 3. R^2 values determined between selected conformational and chemical shift isotropy values of For–L-Phe–NH_2 [a].

	ω_0	ϕ	ψ	ω_1	χ^1	χ^2	$^1H^N$	^{15}N	$^{13}C^\alpha$	$^1H^\alpha$	$^{13}C'$	$^{13}C^\beta$	$^1H^{\beta1}$	$^1H^{\beta2}$
ω_0	1	0.072	0.003	0.329	0.070	0.000	0.000	0.018	0.051	0.113	0.058	0.037	0.038	0.002
ϕ		1	0.019	0.011	0.000	0.005	0.006	0.129	**0.407**	**0.443**	0.031	0.002	0.093	0.046
							(**0.339**)	(**0.375**)	(**0.407**)	(**0.816**)	(0.202)	(0.185)	(0.093)	(0.333)
ψ			1	0.002	0.002	0.000	0.015	0.102	0.333	0.009	0.021	0.042	0.000	0.003
							(**0.408**)	(0.273)	(**0.382**)	(0.145)	(0.208)	(0.139)	(0.065)	(0.083)
ω_1				1	0.002	0.013	0.007	0.010	0.000	0.005	0.019	0.000	0.010	0.035
χ^1					1	0.302	0.002	0.001	0.088	0.027	0.054	0.025	0.180	0.078
							(0.155)	(0.323)	(0.229)	(0.044)	(0.190)	(0.248)	(0.180)	(0.235)
χ^2						1	0.043	0.073	0.006	0.002	0.039	0.018	0.137	0.043
							(0.164)	(0.227)	(0.057)	(0.078)	(0.070)	(0.115)	(0.137)	(0.169)
$^1H^N$							1	0.001	0.024	0.006	0.004	0.129	0.027	0.001
^{15}N								1	0.049	0.003	0.001	0.003	0.053	0.003
$^{13}C^\alpha$									1	0.340	0.111	0.010	0.002	0.064
$^1H^\alpha$										1	0.000	0.047	0.004	0.132
$^{13}C'$											1	0.023	0.000	0.013
$^{13}C^\beta$												1	0.175	0.004
$^1H^{\beta1}$													1	0.372
$^1H^{\beta2}$														1

[a] A total of 27 conformers were used to generate the data. Theoretical shifts refer to level C3 (GIAO-RHF/TZ2P//RHF/6-31+G*) values. All conformational variables (ϕ, ψ, ω_0, ω_1, and χ^1) are scaled between 0° and 360°. Auto-correlation values, *e.g.* $R^2[\phi/\phi]$, are all equal to 1. The average of all R^2 values, $\langle R^2\rangle$, is 0.065 with a standard deviation (σ) of 0.109. All values larger than $\langle R^2\rangle + 2.5\sigma$ ($0.065 + 0.272 = 0.338$) are marked in bold. Selected maximized $R^2_{\max}$ values are given in parentheses, obtained by shifting systematically (with an increments of 5°) the periodic unit of the conformational variables ϕ, ψ, and χ^1.

conformers plus the $\delta_D(g-)$ and $\delta_L(a)$ conformers) and 16 at the 6-31+G(d) RHF level were found to be minima; at the latter level of theory the $\beta_L(g-)$, $\varepsilon_D(g-)$, and $\varepsilon_D(g+)$ structures are no longer minima; (b) conformational variation at the RHF level induced by the increase in the basis set size is marginal (see Fig. 2); (c) dihedral angles describing the fold of these conformers correlate well with data derived from protein structures determined by X-ray diffraction [47]; and (d) the order of *ab initio* calculated energies of the conformers correlates well with the relative natural abundances of conformers derived from experimental data [47]. In this study the 8 and 11 structures of For–L-Phe–NH_2 not found to be minima at the 3-21G RHF and 6-31+G(d) RHF levels, respectively, were optimized with constrained $[\phi, \psi]$ torsion angles (see Tab. 1).

4 Accuracy of computed chemical shifts

In any computational study of chemical shifts one might ask how well the calculated CSs compare with their experimental counterparts. The simplest way to answer this question is to check the accuracy of the appropriate averages of the computed values against the experimental numbers obtained by averaging over the entire conformational range. The experimental average NMR CSs, the random coil values, δ_{rc}, with associated standard deviations for all nuclei of phenylalanine are taken from the BMRB [40] (bottom two lines of Tab. 2B). The computed averages were obtained either by a simple arithmetical averaging or by a Boltzmann-type (energy-weighted) averaging and are presented in Table 2 and in Figure 3.

The agreement between the computed and the experimental CSs is good. The agreement is particularly impressive for *ab initio* data computed at the highest level of theory used in this study (level C3). In general, increase of the theoretical level for the shielding calculation and for structure optimization decreases the deviation between the theoretical and experimental numbers for all nuclei but the amide proton. For example, the difference between the computed and observed $\delta_{rc}(^{15}N)$ can be as small as 1.77 ppm [118.81 ppm (level C3) *vs.* 120.58 ppm (expt.)] (Tab. 2C), while for $\delta_{rc}(^1H^\alpha)$ and $\delta_{rc}(^1H^\beta)$ the difference between the *ab initio* and the experimental values is only ≈0.5 ppm. Reproduction of $\delta^{expt}_{rc}(C^\alpha)$ is not nearly as good with a ≈7.5 ppm or about 13% deviation (Tab. 2C). All these observations are similar to conclusions obtained for For–Gly–NH_2 [24], For–L-Ala–NH_2 [24], and For–L-Val–NH_2 [25]. It is unclear why the error between calculated and observed amide proton ($^1H^N$) CSs is three times larger (≈35–40%) than for any other nucleus. Perhaps some insight is gained by noting that this large error is associated with the most acidic (*i.e.*, deshielded) proton of the molecule. It looks as if deshielding decreases while shielding increases the accuracy of the computed data: $^1H^\alpha$ and $^1H^\beta$ CSs are calculated with considerably higher precision than $^1H^N$. Although the *ab initio* calculated $^1H^N$ and $^{13}C^\alpha$ chemical shift data are significantly different from their experimental counterparts even at higher levels of theory, they remain sensitive toward conformation-induced changes (*vide infra*), which is undoubtedly the most important issue considered in this study.

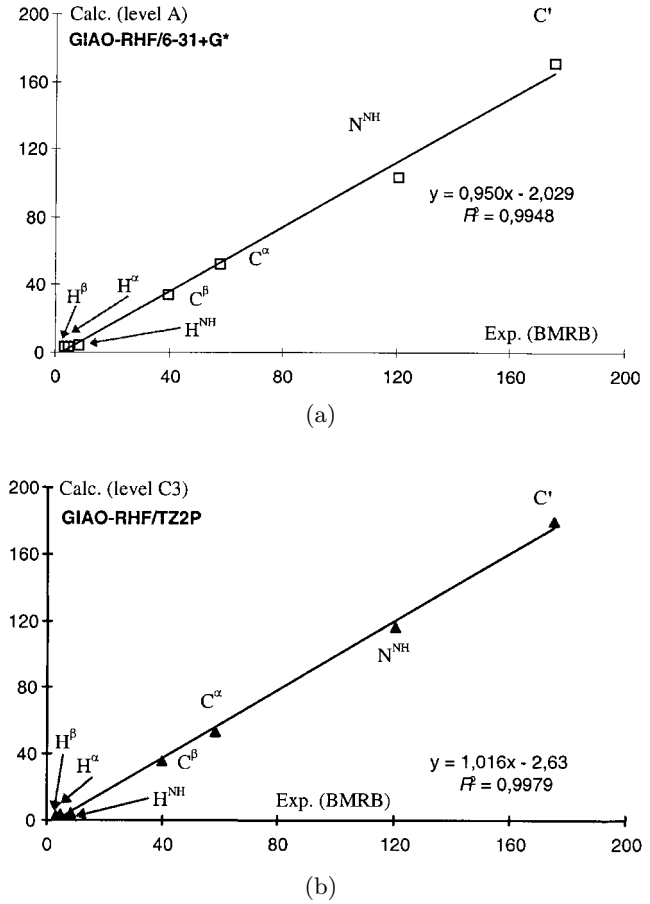


Fig. 3. Correlation between *ab initio* determined and experimentally found average (conformation independent) chemical shifts (in ppm) for selected atoms of For–L-Phe–NH_2 at two levels of theory ([a] GIAO-SCF/6-31+G*//SCF/3-21G (level A) and [b] GIAO-RHF/TZ2P//RHF/6-31+G* (level C3)).

5 Chemical shift – chemical shift correlation maps

From experimental $^1H^N$–^{15}N heterocorrelation studies it is well known [83] that for Gly, Ser, and Thr residues, compared to all other natural amino acids, a characteristic nitrogen upfield shift is observable. This shift, however, is not the type of information we plan to utilize; characteristic up- or downfield shift of selected resonances, induced by sequential and/or side-chain effects, is hoped to be conformationally invariant and thus correctable in a simple manner.

From our point of view the interesting type of CS change is the one which occurs as a function of the main-chain fold. As found for other types of amino acids, the simple rotation of the adjacent amide planes can strongly influence CSs. For example, in the case of Phe the $\gamma_L(a)$ conformer differs from the $\gamma_D(a)$ structure both in its ϕ and ψ values, which implies a close to 180° rotation along both axes: $-85.1°(\phi_{\gamma_L}) \Rightarrow +74.7°(\phi_{\gamma_D})$ and $+89.2°(\psi_{\gamma_L}) \Rightarrow -72.2°(\psi_{\gamma_D})$. As a result, the CS of C^α is altered by 10.6 ppm and that of ^{15}N by 5.1 ppm (Tab. 1). This is definitely a significant shift and fortunately not the only example of such a large-scale change. The structural shift between two conformational neighbors [*e.g.*, $\gamma_L(g+) \Rightarrow \beta_L(g+)$] can also result in a shift of considerable magnitude. Although the latter two backbone conformers differ from each other only by some 75°, at level C3 the relevant conformational shifts are $\Delta^{15}N = 10$ ppm and $\Delta^{13}C^\beta = -2.6$ ppm.

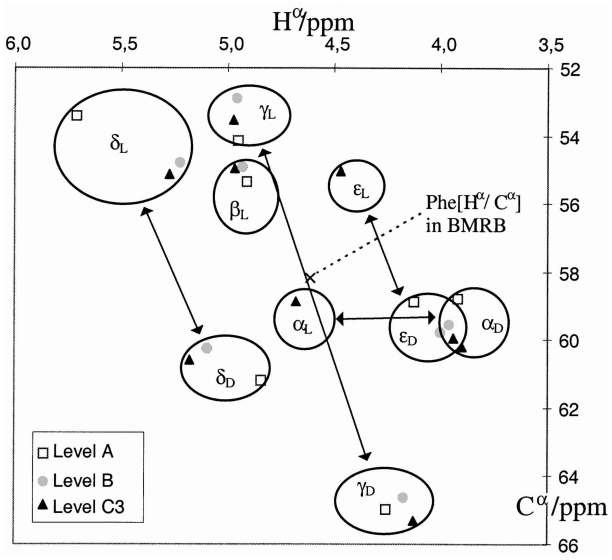
Structure-induced CS changes in the experimentally easily amenable $^1H^N$ shifts are in the range of 0.75 ppm (see Tabs. 1 and 2A). According to computed data, similarly to valine [25], the extended phenylalanine conformer β_L shows a uniquely large $^1H^N$ downfield shift of about 1 ppm relative to the random coil value and ≈1.5 ppm relative to the helical conformer α_L. This backbone conformation dependent $^1H^N$ shift holds to a similar extent for experimental data: $\Delta(^1H^N)_{\beta\Rightarrow\alpha} \approx 1.4$ ppm (see below).

It is straightforward to notice that in 2D cross sections of typical 3D NMR experiments (*e.g.*, $^1H^N$–^{15}N, $^{13}C^\alpha$–^{15}N, and $^{13}C^\alpha$–$^1H^N$ planes of the ^{15}N–$^1H^N$–$^{13}C^\alpha$ surface) the large CS changes mentioned above could be used for characteristic grouping of backbone conformers. This is clearly the case for our computed data. When the amide proton ($^1H^N$) shifts are correlated with their corresponding amide nitrogen (^{15}N) shifts, the frequently employed $^1H^N$–^{15}N HSQC 2D plot emerges. However, the values along the nitrogen dimension can hardly be used to discriminate backbone conformers because, due to an about ±5 ppm deviation induced by the side-chain orientation, the CS changes that depend on the backbone conformation (*e.g.*, $\Delta(^{15}N)_{\beta\Rightarrow\alpha} \approx 2.5$ ppm) largely overlap. Therefore, the case of the $^1H^N$–^{15}N (HSQC-type) 2D plot demonstrates that not only the CS values should differ characteristically from one conformer to another but their standard deviations should also be low enough to allow a successful conformational assignment *via* CS analysis.

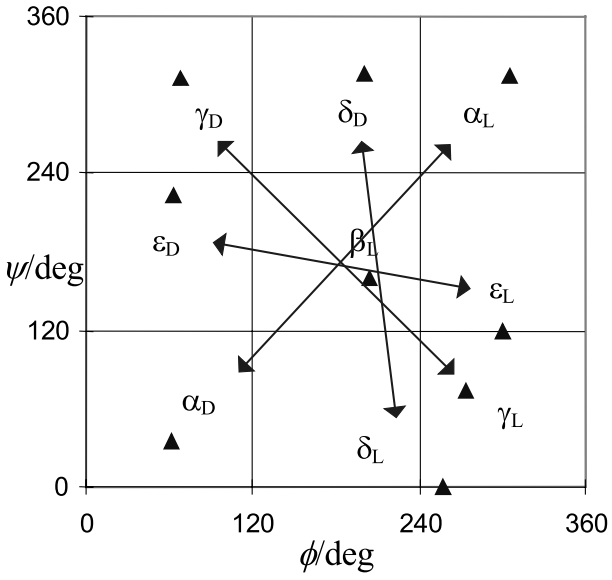
The other type of equally simple and popular heteronuclear NMR experiment involves the correlation of proton CSs with carbons to which they are attached. The $^1H^\alpha$–$^{13}C^\alpha$ region of a ^1H–^{13}C HSQC spectrum is the region where one would mostly expect to recognize the effect of ϕ and ψ torsional angle changes. Nevertheless, care should be exercised in interpreting such results because, as mentioned earlier, the conformation of the individual side chains can significantly affect the observed CS values. In order to "eliminate", at least temporarily, the effect of side-chain conformation induced CS changes, which is expected to be particularly significant for Phe due to the aromatic ring, CS values belonging to the same backbone clusters were averaged (Tab. 2A). Thereby the influence of the individual orientation of side-chains ($g+$, a, or $g-$) on CS values becomes hidden. These average CS values are now associated with a "random"-type side-chain orientation but can be used as typical values associated with a particular backbone conformer (*e.g.*, α-helical, extended, and inverse γ-turn).

As our results show for the present phenylalanine model, in all nine typical backbone orientations the side-chain independent $^{13}C^{\alpha}$ and $^{1}H^{\alpha}$ values differ from each other to a remarkable extent. We note the following: (a) although the level of theory used for the computations does influence the actual CS values on a $^{1}H^{\alpha}$–$^{13}C^{\alpha}$ correlation plot, the conformationally dependent CS values are easily distinguishable in all cases (Fig. 4). (b) At all levels of theory it is straightforward to use the $^{13}C^{\alpha}$ and $^{1}H^{\alpha}$ CS values for recognition of the appropriate main-chain fold. (c) The $^{13}C^{\alpha}$ and $^{1}H^{\alpha}$ side-chain averaged CS values computed here are very similar to those obtained previously for Val and Ala (*e.g.*, compare Fig. 4 of this study to Fig. 5 of Ref. [25]). (d) The average $^{13}C^{\alpha}$ (58.2 ppm) and $^{1}H^{\alpha}$ (4.6 ppm) values retrieved from BMRB for phenylalanine are near the center of the heteronuclear correlation plot. (e) As for valine [25], typical backbone conformations of Phe such as the extended [β_L], polyproline II [ε_L], and inverse γ-turn [γ_L] fold are high-field shifted along the carbon dimension (Fig. 4A). The clear exception, as expected based on experimental data, is the conformational building unit of right-handed alpha helix (the α_L-type backbone orientation). For the latter type of backbone cluster the C^{α} is low-field shifted and therefore widely used as an indicator of this secondary structural element. As mentioned already, conformers of D-type backbone orientation (Figs. 2 and 4b) are less common in proteins and they all have a C^{α} value lower than 59 ppm.

Differences in H^{α} CSs are widely used during structure determinations discriminating α-helices from β-pleated sheets. We wish to investigate here whether it is possible to determine not only the above two but any of the nine typical backbone folds based on their characteristic $^{13}C^{\alpha}$ and $^{1}H^{\alpha}$ CSs. We have compared above selected CS values of $\gamma_L(a)$ and $\gamma_D(a)$ conformers and the two forms of γ-turns, which are conformational mirror images of one another. Figure 4b shows this and the remaining three pairs of enantiomeric backbone conformers ($\gamma_L \Leftrightarrow \gamma_D, \alpha_L \Leftrightarrow \alpha_D, \delta_L \Leftrightarrow \delta_D,$ and $\varepsilon_L \Leftrightarrow \varepsilon_D$) connected by solid lines. Comparing the $^{1}H^{\alpha}$ CS values of these enantiomeric pairs, for each pair the $^{1}H^{\alpha}$ CS values are low-field shifted in the case of L-type structures relative to those of the D-type conformers. The same type of pairwise comparison along the carbon axis shows that $^{13}C^{\alpha}$ CS values of L-type backbone orientations are at high-field region relative to those of D-types. The only exception is found for helices ($\alpha_L \Leftrightarrow \alpha_D$), where the C^{α} CS change is insignificant. However, for the latter pair of backbone conformation the H^{α} CS change is rather distinctive, close to 1 ppm, large on a ^{1}H CS scale. Considering the resolution of modern spectrometers, these CS differences seem to be sufficiently large to distinguish and assign the different backbone types of the same residue in a protein. However, if the side-chain rotation is slow, or if due to structural reasons, as in proteins, rotations about χ^1 is blocked, the assignment of a backbone cluster based on C^{α} and H^{α} CS information is more complicated (Fig. 5). Two out of the three side-chain rotamers is indeed close to each other, but typically the third pair of C^{α} and H^{α}



(a)



(b)

Fig. 4. (a) The C^{α}/H^{α} correlation plot (CSI scale) of the 9 typical backbone conformers determined at three levels of theory: level A {GIAO-RHF/6-31+G*//RHF/3-21G [square//empty symbol]}, level B {GIAO-RHF/6-311++G**//RHF/6-31+G* [circle//filled symbol]} and level C3 {GIAO-RHF/TZ2P//RHF/6-31+G* [triangle//filled symbols]}. Each point is the average of a maximum of three side chain conformers of For–L-Phe–NH_2 (*e.g.* $\beta_L := \beta_L[g+]$, $\beta_L[a]$ and $\beta_L[g-]$). All computed CSI values are referenced (δ scale) and corrected with the difference found between calculated (*ab initio*) and experimentally observed (BMRB) conformational shift values (correction factor: level A [C^{α}/H^{α} = 5.41/0.89], level B [C^{α}/H^{α} = 3.89/0.94] and level C3 [C^{α}/H^{α} = 5.01/0.91]) (see text for more details). (b) The enantiomeric pairs of conformers ($\alpha_L \Leftrightarrow \alpha_D, \gamma_L \Leftrightarrow \gamma_D, \delta_L \Leftrightarrow \delta_D,$ and $\varepsilon_L \Leftrightarrow \varepsilon_D$) connected by solid lines on the Ramachandran surface as determined at the 6-31+G* RHF level of theory.

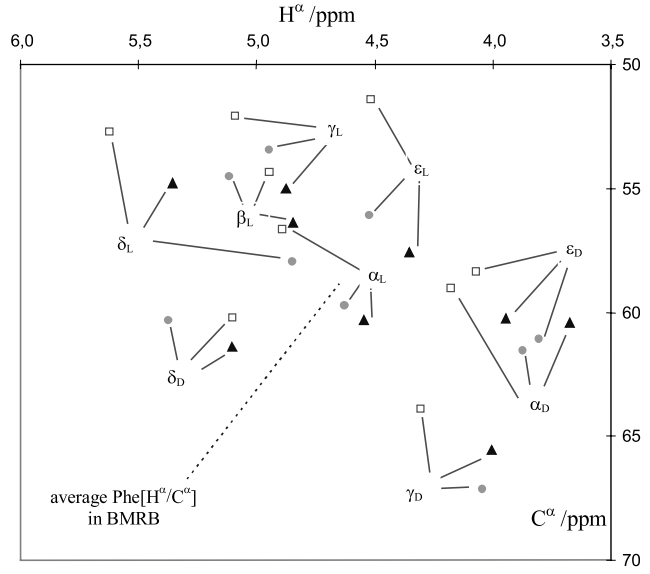


Fig. 5. The C^{α}/H^{α} correlation plot (CSI δ scale) of 27 different conformers of For–L-Phe–NH_2 determined at level C3 {GIAO-RHF/TZ2P//RHF/6-31+G* ([$g+$] square, [a] triangle and [$g-$] circle] (*e.g.* $\beta_L := \beta_L[g+]$, $\beta_L[a]$ and $\beta_L[g-]$}. All computed CSI values are referenced relative to TMS and NH_3 (δ scale), as appropriate. All values are commonly shifted with the difference found between calculated (*ab initio*) and experimentally observed (BMRB) conformational shift values (correction factors: at level C3: $C^{\alpha}/H^{\alpha} = 5.01/0.91$).

CS of Phe is significantly off. This is due to the specific influence of the benzene ring. The backbone clusters are still recognizable but there is some overlap resulting in a more ambiguous picture.

Among the factors that determine the position of the backbone conformers on *ab initio* computed chemical shift–chemical shift correlation maps (*e.g.*, $^1H^{\alpha}$–$^{13}C^{\alpha}$), some are related to the level of computation used, some are to the limitation of the model system. However, most of the backbone conformational clusters of For–L-Phe–NH_2 are clearly recognizable from the $^{13}C^{\alpha}$–$^1H^{\alpha}$ CS plot. It is clear that more than the helical and extended conformers result in different pairs of $^{13}C^{\alpha}$–$^1H^{\alpha}$ CS data. The usefulness of the approach of direct determination of conformations of protein building units from multidimensional NMR experiments seems to depend on whether we can decipher all effects influencing side-chain orientation such as solvation, anisotropic factors and H-bond networks. Therefore, further investigation of other model compounds is mandatory.

6 Chemical shift – structure correlations

In Section 4 we have shown that theoretical and experimental chemical shifts, when averaged over the entire conformational space, are rather similar. Furthermore, in Section 5 we deduced for Phe, as earlier for Ala [24] and Val [25], that CS–CS plots of selected nuclei (*e.g.*, $^1H^{\alpha}$–$^{13}C^{\alpha}$) can serve to identify different secondary structural motifs.

The next task is to select the CS values of nuclei to obtain an unambiguous prediction of main-chain conformations. The result of a systematic analysis is presented here between the most important conformational parameters (ϕ, ψ, ω_0, ω_1, and χ^1) and CS values of all nuclei (^{15}N, $^1H^N$, $^{13}C^{\alpha}$, $^1H^{\alpha}$, $^{13}C^{\beta}$, $^1H^{\beta}$, and $^{13}C'$). A comprehensive analysis is presented for *ab initio* data and experimental values retrieved from proteins. To measure correlations, both the linear correlation (Pearson) coefficient, R, and the standard error, $S_{Y.X}$, is used [84]. A typical correlation matrix, obtained for data computed at level C3, is reported in Table 3. The closer R^2 is to 1 the higher the correlation is. For structure – chemical shift correlation purposes only those R^2 values are considered which measure correlation between conformational parameters and CS data. However, we have to emphasize that torsional variables ξ, where $\xi = \phi, \psi$, etc., are of periodic nature. Thus, either the IUPAC recommendation $[-180° \leq \xi \leq +180°]$ or the classical definition of Ramachandran $[0° \leq \xi \leq +360°]$ can be used. To avoid ambiguity, R^2 values were obtained, unless otherwise noted, using the periodic unit of $0° \leq \xi \leq +360°$ for all torsion angles, reported as $R^2_{[0-360]}$. Note however that the definition used for the periodicity of ξ strongly influences the outcome of the structure–CS data correlation. One can maximize this correlation by simply finding the "optimum" definition of the periodicity of a particular torsional variable. This was done for all torsional variables and the optimum value of a given pair of data ($R^2_{\max}$) is reported in parenthesis in Table 3.

Maximized statistical measures (R) of selected structure–CS and CS–CS pairs ($\phi/^{13}C^{\alpha}$, $\phi/^1H^{\alpha}$, ..., $^1H^{\alpha}/^{13}C^{\alpha}$, etc.) at four theoretical levels, A, B, C1, and C3, are reported in Table 4. These pairs were selected either because they show significant correlation (*e.g.*, $\phi/^{13}C^{\alpha}$, $\phi/^1H^{\alpha}$, $\psi/^{13}C^{\alpha}$, and $^1H^{\alpha}/^{13}C^{\alpha}$) or because of general interest (*e.g.*, $\psi/^1H^{\alpha}$). It can be observed that neither the extension of the basis set applied, nor the improvement of the molecular geometry increase significantly the correlation observed between [ϕ, ψ] conformational values and $[C^{\alpha}, H^{\alpha}]$ CS values (Tab. 5). Except CS values computed for the carbonyl C atom, *ab initio* data at each levels of theory show high resemblance to each other.

Due to their similarity, only *ab initio* data obtained at the highest level of theory (level C3) are discussed in detail. The arithmetical average ($\langle R^2_{[0-360]} \rangle$) of all calculated $R^2_{[0-360]}$ values is close to zero: 0.065 with a standard deviation (σ) of 0.109 (Tab. 3). This indicates that on average there is no linear correlation between conformational parameters and isotropic chemical shifts. (Linear correlation is expected to be the "model free" approach of the problem.) This observation underlines the notion that CS values in general are determined by several factors among which the conformation adopted by the molecule is of little significance. However, for a few pairs of variables $R^2_{[0-360]}$ values can be three to four times higher than the average value close to zero. For example, $R^2_{[0-360]}[\phi/^{13}C^{\alpha}] = 0.407$, $R^2_{[0-360]}[\phi/^1H^{\alpha}] = 0.443$, and

Table 4. Maximized cross-correlation values, R, of selected conformational and chemical shift data of For–L-Phe–NH_2 as determined at four levels of theory, A, B, C1, and C3[a].

	A	B	C1	C3
ϕ – H^{α}	0.807	−0.913	−0.909	−0.903
ϕ – C^{α}	−0.649	−0.771	−0.767	−0.638
ψ – C^{α}	0.686	−0.626	−0.628	−0.618
ψ – H^{α}	0.372	0.359	−0.363	−0.381
ϕ – N	0.496	−0.535	−0.546	−0.612
χ^1 – C'	0.499	0.671	0.666	0.436
H^{α} – C^{α}	−0.652	−0.684	−0.689	−0.583
H^{α} – C^{β}	−0.032	−0.179	−0.189	0.218
H^{N} – N	0.178	0.236	0.218	−0.034

[a] See Scheme 1 for the description of theoretical levels.

Table 5. R^2 of CSI values of selected nuclei of For–L-Phe–NH_2 determined at three levels of theory.

Levels compared[a]	conf.[b]	$^{1}H^{N}$	$^{15}N^{NH}$	$^{13}C^{\alpha}$	$^{1}H^{\alpha}$	$^{13}C'$	$^{13}C^{\beta}$	$^{1}H^{\beta 1}$	$^{1}H^{\beta 2}$
Level A ⇔ Level B	19 ⇔ 16	0.965	0.927	0.919	0.829	0.365	0.933	0.952	0.945
Level A ⇔ Level C3	19 ⇔ 27	0.839	0.865	0.906	0.844	0.341	0.904	0.828	0.920
Level B ⇔ Level C3	16 ⇔ 27	0.993	0.996	0.999	0.999	0.969	0.998	0.999	0.999

[a] See Scheme 1 for the description of theoretical levels.
[b] conf. = number of side-chain conformers used when CSI values (relative to TMS) were determined.

$R^2_{[0-360]}[\psi/^{13}C^{\alpha}] = 0.333$. Furthermore, the optimized correlation coefficients are even higher: $R^2_{max}[\phi/^{13}C^{\alpha}] = 0.407$, $R^2_{max}[\phi/^{1}H^{\alpha}] = 0.816$, and $R^2_{max}[\psi/C^{\alpha}] = 0.382$. Undoubtedly, $\phi/^{1}H^{\alpha}$ provides the most significant correlation between a backbone parameter and a CS value (see Fig. 6). The correlation between $[\phi,\psi]/[^{13}C^{\alpha}, ^{1}H^{\alpha}]$ provides the possibility of "projecting" backbone conformational parameters on the $^{1}H^{\alpha}$–$^{13}C^{\alpha}$ 2D plot and *vice versa*. To the best of our knowledge the fact that these and only these correlations "exist" was first reported by us [24,25] for the Gly, Ala, and Val amino acid residues. The first two studies in this series suggest that, after resonance assignment, simply by using C^{α} and H^{α} CS values significant structural information can be obtained. Although linear correlation of selected chemical shifts with conformational parameters may not be as high as one would have hoped, pairs with large $|R|$ values certainly warrant further investigation. Correlations observed between the $\phi/^{13}C^{\alpha}$, $\phi/^{1}H^{\alpha}$, and $\psi/^{13}C^{\alpha}$ pairs look significant, thus indicating that relative changes in CS of $^{13}C^{\alpha}$ and $^{1}H^{\alpha}$ are coupled with backbone conformational changes.

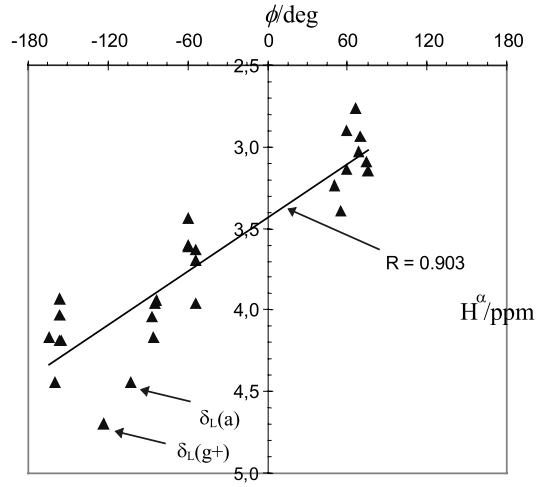


Fig. 6. The maximized fit between $^{1}H^{\alpha}$ and ϕ at level C3 ($R_{max} = 0.903$) (elimination of the $\delta_L(g+)$ and $\delta_L(a)$ conformers increases the maximized correlation to $R_{max} = 0.932$).

The most important conclusions drawn from our new and previous [24,25] data can be summarized as follows. (a) All four elements except $\psi/^{1}H^{\alpha}$ from the $[\phi,\psi]/[^{13}C^{\alpha}, ^{1}H^{\alpha}]$ pairs (a two by two data set) show significant linear correlation (*e.g.*, $\phi/^{1}H^{\alpha}$ can be as high as $|R_{max}| \approx 0.903$, Fig. 6). (b) The CS–CS correlation of $^{1}H^{\alpha}$–$^{13}C^{\alpha}$ is significant at all levels of theory investigated. (c) No linear correlation was found between ψ and CS

values of H^{α} at any theoretical level investigated. (d) The CS values of other nuclei of potential interest (*e.g.*, $^{1}H^{N}$, $^{13}C^{\beta}$, $^{1}H^{\beta}$, and $^{13}C'$) show no linear correlation with conformational parameters. Thus, their CS values cannot be used to monitor structural changes. (e) None of the correlations change markedly as function of the theoretical level used for their determination.

7 Chemical shifts of phenylalanine and tyrosine as function of their backbone conformations in proteins

Several aspects of empirical structure–CS correlation of the α-helix and β-sheet secondary structural elements of proteins have been reported in the literature [26,29,31,34, 48,85]. Further results obtained from *ab initio* electronic structure calculations for model diamide compounds encouraged us to investigate what we can learn from correlating experimental CS data and relevant conformational parameters of aromatic amino acid residues.

Structural and chemical shift data of 48 Phe and 48 Tyr residues extracted from 18 selected and carefully referenced proteins were used for the analysis (see Sect. 2 and Tab. 6). Clearly, the fact that we are analyzing not a single protein but an ensemble of proteins makes this investigation more conclusive than previous ones; however, as seen below, still more experimental information is needed to reach definitive conclusions. Among the 48 experimental conformers of phenylalanine 16 helical ($\alpha_{\rm L}$-), 23 β-pleated sheet ($\beta_{\rm L}$-), 5 polyproline II ($\varepsilon_{\rm L}$-), 2 inverse γ-turns ($\gamma_{\rm L}$-), and 2 $\delta_{\rm D}$-type backbone conformers were assigned (Tab. 6). A very similar distribution of backbone conformers was found for tyrosine.

As mentioned in Section 6, the periodic nature of the torsional variables allow maximalization of the correlations investigated. Table 7 reports $R_{\max}$ values between experimental CS information of nuclei of interest and conformational folds of Phe residues as derived from the 18 proteins. Analysis of the maximized correlation values ($R_{\max}[\phi/^{15}{\rm N}]$, $R_{\max}[\phi/^{1}H^{N}]$, ..., $R_{\max}[\chi^{1}/^{13}C']$) of experimental data results in a coherent picture. The experimental values sample only a fraction of the entire Ramachandran surface: the available NMR data are predominantly for α-helical and β-sheet regions (Tab. 6). For this simple reason, comparing far more restricted experimental data (Tab. 7) with a much larger variety of computed data (Tab. 4) should be done with some care. Nevertheless, the following conclusions appear well established: (a) the $\phi/^{13}C^{\alpha}$, $\phi/^{1}H^{\alpha}$, and $\psi/^{13}C^{\alpha}$ correlations are the most significant ones based on the statistical analysis of both experimental and computed values. (b) The experimental $R^{2}_{\max}[\psi/^{1}H^{\alpha}] = 0.493$ value is clearly much higher than its computed counterpart, 0.145 (level C3). However, if we reduce the statistical analysis of the *ab initio* computed values to helix-like ($\alpha_{\rm L}$) and β-region type conformers ($\beta_{\rm L}$ and $\varepsilon_{\rm L}$), the $R^{2}_{\max}[\psi/^{1}H^{\alpha}]$ correlation coefficient of the computed data increases considerably, to $R^{2}_{\max}[\psi/^{1}H^{\alpha}] = 0.753$ (level C3). Indeed, if only these two major secondary structural elements are included in the database, $R^{2}_{\max}[\psi/^{1}H^{\alpha}]$ becomes as high as $R^{2}_{\max}$ values obtained for $\phi/^{13}C^{\alpha}$, $\phi/^{1}H^{\alpha}$, or $\psi/^{13}C^{\alpha}$. However, we still advocate the use of all typical backbone structural elements in order to obtain a more complete picture of structure–CS correlation. (c) For both experimental and *ab initio* data the CS–CS correlation $^{1}H^{\alpha}$–$^{13}C^{\alpha}$ is found significant (Fig. 7): $R[^{13}C^{\alpha}/^{1}H^{\alpha}] = -0.801$ (expt.) and $R[^{13}C^{\alpha}/^{1}H^{\alpha}] = -0.689$ (level C1) (Tab. 4). The same conclusions are obtained when a similar structure–CS correlation is performed for the Tyr residues. (d) The experimental CS data of all other nuclei (*e.g.*, ^{15}N and $^{1}H^{N}$) show significantly weaker correlation with conformational variables (ϕ, ψ, or χ^{1}). This agrees well with results obtained from the computed data.

Two-dimensional heteronuclear correlation spectra have been used, when available, to check folding properties of macromolecules, especially those of proteins. Significant signal dispersion along both the proton and nitrogen dimensions (*e.g.*, $^{1}H^{N}$–^{15}N HSQC) is considered as an indication of global fold. It is not common, however, to utilize the signal dispersion of the $^{1}H^{\alpha}$–$^{13}C^{\alpha}$ region of an ^{1}H–^{13}C HSQC-type data set. When taking a closer look at proton-carbon correlation maps one observes that while signals correlating H^{β}–C^{β}, H^{γ}–C^{γ}, or H^{δ}–C^{δ} pairs of atoms are closer to the "diagonal" in the case of a folded protein, $^{1}H^{\alpha}$–$^{13}C^{\alpha}$ signals have a larger spread. Thus, the latter type of signal dispersion is related to the backbone conformation of the parent amino acid. Undoubtedly, there are other effects that manifest in the final "location" of a $^{1}H^{\alpha}$–$^{13}C^{\alpha}$ signal; however, we believe that in this case the dominant cause is the backbone fold. Under optimal experimental conditions this could result in the determination of the backbone fold from $^{13}C^{\alpha}$ and $^{1}H^{\alpha}$ chemical shifts.

In the last part of this section we are focusing on the relation between $^{13}C^{\alpha}$ and $^{1}H^{\alpha}$ chemical shift changes and torsional parameters ϕ and ψ (see Tab. 8 and Figs. 7 to 9). Down- and upfield shifts are relative to the random coil values $\delta^{\rm Phe}_{\rm rc}(^{13}C^{\alpha}) = 58.2$ ppm and $\delta^{\rm Phe}_{\rm rc}(^{1}H^{\alpha}) = 4.6$ ppm extracted from BMRB [40]. The relevant results are as follows.

(a) In the case of *right-handed helical* conformations ($\alpha_{\rm L}$-type structure) a characteristic down-field shift of the C^{α} CS can be observed in Figure 9. This shift is especially strong when the side-chain orientation of Phe is *anti*, for the 9 cases the average value is +61.3 ppm, resulting in a $\Delta^{\rm BMRB}\delta_{\alpha{\rm L}(a)}(^{13}C^{\alpha})$ value of +3.1 ppm (*cf.* Tabs. 6 and 8). Similarly large down-field shifts are obtained using *ab initio* computations, $\Delta^{\rm level\ C3}\delta_{\alpha{\rm L}(a)}(^{13}C^{\alpha})$ is +2.1 ppm. Along the other dimension, for $^{1}H^{\alpha}$ an up-field shift is observed both computationally and experimentally. Considering again the same helical conformers of Phe, $\alpha_{\rm L}(a)$ values obtained are $^{1}H^{\alpha}_{\alpha{\rm L}(a)} = 4.5$ ppm (level C3) and $^{1}H^{\alpha}_{\alpha{\rm L}(a)}$(expt.) = 4.3 ppm, which mean $\Delta^{\rm BMRB}\delta_{\alpha{\rm L}(a)}(^{1}H^{\alpha}) = -0.3$ ppm and $\Delta^{\rm level\ C3}\delta_{\alpha{\rm L}(a)}(^{1}H^{\alpha}) = -0.1$ ppm. In the present database containing 18 proteins, ≈13% (6 cases) of

Table 6. Experimental chemical shift isotropy values (δ scale) of 48 phenylalanines extracted from 18 proteins with conformational parameters (ϕ, ψ, χ^1, and χ^2)[a].

N	C^α	C^β	C'	H^N	H^α	*bb*	*sc*	ϕ	ψ	χ^1	χ^2
118.1	60.0	39.4	174.8	8.2	5.1	α_L	a.a	−64.2	−46.2	−171.2	−148.6
123.0	61.3	39.3	174.8	9.3	4.1	α_L	a.g−	−62.9	−42.4	−176.0	−86.5
119.6	62.1		176.1	7.6	3.9	α_L	a.g−	−62.9	−43.7	173.9	−85.0
	61.6			8.1	4.4	α_L	a.g−	−61.1	−43.8	178.4	−85.9
120.3	59.9	38.9		7.5	4.4	α_L	a.g−	−60.8	−44.9	168.8	−94.3
121.4	61.7	39.9	178.0	8.1	4.0	α_L	a.g+	−60.0	−56.8	−172.2	98.0
119.8	61.3	38.9	176.3	7.9	4.4	α_L	a.g+	−61.5	−47.2	175.4	82.1
121.5	60.6	39.0		8.6	4.6	α_L	a.g+	−59.4	−49.5	−179.6	66.1
125.5	63.3	40.6		8.1	3.7	α_L	a.g+	−55.5	−47.4	−157.8	91.9
119.9	57.8	34.1	175.3	6.8	3.8	α_L	g−.g−	−60.0	−44.3	−63.5	−18.0
118.1	62.5	39.8	178.3	8.0	4.2	α_L	g−.g−	−59.7	−41.6	−76.2	−61.4
115.5	59.6		174.4		4.3	α_L	g−.g−	−84.7	−12.5	−65.6	−69.4
112.8	59.6		176.4	8.1	4.5	α_L	g−.g−	−82.7	−37.9	−58.7	−77.3
122.3	60.8	38.1	174.1	7.6	3.8	α_L	g−.g+	−64.8	−46.5	−65.7	89.8
110.9	62.8			7.2	3.9	α_L	g−.g+	−67.5	−42.0	−75.1	94.0
115.9	59.3		176.4	7.6	4.6	α_L	g+.g+	−65.9	−21.2	64.8	85.2
127.1	58.4	40.6	171.6	7.6	4.3	β_L	a.g+	−142.9	109.3	−171.8	84.3
119.6	58.0	44.1	171.9	7.7	4.7	β_L	a.g+	−143.0	133.4	−166.2	73.6
127.6	54.0	41.9	172.5	9.3	5.8	β_L	a.g+	−109.8	121.8	179.3	57.2
122.6	55.8		174.9	9.9	5.1	β_L	g−.g−	−129.7	143.9	−54.4	−85.5
128.2	55.1	41.8	174.9	9.1	5.7	β_L	g−.g−	−133.2	136.4	−59.5	−95.0
120.6	56.1	42.6	174.5	9.3	5.0	β_L	g−.g−	−119.3	142.4	−56.4	−91.8
119.9	56.7			8.2	5.4	β_L	g−.g−	−130.3	132.2	−59.5	−95.0
122.3	55.9		175.0	8.7	5.0	β_L	g−.g−	−125.6	163.9	−63.9	−78.6
	56.7			9.2	4.9	β_L	g−.g−	−105.1	142.6	−75.8	−78.7
125.8	57.2	42.5	173.2	9.7	4.9	β_L	g−.g+	−107.5	142.4	−83.9	85.6
119.4	56.4	41.8	176.0	8.2	4.9	β_L	g−.g+	−119.8	143.6	−70.3	79.1
117.3	55.9	42.8	175.0	9.8	5.7	β_L	g−.g+	−125.3	161.0	−61.3	87.2
120.1	54.8	40.9	175.5	8.6	5.6	β_L	g−.g+	−116.0	140.2	−60.7	97.9
119.0	54.5	41.0	177.3	9.0	5.4	β_L	g−.g+	−111.6	134.9	−53.2	88.8
125.0	56.3		174.5	9.1	4.7	β_L	g−.g+	−115.7	148.3	−77.7	69.6
125.7	56.3		175.6	10.0	5.2	β_L	g−.g+	−112.3	151.4	−58.3	70.5
120.2	54.8		171.1	9.7	5.3	β_L	g+.g−	−133.4	155.7	74.9	−71.6
118.7	56.8	43.0	172.2	8.7	5.5	β_L	g+.g−	−146.7	161.1	63.5	−83.0
119.6	55.9		171.7	9.4	5.1	β_L	g+.g−	−148.1	169.7	68.0	−78.2
113.7	55.8	42.4	172.8	7.9	5.5	β_L	g+.g+	−128.6	147.0	67.9	89.6
116.9	57.1	44.0	172.1	9.0	5.6	β_L	g+.g+	−134.0	151.7	74.8	82.8
122.4	55.3			9.3	5.0	β_L	g+.g+	−128.4	168.8	67.8	93.4
120.3	55.5		172.6	8.3	4.8	β_L	g+.g+	−144.6	152.4	58.6	88.7
	57.7			9.2	4.3	γ_L	a.g−	−96.7	114.1	−172.0	−82.4
128.8	57.3	38.9	172.9	9.3	4.7	γ_L	g−.g−	−111.4	107.7	−63.1	−84.2
112.7	57.5	39.9	176.5	8.1	4.9	δ_D	g−.g−	−104.7	−12.3	−59.1	−58.1
118.1	57.8	41.9	174.5	8.8	5.5	δ_D	g+.g−	−130.2	−25.1	71.6	−82.2
121.0	55.1			8.5	5.1	ε_L	a.g+	−88.6	121.8	174.8	61.3
120.9	57.6	39.9	176.1	7.7	4.9	ε_L	g−.g−	−95.0	153.8	−56.7	−65.2
129.2	58.3	39.5		8.8	5.3	ε_L	g−.g−	−94.7	118.8	−73.7	−95.1
128.5	57.2	41.7		8.9	5.2	ε_L	g−.g−	−101.0	142.6	−74.3	−96.9
126.6	59.5	39.9	175.4	8.8	4.8	ε_L	g−.g+	−78.5	156.2	−71.6	102.3

[a] bb = backbone, sc = sidechain. Chemical shifts in ppm, torsion angles in degrees.

Table 7. Maximized Pearson correlation coefficients, R_{max}, determined between major periodic conformational variables (ϕ, ψ, and χ^1) and chemical shifts assigned for valines in 18 proteins[a].

	^{15}N	$^{1}H^{N}$	$^{13}C^{\alpha}$	$^{1}H^{\alpha}$	$^{13}C^{\beta}$	$^{13}C'$
ϕ	0.256	0.470	**–0.849**	**0.731**	–0.194	0.334
ψ	–0.218	–0.476	**–0.834**	**0.702**	0.241	–0.242
χ^1	–0.307	–0.260	0.485	0.448	–0.226	–0.385

[a] The periodic unit of the three conformational variables ϕ, ψ, and χ^1 were shifted systematically (with an increment of 5°) in order to maximize fitting. Significant correlations are given in bold face.

Table 8. Experimentally determined[a] and *ab initio* (levels A[b] and C3[c]) $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, ϕ, and ψ parameters for phenylalanine conformers occurring frequently in proteins.

		Experiment					Level A1				Level B1			
conf.[d]	abundance[e]	aver. dev.[f]	$^{1}H^{\alpha}$	$^{13}C^{\alpha}$	ϕ	ψ	$^{1}H^{\alpha}$	$^{13}C^{\alpha}$	ϕ	ψ	$^{1}H^{\alpha}$	$^{13}C^{\alpha}$	ϕ	ψ
α_L(g–)	6	17	4.1	60.5	290	323					4.6	59.9	306	315
α_L(g+)	1	9	4.6	59.3	294	339					4.9	56.4	306	315
α_L(a)	9	17	4.3	61.3	299	313					4.5	60.3	306	351
β_L(a)	3	33	5.0	56.8	228	122	4.7	57.9	193	174	4.8	56.4	204	152
β_L(g–)	13	24	5.2	56.0	241	145	5.3	53.6	230	160	5.1	54.3	204	160
β_L(g+)	7	13	5.3	55.9	222	158	4.8	54.6	184	175	4.9	54.2	204	168
δ_D(g+)	1	35	5.5	57.8	230	335	4.9	59.7	174	329	5.1	60.3	196	328
δ_D(g–)	1	34	4.9	57.2	255	348					5.4	60.1	201	315
ε_L(g–)	4	32	5.0	58.1	268	143					4.5	56.2	300	120
ε_L(a)	1	30	5.1	55.1	71	172					4.4	57.6	300	120
γ_L(a)	1	41	4.3	57.7	263	114	4.9	56.2	274	73	4.9	55.0	275	89

[a] Values extracted from the 18 proteins detailed in the section Computational details. [b] Level A = GIAO-RHF/6-31+G*//RHF/3-21G. [Computed chemical shifts for $^{13}C^{\alpha}$ and $^{1}H^{\alpha}$ are uniformly shifted by 5.41 and 0.89 ppm, respectively (*cf.* Tab. 2), to match BMRB experimental radom coil values (Tab. 2).] [c] Level C3 = GIAO-RHF/TZ2P//RHF/6-31+G*. [Computed chemical shifts for $^{13}C^{\alpha}$ and $^{1}H^{\alpha}$ are uniformly shifted by 5.01 and by 0.89 ppm, respectively (Tab. 2), to match BMRB experimental radom coil values (Tab. 2).] [d] Type of phenylalanine conformers found in the present protein data set. [e] Number of cases assigned for the appropriate type of conformation in the 18 proteins. [f] Average deviation (in degrees) between phenylalanine conformers in proteins and the reference ϕ and ψ values.

all helical units adopt a side-chain orientation of *gauche*−. The analysis of $\alpha_L(g-)$-type experimental data reveals both for $^{13}C^{\alpha}$ and for $^{1}H^{\alpha}$ a small but consistent shift: $\Delta^{BMRB}\delta_{\alpha L(g-)}(^{13}C^{\alpha}) = +2.3$ ppm and $\Delta^{BMRB}\delta_{\alpha L(g-)}(^{1}H^{\alpha}) = -0.5$ ppm. The $^{13}C^{\alpha}$ CS change is reproduced reasonably well by *ab initio* calculations: $\Delta^{\text{level C3}}\delta_{\alpha L(g-)}(^{13}C^{\alpha}) = 1.7$ ppm and $\Delta^{\text{level C3}}\delta_{\alpha L(g-)}(^{1}H^{\alpha}) = 0.0$ ppm (see Tab. 8 and Fig. 9), indicating that the CS alteration is due to side-chain conformational changes. The third type of side-chain orientation, $\alpha_L(g+)$, is contained in the database only once; thus, it cannot be the subject of a statistical analysis.

(b) For *β-sheet* backbone conformations experimental data are available for all three types of side-chain orientations: approximately 60% (13/23) have a *gauche*−, $\beta_L(g-)$, 30% (7/23) have a *gauche*+, $\beta_L(g+)$, and 10% (3/23) have an *anti*, $\beta_L(a)$, side-chain orientation (*cf.* Tabs. 6 and 8). Regardless of the type of side-chain orientation, relative to the random coil values there is a significant up-field shift for $^{13}C^{\alpha}$ and a characteristic down-field shift for $^{1}H^{\alpha}$ (*cf.* Tab. 8 and Fig. 9). *Ab initio* CS values follow these changes, supporting the explanation that these CS alterations are due to side-chain conformational changes. Note that in the case of the $\beta_L(g-)$ conformer even the magnitude of the calculated CS, $^{1}H^{\alpha}_{\beta L(g-)}$(level C3) = 5.1 ppm, matches the average experimental CS data ($^{1}H^{\alpha}_{\beta L(g-)}$(expt.) = 5.2 ppm). (Consider that the latter value is an average of 13 $\beta_L(g-)$ Phe conformers retrieved from 18 proteins.) For $\beta_L(a)$ conformers the experimental and theoretical CS values are also very similar along both axes (*e.g.*, $\Delta^{\text{level C3}}\delta_{\beta L(a)}(^{13}C^{\alpha}) = -1.4$ ppm and $\Delta^{BMRB}\delta_{\beta L(a)}(^{13}C^{\alpha}) = -1.8$ ppm).

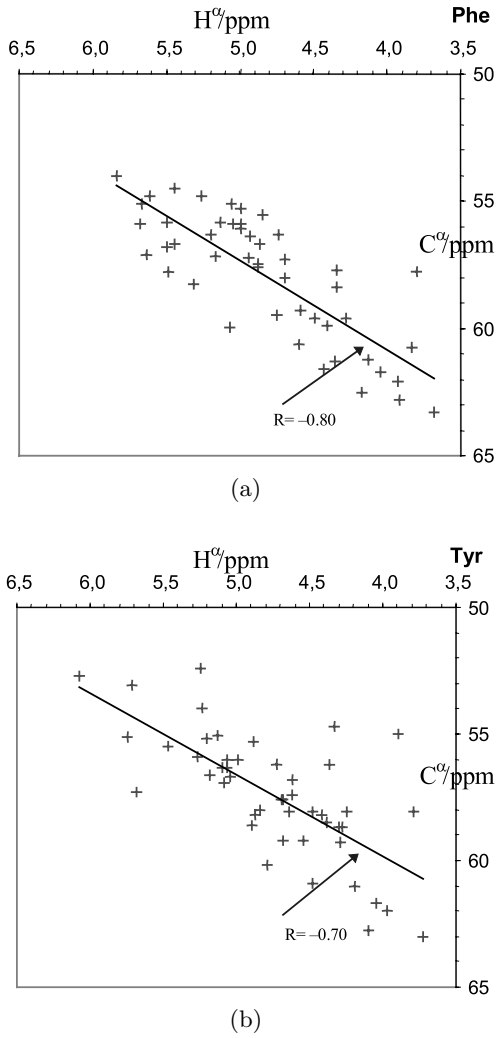


Fig. 7. C^{α}–H^{α} correlation plot as assigned for 48 phenylalanines (a) and 48 tyrosines (b) in 18 experimentally determined proteins.

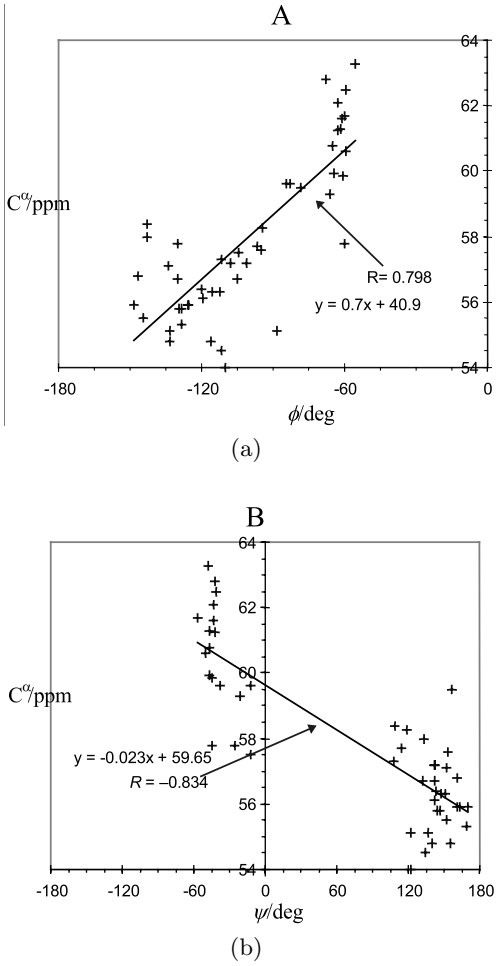


Fig. 8. ϕ–C^{α} [a] and ψ–C^{α} [b] structure-chemical shift plots as assigned for 48 phenylalanines contained in 18 proteins.

(c) The third most frequently assigned backbone conformer in proteins is that of *polyproline II*, built up from ε_L-type backbone conformational units. As observed for helices and for β-sheets, in polyproline II-type structures also the *gauche*− side-chain orientation, $\varepsilon_L(g-)$, is the most frequent one. However, only a total of four (8%) ε_L-type backbone structures were found among the 48 phenylalanine residues retrieved from these 18 proteins. Furthermore, during *ab initio* CS calculations for the For–L-Phe–NH_2 model system, the following ϕ and ψ values were employed: $\phi = -60°$ and $\psi = 120°$. However, the average $[\phi, \psi]$ values obtained for the four experimental structures are off by some 30° from the "text book" values used for the *ab initio* computations ($\phi = -92°$ and $\psi = 143°$). Thus, sizeable differences between experimental and *ab initio* CS values are expected. As established earlier, a few tenth of a ppm discrepancy is expected between $^1H^{\alpha}$(theor.) and $^1H^{\alpha}$(expt.) due to the $\phi/^1H^{\alpha}$ correlation (Tab. 8). In fact, for this conformer the *ab initio* $^1H^{\alpha}$ CS is significantly up-field shifted ($H^{\alpha}_{\varepsilon L(g-)}$(level C3) = 4.5 ppm) when compared to the experimental value of $^1H^{\alpha}_{\varepsilon L(g-)}$(expt.) = 5.0 ppm (Tab. 8 and Fig. 9). The difference between theoretical and experimental values is 0.5 ppm for $^1H^{\alpha}$ CSs and close to 2 ppm for $^{13}C^{\alpha}$ CSs. This signals that backbone changes are indeed coupled both to C^{α} and to H^{α} CS values.

(d) Beyond the α-helix, β-sheet, and polyproline II conformers only a single *inverse* γ-*turn* and two δ_D-*type* backbones were assigned in the 18 proteins. Obviously the occurrence of the last two types of backbone orientations is too small to be used for a meaningful statistical analysis.

8 Conclusions

Understanding correlation between peptide backbone conformations and chemical shifts of selected nuclei has remained a challenge for experimentalists and theoreticians

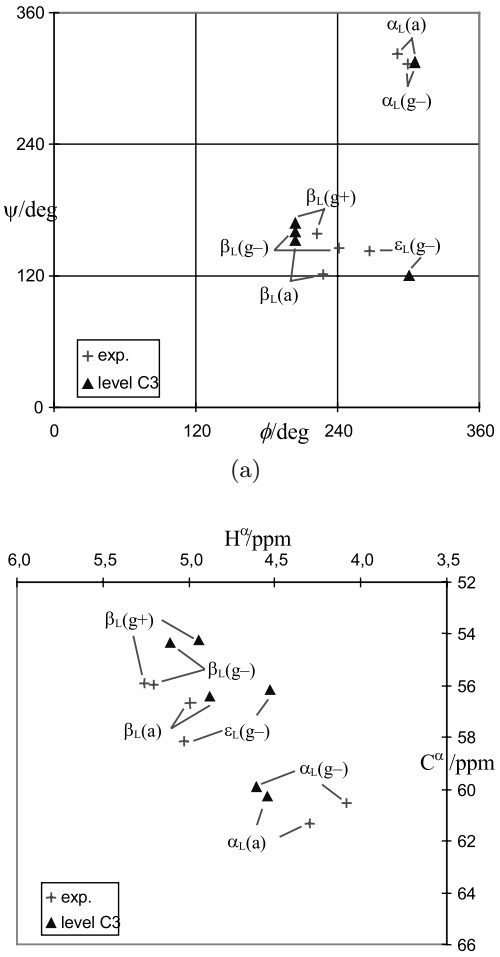


Fig. 9. Experimentally determined and *ab initio* (level C3) computed $^1H^\alpha$ and $^{13}C^\alpha$ chemical shifts of phenylalanine conformers (see also Tab. 6). Conformers assigned 3 times or more are plotted only. (All computed C^α and H^α values are shifted by 5.01 ppm and 0.91 ppm respectively, to match BMRB random coil values.)

alike. Whatever approach is taken limitations are to be faced. When following a computational (in the present case an *ab initio*) strategy for CS determination, uncertainties arise from the size and type of an ideal model system, from the level of electronic structure theory employed, and from the *a priori* selection of the most important structural factors. Thus, specific molecular interactions (*e.g.*, the effect of ring current) are to be handled with care. A critical assessment should reveal the size and the type of the database, the level of homology, and the quality (reliability) of the experimental data used. Knowing most of the limitations, studies like the present one, which consider both computed and carefully selected experimental data, are perhaps the most promising and conclusive.

The findings of this study are as follows.

- Using a conformational library of For–L-Phe–NH_2 as complete as possible, which comprises 27 different *ab initio* structures, a CS database was established. The database contains more than sixty medium- and/or high-level GIAO-RHF results, which allow mapping of NMR shielding properties of aromatic amino acid residues over their entire conformational hyperspace.
- The theoretical levels employed for CS calculations look reasonable, since conformationally averaged CS values from both theoretical and experimental sources match: we observe R^2 values as high as 0.99. As the size of the basis set employed for the shielding calculation increases, deviation between the experimental and theoretical CS values decreases to a few percent for all nuclei but the amide protons.
- Correlating computed $^1H^N$ and ^{15}N, as well as $^1H^\alpha$ and $^{13}C^\alpha$ CS values, useful 2D plots emerge, corresponding to commonly used (H,N)- or (H,C)-type heterocorrelation NMR spectra. Conformational information can be deduced from these 2D plots, independent of the basis set employed for the electronic structure calculation. Averaging out the effect of side-chain perturbation of CS values, backbone conformational data can readily be retrieved from shielding information of carefully selected nuclei.
- The chemical shift values of both α-helices and extended-like structures retrieved from empirical and computational databases correlate well.
- Our comprehensive analysis presented here for phenylalanine, quit similar to data published for valine [25], further supports our strategy based on structure/chemical shift correlation. By now we have compiled data for four different models (For-Xxx-NH_2, Xxx = Gly, Ala, Val and Phe) for all their nine typical backbone conformations. These NMR CS investigations cover most structural subunits of peptides and proteins, matching theoretical data with those derived from experimental data sets. In this paper we present both for our phenylalanine containing diamide model and for phenylalanine residues retrieved from proteins the significant $^1H^\alpha$ and $^{13}C^\alpha$ CS changes as function of the backbone fold.

In summary, *ab initio* CS results presented for the model system For–L-Phe–NH_2 revealed important features of structure-CS interdependence. This theoretical findings encouraged us to investigate a relatively large experimental dataset. Both approaches reveal a coherent picture suggesting the possibility of using CS information directly from multiple-pulse NMR-experiments to obtain other than NOE type structural information.

The authors thank A.K. Füzéry for helpful discussions. This research was partially supported by grants from the Hungarian Scientific Research Fund (OTKA T024044, T032486, and T033074).

References

1. J.H. Prestegard, H. Valafar, J. Glushka, F. Tian, Biochemistry **40**, 8677 (2001)
2. C.A. Fowler, F. Tian, H.M. Al-Hasimi, J.H. Prestegard, J. Mol. Biol. **304**, 447 (2000)
3. G.T. Montelione, Proc. Nat. Acad. Sci. USA **98**, 13488 (2001)
4. (a) L. Müller, J. Am. Chem. Soc. **101**, 4481 (1979); (b) A. Bax, R.H. Griffey, B.L. Hawkins, Magn. Res. **55**, 301 (1983); (c) A.G. Redfield, Chem. Phys. Lett. **96**, 537 (1987)
5. G. Bodenhausen, D.J. Ruben, Chem. Phys. Lett. **69**, 185 (1980)
6. M. Ikura, L.E. Kay, A. Bax, Biochemistry **29**, 4659 (1990)
7. D. Neuhaus, M. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis* (VCH, New York, 1989)
8. K. Wüthrich, in *NMR of Proteins and Nucleic Acids* (Wiley, New York, 1986)
9. K. Wüthrich, Science **243**, 45 (1989)
10. J.A. Smith, L.G. Pease, CRC Crit. Rev. Biochem. Mol. **8**, 315 (1980)
11. M. Nilges, J. Mol. Biol. **245**, 645 (1995)
12. J. Cavanagh, W.J. Fairbrother, A.G. Palmer III, N.J. Skelton, *Protein NMR Spectroscopy. Principles and Practice* (Academic Press, San Diego, 1996)
13. S.W. Fesik, E.R.P. Zuiderweg, Quart. Rev. Biophys. **23**, 97 (1990)
14. G.M. Clore, A.M. Gronenborn, Progr. NMR Spectry. **23**, 43 (1991)
15. A. Bax, S. Grzesiek, Acc. Chem. Res. **26**, 131 (1993)
16. I. Pelczer, B.G. Carter, in *Protein NMR Techniques*, edited by D.G. Reid (Humana Press Totowa, 1997), pp. 71-155
17. F. Tian, H. Valafar, J.H. Prestegard, J. Am. Chem. Soc. **123**, 11791 (2001)
18. N.U. Jain, A. Venot, K. Umemoto, H. Leffler, J.H. Prestegard, Prot. Sci. **10**, 2393 (2001)
19. J.R. Tolman, H.M. Al-Hashimi, L.E. Kay, J.H. Prestegard, J. Am. Chem. Soc. **123**, 1416 (2001)
20. F. Tian, C.A. Fowler, E.R. Zartler, F.A. Jenney, M.W. Adams, J.H. Prestegard, J. Biomol. NMR **18**, 23 (2001)
21. C.A. Fowler, F. Tian, J.H. Prestegard, Biophys. J. **78**, 2827 (2000)
22. J. Gluska, M. Lee, S. Coffin, D. Cowburn, J. Am. Chem. Soc. **111**, 7716 (1989); **112**, 2843 (1990)
23. J.D. Augspurger, C.E. Dykstra, Ann. Rep. NMR Spectr. **30**, 1 (1995)
24. A. Perczel, A.G. Császár, J. Comp. Chem. **21**, 882 (2000)
25. A. Perczel, A.G. Császár, Chem. Eur. J. **7**, 1069 (2001)
26. H. Le, E. Oldfield, J. Phys. Chem. **100**, 16423 (1996)
27. N.R. Luman, M.P. King, J.D. Augspurger, J. Comp. Chem. **22**, 366 (2001)
28. D.S. Wishart, B.D. Sykes, F.M. Richards, J. Mol. Biol. **222**, 311 (1991)
29. J.D. Augspurger, J.G. Pearson, E. Oldfield, C.E. Dykstra, K.D. Park, D. Schwartz, J. Magn. Res. **100**, 342 (1992)
30. D.S. Wishart, B.D. Sykes, Meth. Enzymol. **239**, 363 (1992)
31. L. Szilágyi, Progr. NMR Spectr. **27**, 325 (1995)
32. H. Saito, Magn. Reson. Chem. **24**, 835 (1986) and references therein
33. C.E. Johnson, F.A. Bovey, J. Chem. Phys. **29**, 1012 (1958)
34. (a) S.J. Perkins, K. Wüthrich, Biochim. Biophys. Acta **576**, 409 (1979); (b) C.W. Haigh, R.B. Mallion, Progr. NMR Spectrosc. **13**, 303 (1980); (c) S. Spera, A. Bax, J. Am. Chem. Soc. **113**, 5490 (1991); (d) M.P. Williamson, Biopolymers **29**, 1428 (1990)
35. B. Celda, C. Baimonti, M.J. Arnau, R. Tejero, G.T. Montelione, J. Biomol. NMR **5**, 161 (1995)
36. H. Le, E. Oldfield, J. Biomol. NMR **4**, 341 (1994)
37. (a) A.C. de Dios, J.G. Pearson, E. Oldfield, Science **260**, 1491 (1993); (b) A.C. de Dios, J.G. Pearson, E. Oldfield, J. Am. Chem. Soc. **115**, 9768 (1993); (c) A.C. de Dios, E. Oldfield, J. Am. Chem. Soc. **116**, 5307 (1994); (d) H.-B. Le, J.G. Pearson, A.C. de Dios, E. Oldfield, J. Am. Chem. Soc. **117**, 3800 (1995)
38. M.D. Reily, V. Thanabal, D.O. Omecinsky, J. Am. Chem. Soc. **114**, 6251 (1992)
39. A. Pastore, V. Saudek, J. Magn. Reson. **90**, 165 (1990)
40. (a) B.R. Seavey, E.A. Farr, W.M. Westler, J.L. Markley, J. Biomol. NMR **1**, 217 (1991); (b) J.L. Markley, E.A. Farr, B.R. Seavey, J. Biomol. NMR **1**, 231 (1991); (c) the Biomagnetic Resonance Bank (BMRB) can be reached at `http://www.bmrb.wisc.edu/`
41. W. Gronwald, L. Willard, T. Jellard, R.F. Boyko, K. Rajarathnam, D.S. Wishart, F.D. Sönnichsen, B.D. Sykes, J. Biomol. NMR **12**, 395 (1998)
42. A.G. Császár, A. Perczel, Progr. Biophys. Mol. Biol. **71**, 243 (1999)
43. A.C. deDios, Progr. NMR Spectry. **29**, 229 (1996)
44. D. Jiao, M. Barfield, V.J. Hruby, J. Am. Chem. Soc. **115**, 10883 (1993)
45. N. Asakawa, H. Kurosu, I. Ando, J. Mol. Struct. **323**, 279 (1994)
46. J. Yao, H.J. Dyson, P.E. Wright, FEBS Lett. **419**, 285 (1997)
47. P. Hudáky, I. Jákli, A.G. Császár, A. Perczel, J. Comp. Chem. **23**, 732 (2001)
48. (a) H.M. Sulzbach, P.v.R. Schleyer, H.F. Schaefer III, J. Am. Chem. Soc. **116**, 3967 (1994); (b) H.M. Sulzbach, P.v.R. Schleyer, H.F. Schaefer III, J. Am. Chem. Soc. **117**, 2632 (1995); (c) H.M. Sulzbach, G. Vacek, P.R. Schreiner, J.M. Galbraith, P.v.R. Schleyer, H.F. Schaefer III, J. Comp. Chem. **18**, 126 (1997)
49. D. Jiao, M. Barfield, V.J. Hruby, J. Am. Chem. Soc. **115**, 10883 (1993)
50. J.G. Pearson, J.F. Wang, J.L. Markley, H.-B. Le, E. Oldfield, J. Am. Chem. Soc. **117**, 8823 (1995)
51. R.H. Havlin, H.-B. Le, D.D. Laws, A.C. deDios, E. Oldfield, J. Am. Chem. Soc. **119**, 11951 (1997)
52. J. Heller, D.D. Laws, M. Tomaselli, D.S. King, D.E. Wemmer, A. Pines, R.H. Havlin, E. Oldfield, J. Am. Chem. Soc. **119**, 7827 (1997)
53. R. Ditchfield, Mol. Phys. **27**, 789 (1974)
54. K. Wolinski, J.F. Hinton, P. Pulay, J. Am. Chem. Soc. **112**, 8251 (1990)
55. J. Vaara, J. Kaski, J. Jokisaari, P. Diehl, J. Phys. Chem. A **101**, 5069 (1997)
56. A.C. de Dios, E. Oldfield, J. Am. Chem. Soc. **116**, 11485 (1994)
57. G. Zheng, L.M. Wang, J.Z. Hu, X.D. Zhang, L.F. Shen, C.H. Ye, G.A. Webb, Magn. Res. Chem. **35**, 606 (1997)
58. F. Toma, V. Dive, S. Fermandjian, K. Darlak, Z. Grzonka, Biopolymers **24**, 2417 (1985)
59. C. Baysal, H. Meirovitch, Biopolymers **54**, 416 (2000)

60. E.F. Strittmatter, E.R. Williams, J. Phys. Chem. A **104**, 6069 (2000)
61. T. Yamazaki, K. Nunami, M. Goodman, Biopolymers **31**, 1513 (1991)
62. D.D. Laws, H.-B. Le, A.C. de Dios, R.H. Havlin, E. Oldfield, J. Am. Chem. Soc. **117**, 9542 (1995)
63. J.G. Pearson, H.-B. Le, L.K. Sanders, N. Godbout, R.H. Havlin, E. Oldfield, J. Am. Chem. Soc. **119**, 11941 (1997)
64. P.C. Hariharan, J.A. Pople, Theor. Chim. Acta **28**, 213 (1973)
65. T. Clark, J. Chandrasekhar, G.W. Spitznagel, P.v.R. Schleyer, J. Comp. Chem. **4**, 294 (1983)
66. A. Schafer, C. Huber, R. Ahlrichs, J. Chem. Phys. **100**, 5829 (1994)
67. M.J. Frisch *et al.*, *Gaussian94*, Revision B.2, Gaussian Inc., Pittsburgh PA, 1995
68. M.J. Frisch *et al.*, *Gaussian98*, Revision A.5, Gaussian Inc., Pittsburgh PA, 1998
69. J.L. Markley, A. Bax, Y. Arata, C.W. Hilbers, R. Kaptein, B.D. Sykes, P.E. Wright, K. Wüthrich, Pure Appl. Chem. **70**, 117 (1998)
70. A.G. Császár, W.D. Allen, H.F. Schaefer III, J. Chem. Phys. **108**, 9751 (1998)
71. F.C. Bernstein, T.F. Koetzle, G.J. Williams, E.E. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, J. Mol. Biol. **112**, 535 (1977), the Protein Data Bank (PDB) can be accessed on the Internet at `http://www.pdb.bnl.gov/`
72. C. Ramakrishnan, G.N. Ramachandran, Biophys. J. **5**, 909 (1965)
73. IUPAC-IUB Commission on Biochemical Nomenclature, Biochemistry **9**, 3471 (1970)
74. A. Perczel, J.G. Ángyán, M. Kajtár, W. Viviani, J.-L. Rivail, J.-F. Marcoccia, I.G. Csizmadia, J. Am. Chem. Soc. **113**, 6256 (1991)
75. A. Perczel, M.A. McAllister, P. Császár, I.G. Csizmadia, J. Am. Chem. Soc. **115**, 4849 (1993)
76. A. Perczel, Ö. Farkas, I.G. Csizmadia, J. Am. Chem. Soc. **118**, 7809 (1996)
77. A. Perczel, M. Kajtár, J.F. Marcoccia, I.G. Csizmadia, J. Mol. Struct. (Theochem) **232**, 291 (1991)
78. T. Head-Gordon, M. Head-Gordon, M.J. Frisch, C. Brooks III, J.A. Pople, J. Am. Chem. Soc. **113**, 5989 (1991)
79. A. Perczel, Ö. Farkas, I.G. Csizmadia, J. Am. Chem. Soc. **118**, 7809 (1996)
80. (a) A. Perczel, Ö. Farkas, A.G. Császár, I.G. Csizmadia, Can. J. Chem. **75**, 1120 (1997); (b) I. Jákli, A. Perczel, Ö. Farkas, M. Hollósi, I.G. Csizmadia, J. Mol. Struct. (Theochem) **445**, 303 (1998)
81. P. Hudáky, I. Hudáky, A. Perczel, J. Mol. Struct. (Theochem) **583**, 199 (2002)
82. W. Viviani, J.-L. Rivail, A. Perczel, I.G. Csizmadia, J. Am. Chem. Soc. **115**, 8321 (1993)
83. R.H. Havlin, H.-B. Le, D.D. Laws, A.C. deDios, E. Oldfield, J. Am. Chem. Soc. **119**, 11951 (1997)
84. T.M. Wonnacott, R.J. Wonnacott, *Introductory Statistics*, 5th edn. (Wiley, New York, 1990)
85. H.R. Kricheldorf, D. Müller, Macromolecules **16**, 615 (1983)