

Toward Direct Determination of Conformations of Protein Building Units from Multidimensional NMR Experiments Part II: A Theoretical Case Study of Formyl-L-Valine Amide

András Perczel^{*[a]} and Attila G. Császár^[b]

Abstract: Chemical shielding anisotropy tensors have been determined for all twenty-seven characteristic conformers of For-L-Val-NH₂ using the GIAO-RHF formalism with the 6-31 + G* and TZ2P basis sets. 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 1–3%. Correlated chemical shift plots of selected nuclei, as function of the respective ϕ , ψ , χ_1 , and χ_2 torsional

angles, have been generated. On two-dimensional chemical shift–chemical shift plots, for example, $^1\text{H}^{\text{NH}} - ^{15}\text{N}^{\text{NH}}$ and $^{15}\text{N}^{\text{NH}} - ^{13}\text{C}^\alpha$, regions corresponding to major conformational clusters have been identified, providing a basis for the quantitative identification of conformers from NMR shift data. Experimental NMR resonances of nuclei of valine residues have been deduced from 18 selected proteins, resulting in 93

$^1\text{H}^\alpha - ^{13}\text{C}^\alpha$ chemical shift pairs. These experimental results have been compared to relevant ab initio values revealing remarkable correlation between the two sets of data. Correlations of $^1\text{H}^\alpha$ and $^{13}\text{C}^\alpha$ values with backbone conformational parameters (ϕ and ψ) have also been found for all pairs (e.g. $^1\text{H}^\alpha/\phi$ and $^{13}\text{C}^\alpha/\phi$) but $^1\text{H}^\alpha/\psi$. Overall, the appealing idea of establishing backbone folding of proteins by employing chemical shift information alone, obtained from selected multiple-pulse NMR experiments (e.g. 2D-HSQC, 2D-HMQC, and 3D-HNCA), has received further support.

Keywords: ab initio calculations • chemical shifts • conformational analysis • NMR spectroscopy • structure elucidation

Introduction

The entrenched protocol for determining the three-dimensional structure of proteins utilizing NMR spectroscopy is based primarily on the extraction of structural constraints from nuclear Overhauser exchange spectroscopy (NOESY).^[1] NOEs are first assigned to pairs of atoms (typically protons) and then converted into distances between spatially close pairs of atoms.^[1–7] Data on selected dihedral angles, with information on the hydrogen bonds, could complete the set of constraints employed during structure analysis. Full resonance assignment of ^1H , ^{13}C , and ^{15}N nuclei can be achieved without the use of NOE-type information, through specific sets of 3D experiments (e.g. HNCA, HN(CO)CA, CBCA(CO)NH, HBHA(CO)NH, CC(CO)NH, HCC(CO)NH, and

HCCH-TOCSY),^[6] if and only if doubly-labeled (^{13}C and ^{15}N) protein samples are available. Such resonance assignment strategies, developed for proteins and based on homo- and heteronuclear coupling constants,^[7] bring forth the hope of automated spectrum assignment.^[8] New techniques emerge to directly measure angles between bond vectors^[9a,b] in the solid phase and methods are being developed to determine the secondary structure of proteins in the solid state.^[9c] The use of dipole–dipole cross-correlated relaxation of double- and zero-quantum coherences open new frontiers^[9d] and can provide additional sources of information that can complement NOEs in structure calculation. Nevertheless, even if a full resonance assignment is obtained without NOEs, the most common technique, routinely used at present, to acquire the large number of constraints required for structure elucidation is based primarily on NOE data. Consequently, in one way or another proton–proton distances deduced from dipole–dipole relaxation of protons are quite crucial for the structure determination of peptides and proteins.

Chemical shift changes of a selected nucleus located in the same type of amino acid residue at different sites within a protein are due either to differences in backbone orientations or to the individual molecular environment. If the conforma-

[a] Prof. A. Perczel
Department of Organic Chemistry, Eötvös University
P.O. Box 32, 1518 Budapest 112 (Hungary)
Fax: (+36)-1-3722-620
E-mail: Perczel@para.chem.elte.hu

[b] Prof. A. G. Császár
Department of Theoretical Chemistry, Eötvös University
Budapest (Hungary)

tional effect dominates over other environmental changes, it is reasonable to assume that information on dihedral angles are coded in the chemical shifts themselves. Computer programs developed recently, for example TALOS,^[10] search chemical shift databases for strings of adjacent amino acid residues in order to predict their ϕ, ψ values. Nicely resolved chemical shift data can often be obtained from triple-resonance NMR experiments, since information is spread out in three dimensions. From these 3D data sets useful 2D cuts (e.g., $^1\text{H}-^{15}\text{N}$, $^1\text{H}-^{13}\text{C}$, and $^{15}\text{N}-^{13}\text{C}$) can be extracted and analyzed. An increasing number of chemical shifts have become available for a growing number of proteins investigated by NMR techniques, and they have been deposited in BMRB.^[11] Both these experimental and the complementary computational studies^[12–17] have clearly established a few structure-induced $^{13}\text{C}^\alpha$, $^{15}\text{N}^{\text{NH}}$, and $^1\text{H}^{\text{NH}}$ chemical shift changes in peptides and proteins. On the other hand, due to the still low number of protein structures determined by heteronuclear NMR techniques, correlation between backbone conformers and chemical shifts has been established only for the α -helical and β -sheet regions of the Ramachandran surface.^[18–20] Nevertheless, these results support the appealing idea that analysis of NMR chemical shifts could directly provide vital structural data. However, it is not yet clear whether the analysis of chemical shifts alone provides an effective alternative to the distance-based (NOE) strategy for the elucidation of

the dihedral angle space of proteins. If unambiguous structure–chemical shift correlations were established from shielding data, dependable dihedral angles could be extracted from multiple-pulse experiments (e.g. 2D-HMQC,^[21] 2D-HSQC,^[22] and 3D-HNCA^[23]). To achieve this goal a systematic research correlating NMR shieldings of nuclei with torsional parameters is needed. Today, due to the lack of large number of reliable experimental data, theoretical and as such ab initio computation of conformation-dependent NMR chemical shifts seems to offer the alternative route in establishing and probing such relationships.^[18, 24–31]

The primary goal of ab initio NMR computations on peptides and proteins has been the determination of chemical shielding anisotropy (CSA) tensors and chemical shifts of the ^{13}C , ^{15}N , and ^1H nuclei.^[18, 25–30] Considerable progress has been achieved since the pioneering study of Jiao and co-workers,^[25] who determined the $^{13}\text{C}^\alpha$ CSA tensor of For-Gly-NH₂ as a function of the backbone conformation. It is now well established that $^{13}\text{C}^\alpha$ values in the helical and in the extended backbone conformations are shifted by ≈ 2.3 ppm downfield and ≈ 2.9 ppm upfield, respectively, as compared to the random coil value. Alanine diamide models (e.g. For-L-Ala-NH₂)^[18c, 26, 31] proved to be particularly popular for such studies. In order to investigate the effect of intramolecular hydrogen bonds on shieldings of different nuclei, for example on ^{13}C , GIAO-RHF calculations have been performed both on small model systems (e.g. *N*-methylacetamide interacting with formamide^[32, 33]) and on larger molecules such as For-(Ala)₅-NH₂.^[33] The most significant perturbation caused by hydrogen bonding was observed on carbonyl carbons.^[34] Although alanine-containing peptide models are of importance to decipher NMR shielding properties of peptides, additional amino acid residues must be investigated to understand the influence of different side chains and conformations on the NMR properties of nuclei forming the backbone of proteins. In the present study we attempt to provide a full NMR description of a peptide model containing valine, namely For-L-Val-NH₂.

Although longer polypeptides and proteins exhibit unique and more or less stable folds, their 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 the library of all accessible conformers for any amino acid residue found in these macromolecules. Therefore, theoretical conformational analysis of computationally accessible model systems, such as For-L-Val-NH₂, are of considerable utility^[35] and nowadays peptide models of the size of For-L-Val-NH₂ can be routinely studied by ab initio techniques. Nevertheless, results from ab initio calculations are occasionally ignored or downplayed by skeptics, questioning particularly the relevance of “environment-free” conformers when deciphering protein structures. Therefore, we must note that in proteins most valines ($\geq 80\%$) are buried, often contributing to the hydrophobic core region of the globular system, where water is mostly excluded. For this reason we consider ab initio results obtained for hydrophobic amino acid residues as adequate for “real” polypeptide units

Abstract in Hungarian: *A For-L-Val-NH₂ molekula mind a 27 lehetséges konformerére kiszámítottuk az összes atom NMR kémiai eltolódásának értékeit a GIAO-RHF formalizmus valamint a 6-31 + G* és a TZ2P bázisok. A számított értékeket, valamint azok konformációs átlagait rendre összevetettük a “Bio Magnetic Resonance Bank”-ban (BMRB) található kísérleti adatokkal. A számításoknál alkalmazott legmagasabb elméleti szint esetén—az amid protonoktól eltekintve—az összes atomtípusra igaz, hogy a meghatározott kísérleti és elméleti adatok különbsége mindössze 1–3 relatív százalék. A ϕ , ψ , χ_1 és χ_2 torziószögek függvényében elkészítettük a fontosabb atomtípusok korrelációs térképeit. A peptidgerinc főbb konformációs családjai jól meghatározott és egymástól elkülönülő alcsoportokat alkotnak olyan kémiai eltolódás–kémiai eltolódás kétdimenziós térképeken, mint például a $^1\text{H}^{\text{NH}}-^{15}\text{N}^{\text{NH}}$ vagy a $^{15}\text{N}^{\text{NH}}-^{13}\text{C}^\alpha$ felületek. Ez a megfigyelés elvi alapként szolgál a molekuláris térszerkezet NMR-eltolódásértékeken alapuló meghatározásának kvantitatív kidolgozásához. Tizennyolc gondosan kiválogatott fehérje NMR-adatainak analízise a valinra vonatkozóan összesen 93 darab $^1\text{H}^\alpha-^{13}\text{C}^\alpha$ adatpárt határozott meg, és ezek a kísérleti értékek az elméleti számítások során meghatározott adatokkal nagyfokú egyezést mutatnak. A $^1\text{H}^\alpha$ és $^{13}\text{C}^\alpha$ valamint a ϕ és ψ adatok korrelációja (pl. $^1\text{H}^\alpha/\phi$, $^{13}\text{C}^\alpha/\phi$) a $^1\text{H}^\alpha/\psi$ adatpár kivételével rendre szignifikáns. A fenti eredmények tovább erősítik azt az elképzelést, hogy a fehérjék gerinckonformációit csupán a megfelelő többdimenziós NMR kísérletekből (pl. 2D-HSQC, 2D-HMQC, 3D-HNCA) meghatározható kémiai eltolódásértékek alapján is meglehetősen határozni.*

folded in aqueous media. We also note, especially in this respect, that one of our recent studies^[36] indicates that ab initio conformers of For-L-Val-NH₂ show great similarity to -Val- conformations found in globular proteins and that computed relative energies and natural occurrences of relevant conformers correlate remarkably well.

Using the gauge-including atomic orbital restricted Hartree–Fock (GIAO-RHF) method,^[37, 38] the ¹³C CSA tensors of For-L-Val-NH₂^[18c] have been determined as a function of the main-chain fold. It has been concluded that a) substitution on C^α consistently induces larger shielding shift on C^β than on C^α,^[18c] and b) the ϕ , ψ , and χ_1 conformational parameters have a significant influence on chemical shifts. By careful analyses of chemical shift changes of Ala and Val residues in proteins such as calmodulin and nuclease the applicability of a direct strategy for structure elucidation has been probed,^[18c] although the NMR shieldings have been investigated^[28] only with respect to the helical and β -sheet conformations. For all residues studied ¹³C^α showed the expected ≈ 5 ppm increase in the isotropic shielding of the β -conformation over the helical structure. Furthermore, the diagonal tensor elements were all found to be sensitive to changes in the ϕ , ψ , and χ_1 torsion angles. Pearson and co-workers^[27] have found that in a protein-like nuclease the $[\phi, \psi]$ values of Val residues can be estimated using chemical shifts, although the results obtained are less dependable than those derived from NOE values or from *J*-coupling constraints. In summary, all calculations indicate that it is possible to deduce certain backbone and side chain orientations and to some extent even their torsion angles from ¹³C^α shielding tensors. Laws and co-workers^[39] have conducted a research project to determine the lowest appropriate theoretical level (*n*-particle space) and basis set (one-particle space) to calculate CSA tensors of peptides. Somewhat unfortunately, only substandard and split-valence basis sets [STO-3G, 3-21G, 4-31G, and 6-311 + G(2d)] have been tested on Ala and Val amino acid residues. Nevertheless, it has been found that chemical shifts of these residues determined using smaller basis sets correlate remarkably well with results obtained with larger basis sets, even when the underlying reference geometries are slightly different. This opens up the possibility of scaling of chemical shifts computed under rather different conditions. In a further study of relevance,^[40] by using empirical chemical shift surfaces Pearson et al. attempted to predict ¹³C NMR shifts of valine residues in calmodulin, nuclease, and ubiquitin, utilizing their X-ray structures. Most RHF results and experimental values showed worse agreement than expected, which improved slightly by using density functional theory-type calculations. Pearson and co-workers^[39] have thus concluded that for accurate chemical shift computations, geometry optimization and the inclusion of electron correlation in the theoretical treatment appear to be important.

Table 1. Calculations employed for the For-L-Val-NH₂ model system.

Level	NMR computation	Geometry optimization	No. of structures	Comment
A1	GIAO-RHF/6-31 + G*	RHF/3-21G	27	full opt. plus $[\phi, \psi]$ constraints
A2	GIAO-RHF/6-31 + G*	RHF/3-21G	20	full optimization
B1	GIAO-RHF/TZ2P	B3LYP/6-311 ++ G**	27	full opt. plus $[\phi, \psi]$ constraints
B2	GIAO-RHF/TZ2P	B3LYP/6-311 ++ G**	18	full optimization

In this report a library of calculated geometric parameters and chemical shifts is established for the model peptide For-L-Val-NH₂. Partly because of the lack of detailed experimental information, it is not obvious what is the best use of the large number of quantum chemical data in the library. We feel that a thorough statistical analysis, even in its simplest, linearized form, offers the best way to confirm existing structure–structure, structure–chemical shift, and chemical shift–chemical shift correlations and derive new ones. Therefore, we compare direct and statistically averaged chemical shifts, obtained from BMRB, to their theoretically determined counterparts. Furthermore, an attempt is made to correlate calculated isotropic NMR shieldings and chemical shifts with *all* characteristic backbone conformations^[41] of For-L-Val-NH₂, concentrating not only on ¹³C but also on ¹⁵N and ¹H NMR shifts. Finally, an effort is made, utilizing experimental and ab initio results determined as part of this study, to understand valine chemical shifts found in 18 selected proteins.

Experimental Section

The For-L-Val-NH₂ model system is depicted on Figure 1. Average chemical shifts of all nuclei forming the amino acid core are also reported therein, as taken from the BioMagnetic Resonance Bank (BMRB).^[11] Computation of NMR chemical shielding anisotropy (CSA) tensors, performed at the GIAO-RHF (gauge including atomic orbitals restricted Hartree–Fock)^[37, 38] level employing basis sets of 6-31 + G*^[42] and TZ2P^[43] quality, utilized the Gaussian94^[44] and Gaussian98^[45] program packages.

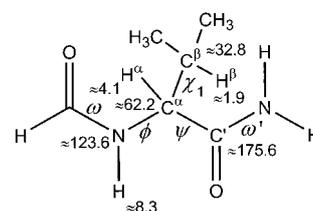


Figure 1. The For-L-Val-NH₂ model; approximate chemical shifts of valine residues within proteins (δ values taken from BMRB) are indicated next to certain nuclei.

Reference geometries employed for CSA computations have been determined at the 3-21G RHF and 6-311 ++ G** DFT(B3LYP) levels. The DFT level was employed extensively for geometry optimizations as it proved to be especially reliable not only for peptides (e.g., For-Gly-NH₂,^[31] For-L-Ala-NH₂,^[31] and For-L-Val-NH₂^[36]) but also for neutral amino acids, such as glycine^[46, 47] and α -alanine.^[48] Consequently, geometric results obtained at the DFT level are considered to be much more reliable than the 3-21G RHF ones. Our goal has been to obtain a library of chemical shifts covering all twenty-seven characteristic conformers of For-L-Val-NH₂. Therefore, in addition to full geometry optimizations, for selected conformers *constrained* geometry optimizations had to be performed keeping the ϕ and ψ torsional angles constant at values characteristic for the related catchment regions. Different combinations of full and constrained geometry optimizations and CSA calculations are designated as levels A1 to B2, as given on Table 1.

Selected geometric parameters (including optimized/fixed ϕ and ψ values), as well as energies and isotropic chemical shielding values (σ scale) are reported in Table 2–3. When relative chemical shifts (δ scale) are used, the appropriate isotropic chemical shielding values of ^1H , ^{13}C , and ^{15}N are referenced to ^1H and ^{13}C of tetramethylsilane (TMS) and to ^{15}N of NH_3 .

(The reference geometry chosen for NH_3 corresponds to the aug-cc-pVTZ CCSD(T) optimized geometry,^[49] while the geometry of TMS has been optimized at the 3-21G RHF level.)

Average (or random) chemical shifts and associated standard deviations for Val are taken from data deposited in BMRB^[11] (see Figure 1 and Table 4).

Table 2. Selected conformational parameters and ab initio chemical shift (δ) values^[a] of For-L-Val- NH_2 determined at the GIAO-RHF/TZ2P//DFT(B3LYP)/6-311++G** (B1) and at the GIAO-RHF/6-31+G**/RHF/3-21G (A1) levels of theory.

Conf. ^[b]	PDB conf. ^[c]	Level ^[d]	ϕ, ψ ^[e]	ϕ	ψ	χ_1	$^{15}\text{N}^{\text{NH}}$	$^1\text{H}^{\text{NH}}$	$^{13}\text{C}^\alpha$	$^1\text{H}^\alpha$	$^{13}\text{C}^\beta$	$^1\text{H}^\beta$	$^{13}\text{C}^\gamma$
$\alpha_{\text{D}}(g+)$	(a)	A1	opt.	60.3	40.9	76.4	101.24	4.36	55.45	2.45	24.05	2.76	170.40
		B1		63.2	39.1	83.2	120.15	4.87	59.43	2.65	27.06	2.85	182.71
$\alpha_{\text{D}}(a)$	(g-)	A1	opt.	47.3	44.6	178.7	95.37	4.20	54.12	2.80	27.53	2.27	168.64
		B1		51.7	33.7	-179.8	115.12	4.73	60.00	3.06	29.93	2.77	182.20
$\alpha_{\text{D}}(g-)$	(g+)	A1	opt.	50.0	43.1	-44.9	106.99	4.28	52.83	2.88	34.65	1.31	168.82
		B1		51.2	39.6	-42.8	127.78	4.77	58.39	3.12	38.56	1.51	181.86
$\alpha_{\text{L}}(g+)$	(a)	A1	constr.	-54.0 ^[f]	-45.0	61.5	104.14	4.19	56.05	3.25	26.74	1.21	170.77
		B1		-54.0	-45.0	54.8	125.65	4.72	60.20	3.55	30.58	1.40	183.40
$\alpha_{\text{L}}(a)$	(g-)	A1	constr.	-54.0	-45.0	176.0	97.76	4.01	54.56	3.68	24.87	1.88	170.88
		B1		-54.0	-45.0	176.6	117.57	4.68	59.16	3.97	28.59	2.14	184.38
$\alpha_{\text{L}}(g-)$	(g+)	A1	constr.	-54.0	-45.0	-54.8	105.73	3.97	54.08	3.72	26.14	1.34	169.22
		B1		-54.0	-45.0	-54.2	126.36	4.56	58.77	3.98	30.85	1.55	182.29
$\beta_{\text{L}}(g+)$	(a)	A1	opt.	-137.5	143.5	66.5	104.94	5.45	52.21	3.74	30.02	1.46	174.17
		B1		-118.8	125.8	59.2	124.64	5.36	54.37	3.97	31.77	1.77	185.07
$\beta_{\text{L}}(a)$	(g-)	A1	opt.	-142.4	163.5	172.9	97.04	5.64	50.02	4.12	30.86	1.28	173.59
		B1		-131.1	162.2	178.8	114.21	6.09	52.81	4.42	33.61	1.63	184.58
$\beta_{\text{L}}(g-)$	(g+)	A1	opt.	-163.2	157.4	-55.2	103.11	5.85	52.16	3.75	24.81	2.37	171.88
		B1		-151.9	157.6	-49.6	121.65	6.56	54.81	4.12	31.23	2.20	182.95
$\delta_{\text{D}}(g+)$	(a)	A1	opt.	-136.7	-59.9	67.2	103.26	4.16	54.43	4.17	31.02	0.98	171.11
		B1		-126.9	-68.0	60.0	122.49	4.62	57.59	4.46	34.10	1.16	182.16
$\delta_{\text{D}}(a)$	(g-)	A1	opt.	-170.4	-46.6	141.4	100.63	4.03	56.21	3.64	28.14	1.50	172.36
		B1	constr.	-144.0	-54.0	169.0	115.67	4.68	58.41	4.68	31.92	1.78	182.58
$\delta_{\text{D}}(g-)$	(g+)	A1	opt.	-175.5	-34.6	-55.4	100.64	3.96	56.62	3.87	22.49	3.20	171.09
		B1		-159.3	-40.1	-55.1	119.07	4.91	60.18	4.35	28.10	2.76	182.99
$\delta_{\text{L}}(g+)$	(a)	A1	opt.	-130.0	30.0	80.8	109.38	3.85	49.90	4.33	27.96	1.68	173.23
		B1		-130.0	30.0	83.7	128.03	4.34	53.38	4.46	30.51	1.96	184.17
$\delta_{\text{L}}(a)$	(g-)	A1	constr.	-125.5	28.9	178.9	98.66	4.06	49.37	4.64	24.87	2.47	171.84
		B1		-114.0	12.1	177.2	115.97	4.67	54.19	4.68	27.48	2.90	183.65
$\delta_{\text{L}}(g-)$	(g+)	A1	opt.	-137.0	36.1	-49.2	109.42	4.35	48.33	4.58	30.59	1.39	171.95
		B1		-112.1	4.3	-43.0	128.25	4.86	53.08	4.74	35.68	1.52	184.04
$\varepsilon_{\text{D}}(g+)$	(a)	A1	opt.	75.2	152.6	61.9	98.34	3.68	59.72	2.35	24.77	3.53	170.70
		B1		85.1	133.3	61.4	119.22	4.03	64.50	2.51	27.61	3.56	183.42
$\varepsilon_{\text{D}}(a)$	(g-)	A1	opt.	70.5	170.3	-152.6	96.23	3.72	55.29	3.02	27.11	1.47	169.25
		B1	constr.	84.1	152.0	-158.1	119.23	4.02	61.47	3.22	31.49	1.79	182.57
$\varepsilon_{\text{D}}(g-)$	(g+)	A1	opt.	75.4	162.6	-17.6	102.56	3.86	57.15	2.58	28.09	2.15	169.63
		B1		83.7	160.6	-18.7	122.83	4.19	61.59	2.96	30.09	2.57	181.68
$\varepsilon_{\text{L}}(g+)$	(a)	A1	constr.	-60.0 ^[g]	120.0	60.6	109.83	4.27	55.40	2.96	25.96	1.77	174.96
		B1		-60.0	120.0	60.2	128.63	4.73	57.96	3.29	29.49	1.85	188.39
$\varepsilon_{\text{L}}(a)$	(g-)	A1	constr.	-60.0	120.0	172.8	106.40	4.66	51.75	3.34	25.82	1.53	174.58
		B1		-60.0	120.0	172.0	117.95	5.08	54.81	3.62	28.91	1.85	188.38
$\varepsilon_{\text{L}}(g-)$	(g+)	A1	constr.	-60.0	120.0	-62.6	100.30	4.49	51.88	3.30	25.51	1.66	175.67
		B1		-60.0	120.0	-58.6	125.07	5.30	54.91	3.67	30.12	1.75	187.47
$\gamma_{\text{D}}(g+)$	(a)	A1	opt.	74.2	-61.5	58.0	102.75	4.48	62.56	2.65	22.93	2.70	172.19
		B1		72.8	-63.2	63.3	122.51	5.09	66.86	2.87	27.12	2.75	183.13
$\gamma_{\text{D}}(a)$	(g-)	A1	opt.	59.5	-38.5	-168.8	101.19	4.10	61.48	3.17	29.66	1.94	172.39
		B1		57.2	-31.0	-172.1	121.46	4.70	65.77	3.34	33.33	2.27	184.06
$\gamma_{\text{D}}(g-)$	(g+)	A1	opt.	62.8	-39.1	-35.8	110.56	4.45	60.76	3.07	33.47	1.57	171.81
		B1		61.5	-39.0	-36.8	131.69	5.04	65.31	3.32	37.94	1.77	182.27
$\gamma_{\text{L}}(g+)$	(a)	A1	opt.	-86.6	71.4	65.9	113.02	4.27	51.79	3.57	24.30	1.91	172.18
		B1		-83.9	82.4	65.8	130.78	4.82	54.94	3.79	27.04	2.13	183.12
$\gamma_{\text{L}}(a)$	(g-)	A1	opt.	-84.9	62.7	173.2	103.07	4.54	49.31	3.83	24.16	2.06	173.24
		B1		-83.2	62.0	170.4	121.84	5.23	53.06	4.06	26.67	2.39	183.46
$\gamma_{\text{L}}(g-)$	(g+)	A1	opt.	-85.3	65.6	-56.9	110.77	4.76	48.22	3.88	27.96	1.36	172.77
		B1		-83.8	74.1	-57.8	128.49	5.51	51.53	4.13	31.36	1.61	182.87

[a] All chemical shifts reported are relative to the appropriate TMS and NH_3 isotropic chemical shieldings, determined at ab initio levels A1 and B1, respectively. [b] Symbols employed for the relevant conformation type α_{L} , β_{L} , ε_{L} etc. describe backbone conformations, while $g+$, a and $g-$ describe side chain (χ_1) orientation using the $\text{N}-\text{C}^\alpha-\text{C}^\beta-\text{H}^\beta$ torsional angle. [c] Description of side chain conformations according to the convention used in PDB. [d] See Table 1. [e] Opt.: all $(3n-6)$ internal coordinates are optimized; constr.: all internal coordinates but the $[\phi, \psi]$ torsions, i.e., $(3n-8)$ coordinates altogether, are optimized. [f] The torsional angles $\phi = -54^\circ$ and $\psi = -45^\circ$ are typical values for helical secondary structural elements found in globular proteins. [g] The torsional angles $\phi = -60^\circ$ and $\psi = 120^\circ$ are typical values for poly-proline II secondary structural elements found in globular proteins.

Table 3. Unsigned differences^[a] between GIAO-RHF/TZ2P//DFT(B3LYP)/6-311++G** and GIAO-RHF/6-31+G**/RHF/3-21G conformational and chemical shift (δ) parameters of For-L-Val-NH₂.^[b]

Conf. ^[c]	PDB conf. ^[d]	ϕ, ψ ^[e]	$\Delta\phi$	$\Delta\psi$	$\Delta\chi_1$	$\Delta^{15}\text{N}$	$\Delta^1\text{H}$	$\Delta^{13}\text{C}^\alpha$	$\Delta^1\text{H}^\alpha$	$\Delta^{13}\text{C}^\beta$	$\Delta^1\text{H}^\beta$	$\Delta^{13}\text{C}^\gamma$
$\alpha_{\text{D}}(g+)$	(a)	opt.	2.88	1.76	6.76	18.91	0.51	3.97	0.20	3.01	0.10	12.31
$\alpha_{\text{D}}(a)$	(g-)	opt.	4.45	10.84	1.12	19.74	0.53	5.88	0.26	2.39	0.50	13.56
$\alpha_{\text{D}}(g-)$	(g+)	opt.	1.17	3.52	2.15	20.79	0.50	5.55	0.25	3.91	0.20	13.04
$\alpha_{\text{L}}(g+)$	(a)	constr.	0.00	0.00	6.79	21.51	0.53	4.15	0.30	3.84	0.19	12.63
$\alpha_{\text{L}}(a)$	(g-)	constr.	0.00	0.00	0.62	19.81	0.67	4.60	0.29	3.72	0.26	13.50
$\alpha_{\text{L}}(g-)$	(g+)	constr.	0.00	0.00	0.54	20.63	0.59	4.69	0.26	4.71	0.21	13.07
$\beta_{\text{L}}(g+)$	(a)	opt.	18.63	17.66	7.23	19.70	0.09	2.16	0.23	1.74	0.32	10.90
$\beta_{\text{L}}(a)$	(g-)	opt.	11.26	1.34	5.95	17.16	0.45	2.79	0.31	2.75	0.35	10.99
$\beta_{\text{L}}(g-)$	(g+)	opt.	11.36	0.19	5.67	18.54	0.71	2.65	0.37	6.42	0.17	11.07
$\delta_{\text{D}}(g+)$	(a)	opt.	9.79	8.08	7.18	19.23	0.46	3.16	0.29	3.08	0.18	11.04
$\delta_{\text{D}}(a)$	(g-)	opt.-constr.	26.38	7.45	27.53	15.04	0.65	2.20	1.04	3.77	0.28	10.21
$\delta_{\text{D}}(g-)$	(g+)	opt.	16.24	5.52	0.33	18.43	0.95	3.56	0.48	5.61	0.44	11.90
$\delta_{\text{L}}(g+)$	(a)	opt.	0.00	0.00	2.88	18.65	0.50	3.48	0.13	2.55	0.28	10.93
$\delta_{\text{L}}(a)$	(g-)	constr.	11.55	16.81	1.72	17.31	0.61	4.82	0.04	2.61	0.43	11.80
$\delta_{\text{L}}(g-)$	(g+)	opt.	24.84	31.71	6.18	18.83	0.51	4.75	0.16	5.09	0.13	12.09
$\varepsilon_{\text{D}}(g+)$	(a)	opt.	9.90	19.32	0.48	20.88	0.35	4.78	0.16	2.84	0.03	12.72
$\varepsilon_{\text{D}}(a)$	(g-)	opt.-constr.	13.57	18.36	5.52	23.00	0.30	6.18	0.20	4.38	0.32	13.31
$\varepsilon_{\text{D}}(g-)$	(g+)	opt.	8.23	2.02	1.13	20.26	0.34	4.44	0.38	2.00	0.42	12.05
$\varepsilon_{\text{L}}(g+)$	(a)	constr.	0.00	0.00	0.48	18.79	0.46	2.56	0.33	3.53	0.08	13.43
$\varepsilon_{\text{L}}(a)$	(g-)	constr.	0.00	0.00	0.72	11.55	0.42	3.06	0.28	3.09	0.32	13.80
$\varepsilon_{\text{L}}(g-)$	(g+)	constr.	0.00	0.00	3.99	24.76	0.81	3.03	0.37	4.61	0.09	11.80
$\gamma_{\text{D}}(g+)$	(a)	opt.	1.34	1.75	5.28	19.76	0.61	4.29	0.23	4.19	0.05	10.94
$\gamma_{\text{D}}(a)$	(g-)	opt.	2.30	7.55	3.30	20.27	0.60	4.29	0.17	3.67	0.33	11.67
$\gamma_{\text{D}}(g-)$	(g+)	opt.	1.27	0.09	0.98	21.12	0.60	4.55	0.25	4.47	0.20	10.46
$\gamma_{\text{L}}(g+)$	(a)	opt.	2.65	11.02	0.03	17.76	0.55	3.15	0.22	2.74	0.23	10.94
$\gamma_{\text{L}}(a)$	(g-)	opt.	1.76	0.68	2.83	18.77	0.69	3.75	0.23	2.50	0.33	10.22
$\gamma_{\text{L}}(g-)$	(g+)	opt.	1.55	8.46	0.90	17.72	0.75	3.31	0.25	3.40	0.25	10.10
average unsigned difference						19.2	0.5	3.9	0.3	3.6	0.2	11.9
standard deviation from average						2.4	0.2	1.1	0.2	1.1	0.1	1.2

[a] All chemical shifts reported are relative to TMS or NH₃, determined at ab initio levels A1 or B1 (see text). [b] Values reported in Table 2 have been used when subtracting level B1 data from level A1 data. See Table 1 for the definition of the theoretical levels. [c] See footnote b of Table 2. [d] Description of side chain conformations according to convention used in PDB. [e] Opt. = all (3*n* - 6) internal coordinates are optimized; constr. = all internal coordinates but ϕ and ψ are optimized.

Table 4. Selected structural and relative chemical shift (δ) values of all conformers of For-L-Val-NH₂ obtained by simple arithmetical averaging.

Conf. ^[a]	Level ^[b]	sc. ^[c]	ϕ	ψ	ΔE ^[d]	$p(i)/\sum p(i)$ ^[e]	N ^{NH}	H ^{NH}	C ^{α}	H ^{α}	C ^{β}	H ^{β}	C ^{γ}
α_{L}	A1	3	-54.0	-45.0	5.68	0.00	102.54	4.06	54.90	3.55	25.92	1.48	170.29
	B1	3	-54.0	-45.0	5.35	0.00	123.19	4.65	59.37	3.83	30.00	1.70	183.35
α_{D}	A1	3	52.5	42.9	7.26	0.00	101.20	4.28	54.13	2.71	28.74	2.11	169.29
	B1	3	55.4	37.5	6.74	0.00	121.01	4.79	59.27	2.94	31.85	2.38	182.25
β_{L}	A1	3	-147.7	154.8	0.21	0.58	101.70	5.65	51.46	3.87	28.56	1.70	173.21
	B1	3	-133.9	148.5	0.24	0.60	120.16	6.00	53.99	4.17	32.20	1.87	184.20
γ_{L}	A1	3	-85.6	66.6	0.86	0.35	108.95	4.52	49.77	3.76	25.47	1.78	172.73
	B1	3	-83.6	72.9	0.99	0.33	127.03	5.19	53.17	3.99	28.35	2.04	183.15
γ_{D}	A1	3	65.5	-46.4	4.33	0.02	104.83	4.34	61.60	2.96	28.69	2.07	172.13
	B1	3	63.9	-44.4	4.62	0.01	125.22	4.94	65.98	3.18	32.79	2.26	183.15
δ_{L}	A1	3	-130.8	31.6	2.88	0.03	105.82	4.09	49.20	4.52	27.80	1.85	172.34
	B1	3	-118.7	15.5	3.31	0.01	124.08	4.62	53.55	4.63	31.22	2.13	183.95
δ_{D}	A1	3	-160.9	-47.0	7.26	0.00	101.51	4.05	55.75	3.90	27.22	1.89	171.52
	B1	3	-143.4	-54.0	6.82	0.00	119.07	4.74	58.72	4.50	31.37	1.90	182.57
ε_{L}	A1	3	-60.0	120.0	2.78	0.03	105.51	4.47	53.01	3.20	25.76	1.65	175.07
	B1	3	-60.0	120.0	2.26	0.05	123.88	5.04	55.89	3.53	29.50	1.82	188.08
ε_{D}	A1	3	73.7	161.9	9.98	0.00	99.04	3.75	57.39	2.65	26.65	2.38	169.86
	B1	3	84.3	148.6	9.96	0.00	120.42	4.08	62.52	2.90	29.73	2.64	182.55

[a] Conf. = backbone conformation (see text). [b] See Table 1. [c] sc. = number of side chain conformers used when chemical shift values have been averaged (for relevant CSA data see Table 2). [d] Averages of relative energies within a given backbone catchment region. Typically three (in some cases less) side chain conformations have been determined at the 6-31+G* (Level A1) or TZ2P (Level B1) RHF levels. [e] Relative populations are calculated based on relative energies as $\exp(-\Delta E/kT)/\sum \exp(-\Delta E/kT)$, where $kT = 0.595371$ [kcal mol⁻¹] at $T = 300$ K.

Chemical shift data used for statistical purposes has to be referenced rather carefully to a common standard. This was done by Dr. David Wishart and the results were kindly provided for us for the following 18 proteins: 1BPI, 3LZM, 1LZ1, 2RN2, 2RNT, 1SNC, 1HCB, 1UBQ, 1CEX, 1GZI, 5P21, 1ROP, 1ICM, 192L, 1IGD, 3RN3, 2TRX, and 1A2P. The relevant

experimental structures of these proteins were retrieved from the Protein Data Bank (PDB).^[50] Data from the above two sources (BMRB^[11] and PDB^[50]) were aligned and employed as an experimental database for testing theoretically predicted chemical shift values. One has to note that no C ^{α} chemical shifts are available for proteins 1LZ1, 2RNT, 1GZI, 1A2P,

and 192L. Furthermore, for 2TRX there are two, while for 1A2P there are three relevant PDB files, with slightly different coordinates. By removing all questionable entries a total of 93 unambiguous Val residues remained, where both proton and carbon chemical shifts as well as dihedral parameters were available. Although the 93 valines form a significantly larger database than any previous one, even more experimental data is needed to draw an unambiguous statistical picture.

Results and Discussion

Structures and energetics: Traditionally, the $[\phi, \psi]_i$ torsional angle pair is used to describe the backbone conformation of residue i within a protein.^[51, 52] The torsional potentials along both ϕ and ψ are expected to have a maximum of three minima for an alpha amino acid residue. Consequently, we expect nine characteristic conformers for each “peptide unit” (-NH-CHR-CO-).^[53, 54, 35, 55] These nine conformers are named, according to our established convention,^[53, 35] as follows: α_L , α_D , β_L , γ_L , γ_D , δ_L , δ_D , ε_L , and ε_D (see also Figure 2 for the Ramachandran surface of For-L-Val-NH₂ determined at two levels of theory).

The $[\phi, \psi]$ values of residues found in hundreds of non-homologous proteins, in which the structures have been investigated by X-ray diffraction, revealed the existence of all nine backbone conformer types, though with very different

abundances. Conformers such as β_L , γ_L , δ_L , and ε_L , forming the broad β region, or structures such as α_L , symbolizing the helical subspace (3_{10} - and/or 4_{13} -helices) of the Ramachandran surface, are observed most often. On the other hand, D-type conformers (α_D , γ_D , δ_D , and ε_D) are rare for amino acid residues of S chirality.

Systematic ab initio calculations carried out on amino acid diamide models For-Xaa-NH₂ and Ac-Xaa-NHMe, where Xaa = Gly,^[31] Ala,^[31, 56] Ser,^[55, 57] Phe,^[58] Val,^[36, 59] resulted in considerably fewer minima than the maximum number allowed. For example, at the 3-21G(6-311++G**) RHF level two (three) conformers could not be located for For-L-Ala-NH₂. These calculations showed that the α_L , ε_L , and ε_D conformers are missing most often in dipeptides of various amino acids. Nevertheless, existence of these “missing” minima has been verified by ab initio techniques for larger peptide models.

The hydrophobic and rather compact side chain of Val is a typical representative of those amino acid residues (Val, Ile, and Thr) which have a single H ^{β} . Formally, the isopropyl side chain of Val, -CH(CH₃)₂, is derived from the methyl side chain of Ala: the two H ^{β} s are replaced by two methyl groups. Both methyl groups adopt typically a *gauche*+ -like orientation ($\chi_2 = 60^\circ$ and $\chi_2' = 60^\circ$). Consequently, it is important to restrict the systematic conformational analysis of a valine dipeptide model (HCO-L-Val-NH₂ or CH₃CO-L-Val-NHCH₃) to three torsional variables: $E = E(\phi, \psi, \chi_1)$, with χ_2 and χ_2' both around 60° . The resulting potential energy hypersurface is expected to have 27 characteristic conformers (three χ_1 orientations for each of the nine typical backbone conformers). Twenty out of the twenty-seven molecular structures were indeed determined previously at the 3-21G RHF level.^[59] When the side chain (χ_1) has a *gauche*+ or *gauche*- orientation, all seven backbone conformers (α_D , β_L , γ_L , γ_D , δ_L , δ_D , and ε_D) typical of For-L-Ala-NH₂ have also been assigned for For-L-Val-NH₂. However, if χ_1 has an *anti* orientation, not only the expected α_L and ε_L but also the δ_L minimum vanishes.

In this study we do not intend to discuss structural and energetic features of For-L-Val-NH₂ in detail, since this has been done both at the 3-21G RHF^[59] and 6-311++G** DFT(B3LYP) levels.^[36] Selected relevant conclusions of these studies are nevertheless repeated here as follows: a) Out of the 27 expected characteristic structures, at the RHF level only 20, while at the DFT level only 18 conformers correspond to minima. As mentioned in the previous section, in the present study the remaining 7(9) structures of For-L-Val-NH₂ have also been optimized, using constrained $[\phi, \psi]$ torsion angles; b) the $[\phi, \psi]$ torsional angles do not change significantly when the 3-21G RHF geometries are reoptimized at the DFT level (compare panels A and B of Figure 2); c) the dihedral angles of the conformers correlate well with those derived from proteins;^[36] and d) the calculated energy order of the ab initio conformers shows remarkable correlation with the relative abundances of conformers found in proteins.^[36]

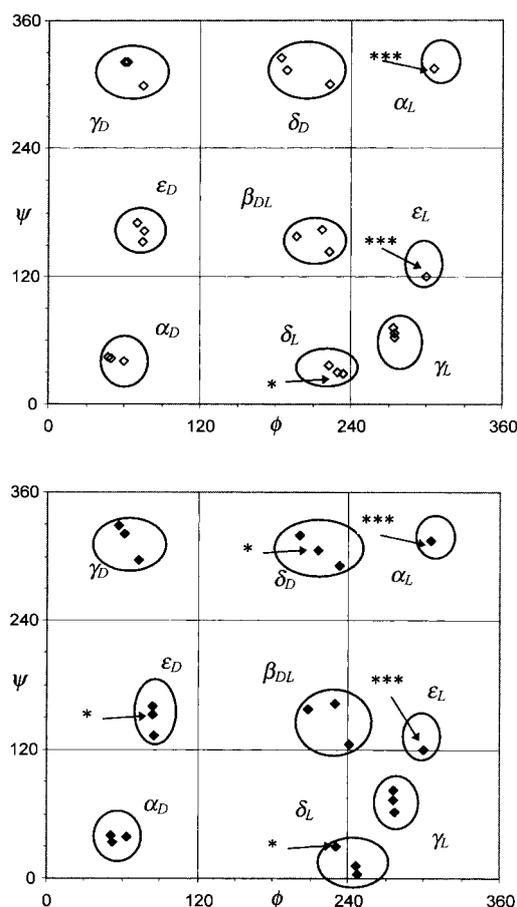


Figure 2. 3-21G RHF (\diamond , top) and 6-311++G** B3LYP (\blacklozenge , bottom) locations of the different conformers of For-L-Val-NH₂ on a Ramachandran surface. (* stands for conformers with constrained $[\phi, \psi]$ torsional angles, three stars (e.g. in the case of α_L) means that all three side chain rotamers ($g+$, a , and $g-$) were calculated with constrained backbone parameters.)

Accuracy of computed chemical shifts: The average NMR chemical shifts (and their standard deviations) for all nuclei of the valine residue have been extracted from BMRB^[11] and are

reported in Table 5 and Table 6. Utilizing these data we have attempted to answer the question of how well the calculated chemical shifts compare with their experimental counterparts.

The conformationally averaged experimental shifts of all nuclei of interest ($^{15}\text{N}^{\text{NH}}$, $^{13}\text{C}^{\alpha}$, $^{13}\text{C}^{\beta}$, $^{13}\text{C}'$, $^1\text{H}^{\text{NH}}$) retrieved from the database can be compared either to a simple arithmetical or to a Boltzmann-type (energy-weighted) average of the ab initio shifts of the nuclei of individual conformers. The corresponding data are presented in Table 5 and on Figure 3. Overall, the agreement between the average theoretical and experimental shifts is rather impressive, especially for those ab initio data obtained using the higher level of theory (level B1). As the level of ab initio theory employed for geometry optimization increases, that is as the basis changes from 3-21G to 6-311++G** and the method changes from RHF to DFT(B3LYP), the deviation between theoretically and experimentally determined average shifts becomes small for all nuclei but the amide proton. At the highest level of theory (level B2), neglecting the $^1\text{H}^{\text{NH}}$ shifts, the error for all nuclei is down to a few percent. For example, using the energy-weighted sum of the theoretical values (Table 6), the difference between the best calculated (level B1) and experimentally determined shifts for $^{15}\text{N}^{\text{NH}}$ is only 1.03 ppm ($\delta = 122.60$ vs 123.63). For the H^{α} and H^{β} protons the difference between ab initio and experimental average shifts is less than 0.1 ppm. For the C^{α} values the difference is larger, 7.94 ppm, which represents an error of 13% (Table 6). These observations are very similar to those found for For-L-Ala-NH₂ and For-Gly-NH₂.^[31] It is not clear why the error between the calculated and observed amide proton ($^1\text{H}^{\text{NH}}$) shifts is so much larger (33%) than for any other nucleus. Although deviation of the calculated $^1\text{H}^{\text{NH}}$ shift from the experimental value decreases when higher levels of theory are applied, it remains substantial. One has to note that this large error is associated with the most acidic (i.e., deshielded) proton, H^{NH} , of the molecule. As deshielding decreases, accuracy of the computed

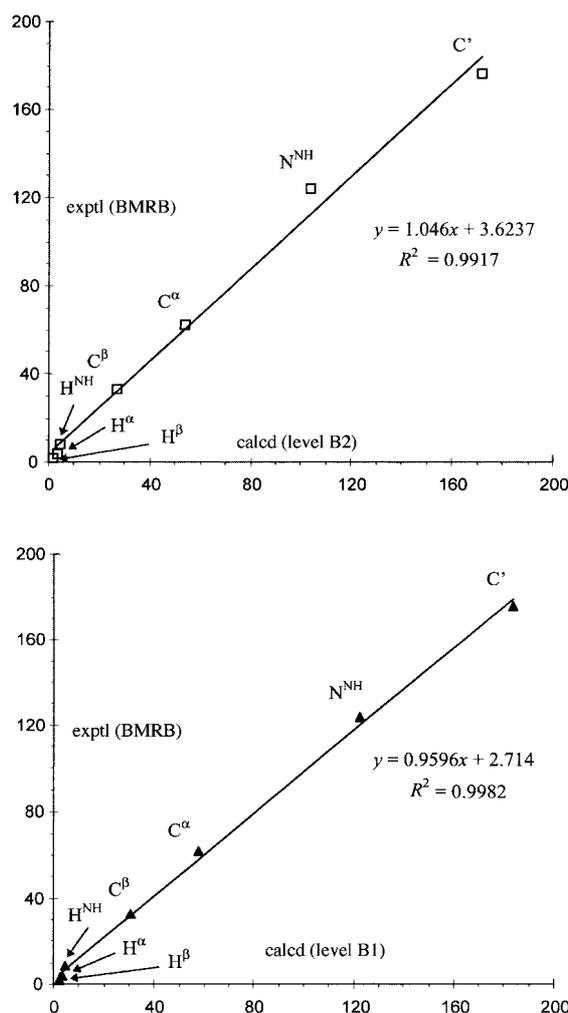


Figure 3. Correlation between ab initio (top: GIAO-RHF/6-31++G**//RHF/3-21G and bottom: GIAO-RHF/TZ2P//B3LYP/6-311++G**) and experimentally determined average (conformation independent) chemical shifts for all nuclei of For-L-Val-NH₂.

Table 5. Statistical analysis of chemical shifts averaged over all backbone conformations using simple arithmetical averaging.^[a]

Level	Conf. ^[b]	$^{15}\text{N}^{\text{NH}}$			$^1\text{H}^{\text{NH}}$			$^{13}\text{C}^{\alpha}$			$^1\text{H}^{\alpha}$			$^{13}\text{C}^{\beta}$			$^1\text{H}^{\beta}$			$^{13}\text{C}'$		
		av.	std.	acc.	std.	acc.	av.	std.	acc.	av.	std.	acc.	av.	std.	acc.	av.	std.	acc.	av.	std.	acc.	
A1	27	103.46	4.86	16	4.36	0.54	47	54.13	3.94	13	3.46	0.63	15	27.20	3.10	17	1.88	0.63	3	171.83	1.86	2
A2	20	102.99	5.01	17	4.41	0.60	47	54.40	4.40	12	3.44	0.69	16	27.57	3.50	16	1.98	0.69	2	171.50	1.54	2
B1	27	122.68	5.01	1	4.90	0.55	41	58.05	4.31	7	3.74	0.65	9	30.78	3.18	6	2.08	0.57	7	183.70	1.80	5
B2	18	122.67	5.15	1	5.00	0.60	39	58.24	4.98	6	3.70	0.72	10	31.04	3.83	5	2.23	0.64	15	183.12	0.93	4
exptl.		123.63	27.04		8.25	0.73		62.17	2.95		4.09	0.77		32.82	1.81		1.94	0.62		175.64	2.00	

[a] See Table 1 for the levels of theory applied. av. = average (in ppm); std. = standard deviation (in ppm); acc. = accuracy (in %), defined as $(\text{CSA}^{\text{exptl}} - \text{CSA}^{\text{calcd}})/\text{CSA}^{\text{exptl}}$. Exptl = experimental average shifts and their standard deviations taken from BMRB. [b] Number of conformers.

Table 6. Statistical analysis of chemical shifts averaged over all backbone conformations using energy-weighted (Boltzmann) averaging.^[a]

Level	Conf. ^[b]	$^{15}\text{N}^{\text{NH}}$			$^1\text{H}^{\text{NH}}$			$^{13}\text{C}^{\alpha}$			$^1\text{H}^{\alpha}$			$^{13}\text{C}^{\beta}$			$^1\text{H}^{\beta}$			$^{13}\text{C}'$		
		av.	std.	acc.	std.	acc.	av.	std.	acc.	av.	std.	acc.	av.	std.	acc.	av.	std.	acc.	av.	std.	acc.	
A1	27	104.16		16	5.09		38	51.34		17	3.81		7	27.12		17	1.77		9	172.96		2
A2	20	103.99		16	5.12		38	51.23		18	3.83		6	27.16		17	1.77		9	172.90		2
B1	27	122.60		1	5.55		33	54.23		13	4.06		1	30.57		7	1.93		1	184.14		5
B2	18	122.32		1	5.59		32	54.05		13	4.10		0	30.63		7	1.93		0	183.93		5
exptl.		123.63	27.04		8.25	0.73		62.17	2.95		4.09	0.77		32.82	1.81		1.94	0.62		175.64	2.00	

[a] See footnote [a] to Table 5.

shielding increases: The less acidic H^{α} and H^{β} protons are calculated with an impressive precision. Although both the $^1H^{NH}$ and $^{13}C^{\alpha}$ calculated shifts are considerably off on the NMR scale, they remain sensitive to conformational changes (see below). Until now, theoretical studies have concentrated only on ^{13}C and ^{15}N shifts of the α -helical and β -sheet conformers of the peptide models. It is demonstrated here that for all nuclei but perhaps one, considering all typical backbone orientations, the average chemical shifts can be determined with remarkable accuracy.

Chemical shift–chemical shift correlation maps: Chemical shifts of backbone nuclei are modified not only by the composition of the side chain (note, for example, the characteristic $^{15}N^{NH}$ upfield shift of Gly, Ser, and Thr residues^[60]) but for the same residue significant up- or downfield shifts are expected due to changes in the molecular environment and/or the backbone and side chain orientations. Simple rotation of the adjacent amide planes can have a remarkable effect on the chemical shifts of selected nuclei. For example, the $\gamma_L(a) \Rightarrow \gamma_D(a)$ (at level A1 $\{\phi_{\gamma_L} = -84.9^\circ, \psi_{\gamma_L} = 62.7^\circ\} \Rightarrow \{\phi_{\gamma_D} = 59.5^\circ, \psi_{\gamma_D} = -38.5^\circ\}$) conformational change is followed by a 12.2 ppm change for the $^{13}C^{\alpha}$ and a 5.5 ppm change for the $^{13}C^{\beta}$ chemical shifts (Table 2). Furthermore, interconversion between two conformational neighbors, for example that of $\gamma_L(g+) \Rightarrow \beta_L(g+)$, can easily be monitored through changes of selected chemical shifts (at level B1 the appropriate changes are $\Delta^{15}N^{NH} = 6.14$ ppm and $\Delta^{13}C^{\beta} = -4.73$ ppm) (see Table 2).

Such shifts are clearly noticeable on correlated 3D-plots (e.g., on the $^{15}N^{NH}-^1H^{NH}-^{13}C^{\alpha}$ plot of an HNCA- or HN(CO)CA-type triple-resonance NMR experiment) (Figure 4). Figure 4, as well as any of the related 2D plots, referring to the $^{15}N^{NH}-^1H^{NH}$, $^{13}C^{\alpha}-^{15}N^{NH}$, and $^{13}C^{\alpha}-^1H^{NH}$ planes (not shown), reveal different locations for the nine characteristic backbone conformations of For-L-Val-NH₂. Structure-induced changes in the experimentally easily amenable $^1H^{NH}$ shifts are relatively small. However, if the amide proton ($^1H^{NH}$) shifts are correlated with those of amide nitrogens ($^{15}N^{NH}$), a seemingly useful (and popular HSQC-type) 2D plot emerges, where all characteristic backbone orientations separate clearly ($H^{NH}-N^{NH}$ 2D planes on Figure 4). All L-type conformers but β_L occupy the left part of this 2D plot. A similarly useful 2D plot is that of the $^{13}C^{\alpha}-^1H^{\alpha}$ correlation map, presented on Figure 5. When chemical shift averaging is performed according to the backbone clusters, the influence of the different side chain orientations ($g+$, a , or $g-$) on chemical shifts are hidden. These average values, representative for the different backbone clusters, differ from each other to a great extent. Furthermore, the relative positions of these shifts on a $^{13}C^{\alpha}-^1H^{\alpha}$ correlation map are similar to those calculated for For-L-Ala-NH₂.^[31] It seems to be straightforward to recognize and assign the backbone conformations of the amino acid residues from these relative chemical shifts.

At all levels of theory the L-type and the D-type conformers cluster in two distinct regions (Figure 5/A and 5/C). All enantiomeric pairs of conformers ($\alpha_L \Leftrightarrow \alpha_D$, $\gamma_L \Leftrightarrow \gamma_D$, $\delta_L \Leftrightarrow \delta_D$ and $\epsilon_L \Leftrightarrow \epsilon_D$) are connected by a solid line on Figure 5.

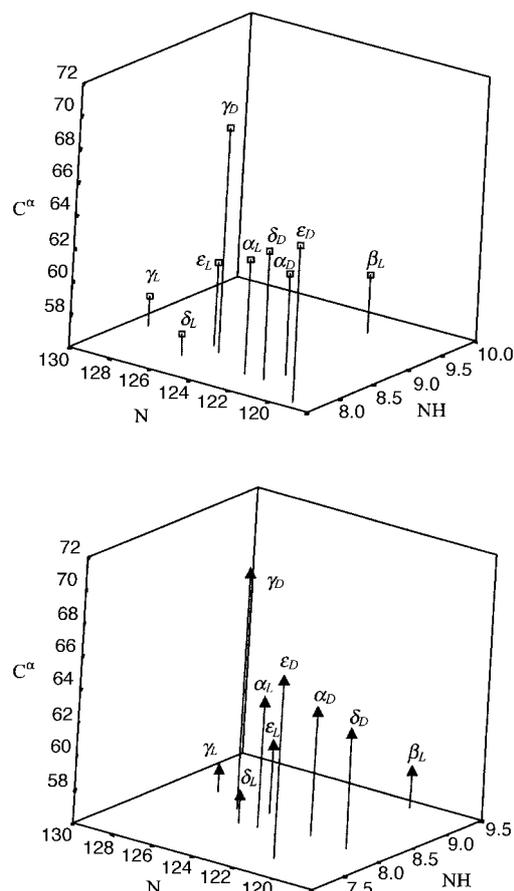


Figure 4. $^{15}N^{NH}-^1H^{NH}-^{13}C^{\alpha}$ correlated 3D-plots for For-L-Val-NH₂ at two levels of theory: Top: GIAO-RHF/6-31+G*/RHF/3-21G and bottom: GIAO-RHF/TZ2P/B3LYP/6-311++G**. Chemical shifts for each backbone type were obtained by averaging all three side chain rotamers.

Compared to the relevant L-type structures all D-type conformers but α_D are markedly downfield shifted along the carbon axes and slightly upfield shifted along the proton axes. In the case of $\alpha_L \Leftrightarrow \alpha_D$, while the C^{α} shift is insignificant, the H^{α} upfield shift is almost 1 ppm. Considering the resolution of modern spectrometers, these chemical shift differences seem to be sufficient to distinguish and assign the different backbone conformer types of the same residue in a protein. However, if the side chain rotation is slow, or if due to structural reasons individual χ_1 rotamers are to be expected, the assignment of the residue to a backbone cluster based on $^{13}C^{\alpha}-^1H^{\alpha}$ chemical shifts may prove to be cumbersome.

There are several factors determining the position of the backbone conformers on ab initio chemical shift–chemical shift correlation maps. Some of them are related to computations, such as the effects of one-particle basis set deficiency, the extent of electron correlation considered, and the effect of geometry optimization. Results reported in Table 2 and Table 3, as well as data shown on Figure 4 and Figure 5, reveal that the backbone conformational clusters of For-L-Val-NH₂ are clearly distinguishable on the correlated plots, especially on the $^{13}C^{\alpha}-^1H^{\alpha}$ plot, independent of the basis set employed (6-31+G* or TZ2P) for the calculation of the chemical shieldings. Although the choice of whether the

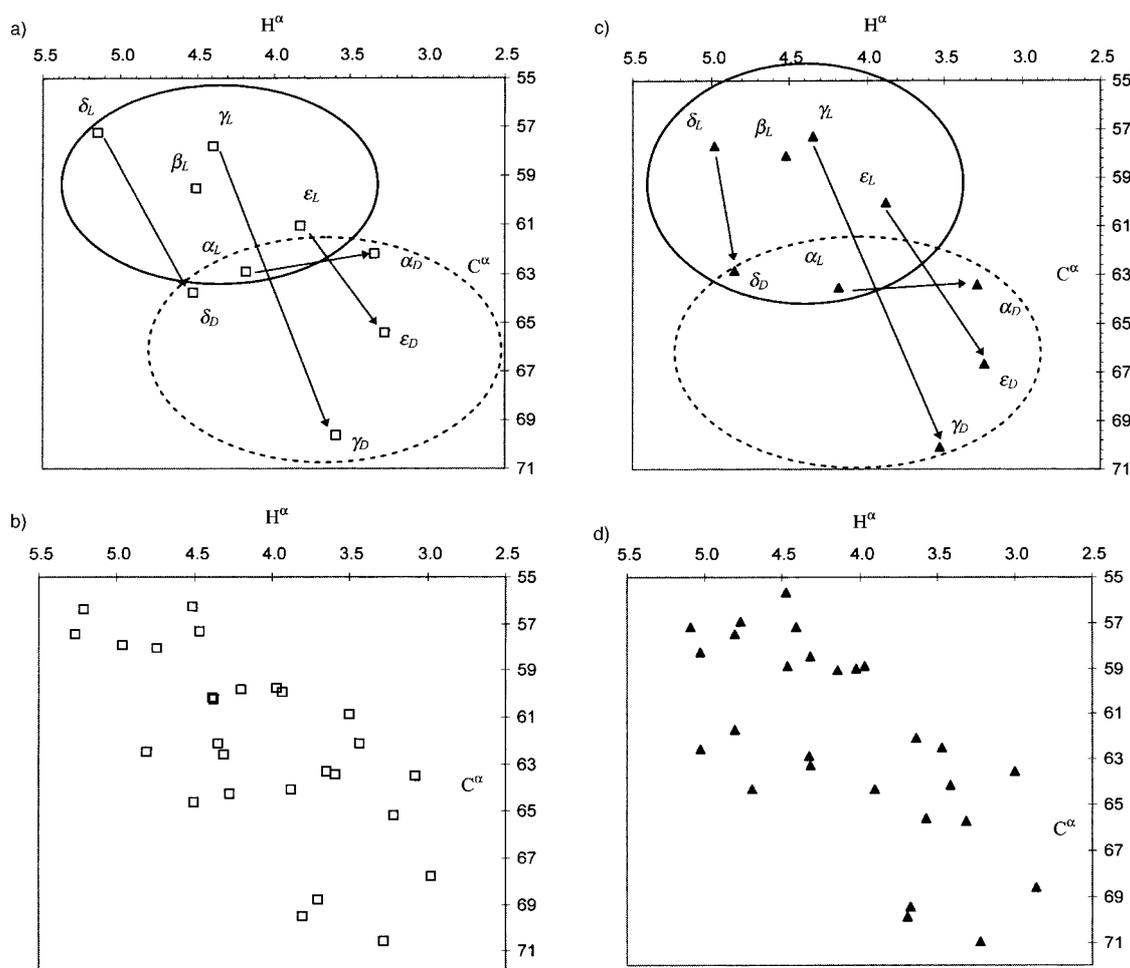


Figure 5. $^1\text{H}^\alpha$ – $^{13}\text{C}^\alpha$ correlated 2D-plots for For-L-Val-NH₂. Panels B and D compile all 27 conformers, while data reported on panels A and C were calculated after averaging the effect of side chain orientation. On panels A and C the conformational mirror images (x_L and x_D) are connected by arrows, the L-type backbone conformers are circled by solid lines while the D-type conformers by broken lines, each associated with three side chain conformers. All ab initio chemical shifts were corrected, the global average of all calculated and all experimentally observed conformational shifts were shifted to a common value (see text for details). Two computational levels have been employed: GIAO-RHF/6-31 + G**//RHF/3-21G [square//empty symbol] (A, B) and GIAO-RHF/TZ2P//B3LYP/6-311 ++ G** [triangle//filled symbols] (C, D).

GIAO-RHF calculations were performed at constrained (levels A1 and B1) or at fully optimized (levels A2 and B2) reference geometries does have an effect on the calculated shifts, the separation of the conformers is clearly not affected. Unlike in the case of For-Gly-NH₂,^[31] the effect of electron correlation on CSA has not been determined directly by performing, for example, GIAO-MP2 calculations for the present valine model. Nevertheless, based on data found for the glycine model For-Gly-NH₂,^[31] it is not expected that electron correlation will alter profoundly the following qualitative picture: The characteristic peptide conformers appear at different regions on certain chemical shift–chemical shift correlation plots. Therefore, the usefulness of the approach of direct determination of conformations of protein building units from multidimensional NMR experiments seems to depend on what effect the side chains, solvation, anisotropic factors, and inter- and intramolecular hydrogen bonds have on the relative chemical shifts of the selected nuclei. As to the future, detailed theoretical investigations of more model compounds and more correlated-level chemical shielding plots are needed, while it is also hoped that our

present and previous^[31] theoretical results will encourage experimental work in order to establish the magnitude of these effects.

Chemical shift–structure correlations: One principal goal of our research has been the investigation of the efficacy of the use of chemical shift information to predict the main chain fold of peptides and proteins. So far, we have demonstrated that a) theoretical and experimental chemical shifts, when averaged over the entire conformational space, are rather similar); and b) chemical shifts form distinct groups on chemical shift–chemical shift correlation plots as a function of the $[\phi, \psi]$ conformational clusters. Prompted by these results, we decided to perform a systematic linear correlation analysis between all important conformational parameters (ϕ , ψ , ω_0 , ω , and χ_1) and all chemical shifts ($^{15}\text{N}^{\text{NH}}$, $^1\text{H}^{\text{NH}}$, $^{13}\text{C}^\alpha$, $^1\text{H}^\alpha$, $^{13}\text{C}^\beta$, $^1\text{H}^\beta$, and $^{13}\text{C}^\gamma$) of interest of For-L-Val-NH₂. We repeated these studies using both sets of ab initio data determined. The linear correlation (Pearson) coefficient, R , and the related standard error, $S_{Y,X}$, have been employed during these correlation studies.^[61]

Typical correlation result are shown in Table 7, obtained at level B1 and using R^2 values as a measure for the goodness of the regression. For structure–shift correlation only those R^2 values are relevant which relate conformational parameters with chemical shifts, as given in the box of broken lines in Table 7. (In this respect all other R^2 data, reported in italics in Table 7, are irrelevant). A technical problem arises when correlating a set of chemical shifts with torsional variables, ξ , where $\xi = \phi, \psi$, etc., since torsions have a periodic nature. Both the IUPAC [$-180^\circ \leq \xi \leq +180^\circ$] and the classical definition of Ramachandran [$0^\circ \leq \xi \leq +360^\circ$] used for defining the periodicity of the torsions influence the outcome of the correlation. To avoid this ambiguity, R^2 values are calculated, unless noted otherwise, using the periodic unit $0^\circ \leq \xi \leq +360^\circ$ for all torsion angles, reported as $R_{[0-360]}^2$. It is evident that often the goodness of the linear correlation can be maximized by simply finding the “optimum” definition of the periodicity of a particular torsional variable. When reported, the optimum correlation between relevant pairs is described as R_{\max}^2 .

The arithmetical average ($R_{[0-360]}^2$) of all calculated $R_{[0-360]}^2$ values at level B1 is rather low: 0.087 with a standard deviation (σ) of 0.14 (Table 7). This means that on average there is no meaningful linear correlation between conformational parameters and chemical shifts. Nevertheless, for a few pairs of variables $R_{[0-360]}^2$ values can be three to four times higher than the above stated average. For example, $R_{[0-360]}^2[\phi/^{13}\text{C}^\alpha] = 0.473$, $R_{[0-360]}^2[\phi/^1\text{H}^\alpha] = 0.418$, and $R_{[0-360]}^2[\psi/^{13}\text{C}^\alpha] = 0.345$. All significant correlations are printed in bold in Table 7. The squares of the optimized Pearson correlation coefficients between these variables are usually even higher: $R_{\max}^2[\phi/^{13}\text{C}^\alpha] = 0.557$, $R_{[0-360]}^2[\phi/^1\text{H}^\alpha] = 0.818$, and $R_{[0-360]}^2[\psi/^{13}\text{C}^\alpha] = 0.345$. Undoubtedly, $\phi/^1\text{H}^\alpha$ provides the most significant correlation between a backbone parameter and a chemical shift (see Figure 6). The correlations between $[\phi, \psi]/[^{13}\text{C}^\alpha, ^1\text{H}^\alpha]$ provide the possibility to “project” backbone conformational parameters on the $^1\text{H}^\alpha$ – $^{13}\text{C}^\alpha$ 2D plot and vice versa. This means that even a simple ^1H – ^{13}C HSQC-type spectrum, once resonances are assigned to the different amino acid residues, could provide a handful of structural information. This is discussed further below.

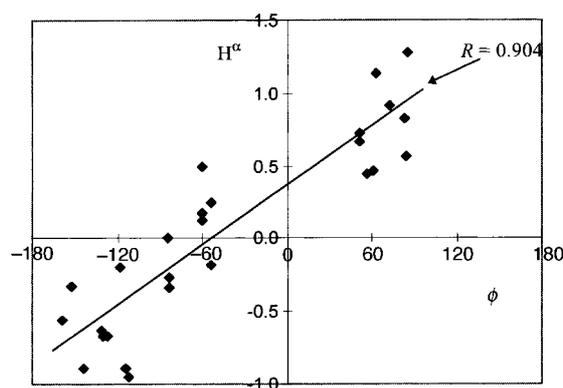


Figure 6. The maximized fit between $^1\text{H}^\alpha$ and ϕ at level B1 ($R^2 = 0.818$).

Statistical measures (R and R^2/σ) of selected structure/shift and shift/shift pairs ($\phi/^{13}\text{C}^\alpha$, $\phi/^1\text{H}^\alpha$, ..., $^1\text{H}^\alpha/^{13}\text{C}^\alpha$, etc.) for the four theoretical levels, A1 through B2, are reported in Table 8. The variable pairs were selected either because the correlation found was high enough to consider (e.g., $\phi/^{13}\text{C}^\alpha$, $\phi/^1\text{H}^\alpha$, $\psi/^{13}\text{C}^\alpha$, and $^1\text{H}^\alpha/^{13}\text{C}^\alpha$) or because the type of correlation must be discussed regardless of the goodness of the correlation (e.g., $\psi/^1\text{H}^\alpha$). The most important conclusions which can be drawn from the data presented in Table 7 and 8 are as follows: a) All elements of the two by two data set $[\phi, \psi]/[^{13}\text{C}^\alpha, ^1\text{H}^\alpha]$ but $\psi/^1\text{H}^\alpha$ show significant correlation; for example, $\phi/^1\text{H}^\alpha$ can be as high as $|R_{\max}| \approx 0.904$ (Figure 6). b) The $^1\text{H}^\alpha/^{13}\text{C}^\alpha$ shift–shift correlation is significant with $R^2/\sigma \geq 3$ at all investigated levels of theory. c) We do not find correlation between ψ and $^1\text{H}^\alpha$ at any of the levels investigated. d) The $^{13}\text{C}^\alpha$ shift of For-L-Val-NH₂ correlates with ψ almost as well as with ϕ . e) None of the shifts of the remaining nuclei ($^1\text{H}^{\text{NH}}$, $^{13}\text{C}^\beta$, $^1\text{H}^\beta$, and $^{13}\text{C}^\gamma$), except for that of $^{15}\text{N}^{\text{NH}}$, correlate with any of the structural parameters. Nevertheless, if R^2 is maximized, the $\chi_1/^{15}\text{N}^{\text{NH}}$ correlation becomes significant with $|R_{\max}| = 0.657$. f) None of the correlations mentioned become significantly different if data are calculated at a higher level of theory, the overall qualitative picture remains the same.

Although linear correlation of selected chemical shifts with conformational parameters is not as high as one would hope,

Table 7. R^2 values determined between selected conformational and chemical shift values of For-L-Val-NH₂.^[a]

	ω_0	ϕ	ψ	ω_1	χ_1	$^{15}\text{N}^{\text{NH}}$	$^1\text{H}^{\text{NH}}$	$^{13}\text{C}^\alpha$	$^1\text{H}^\alpha$	$^{13}\text{C}^\beta$	$^1\text{H}^\beta$	$^{13}\text{C}^\gamma$
ω_0	1	0.000	0.016	0.668	0.006	0.006	0.262	0.002	0.001	0.023	0.036	0.130
ϕ		1	0.000	0.002	0.010	0.020(0.220)	0.069(0.213)	0.473(0.557)	0.418(0.818)	0.056(0.104)	0.198(0.260)	0.267(0.267)
ψ			1	0.007	0.000	0.007(0.137)	0.003(0.116)	0.345(0.345)	0.000(0.127)	0.014(0.089)	0.027(0.077)	0.026(0.272)
ω_1				1	0.000	0.010	0.127	0.015	0.010	0.006	0.012	0.056
χ_1					1	0.015(0.431)	0.048(0.144)	0.006(0.155)	0.024(0.145)	0.190(0.297)	0.026(0.298)	0.050(0.085)
N^{NH}						1	0.001	0.005	0.003	0.114	0.180	0.000
H^{NH}							1	0.204	0.097	0.013	0.042	0.038
C^α								1	0.457	0.001	0.141	0.096
H^α									1	0.022	0.188	0.004
C^β										1	0.412	0.042
H^β											1	0.020
C^γ												1

[a] A total of 27 conformers are involved; theoretical shifts refer to level B1 values. All conformational variables (ϕ , ψ , ω_0 , ω_1 , and χ_1) are scaled between 0° and 360° . Auto-correlation values, that is $R^2[\phi/\phi]$, are all equal to 1. The average of all R^2 values, $\langle R^2 \rangle$, is 0.087, with a standard deviation (σ) of 0.140. All values larger than $R^2 + 1.5\sigma$ ($0.087 + 0.21 = 0.297$) are marked bold. Selected maximized R_{\max}^2 values are given in parentheses, obtained by shifting systematically (with an increment of 5°) the periodic unit of the conformational variables ϕ , ψ , and χ_1 .

Table 8. Selected statistical parameters of the correlation analysis between conformational and chemical shift values of For-L-Val-NH₂ determined at four levels of theory.^[a]

	A1		A2		B1		B2	
	<i>R</i>	<i>R</i> ² / σ						
ϕ / ¹ H ^{α}	-0.60	3.12	-0.84	4.42	-0.65	2.98	-0.85	4.25
ψ / ¹ H ^{α}	0.10	0.08	0.10	0.06			0.08	0.04
ϕ / ¹³ C ^{α}	0.58	2.97	0.74	3.42	0.69	3.37	0.84	4.16
ψ / ¹³ C ^{α}	-0.67	3.94	-0.72	3.24	-0.59	2.46	-0.62	2.28
ϕ / ¹⁵ N	-0.31	0.83	-0.35	0.75	-0.14	0.14	-0.10	0.06
χ_1 / ¹³ C ^{γ}	0.22	0.41	0.24	0.36	0.22	0.35	0.34	0.67
¹ H ^{α} / ¹³ C ^{α}	-0.69	4.08	-0.69	3.03	-0.68	3.26	-0.73	3.15
¹ H ^{α} / ¹³ C ^{β}	0.10	0.09	0.10	0.06	0.15	0.16	0.15	0.14
¹ NH/ ¹⁵ NH	0.10	0.09	0.17	0.18	-0.03	0.00	-0.03	0.01

[a] See Table 1 for details about the levels of theory (A1, A2, B1, and B2) applied. *R* = Pearson coefficient, σ = standard error. The average *R*²/ σ values, obtained after considering all possible correlations between conformational and chemical shift values are 0.115, 0.159, 0.140, and 0.170 for levels A1, A2, B1, and B2, respectively. All conformational variables have been defined using the range [0°, 360°].

pairs with large $|R|$ values certainly warrant further investigation. All available results suggest that relative changes in the chemical shifts of ¹³C ^{α} and ¹H ^{α} are coupled with backbone conformational changes: correlation between ϕ /¹³C ^{α} , ϕ /¹H ^{α} , and ψ /¹³C ^{α} are significant. Therefore, we decided to investigate the dependence of the calculated shifts of H ^{α} and C ^{α} nuclei as a function of the applied level of theory (not detailed here), as performed previously for For-L-Ala-NH₂.^[31] We observed that extension of the basis set or improvement in the quality of the reference geometry do not increase appreciably the reliability of predicting $[\phi, \psi]$ values from calculated ¹³C ^{α} and ¹H ^{α} values.

Chemical shifts of valine as function of its conformation in proteins: The two predominant secondary structural elements of proteins are the α -helix and the β -sheet. Therefore, accurate description of the NMR characteristics of these two conformational building units has been of

Table 9. Selected *R*_{max}² values determined between major periodic conformational variables and chemical shifts assigned for valines in 18 proteins.^[a]

	¹⁵ N ^{NH}	¹ H ^{NH}	¹³ C ^{α}	¹ H ^{α}	¹³ C ^{β}	¹³ C ^{γ}
ϕ		0.128	0.054	0.648	0.555	0.569
ψ		0.182	0.08	0.669	0.434	0.363
χ_1		0.201	0.045	0.194	0.085	0.100

[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. A total of 18 proteins was analyzed for the torsional and NMR properties of valine residues.

primary interest for experimentalists and theoreticians alike.^[12, 13, 17, 18, 24, 62] In this section the correlation between experimentally determined chemical shifts of selected nuclei and the backbone fold of 93 Val residues in 18 selected proteins is discussed. The analysis is more detailed than previous ones in the sense that *all* backbone conformational types are investigated.

As detailed in the last Section, correlation between torsional variables ($\xi = \phi, \psi,$ and χ_1) and NMR chemical shifts (¹⁵N^{NH}, ¹H^{NH}, ¹³C ^{α} , ¹H ^{α} , ¹³C ^{β} , and ¹³C ^{γ}) can be maximized. This has been done for the experimental data and selected results are reported in Table 9 and Table 10. Experimental *R*_{max}² data sets (Table 9) and comparison with their theoretical counterparts (Table 7) results in the following conclusions: a) Neither ¹⁵N^{NH} nor ¹H^{NH} show correlation with conformational variables $\phi, \psi,$ or χ_1 ; b) the ϕ /¹³C ^{α} , ϕ /¹H ^{α} , and ψ /¹³C ^{α} correlations are among the most significant ones; c) ¹³C ^{α} /¹H ^{α} data from both sources show significant correlation (Figure 7) with $R_{\text{exptl}}[\sup{13}\text{C}^{\alpha}/\sup{1}\text{H}^{\alpha}] = -0.83$ and $R_{\text{Level B2}}[\sup{13}\text{C}^{\alpha}/\sup{1}\text{H}^{\alpha}] = -0.73$; and d) the only obvious discrepancy between “theory and experiment” concerns *R*_{max}² $[\phi/\sup{13}\text{C}^{\beta}]$ and *R*_{max}² $[\psi/\sup{13}\text{C}^{\beta}]$. While for experimental data these *R*² values are relatively large with $R_{\text{exptl(max)}}^2[\phi/\sup{13}\text{C}^{\beta}] = 0.57$ and $R_{\text{exptl(max)}}^2[\psi/\sup{13}\text{C}^{\beta}] = 0.36$ (Table 9), for data determined by ab initio techniques they are typically low, $R_{\text{Level B1(max)}}^2[\phi/\sup{13}\text{C}^{\beta}] = 0.10$ and $R_{\text{Level B1(max)}}^2[\psi/\sup{13}\text{C}^{\beta}] = 0.09$ (Table 7).

Table 10. Experimentally determined^[a] and ab initio (level A1^[b] and B1^[c]) ¹H ^{α} , ¹³C ^{α} , ϕ , and ψ parameters for valine conformers occurring frequently in proteins.

conf. ^[d]	Experiment						Level A1				Level B1			
	abundance ^[e]	av. dev. ^[f]	¹ H ^{α}	¹³ C ^{α}	ϕ	ψ	¹ H ^{α}	¹³ C ^{α}	ϕ	ψ	¹ H ^{α}	¹³ C ^{α}	ϕ	ψ
$\alpha_L(g-)$	5	18	4.28	63.5	-78.0	-25.0	4.31	62.6	-54.0	-45.0	4.32	63.3	-54.0	-45.0
$\alpha_L(g+)$	1	12	3.6	65.2	-53.0	-31.0	4.35	62.1	-54.0	-45.0	4.33	62.9	-54.0	-45.0
$\alpha_L(a)$	22	18	3.51	66.1	-66.0	-42.0	3.88	64.1	-54.0	-45.0	3.90	64.3	-54.0	-45.0
$\beta_L(a)$	27	30	4.66	60.9	-120.0	131.0	4.37	60.2	-137.5	143.5	4.32	58.5	-118.8	125.8
$\beta_L(g-)$	13	21	4.83	58.8	-133.0	157.0	4.75	58.1	-142.4	163.5	4.77	56.9	-131.1	162.2
$\beta_L(g+)$	2	20	4.95	58.9	-151.0	141.0	4.38	60.2	-163.2	157.4	4.47	58.9	-151.9	157.6
$\delta_D(a)$	1	27	3.76	63.3	-109.0	-55.0	4.81	62.5	-136.7	-59.9	4.81	61.7	-126.9	-68.0
$\delta_L(g-)$	2	26	4.48	60.2	-116.0	-8.0	5.27	57.4	-125.5	28.9	5.03	58.3	-114.0	12.1
$\epsilon_L(g+)$	2	14	4.04	63.3	-60.0	136.0	3.93	59.9	-60.0	120.0	4.02	59.0	-60.0	120.0
$\epsilon_L(g-)$	4	32	3.96	61.0	-86.0	153.0	3.97	59.8	-60.0	120.0	3.97	58.9	-60.0	120.0
$\epsilon_L(a)$	8	34	4.27	62.3	-89.0	124.0	3.59	63.4	-60.0	120.0	3.64	62.1	-60.0	120.0
$\gamma_L(a)$	6	37	4.00	62.3	-99.0	107.0	4.20	59.8	-86.6	71.4	4.14	59.1	-83.9	82.4

[a] Values extracted from the 18 proteins given in the Experimental Section. [b] Level A1 = GIAO-RHF/6-31 + G**/RHF/3-21G. [Computed chemical shifts for ¹³C ^{α} and ¹H ^{α} are uniformly shifted by 8.04 and 0.63 ppm, respectively (cf. Table 5), to match BMRB experimental averages (Table 5).] [c] Level B1 = GIAO-RHF/TZ2P/B3LYP/6-311 ++ G***. [Computed chemical shifts for ¹³C ^{α} and ¹H ^{α} are uniformly shifted by 4.12 and 0.35 ppm, respectively (Table 5), to match BMRB experimental averages (Table 5).] [d] Type of valine 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 valine conformers in proteins and the reference $\phi, \psi,$ and χ_1 values.

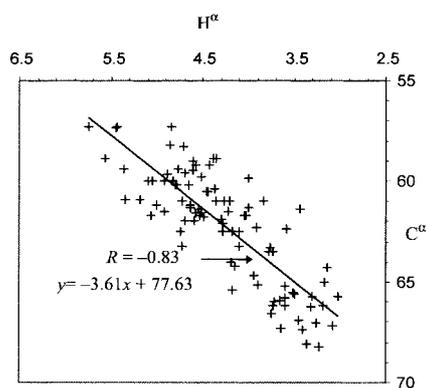


Figure 7. $^{13}\text{C}^\alpha$ – $^1\text{H}^\alpha$ chemical shift–chemical shift correlation plot for 93 valines found in 18 proteins.

On experimental proton/carbon correlation maps (e.g., $^1\text{H}/^{13}\text{C}$ HSQC) it is easy to see that the $^{13}\text{C}^\alpha/{}^1\text{H}^\alpha$ region shows higher conformational dependence than the rest of the spectrum: dots characterizing $^{13}\text{C}^\alpha/{}^1\text{H}^\alpha$ correlations are much more spread out than $^{13}\text{C}^\beta/{}^1\text{H}^\beta$ or $^{13}\text{C}^\gamma/{}^1\text{H}^\gamma$ values. The level of correlation found for ϕ/C^α , $\phi/{}^1\text{H}^\alpha$, ψ/C^α , and $\psi/{}^1\text{H}^\alpha$ allows for prediction of backbone conformations from selected NMR shielding values. Under optimal conditions this could result in determination of the backbone fold solely from $^{13}\text{C}^\alpha$ and $^1\text{H}^\alpha$ chemical shifts and vice versa. Therefore, in this study we focused primarily on understanding the nature of $^{13}\text{C}^\alpha$ and $^1\text{H}^\alpha$ chemical shift changes and their relation to parameters $[\phi, \psi]$ (Table 10 and Figure 8). Down- and upfield shifts are relative

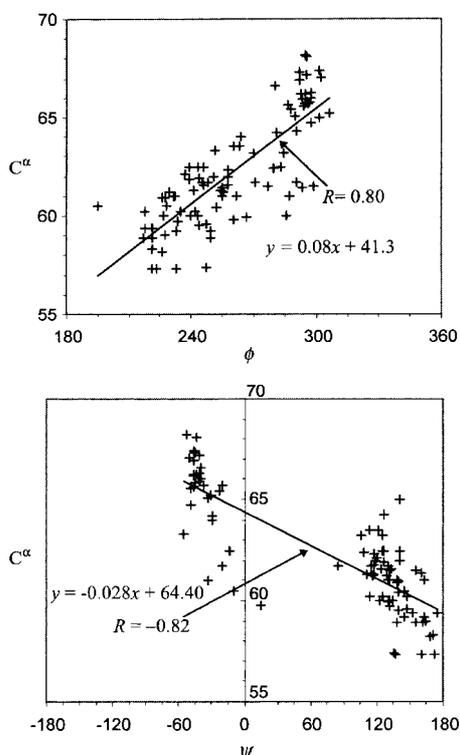


Figure 8. $^{13}\text{C}^\alpha/\phi$ (top) and $^{13}\text{C}^\alpha/\psi$ (bottom) chemical shift conformation correlation plot for 93 valines found in 18 proteins.

to the following average values extracted from BMRB for valine residues: $\delta(^{13}\text{C}^\alpha_{\text{BMRD}}) = 62.17$ and $\delta(^1\text{H}^\alpha_{\text{BMRD}}) = 4.09$. The results obtained can be summarized as follows:

- There is a characteristic downfield shift for the $^{13}\text{C}^\alpha$ chemical shifts associated with the right-handed helix conformation (α_L -type structure). The $^{13}\text{C}^\alpha$ downfield shift of the helical building unit is especially significant for the most populated *anti* side chain orientation $\alpha_L(a-)$, $^{13}\text{C}^\alpha_{\alpha_L(a)} \text{exptl} = 66.1$ ppm. The experimentally observed downfield shift is reproduced at both ab initio levels, although its magnitude is slightly smaller: $^{13}\text{C}^\alpha_{\alpha_L(a)} \text{Level A1} = 64.1$ ppm and $^{13}\text{C}^\alpha_{\alpha_L(a)} \text{Level B1} = 64.3$ ppm. Along the proton direction, both the calculated and the experimental $^1\text{H}^\alpha$ chemical shifts of the $\alpha_L(a-)$ conformer are shifted upfield: $^1\text{H}^\alpha_{\alpha_L(a)} \text{Level A1} = 3.88$ ppm, $^1\text{H}^\alpha_{\alpha_L(a)} \text{Level B1} = 3.90$ ppm, and $^1\text{H}^\alpha_{\alpha_L(a)} \text{exptl} = 3.51$ ppm. Within the helical parts of the 18 proteins analyzed here, valines adopt a *gauche*–orientation in only 18% of all cases, the *anti* side chain orientation is much more frequent. Both the $^{13}\text{C}^\alpha$ and the $^1\text{H}^\alpha$ chemical shifts associated with the $\alpha_L(g-)$ conformers show small but significant downfield shifts: $^{13}\text{C}^\alpha_{\alpha_L(g-)} \text{exptl} = 63.5$ ppm and $^1\text{H}^\alpha_{\alpha_L(g-)} \text{exptl} = 4.28$ ppm. These shifts are reproduced at both ab initio levels: $^{13}\text{C}^\alpha_{\alpha_L(g-)} \text{Level A1} = 62.6$ ppm, $^{13}\text{C}^\alpha_{\alpha_L(g-)} \text{Level B1} = 63.3$ ppm and $^1\text{H}^\alpha_{\alpha_L(g-)} \text{Level A1} = 4.31$ ppm, $^1\text{H}^\alpha_{\alpha_L(g-)} \text{Level B1} = 4.32$ ppm (see also Table 10 and Figure 9). Since for the third type of side chain orientation, $\alpha_L(g+)$, only a single experimental case was found, analysis of properties of this conformation is not possible. In summary, the sign of the shifts characteristic for helical backbone conformations are reproduced by calculations for both side chain orientations *a*– and *g*–. Furthermore, for $\alpha_L(g+)$ ab initio calculations reproduce the carbon and proton experimental shifts remarkably well.
- Among all valine residues found in β -pleated sheet conformation approximately 1/3 (13/42) have a *gauche*–, $\beta_L(g-)$, and 2/3 (27/42) have an *anti*, $\beta_L(a)$,

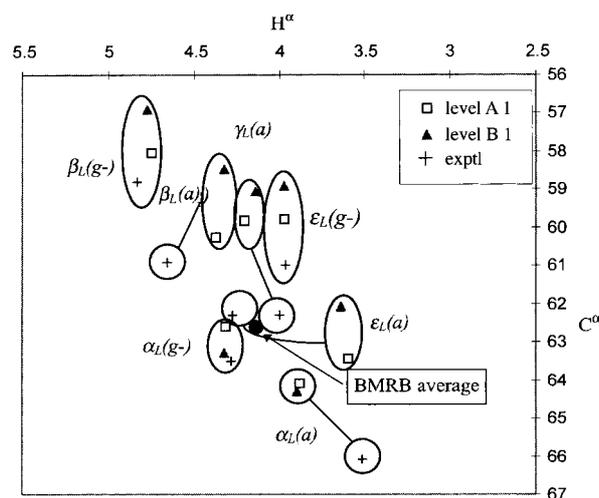


Figure 9. Experimentally determined (in 18 proteins) and ab initio calculated (level A1 and B1) $^1\text{H}^\alpha$ and $^{13}\text{C}^\alpha$ chemical shifts of valine conformers (see also Table 8). Only those conformers are plotted which were found in the 18 proteins more than twice. (Computed values are uniformly shifted by 8.04 ppm [$^{13}\text{C}^\alpha$] and 0.63 ppm [$^1\text{H}^\alpha$] at level A1, and by 4.12 ppm [$^{13}\text{C}^\alpha$] and 0.35 ppm [$^1\text{H}^\alpha$] at level B1 to match BMRB experimental average values $\delta(^{13}\text{C}^\alpha) = 62.17$ and $\delta(^1\text{H}^\alpha) = 4.09$.)

side chain orientation (Table 10). For both types of side chain orientations there is a significant upfield shift for $^{13}\text{C}^\alpha$ and a characteristic downfield shift for $^1\text{H}^\alpha$ (Table 10 and Figure 9). Once again, our ab initio calculations nicely predict the sign of these changes relative to the average value deduced from BMRB. For the $\beta_{\text{L}}(g-)$ conformer even the magnitudes of the calculated proton chemical shifts, $^1\text{H}^{\alpha_{\beta(g-)} \text{Level A1}} = 4.75$ ppm and $^1\text{H}^{\alpha_{\beta(g-)} \text{Level B1}} = 4.77$ ppm, are very close to the average value ($^1\text{H}^{\alpha_{\beta(g-)} \text{exptl}} = 4.83$ ppm) of the 13 $\beta_{\text{L}}(g-)$ conformers assigned in the present protein database. Similarly to α -helices, valines in β -sheets having a $g+$ side chain orientation are rather rare, only two cases have been found out of a total of 42.

- c) Besides the helical (α_{L}) and the extended (β_{L}) conformations discussed already, the third most frequently assigned backbone conformer of proteins is that of poly-proline II, built up from ε_{L} -type conformational units. As for helices and β -sheets, in poly-proline II structures the *anti* side chain orientation is the more frequent one, $\varepsilon_{\text{L}}(a) \approx 60\%$, while *gauche* $-$ is found in about 30% of the cases. The $g+$ side chain orientation with ε_{L} -type backbone structure was found only twice among the 93 valines analyzed. During ab initio calculations for the For-L-Val-NH₂ model system, the following typical poly-proline II ϕ and ψ values were employed: $\phi = -60^\circ$ and $\psi = 120^\circ$. For all three types of ε_{L} conformers the experimental average [ϕ , ψ] values obtained from X-ray data deviate significantly, $\approx 25^\circ$, from these typical backbone values, resulting in the possibility of sizeable differences between experimental and ab initio chemical shifts. Due to the generally established $\phi^1\text{H}^\alpha$ correlation (Table 9), a few tenth of ppm discrepancy is expected between $^1\text{H}^{\alpha_{\text{theor}}}$ and $^1\text{H}^{\alpha_{\text{exptl}}}$. Indeed, for this conformer the ab initio $^1\text{H}^\alpha$ chemical shift is significantly upfield shifted ($^1\text{H}^{\alpha_{\varepsilon_{\text{L}}(a)} \text{Level A1}} = 3.59$ ppm and $^1\text{H}^{\alpha_{\varepsilon_{\text{L}}(a)} \text{Level B1}} = 3.64$ ppm) when compared to the experimental value of $^1\text{H}^{\alpha_{\varepsilon_{\text{L}}(a)} \text{exptl}} = 4.27$ ppm (Table 10 and Figure 9). Nevertheless, for the $\varepsilon_{\text{L}}(g-)$ structures the ab initio values agree well with experimental data, especially for the $^1\text{H}^\alpha$ chemical shifts ($^1\text{H}^{\alpha_{\varepsilon_{\text{L}}(g-)} \text{Level A1}} = 3.97$ ppm, $^1\text{H}^{\alpha_{\varepsilon_{\text{L}}(g-)} \text{Level B1}} = 3.97$ ppm, and $^1\text{H}^{\alpha_{\varepsilon_{\text{L}}(g-)} \text{exptl}} = 3.96$ ppm).
- d) In addition to the α -helix, β -sheet, and poly-proline II conformers six inverse γ -turns were assigned in the 18 proteins investigated, all having an *anti* side chain orientation, denoted as $\gamma_{\text{L}}(a)$. As for ε_{L} , the torsional parameters of γ_{L} , especially ψ , are off by some 25° ; causing expected differences between calculated and observed chemical shifts. As reported in Table 10 and Figure 9, this difference is indeed present, suggesting that both $^1\text{H}^\alpha$ and $^{13}\text{C}^\alpha$ are sensitive to conformational mismatches.

Conclusion

Establishing correlations between peptide backbone conformations and chemical shifts of relevant nuclei has provided a continuous challenge for experimentalists and theoreticians alike. When performing ab initio chemical shift calculations

uncertainties arise from the problem of ideal size and type of peptide model to be used, from the required level of quantum theory, and from the incorporation of important structural and environmental factors (e.g., non-planarity of the amide groups and solvation effects). To consider all these effects at once is an almost formidable task. An important challenge for experimentalists is to provide a much larger, non-homologous, and carefully revised protein NMR database, something similar to BMRB. In the meantime, studies such as the present one incorporating both selected ab initio results and carefully chosen values from experimental databases provide perhaps the most promising approach. Important findings of our study are detailed as follows:

- For the first time the entire conformational library of For-L-Val-NH₂, a typical representative of hydrophobic but non-aromatic side chain containing amino acids, is presented. The ab initio data set comprises 27 different structures optimized at two levels of theory. All α_{L} , ε_{L} , and some of the ε_{D} and δ_{D} conformers are found not to correspond to local minima at the 6-311++G** DFT(B3LYP) level of theory. For these structures partially constrained geometry optimizations were performed. As many as 54 medium- and/or high-level GIAO-RHF NMR shielding calculations were performed to map the conformational hyperspace of valines in proteins.
- Agreement between theoretical and experimental chemical shifts averaged over all conformers for the different types of proton, nitrogen and carbon nuclei is impressive, R^2 values as high as 0.99 are observed. As the size of the basis set employed for the shielding calculation increases, deviation between the experimental and theoretical shifts decreases to a few percent for all nuclei but the amide proton.
- In correlating $^1\text{H}^{\text{NH}}$ and $^{15}\text{N}^{\text{NH}}$ chemical shifts, as well as $^1\text{H}^\alpha$ and $^{13}\text{C}^\alpha$ chemical shifts, useful 2D plots emerge independently of the basis set employed. In the case of our valine model peptide, after averaging the chemical shift modifying effect of the side chain orientation, the positions of the correlated shifts of the various backbone conformers differ from each other to such an extent that it is straightforward to recognize and assign the appropriate backbone conformations from their relative chemical shifts. The $^{15}\text{N}^{\text{NH}} - ^1\text{H}^{\text{NH}}$ and/or the $^1\text{H}^\alpha$ and $^{13}\text{C}^\alpha$ planes, which correspond to results of HSQC (N,H)-type NMR experiments, reveal that all nine typical backbone conformations separate clearly.
- The empirical structure–chemical shift correlations observed by Wishart,^[12, 13] Oldfield,^[18] Bax^[8] and others have been successfully applied in structural determinations of helical and extended-like secondary structures. Data presented in this paper for valine give further justification to these structure–chemical shift correlations and shows that they hold for all typical backbone conformers. Our present comprehensive analysis, augmented by our previous results^[31] for the For-Gly-NH₂ and For-L-Ala-NH₂ models, reveals the intrinsic correlation of *all nine* characteristic backbone conformations of peptides and their chemical shift information (especially $^1\text{H}^\alpha$ and $^{13}\text{C}^\alpha$).
- Our detailed NMR shift investigation, covering the entire conformational hyperspace available for a valine diamide

model and all experimental valine chemical shifts extracted from 18 carefully matched proteins, reveals significant conformational dependence of $^1\text{H}^\alpha$ and $^{13}\text{C}^\alpha$ chemical shifts. However, more model compounds, for example For-Ser-NH₂ (a model for hydrophilic amino acids) and For-Phe-NH₂ (a model of amino acids with aromatic side chains), should be studied to obtain a conclusive and general answer as to whether chemical shift information in general is good enough to use for structure determination of peptides and proteins.

In summary, *ab initio* isotropic NMR shielding results presented in this paper for the model system For-L-Val-NH₂ facilitate and encourage the application of correlated relative chemical shift information from $\{^1\text{H}-^{15}\text{N}\}$ HSQC, $\{^1\text{H}-^{13}\text{C}\}$ HSQC, HNCA, HNCB, and other multiple-pulse NMR experiments to extract structural information *directly* from these measurements, thus opening an alternative route to NOEs in deriving structures of proteins from their NMR spectra.

Acknowledgement

The authors thank P. Hudáky and I. Jáklí for helpful discussions. This research was partially supported by grants from the Hungarian Scientific Research Fund (OTKA T017604, T024044, T032486, T033074, and T030841) and by a grant from the Hungarian Academy of Sciences (AKP96/2-427 2.4).

- [1] D. Neuhaus, M. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VCH, New York, **1989**.
- [2] K. Wüthrich in *NMR of Proteins and Nucleic Acids*, Wiley, New York, **1986**.
- [3] K. Wüthrich, *Science* **1989**, *243*, 45–50.
- [4] J. A. Smith, L. G. Pease, *CRC Crit. Rev. Biochem. Mol.* **1980**, *8*, 315–399.
- [5] M. Nilges, *J. Mol. Biol.* **1995**, *245*, 645–660.
- [6] J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, N. J. Skelton, *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, San Diego, **1996**.
- [7] a) S. W. Fesik, E. R. P. Zuiderweg, *Q. Rev. Biophys.* **1990**, *23*, 97–131; b) G. M. Clore, A. M. Gronenborn, *Prog. Nucl. Magn. Reson. Spectrosc.* **1991**, *23*, 43–92; c) A. Bax, S. Grzesiek, *Acc. Chem. Res.* **1993**, *26*, 131–138.
- [8] I. Pelczar, B. G. Carter in *Protein NMR Techniques* (Ed.: D. G. Reid), Humana Press, Totowa, **1997**, pp. 71–155.
- [9] a) B. Reif, M. Henning, C. Griesinger, *Science* **1997**, *276*, 1230–1233; b) N. Tjandra, A. Bax, *Science* **1997**, *278*, 1111–1114; c) J. D. van Beek, L. Beaulieu, H. Schafer, M. Demura, T. Asakura, B. H. Meier, *Nature* **2000**, *405*, 1077–1079; d) J. H. Prestegard, *Nat. Struct. Biol.* **1998**, *517*–522.
- [10] G. Cornilescu, F. Delaglio, A. Bax, *J. Biomol. NMR* **1999**, *13*, 289–302.
- [11] B. R. Seavey, E. A. Farr, W. M. Westler, J. L. Markley, *J. Biomol. NMR* **1991**, *1*, 217.
- [12] D. S. Wishart, B. D. Sykes, F. M. Richards, *J. Mol. Biol.* **1991**, *222*, 311–333.
- [13] D. S. Wishart, B. D. Sykes, *Meth. Enzymol.* **1994**, *239*, 363–392.
- [14] L. Szilágyi, *Prog. Nucl. Magn. Reson. Spectrosc.* **1995**, *27*, 325–443.
- [15] R. Tabeta, H. Saito, *Biochemistry* **1985**, *24*, 7696–7702.
- [16] C. E. Johnson, F. A. Bovey, *J. Chem. Phys.* **1958**, *29*, 1012.
- [17] a) S. J. Perkins, K. Wüthrich, *Biochim. Biophys. Acta* **1979**, *576*, 409–423; b) C. W. Haigh, R. B. Mallion, *Prog. Nucl. Magn. Reson. Spectrosc.* **1980**, *13*, 303–344; c) S. Spera, A. Bax, *J. Am. Chem. Soc.* **1991**, *113*, 5490–5492; d) M. P. Williamson, *Biopolymers* **1990**, *29*, 1428–1431.
- [18] a) A. C. de Dios, J. G. Pearson, E. Oldfield, *Science* **1993**, *260*, 1491–1496; b) A. C. de Dios, J. G. Pearson, E. Oldfield, *J. Am. Chem. Soc.* **1993**, *115*, 9768–9773; c) A. C. de Dios, E. Oldfield, *J. Am. Chem. Soc.* **1994**, *116*, 5307–5314; d) H.-B. Le, J. G. Pearson, A. C. de Dios, E. Oldfield, *J. Am. Chem. Soc.* **1995**, *117*, 3800–3807.
- [19] M. D. Reily, V. Thanabal, D. O. Omeccinsky, *J. Am. Chem. Soc.* **1992**, *114*, 6251–6252.
- [20] A. Pastore, V. Saudek, *J. Magn. Reson.* **1990**, *90*, 165–176.
- [21] a) L. Müller, *J. Am. Chem. Soc.* **1979**, *101*, 4481–4484; b) R. H. Griffey, C. D. Poulter, A. Bax, B. L. Hawkins, A. Yamaizumi, S. Nishimura, *Proc. Natl. Acad. Sci. USA.* **1983**, *80*, 5895–5897.
- [22] G. Bodenhausen, D. J. Ruben, *Chem. Phys. Lett.* **1980**, *69*, 185–189.
- [23] M. Ikura, L. E. Kay, A. Bax, *Biochemistry* **1990**, *29*, 4659–4667.
- [24] a) H. M. Sulzbach, P. v. R. Schleyer, H. F. Schaefer III, *J. Am. Chem. Soc.* **1994**, *116*, 3967–3972; b) H. M. Sulzbach, P. v. R. Schleyer, H. F. Schaefer III, *J. Am. Chem. Soc.* **1995**, *117*, 2632–2637; c) H. M. Sulzbach, G. Vacek, P. R. Schreiner, J. M. Galbraith, P. v. R. Schleyer, H. F. Schaefer III, *J. Comp. Chem.* **1997**, *18*, 126–138.
- [25] D. Jiao, M. Barfield, V. J. Hruby, *J. Am. Chem. Soc.* **1993**, *115*, 10883–10887.
- [26] N. Asakawa, H. Kurosu, I. Ando, *J. Mol. Struct.* **1994**, *323*, 279–285.
- [27] J. G. Pearson, J. F. Wang, J. L. Markley, H.-B. Le, E. Oldfield, *J. Am. Chem. Soc.* **1995**, *117*, 8823–8829.
- [28] R. H. Havlin, H.-B. Le, D. D. Laws, A. C. de Dios, E. Oldfield, *J. Am. Chem. Soc.* **1997**, *119*, 11951–11958.
- [29] J. Heller, D. D. Laws, M. Tomaselli, D. S. King, D. E. Wemmer, A. Pines, R. H. Havlin, E. Oldfield, *J. Am. Chem. Soc.* **1997**, *119*, 7827–7831.
- [30] A. C. de Dios, E. Oldfield, *Solid State Nucl. Magn. Reson.* **1996**, *6*, 101–125.
- [31] A. Perczel, A. G. Császár, *J. Comp. Chem.* **2000**, *21*, 882–900.
- [32] J. Vaara, J. Kaski, J. Jokisaari, P. Diehl, *J. Phys. Chem. A* **1997**, *101*, 5069–5081.
- [33] A. C. de Dios, E. Oldfield, *J. Am. Chem. Soc.* **1994**, *116*, 11485–11488.
- [34] G. Zheng, L. M. Wang, J. Z. Hu, X. D. Zhang, L. F. Shen, C. H. Ye, G. A. Webb, *Magn. Reson. Chem.* **1997**, *35*, 606–608.
- [35] A. G. Császár, A. Perczel, *Prog. Biophys. Mol. Biol.* **1999**, *71*, 243–309.
- [36] P. Hudáky, I. Jáklí, A. G. Császár, A. Perczel, *J. Comp. Chem.*, in press.
- [37] R. Ditchfield, *Mol. Phys.* **1974**, *27*, 789.
- [38] K. Wolinski, J. F. Hinton, P. Pulay, *J. Am. Chem. Soc.* **1990**, *112*, 8251–8260.
- [39] D. D. Laws, H.-B. Le, A. C. de Dios, R. H. Havlin, E. Oldfield, *J. Am. Chem. Soc.* **1995**, *117*, 9542–9546.
- [40] J. G. Pearson, H.-B. Le, L. K. Sanders, N. Godbout, R. H. Havlin, E. Oldfield, *J. Am. Chem. Soc.* **1997**, *119*, 11941–11950.
- [41] D. D. Laws, A. C. de Dios, E. Oldfield, *J. Biomol. NMR* **1993**, *3*, 607–612.
- [42] P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta* **1973**, *28*, 213.
- [43] A. Schafer, C. Huber, R. Ahlrichs, *J. Chem. Phys.* **1994**, *100*, 5829–5835.
- [44] *Gaussian94*, Revision B.2, M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, A. Nanavakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, J. A. Pople, Gaussian, Inc., Pittsburgh PA, **1995**.
- [45] *Gaussian98*, Revision A.5, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B.

- Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, J. A. Pople, Gaussian, Inc., Pittsburgh PA, **1998**.
- [46] A. G. Császár, *J. Am. Chem. Soc.* **1992**, *114*, 9568–9575.
- [47] V. Barone, C. Adamo, F. Lelj, *J. Chem. Phys.* **1995**, *102*, 364.
- [48] A. G. Császár, *J. Phys. Chem.* **1996**, *100*, 3541–3551.
- [49] A. G. Császár, W. D. Allen, H. F. Schaefer III, *J. Chem. Phys.* **1998**, *108*, 9751–9764.
- [50] 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.* **1977**, *112*, 535–542.
- [51] C. Ramakrishnan, G. N. Ramachandran, *Biophys. J.* **1965**, *5*, 909–933.
- [52] IUPAC-IUB, Commission on Biochemical Nomenclature, *Biochemistry* **1970**, *9*, 3471.
- [53] A. Perczel, J. G. Ángyán, M. Kajtár, W. Viviani, J.-L. Rivail, J.-F. Marcocchia, I. G. Csizmadia, *J. Am. Chem. Soc.* **1991**, *113*, 6256–6265.
- [54] A. Perczel, M. A. McAllister, P. Császár, I. G. Csizmadia, *J. Am. Chem. Soc.* **1993**, *115*, 4849–4858.
- [55] A. Perczel, Ö. Farkas, I. G. Csizmadia, *J. Am. Chem. Soc.* **1996**, *118*, 7809–7817.
- [56] T. Head-Gordon, M. Head-Gordon, M. J. Frisch, C. Brooks III, J. A. Pople, *J. Am. Chem. Soc.* **1991**, *113*, 5989–5997.
- [57] A. Perczel, I. G. Csizmadia in *The Amide Linkage* (Eds.: A. Greenberg, C. M. Breneman, J. F. Liebman), Wiley, New York, **2000**, pp. 409–463.
- [58] A. Perczel, Ö. Farkas, A. G. Császár, I. G. Csizmadia, *Can. J. Chem.* **1997**, *75*, 1120–1130.
- [59] W. Viviani, J.-L. Rivail, A. Perczel, I. G. Csizmadia, *J. Am. Chem. Soc.* **1993**, *115*, 8321–8329.
- [60] R. H. Havlin, H.-B. Le, D. D. Laws, A. C. de Dios, E. Oldfield, *J. Am. Chem. Soc.* **1997**, *119*, 11951–11958.
- [61] T. M. Wonnacott, R. J. Wonnacott, *Introductory Statistics*, 5th ed., Wiley, New York, **1990**.
- [62] H. R. Kricheldorf, D. Müller, *Macromolecules* **1983**, *16*, 615–623.

Received: July 5, 2000 [F2584]