

On NMR isotropic chemical shift surfaces of peptide models

Eszter Czinki, Attila G. Császár*

Department of Theoretical Chemistry, Eötvös University, P.O. Box 32, H-1518 Budapest 112, Hungary

Abstract

Nuclear magnetic resonance isotropic chemical shift (ICS) surfaces have been generated for all the atoms of the central Ala residue of the peptide models For-(L-Ala)_n-NH₂, $n = 1, 3$, and 5, at the GIAO-B3LYP/TZ2P level at a grid of 10° in the backbone dihedral angles ϕ and ψ for $n = 1$ and at a grid of 30° for the larger models. The resulting periodic 3D ICS(ϕ, ψ) surfaces were fitted employing a number of suitable mathematical functions. A cosine expansion provided the most accurate and compact representation of the highly structured ICS surfaces. At 10th order, involving 66 coefficients, the mean errors in the ¹³C^α and ¹H^α ICS predictions for the For-L-Ala-NH₂ model are 0.33 and 0.034 ppm, respectively. Similar fitting residues were found for other nuclei, as well. The computed surfaces are compared to results corresponding to an experimental database. The changes in chemical shifts due to the increase in the model size are investigated in detail. © 2004 Elsevier B.V. All rights reserved.

Keywords: Nuclear magnetic resonance spectroscopy; Isotropic chemical shifts; Peptide models

1. Introduction

The important and almost unique role nuclear magnetic resonance (NMR) spectroscopy plays in the determination of 3D structures of biomolecules, especially of proteins, is now firmly established. In recent years significant advances have been made both on the experimental [1–5] and theoretical [6–14] fronts, resulting in an improved understanding not only of the technique itself but also of the quality of the results obtained.

From a theoretical perspective, a particularly important challenge is to understand the dependence of NMR chemical shielding tensors on structural parameters. While, of course, this is also an overwhelmingly important problem for experimentalists, one has to realize (vide infra) that sufficiently detailed experimental data have been generated only for rather limited ranges of the Ramachandran map. First-principles investigation of suitable peptide models, focusing on the backbone dihedral angles ϕ and ψ and the side-chain angle χ (see Fig. 1 and Ref. [15] for definition of these angles) confirmed and extended [6–12, 16–23] experimental findings that these structural parameters are intimately related to the shielding properties of the backbone and H^α atoms, for which most of the experimental NMR chemical shift information is available.

To make any modeling study tractable and of practical relevance at the same time, it is usual to explore

the shieldings or chemical shifts of smaller model compounds as a function of the dihedral angles ϕ and ψ . While surfaces depending on any other structural variable can also be generated straightforwardly from the results of the same computations, at the moment these do not seem to be of value for the practising NMR spectroscopist. It is also common practice to focus on isotropic chemical shifts (ICS), thereby neglecting the anisotropy of chemical shielding tensors, the intrinsic results of first-principles computations of NMR properties of nuclei. Understanding of the characteristics of (ϕ, ψ)-dependent ICS surfaces can be achieved either by focusing on only the most relevant (ϕ, ψ) regions supporting well-known conformational types [16], or by doing a full surface analysis [12,21–23]. Previously we [16–20] have followed the first route and found seemingly useful chemical shift—chemical shift and chemical shift—structure correlations indicating that shifts from multidimensional NMR spectroscopy can provide directly for structure analysis. The second route, resulting in the availability of chemical shift surfaces constitutes an important step toward a deeper understanding of NMR determination of 3D structures of biomolecules though certain regions of such surfaces may be thought of as irrelevant. Naturally, it is worth exploring useful combinations of the two approaches.

Whether one follows basically the first or the second route, it is of utmost importance to determine which atoms change their shielding properties based on local changes in their environment, giving pronounced torsional dependence,

* Corresponding author.

E-mail address: csaszar@chem.elte.hu (A.G. Császár).

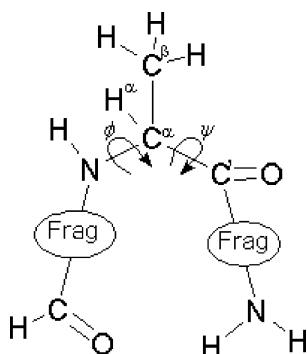


Fig. 1. Scheme of For-(L-Ala)_n-NH₂, *n* = 1, 3, 5 models, where Frag = CO-CH(CH₃)-NH and (CO-CH(CH₃)-NH)₂ for *n* = 3 and 5, respectively. The dihedral angles ϕ and ψ are defined as $\phi = C_{i-1}^{\alpha}-N^H-C^{\alpha}-C'$; $\psi = N^H-C^{\alpha}-C'-N_{i+1}^H$.

and which are affected the most by non-local effects, deteriorating the quality of simple modeling efforts. In particular, when the goal is to obtain ICS(ϕ, ψ) surfaces, important effects limiting the modeling accuracy must be considered, namely capping and neighboring effects, primarily in case of atoms on the 'edge' of the peptide residue (such as N^H and H^N), dependence of the shifts on the χ torsional angle, mostly affecting the shielding of the side-chain atoms (e.g. C^B) and the carbonyl C (C'), and the effects of inter- and intramolecular as well as inter- and intraresidual H-bonds. These effects can be explored and, in favorable cases, can be modeled [22].

While the accuracy of the computation of NMR chemical shifts at levels affordable for protein models cannot be considered fully satisfactory, our own studies [16–20], along with those of others [13,14,22] indicate reasonable precision, certainly sufficient for the present modeling study. Computations have several advantages.

Most importantly, they allow one to obtain information about not only regions of the potential energy surface (PES) which would be populated based on simple energy arguments (a considerable problem of computational studies is that energy surfaces of simple peptide models seem to be drastically different from those of larger models though, fortunately, they support the same structural types [16]) but about the whole surface including all conformers and conformations. The availability of more and more powerful and inexpensive computers allows ab initio computation of chemical shifts for larger and larger molecules, and also allows one to use larger basis sets and, perhaps in the near future, to incorporate electron correlation effects [20,24–26].

In this study we chose alanine (Ala) as our residue model since it offers several advantages: (a) it introduces no pronounced side-chain effects; (b) allows relative ease of computation, even of larger models; (c) larger models can be built straightforwardly without undue effort; and (d) when full surfaces are considered Ala serves as a good model of several residues. The last point should be true if the first-order effect of side-chain—backbone interactions is simply to force the residue to adopt a certain conformation but shieldings of most of the atoms is determined by their local environment, notably by the (ϕ, ψ) angles.

The raw ab initio chemical shift data generated for the Ala models can be fitted to periodic functions of ϕ, ψ (and perhaps χ) in a number of ways [12,21,27] yielding ICS(ϕ, ψ) surfaces. While tremendous amount of work has been devoted to fitting PESs [28–32] and dipole moment surfaces [30] of small molecules, there is comparatively little evidence about the best functional forms applicable to ICS surfaces. This aspect of the problem has been dealt with at considerable detail in Section 3 of the present study.

Table 1
Functions most suitable to reproduce periodic ICS(ϕ, ψ) surfaces

Function	Formula
<i>N</i> th-order cosine series ^a	$\text{ICS}(\phi, \psi) = a + \sum_{n=1}^N [b_n \cos(n\phi') + c_n \cos(n\psi')] + \sum_{n=1}^{N-1} \sum_{m=1}^{N-n} d_{nm} \cos(n\phi') \cos(m\psi')$
2 × <i>N</i> th-order Fourier series ^a	$\text{ICS}(\phi, \psi) = a + \sum_{n=1}^N \sum_{w=\phi, \psi} [b_n \cos(nw') + c_n \sin(nw')] + \sum_{n=1}^{N-1} \sum_{m=1}^{N-n} [d_{nm} \cos(n\phi') \cos(m\psi') + e_{nm} \sin(n\phi') \sin(m\psi')] + \sum_{n=1}^{N-1} \sum_{m=1}^{N-n} [f_{nm} \cos(n\phi') \sin(m\psi') + g_{nm} \sin(n\phi') \cos(m\psi')]$
<i>N</i> th-order Chebyshev polynomial ^b	$\text{ICS}(\phi, \psi) = a + \sum_{n=1}^N [b_n T_n(\phi') + c_n T_n(\psi')] + \sum_{n=1}^{N-1} \sum_{m=1}^{N-n} d_{nm} T_n(\phi') T_m(\psi')$
<i>N</i> th-order sigmoid series ^c	$\text{ICS}(\phi, \psi) = a + \sum_{n=1}^N [b_n S_n(\phi') + c_n S_n(\psi')] + \sum_{n=1}^{N-1} \sum_{m=1}^{N-n} d_{nm} S_n(\phi') S_m(\psi')$

^a $\phi' = \phi$ scaled to $[0, \pi]$ and $\psi' = \psi$ scaled to $[0, \pi]$.

^b $T_n(x') = \cos(n \arccos(x'))$; $\phi' = \phi$ scaled to $[-1, +1]$ and $\psi' = \psi$ scaled to $[-1, +1]$.

^c $S_1(x') = x'$, $S_{i=2, \dots, n}(x') = -1 + \frac{2}{1 + \exp(\frac{-x'(1 - (i-1)(2/n))}{0.12})}$, $x' = x$ scaled to $[-1, +1]$, and $x = \phi$ or ψ .

Table 2

Coefficients of best fitted surfaces for the 10th-order cosine series for the $^{13}\text{C}^\alpha$, $^1\text{H}^\alpha$, $^{13}\text{C}^\beta$, $^{15}\text{N}^\text{H}$, $^1\text{H}^\text{N}$ nuclei of the For–Ala–NH₂ model

Coeff.	$^{13}\text{C}^\alpha$	$^1\text{H}^\alpha$	$^{13}\text{C}^\beta$	$^{13}\text{C}'$	$^{15}\text{N}^\text{H}$	$^1\text{H}^\text{N}$
<i>a</i>	58.3169	4.3081	20.9755	178.6910	154.0894	4.8361
<i>b</i> ₁	−4.1053	0.5193	0.7473	−0.0053	4.1569	0.0256
<i>b</i> ₂	−1.5757	0.2805	−1.8738	−0.8353	0.3256	0.4592
<i>b</i> ₃	1.6091	−0.2812	0.1972	−0.3074	−1.7499	0.1984
<i>b</i> ₄	1.3266	−0.0131	1.5328	−0.3341	−0.8947	−0.1214
<i>b</i> ₅	1.2674	−0.1343	−0.7805	0.0768	−0.8781	−0.1250
<i>b</i> ₆	0.1331	−0.1363	−1.2824	0.2156	−0.2177	0.1929
<i>b</i> ₇	0.6996	−0.0538	−0.0059	0.0823	−0.6190	−0.0457
<i>b</i> ₈	0.0119	0.0009	−0.0426	−0.0927	0.6172	−0.0202
<i>b</i> ₉	0.1268	−0.0092	−0.0979	0.1078	−0.2869	−0.0151
<i>b</i> ₁₀	−0.0240	0.0060	−0.0622	−0.0274	−0.2217	−0.0189
<i>c</i> ₁	1.1493	0.0507	−0.3649	−0.3540	−0.7377	0.0912
<i>c</i> ₂	−2.3026	−0.0434	0.0374	−0.1648	−3.6889	0.1343
<i>c</i> ₃	−1.8350	0.0363	−1.6281	−0.5189	0.4971	−0.0091
<i>c</i> ₄	0.0506	0.0492	1.0937	0.7786	−0.9549	0.1169
<i>c</i> ₅	0.4255	−0.0655	0.8235	0.4766	−0.0671	−0.0259
<i>c</i> ₆	0.2974	0.0492	0.1666	−0.1866	−0.7882	0.0722
<i>c</i> ₇	0.4209	−0.0013	0.2796	0.1800	−0.1224	0.0015
<i>c</i> ₈	−0.1465	0.0164	0.1728	−0.4131	−0.3317	−0.0137
<i>c</i> ₉	−0.1065	−0.0037	0.3130	0.1134	0.1709	−0.0174
<i>c</i> ₁₀	0.0007	0.0004	−0.0245	−0.0394	−0.0342	0.0092
<i>d</i> ₁₁	0.3502	0.0386	0.3679	−0.4791	−1.0529	−0.3898
<i>d</i> ₁₂	0.1706	0.0079	−0.0124	0.1791	0.0937	0.0364
<i>d</i> ₁₃	0.5250	0.0465	0.0665	−0.4362	2.0508	−0.0770
<i>d</i> ₁₄	0.6174	−0.0443	0.5192	0.1571	−0.0802	−0.0175
<i>d</i> ₁₅	−0.2590	−0.0301	−0.1444	0.7828	−0.7037	0.1608
<i>d</i> ₁₆	0.1303	−0.0085	0.1737	0.0782	0.0312	−0.0005
<i>d</i> ₁₇	−0.3217	−0.0077	−0.0464	−0.0862	0.1844	0.0963
<i>d</i> ₁₈	−0.1670	0.0156	0.0110	0.0173	0.2048	0.0116
<i>d</i> ₁₉	−0.0441	−0.0167	−0.0315	−0.0304	−0.2304	0.0619
<i>d</i> ₂₁	−0.1025	−0.0516	0.1577	0.0638	−0.0332	0.0227
<i>d</i> ₂₂	0.0864	0.0007	0.2249	0.1040	2.2891	0.6617
<i>d</i> ₂₃	0.2308	−0.0278	−0.0603	0.3436	0.4257	0.0079
<i>d</i> ₂₄	−0.8060	−0.0298	−0.0617	1.0901	−0.6357	0.2755
<i>d</i> ₂₅	0.2127	0.0167	0.1302	−0.1174	−0.3975	0.0268
<i>d</i> ₂₆	−0.3291	−0.0139	0.0810	−0.5148	0.1942	0.1250
<i>d</i> ₂₇	−0.2815	0.0401	−0.0989	−0.1989	−0.1626	−0.0085
<i>d</i> ₂₈	−0.1242	−0.0097	−0.0224	−0.0360	−0.0505	0.0254
<i>d</i> ₃₁	0.1795	−0.0716	−0.6430	1.4682	1.7368	0.3494
<i>d</i> ₃₂	0.3252	−0.0966	0.1997	−0.1269	−0.1686	0.0076
<i>d</i> ₃₃	−0.9351	−0.0402	0.0089	0.2077	−0.3668	−0.0237
<i>d</i> ₃₄	−0.2512	0.0421	−0.2964	−0.2631	0.0214	−0.0553
<i>d</i> ₃₅	0.4794	0.0215	0.3433	−0.8793	−0.3250	−0.1020
<i>d</i> ₃₆	−0.0778	0.0148	−0.2101	−0.0555	−0.2225	0.0067
<i>d</i> ₃₇	0.0627	0.0307	0.1076	−0.4167	−0.4601	−0.0794
<i>d</i> ₄₁	0.5306	−0.0604	−0.0529	0.2147	−0.5072	0.0223
<i>d</i> ₄₂	−0.6915	−0.1136	−0.6860	−0.2623	−1.3647	0.1552
<i>d</i> ₄₃	−0.0978	0.0380	−0.3347	−0.1639	0.0283	−0.0001
<i>d</i> ₄₄	0.2983	0.0252	0.2779	−0.9395	0.0751	0.0719
<i>d</i> ₄₅	−0.2485	0.0152	−0.0353	−0.0538	−0.1127	−0.0035
<i>d</i> ₄₆	−0.1269	0.0299	−0.0200	−0.1544	−0.3506	0.0366
<i>d</i> ₅₁	−0.1653	−0.0197	0.1431	−0.7374	−0.7350	0.0215
<i>d</i> ₅₂	0.0528	0.0111	0.0317	0.0890	0.5844	−0.0927
<i>d</i> ₅₃	0.4144	0.0353	0.2761	−0.0719	−0.7103	−0.0558
<i>d</i> ₅₄	−0.2657	0.0027	−0.0577	0.0762	0.2123	0.0137
<i>d</i> ₅₅	−0.2788	−0.0060	−0.2622	0.6347	0.6454	0.0444
<i>d</i> ₆₁	−0.1929	0.0198	0.2733	0.0955	0.2437	−0.0290
<i>d</i> ₆₂	0.3499	0.0186	−0.0566	−0.2208	0.4745	−0.0438
<i>d</i> ₆₃	−0.1576	0.0014	0.0628	0.0113	0.3101	−0.0337
<i>d</i> ₆₄	−0.2112	−0.0148	−0.1158	0.6270	0.2352	0.0356
<i>d</i> ₇₁	−0.0643	0.0140	−0.2049	−0.1082	−0.4059	0.0425

Table 2 (continued)

Coeff.	$^{13}\text{C}^\alpha$	$^1\text{H}^\alpha$	$^{13}\text{C}^\beta$	$^{13}\text{C}'$	$^{15}\text{N}^\text{H}$	$^1\text{H}^\text{N}$
<i>d</i> ₇₂	−0.2662	0.0526	−0.1032	−0.0702	−0.3968	0.0305
<i>d</i> ₇₃	−0.2139	−0.0252	−0.1557	0.3184	−0.0668	0.0694
<i>d</i> ₈₁	−0.0350	0.0096	−0.0435	−0.0057	−0.1881	0.0151
<i>d</i> ₈₂	−0.4073	0.0377	0.1699	0.0592	0.2604	−0.0633
<i>d</i> ₉₁	−0.1331	0.0185	0.1494	−0.0021	0.4035	−0.0339

The coefficients in this table correspond to the function defined in Table 1. Fitting employed the filtered computed data points (see text) in case of $^{13}\text{C}^\alpha$.

2. Computational details

Peptide models containing alanine, in form of For–(Ala)_{*n*}–NH₂, *n* = 1, 3, 5 (Fig. 1), were built for exploring chemical shift surfaces. In order to generate the surfaces as a function of backbone dihedral angles ϕ and ψ , a 36×36 grid in the $[-180^\circ, +180^\circ]$ conformational space has been used for the smallest model, For–Ala–NH₂. This unusually dense grid of 10° results in 1296 structures. The dihedral angles of the input structures were kept fixed while all the other geometrical parameters were optimized at the DFT(B3LYP)/6-31 + G* level using the program system GAUSSIAN98 [33]. For the larger alanine models, For–(Ala)₃–NH₂ and For–(Ala)₅–NH₂, a 30° grid in ϕ and ψ was used for the dihedral angles of the amino acid residue in the middle of the structure, while the ϕ and ψ angles of the other residues were kept fixed at $\phi = -60^\circ$ and $\psi = -50^\circ$, as if they were part of α -helices. Note that in case of the *n* = 3 and 5 models not all of the 144 points could be optimized to a realistic peptide fragment geometry, we excluded those structures from further analysis. Thus, for example, in case of For–(Ala)₃–NH₂ we obtained just 137 structures.

The B3LYP (Becke3–Lee–Yang–Parr) [34,35] functional of Density Functional Theory (DFT) and the Gauge-Including Atomic Orbital (GIAO) method [36,37], as implemented in GAUSSIAN98 [33], were employed for computation of NMR shielding tensors at the structures resulting from the constrained optimizations. Since the importance of employing relatively large basis sets for NMR chemical shielding calculations is well established, we used a TZ2P [38] basis set, especially applicable for NMR shielding calculations. Relative chemical shifts (δ -scale) were calculated using ^1H and ^{13}C isotropic chemical shielding values of tetramethylsilane (TMS), and ^{15}N isotropic chemical shielding values of NH₃, as references. The reference geometry of NH₃ was optimized at the all-electron CCSD(T)aug-cc-pVTZ level of theory [39], while the reference geometry chosen for TMS corresponds to the B3LYP/6-311++G** level.

Given the large number of structures and shieldings to be handled, 1296 reference points for mono-, and about 144 points for tri- and penta-Ala models, automatization of data

handling becomes mandatory. Therefore, we employed Python [40] scripts extensively for input generation, data handling, output testing, and data extraction. The large amount of data extracted was entered into a database for data mining. We used MySQL [41] for data storage, and a Python-MySQL interface for the seamless deposition and retrieval of data.

3. Surface fitting

In order to efficiently describe ICS as a function of geometric parameters, such as ϕ and ψ , it is necessary to construct an interpolating surface from the points corresponding to a given grid. Nevertheless, fitting of a periodic surface of complex shape, such as the ICS surfaces of all nuclei studied, is not a trivial task. Although generation of shielding surfaces was part of previous studies [12,21,27], no detailed description of the functions, the parameters, and the problems arising during the fitting seems to exist.

It is clear that certain function types (e.g. simple polynomial and rational functions) are not well suited for the task. This has also been born out of our own investigations. On the other hand, expansion in trigonometric series is highly suitable. In fact, of the many functional forms we tried the cosine series, the (multi-dimensional) Fourier series, the Chebyshev polynomials, and the sigmoid series proved to be the most suitable. These functions are given in detail in Table 1. One of the principal practical conclusions of this study is that all these approaches give approximately the same result using a comparable number of fitting parameters. Goodness of fit, r^2 [42], values larger than 0.95 can be achieved with cosine series of order 7 (and with larger orders), with sigmoid series of order 8 (and larger), with Chebyshev polynomials of order 8 (and larger), and with Fourier series of order 2×4 (and larger). Nevertheless, the 10th-order cosine series appears as the best overall choice.

Due to the complex, non-symmetric dependence of chemical shifts on ϕ and ψ , incorporation of cross-terms in the expansion is essential. For example, for the otherwise well-fitted $\delta(^{13}\text{C}^\alpha)$ surface the highest r^2 value at 10th order (41 parameters) is only 0.92 if the Fourier expansion lacks

cross-terms. The importance of cross-terms can be clearly seen in Table 2, which contains the parameters of the best fitted surfaces, obtained using the 10th-order cosine series, for the $^{13}\text{C}^\alpha$, $^1\text{H}^\alpha$, $^{13}\text{C}^\beta$, $^{13}\text{C}'$, $^{15}\text{N}^\text{H}$, and $^1\text{H}^\text{N}$ nuclei. Note, in particular, how large some of the higher-order d_{nm} coefficients are.

The chemical shift surfaces of atoms $^{13}\text{C}^\alpha$ and $^1\text{H}^\alpha$ can be obtained with $r^2 \cong 0.99$, the fitted chemical shift surfaces of $^{13}\text{C}^\beta$, $^{15}\text{N}^\text{H}$, and $^1\text{H}^\text{N}$ are characterized by $r^2 \cong 0.97$. The chemical shift surface of $^{13}\text{C}'$ can be fitted worst, the r^2 value obtained is only about 0.82. It seems that the more complex the expected non-local behavior is the more complex the ICS surface of that nucleus becomes, resulting in smaller r^2 values. Table 3 contains the average errors in the reproduction of chemical shifts of the nuclei discussed.

The r^2 value of the fit deteriorates if the data set contains outlying points seriously different from their neighbors. Instead of applying complex statistical measures for filtering out the troublesome points we used the following simple procedure on the raw data: if a given point differs from the arithmetic average of its neighboring points ($\phi_{i\pm\{0,1\}}$, $\psi_{j\pm\{0,1\}}$) by more than ε (e.g. 1.0 ppm in case of C^α), that point is dropped from the set to be fitted. Following this procedure, we dropped some 20 points in case of C^α for the For–Ala–NH₂ model. Since the full raw dataset contains 1296 points for this model, the slight decrease in the member of points results in no deterioration in the quality of the surface. The fit became slightly better after removal of these points, the lonely big residuals corresponding to the severely outlying points disappeared. Naturally, the fit still produces some (relatively) sizeable residuals. They correspond to small regions of complex shape even the best functional form employed has problems describing. Because all surfaces investigated possess rather complex shapes, this is acceptable.

In order to derive a smoothed surface from measured or calculated data, one can try simple smoothing formulas instead of explicitly fitting functions. Spera and Bax [27] as well as Le et al. [12] followed this route for constructing a surface based on their experimental data. Since the two formulas mentioned yield approximately the same surfaces, the results discussed below correspond to the very slightly

Table 3

Average errors in the reproduction of isotropic chemical shift surfaces of For–Ala–NH₂ using a 10th-order cosine expansion

	$\delta(^{13}\text{C}^\alpha)^a$	$\delta(^1\text{H}^\alpha)$	$\delta(^{13}\text{C}^\beta)$	$\delta(^{13}\text{C}')$	$\delta(^{15}\text{N}^\text{H})$	$\delta(^1\text{H}^\text{N})$
MAD of residual	0.3331	0.0334	0.3393	0.6460	0.6636	0.0741
MAD of residual%	0.5735	0.7849	1.5831	0.3620	0.4349	1.5797
Std of residuals ^b	0.2878	0.0289	0.3223	0.5762	0.6499	0.0603
UMD of residuals ^c	1.8440	0.2441	3.3076	2.7257	5.5115	0.4224

The coefficients correspond to the function defined in Table 1. All values in ppm. Residual = ICS(ϕ_i , ψ_i)—directly computed chemical shift (ϕ_i , ψ_i); residual% = (residual/calculated value) \times 100; MAD, mean absolute deviation.

^a Fitting employed the filtered data points (see text).

^b Std, standard deviation of values.

^c UMD, unsigned maximum deviation.

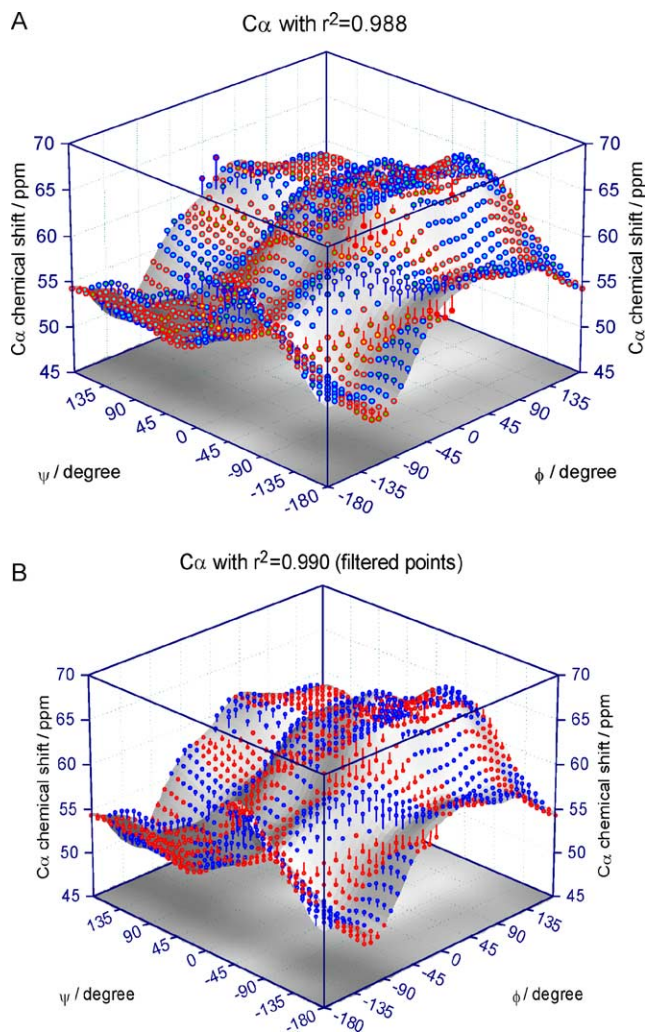


Fig. 2. $^{13}C_{\alpha}$ isotropic chemical shift surface of For–(L-Ala)–NH₂. The 10th-order cosine fitting employed the full (A) and the filtered (B) set of computed data points.

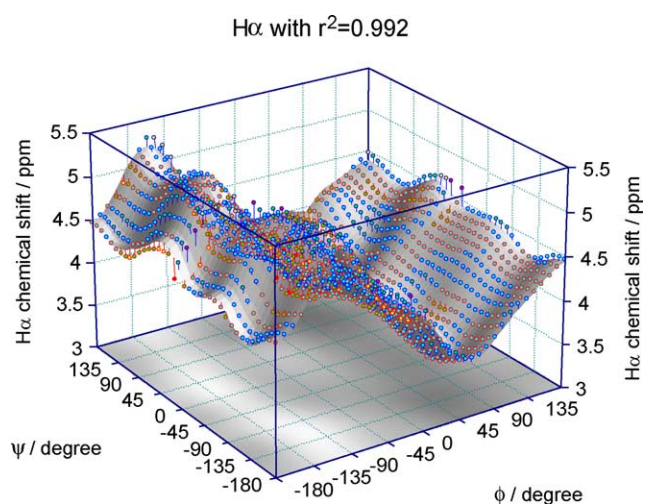


Fig. 3. $^1H_{\alpha}$ isotropic chemical shift surface of For–(L-Ala)–NH₂.

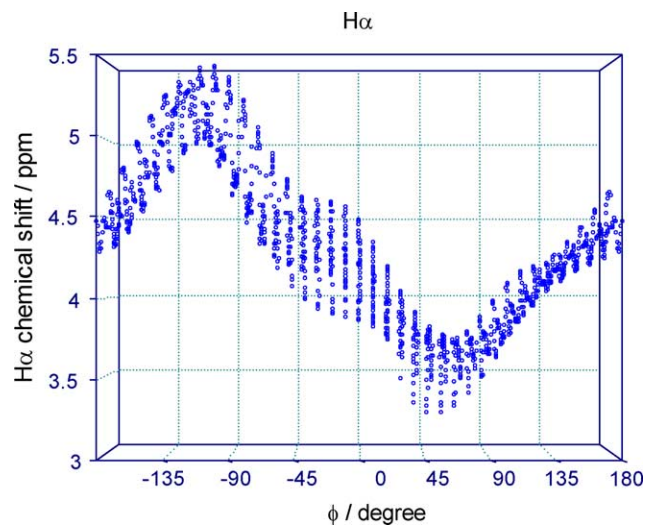


Fig. 4. The relatively simple dependence of calculated $^1H_{\alpha}$ isotropic chemical shifts of For–(L-Ala)–NH₂ on the dihedral angle ϕ .

more favorable formula of Le et al., given as follows

$$S(x, y) = \frac{\sum_i P_i(\phi_i, \psi_i) \exp\left(\frac{-\left(\sin^2\left(\frac{\phi_i-x}{2}\right) + \sin^2\left(\frac{\psi_i-y}{2}\right)\right)}{0.03}\right)}{\sum_i \exp\left(\frac{-\left(\sin^2\left(\frac{\phi_i-x}{2}\right) + \sin^2\left(\frac{\psi_i-y}{2}\right)\right)}{0.03}\right)}, \quad (1)$$

where $P_i(\phi_i, \psi_i)$ is the ICS given at the i th point, and the factor 0.03 was taken from Ref. [12].

Employing this smoothing function with the calculated $^{13}C_{\alpha}$ chemical shifts for the For–Ala–NH₂ model, with all values at the whole Ramachandran surface corresponding to a grid of 10°, yield average deviations nearly as small as was found for the fitted surface.

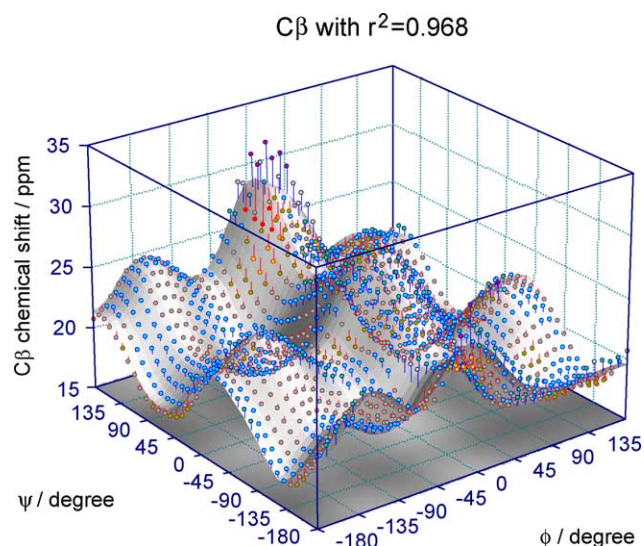


Fig. 5. $^{13}C_{\beta}$ isotropic chemical shifts surface of For–(L-Ala)–NH₂.

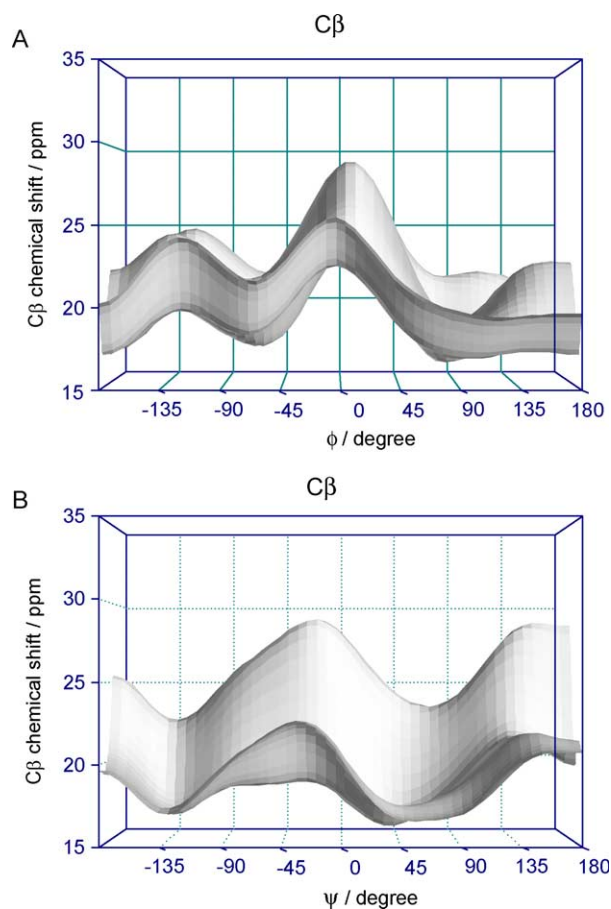


Fig. 6. Dependence of the calculated $^{13}\text{C}^\beta$ isotropic chemical shifts of For-(L-Ala)- NH_2 on dihedral angles ϕ (A) and ψ (B).

The ICS surfaces of all nuclei of interest resulting from the fitting procedure are given on Figs. 2–9. On these figures, the fitted function is represented as a surface as well as a contour plot. The balls of the figures represent the raw data points.

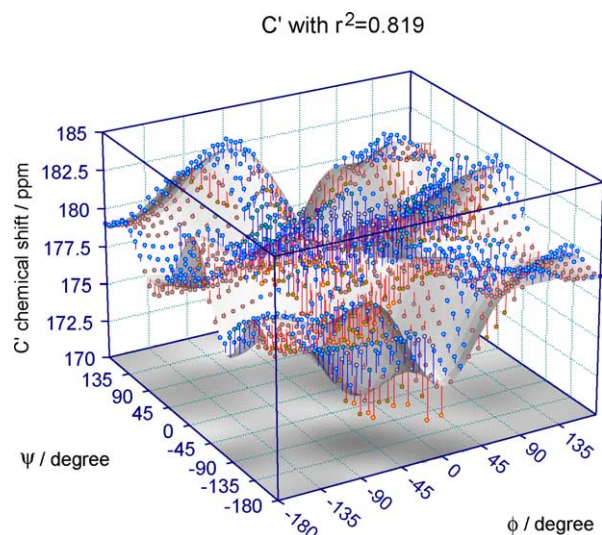


Fig. 7. $^{13}\text{C}'$ isotropic chemical shifts surface of For-(L-Ala)- NH_2 .

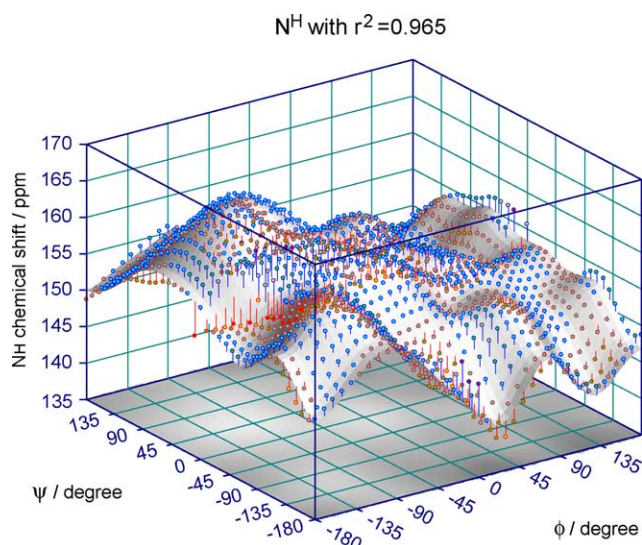


Fig. 8. $^{15}\text{N}^{\text{H}}$ isotropic chemical shifts surface of For-(L-Ala)- NH_2 .

By closer inspection of the ICS surfaces, one can observe that the chemical shifts of $^1\text{H}^\alpha$ mainly depend on ϕ . This dependence can be described with approximately the same functional forms for different dihedral angles ψ , the minimum of these 1D cuts through the surface occur around $\phi = 60^\circ$ (Fig. 4). The chemical shifts of $^{13}\text{C}^\beta$ considerably depend on both ϕ and ψ , but when 1D cuts of the surface taken, the shape of these cuts is approximately the same for each $\delta(^{13}\text{C}^\beta) - \phi$ and $\delta(^{13}\text{C}^\beta) - \psi$ functions (Fig. 6).

4. Calculated vs experimental shifts

The fitted ICS surfaces of our For-Ala- NH_2 model can be tested against experimental chemical shift values. In a future publication we are planning to test the Z-surface method [12] employing our ICS surfaces. Nevertheless, for now we just simply test our best fitted

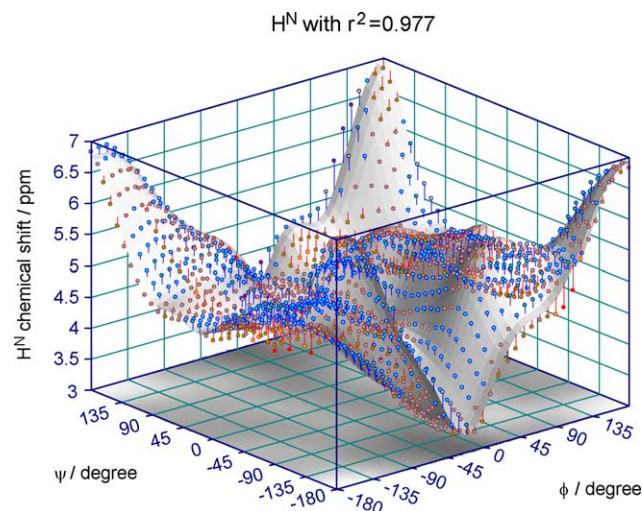


Fig. 9. $^1\text{H}^{\text{N}}$ isotropic chemical shifts surface of For-(L-Ala)- NH_2 .

Table 4

Comparison of chemical shifts of For–Ala–NH₂ (referred as ‘mono’) and For–(Ala)₃–NH₂ (referred as ‘tri’) models with BMRB average values

	$\delta(^{13}\text{C}^\alpha)$		$\delta(^1\text{H}^\alpha)$		$\delta(^{13}\text{C}^\beta)$		$\delta(^{13}\text{C}')$		$\delta(^1\text{H}^\text{N})$		$\delta(^{15}\text{N}^\text{H})$	
	Avg	Std	Avg	Std	Avg	Std	Avg	Std	Avg	Std	Avg	Std
Mono	58.326	4.443	4.308	0.509	20.976	2.766	178.691	2.042	4.836	0.656	154.089	5.062
Tri	59.476	3.941	4.269	0.495	20.237	2.831	177.828	2.487	5.679	0.550	141.999	5.531
Mono–tri	1.389	0.955	0.097	0.077	0.736	0.728	1.555	1.665	0.801	0.249	12.653	2.535
BMRB	53.160	2.060	4.260	0.420	18.900	1.850	177.800	2.180	8.200	0.600	123.230	3.700
Mono–BMRB	5.166		0.048		2.076		0.891		–3.364		30.859	
Tri–BMRB	6.316		0.009		1.337		0.028		–2.521		18.769	

Avg, mean value of all chemical shifts computed at the whole Ramachandran surface. In case of For–Ala–NH₂: 1296, For–(Ala)₃–NH₂: 137 points. Std, standard deviation of values. BMRB values were retrieved at 2003. 02. 18 from Ref. [45].

functions against an experimental database. We obtained the experimental chemical shifts of Ala residues from Wishart’s re-referenced protein chemical shift database, RefDB [43] and from BMRB [44,45]. RefDB contains experimental chemical shifts for ¹³C^α, ¹H^α, ¹³C^β, ¹H^N, and ¹⁵N^H nuclei. Note, however, that not all data are available for all residues.

For the calculated surfaces the range of calculated $\delta(^{13}\text{C}^\alpha)$ shifts is 49.11–68.98 ppm, exhibiting a spread of 19.87 ppm with a standard deviation of 4.46 ppm. The experimental $\delta(^{13}\text{C}^\alpha)$ shifts are in the range of 48.06–63.69 ppm, a spread of 15.63 ppm with a standard deviation of 2.03 ppm. The larger computational spread and its larger standard deviation are due to the fact that the computed surface is complete over the (ϕ , ψ) range while the experimental one is restricted basically to the smoother α_L (approx. 50% of data), β_L (approx. 15% of data), and ϵ_L (approx. 15% of data) regions.

Using the best ICS surface for ¹³C^α, obtained with the 10th-order cosine series (Tables 1 and 2), the calculated $\delta(^{13}\text{C}^\alpha)$ values differ from the smoothed experimental ones by 2.19 ppm on average with a standard deviation of 1.11 ppm. (The smoothed experimental surface was calculated with Eq. (1).) The maximum difference was rather large, 11.58 ppm. Therefore, the computed results are not only shifted by a constant amount compared to the experimental ones but appear to be distributed somewhat randomly, most likely due to non-local effects present in the different proteins. Although the average deviation is only about 1/20th of the full range of available values, these sizable deviations limit our ability to predict chemical shift–dihedral angle correlations.

One possible contributing factor to deviations between our computed ICS surfaces and the experimental database is the short length of our model system. The For–Ala–NH₂ model is only able to model local effects, whereas the experimental dataset clearly contains shifts due to non-local interactions. Therefore, chemical shift values have also been generated for the relevant atoms of the middle residue of For–(Ala)_n–NH₂, $n = 3$ and 5. As can be seen in Table 4, for the nuclei least amenable to non-local effects, namely for C^α, H^α and C^β, and to some extent even for C^γ,

the difference between the $n = 1$ and 3 models is relatively small. As far as the ‘edge’ atoms are concerned, where capping effects are most significant, the difference between chemical shifts of the two models is considerably larger. Especially pronounced is the large drop in the average

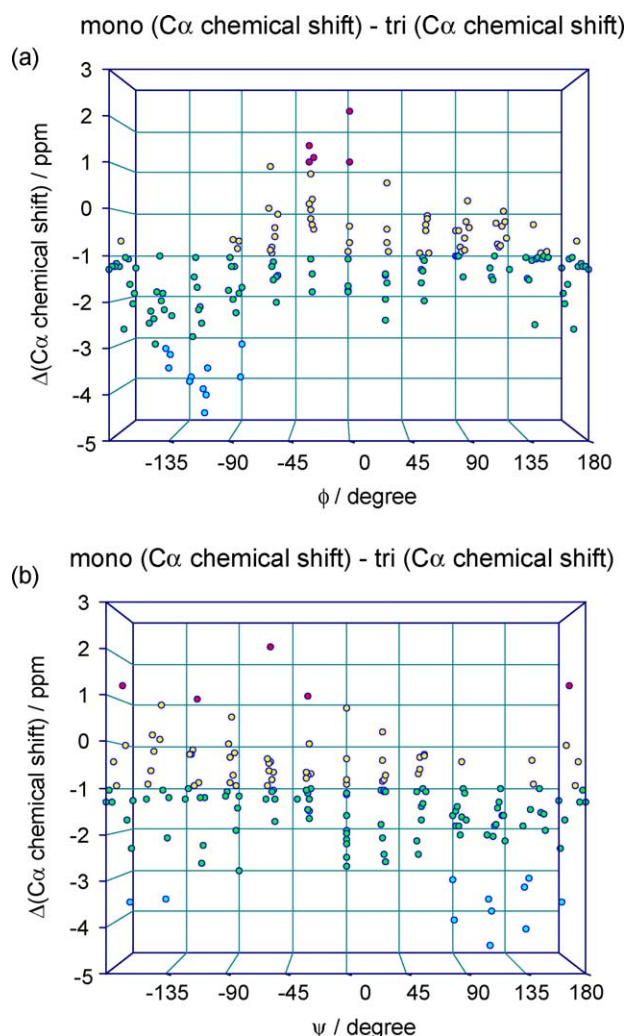


Fig. 10. Dependence of computed ¹³C^α isotropic chemical shifts on the model size [$\Delta(^{13}\text{C}^\alpha) = ^{13}\text{C}^\alpha(\text{For-L-Ala-NH}_2) - ^{13}\text{C}^\alpha(\text{For-(L-Ala)}_3\text{-NH}_2)$] as a function of dihedral angles ϕ (A) and ψ (B).

$\delta(^1\text{H}^{\text{N}})$ value toward that in the BMRB database. For all nuclei the standard deviations of the differences between the For-Ala-NH₂ and For-(Ala)₃-NH₂ models are smaller than the standard deviations characteristic of the nuclei of either model.

Comparison of results for the For-(Ala)_n-NH₂ models, $n = 1, 3, 5$, in the most interesting α_{L} region shows that

the chemical shifts of ¹³C^α, ¹H^α, and ¹³C^β, in contrast to those of ¹H^N and ¹⁵N^H, do not change considerably when the model is enlarged. The percentage error of calculated chemical shifts, compared to adequate experimental (RefDB) values, of these nuclei is below 20% in all models investigated. For these nuclei in the middle of the residue, capping effects are not particularly strong, making modeling

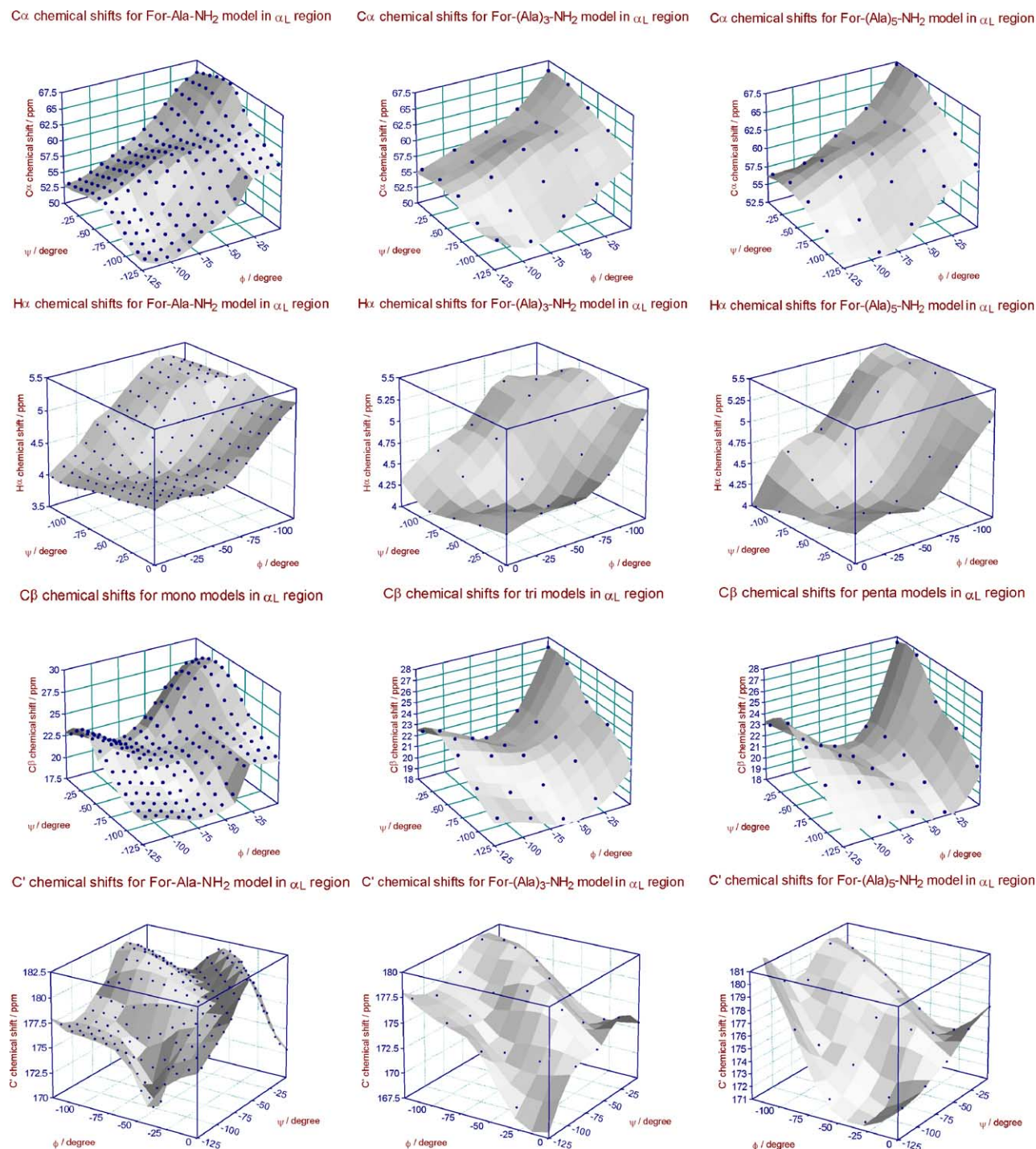


Fig. 11. Calculated ¹³C^α (first row), ¹H^α (2nd row), ¹³C^β (3rd row), ¹³C' (4th row), ¹H^N (5th row), and ¹⁵N^H (6th row) chemical shifts for the models For-(L-Ala)_n-NH₂, $n = 1$ (first column), 3 (second column), 5 (third column) in the α_{L} region ($\phi = [-120, 0]^\circ$, $\psi = [-120, 0]^\circ$).

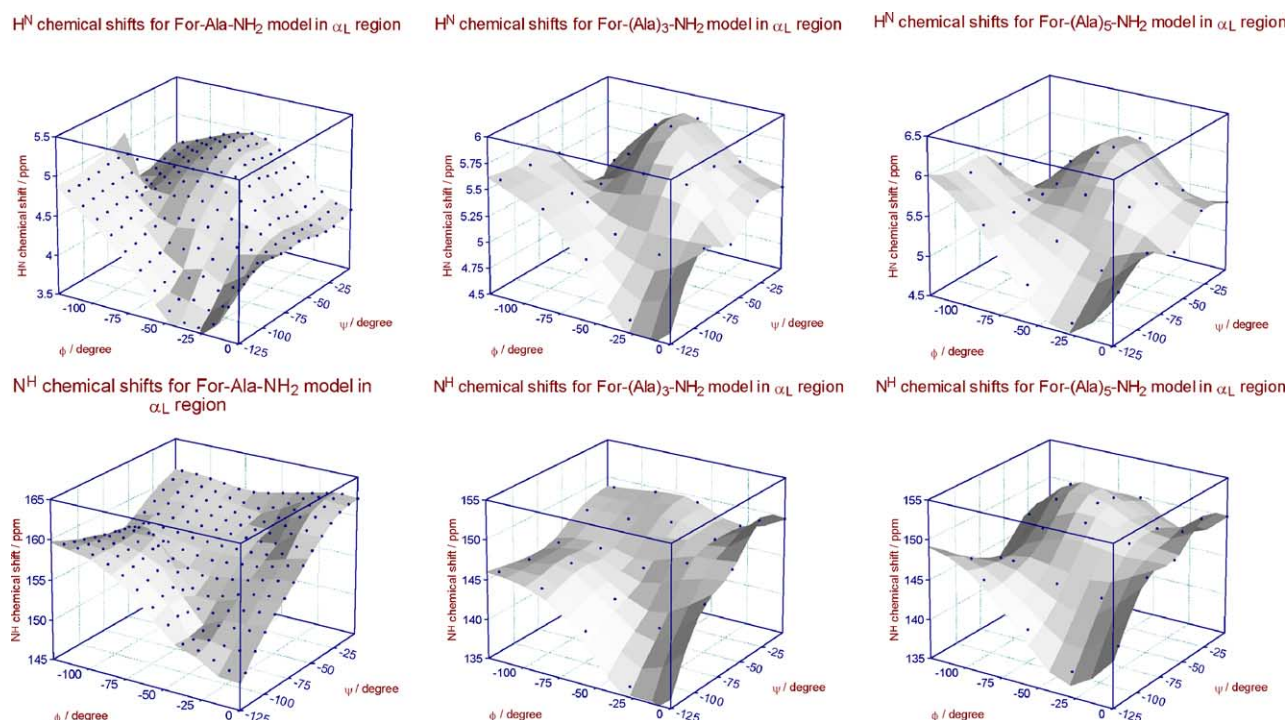


Fig. 11 (continued)

studies for these nuclei meaningful and feasible. When the full surface is investigated, for the relatively buried C^{α} nucleus the $^{13}\text{C}^{\alpha}$ shift depends on the size of the model, as shown on Fig. 10. Dependence of the differences on the dihedral angles ϕ and ψ does not show any simple pattern. On the other hand, within any given region, for example for the α_L region, the model size does not affect significantly the shape of the ICS surfaces. For all nuclei investigated the ICS surfaces change smoothly within the α_L region, as shown on Fig. 11, and correlate well with the dihedral angles.

Deviation between calculated and experimental $\delta(^1\text{H}^{\text{N}})$ and $\delta(^{15}\text{N}^{\text{H}})$ values is substantial. The calculated chemical shifts of $^1\text{H}^{\text{N}}$ and $^{15}\text{N}^{\text{H}}$ change substantially toward the experimental values when the model size is enlarged from For-Ala-NH₂ to For-(Ala)₃-NH₂. Nevertheless, the change is small when the model is further enlarged to For-(Ala)₅-NH₂. This proves that in case of $^1\text{H}^{\text{N}}$ and $^{15}\text{N}^{\text{H}}$ the capping effect causes the largest error in the computed chemical shifts. This observation is in line with our previous results [20] that not the lack of electron correlation causes the deviation between experiment and theory. Still, even in the larger models the percentage error in calculated chemical shifts of these nuclei as compared to experiment is larger than it is in the shifts of the nuclei in the middle of the residue. Note that H-bonding can cause substantial shifts in the $\delta(^1\text{H})$ and $\delta(^{15}\text{N})$ values. These H-bonds are often present in the peptides experimentally investigated, but not in our models. This could be one of the reasons for the larger computational error.

5. Conclusions

In this study isotropic NMR chemical shift (ICS) surfaces, computed at the TZ2P GIAO-B3LYPTZZP level, have been presented for the $^{13}\text{C}^{\alpha}$, $^1\text{H}^{\alpha}$, $^{13}\text{C}^{\beta}$, $^{13}\text{C}'$, $^1\text{H}^{\text{N}}$, and $^{15}\text{N}^{\text{H}}$ nuclei of the central Ala residue of the peptide models For-(Ala)_n-NH₂, $n = 1, 3, \text{ and } 5$.

In order to describe ICS as a function of geometric parameters, it is most efficient to construct an interpolating surface using the computed points. Several suitable mathematical functions have been investigated. Overall, a 10th-order cosine expansion turned out to be the best for fitting surfaces of such complex shape. As expected, it is necessary to involve cross-terms in the expansion. For all nuclei investigated, the 10th-order cosine expansion, with its 66 parameters, could efficiently describe the highly structured ICS surfaces with goodness of fit, r^2 , values of 0.97–0.99 for the $^{13}\text{C}^{\alpha}$, $^1\text{H}^{\alpha}$, $^{13}\text{C}^{\beta}$, $^1\text{H}^{\text{N}}$, and $^{15}\text{N}^{\text{H}}$ nuclei.

The effect of model size on chemical shifts and on the shape of the resulting surfaces was also investigated. Nuclei in the middle of a residue, namely C^{α} , H^{α} , and C^{β} , experience rather small changes upon increase of model size, proving that capping effects are not significant for these nuclei. The change in chemical shifts is relatively large for the H^{N} and N^{H} nuclei. For these nuclei enlargement of the model brings the computed values into closer agreement with their experimental counterparts. Nevertheless, it is shown that the shape of the ICS surfaces within a given region is not affected significantly by the enlargement of the peptide model investigated.

Acknowledgements

The authors wish to thank A. Czajlik for providing the experimental database. The research described was supported by a grant from the Hungarian Scientific Research Fund (OTKA T033074).

References

- [1] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.
- [2] J.N.S. Evans, *Biomolecular NMR Spectroscopy*, Oxford University Press, New York, 1995.
- [3] D.S. Wishart, B.D. Sykes, *Methods in Enzymology*, Academic Press, New York, 1992.
- [4] H. Saito, *Magn. Reson. Chem.* 24 (1986) 835.
- [5] C. Santiveri, M. Rico, M.A. Jimenez, *J. Biomol. NMR* 19 (2001) 331.
- [6] A.C. de Dios, J.G. Pearson, E. Oldfield, *Science* 260 (1993) 1491.
- [7] A.C. de Dios, J.G. Pearson, E. Oldfield, *J. Am. Chem. Soc.* 115 (1993) 9768.
- [8] D. Jiao, M. Barfield, V.J. Hruby, *J. Am. Chem. Soc.* 115 (1993) 10883.
- [9] H.B. Le, E. Oldfield, *J. Biomol. NMR* 4 (1994) 341.
- [10] A.C. de Dios, E. Oldfield, *J. Am. Chem. Soc.* 116 (1994) 5307.
- [11] L. Szilágyi, *Prog. Nucl. Magn. Res. Spectrosc.* 27 (1995) 325.
- [12] H.B. Le, J.G. Pearson, A.C. de Dios, E. Oldfield, *J. Am. Chem. Soc.* 117 (1995) 3800.
- [13] X.P. Xu, D.A. Case, *J. Biomol. NMR* 21 (2001) 321.
- [14] A.C. de Dios, *Prog. Nucl. Magn. Res. Spectrosc.* 29 (1996) 229.
- [15] IUPAC IUB Commission on Biochemical Nomenclature, *Biochemistry* 9 (1970) 3471.
- [16] A.G. Császár, A. Perczel, *Prog. Biophys. Mol. Biol.* 71 (1999) 243.
- [17] A. Perczel, A.G. Császár, *J. Comput. Chem.* 21 (2000) 882.
- [18] A. Perczel, A.G. Császár, *Chem.-Eur. J.* 7 (2001) 1069.
- [19] A. Perczel, A.G. Császár, *Eur. Phys. J. D* 20 (2002) 513.
- [20] E. Czinki, A.G. Császár, A. Perczel, *Chem.-Eur. J.* 9 (2003) 1182.
- [21] C.W. Swalina, R.J. Zauhar, M.J. DeGrazia, G. Moyna, *J. Biomol. NMR* 21 (2001) 49.
- [22] H.M. Sulzbach, G. Vacek, P.R. Schreiner, J.M. Galbraith, P.V.R. Schleyer, H.F. Schaefer III, *J. Comput. Chem.* 18 (1997) 126.
- [23] H.H. Sun, L.K. Sanders, E. Oldfield, *J. Am. Chem. Soc.* 124 (2002) 5486.
- [24] J. Gauss, *J. Chem. Phys.* 99 (1993) 3629.
- [25] J. Gauss, J.F. Stanton, *J. Chem. Phys.* 102 (1995) 251.
- [26] J. Gauss, J.F. Stanton, *J. Chem. Phys.* 104 (1996) 2574.
- [27] S. Spera, A. Bax, *J. Am. Chem. Soc.* 113 (1991) 5490.
- [28] J.N. Murrell, S. Carter, S.C. Farantos, P. Huxley, A.J.C. Varandas, *Molecular Potential Energy Surfaces*, Wiley, New York, 1984.
- [29] P.G. Mezey, *Potential Energy Hypersurfaces*, Elsevier, New York, 1987.
- [30] H. Partridge, D.W. Schwenke, *J. Chem. Phys.* 106 (1997) 4618.
- [31] A.G. Császár, W.D. Allen, Y. Yamaguchi, H.F. Schaefer III, in: P. Jensen, P.R. Bunker (Eds.), *Computational Molecular Spectroscopy*, Wiley, Chichester, 2000.
- [32] O.L. Polyansky, A.G. Császár, S.V. Shirin, N.F. Zobov, P. Barletta, J. Tennyson, D.W. Schwenke, P.J. Knowles, *Science* 299 (2003) 539.
- [33] M.J. Frisch, et al. *GAUSSIAN98*, Gaussian Inc, Pittsburgh, PA, 1998.
- [34] A.D. Becke, *Phys. Rev. A* 38 (1988) 3098.
- [35] C. Lee, W. Yang, R.G. Parr, *Phys. Rev. B* 37 (1988) 785.
- [36] R. Ditchfield, *Mol. Phys.* 27 (1974) 789.
- [37] K. Wolinski, J.F. Hinton, P. Pulay, *J. Am. Chem. Soc.* 111 (1990) 8251.
- [38] A. Schafer, C. Huber, R. Ahlrichs, *J. Chem. Phys.* 100 (1994) 5829.
- [39] A.G. Császár, W.D. Allen, H.F. Schaefer III, *J. Chem. Phys.* 108 (1998) 9751.
- [40] See www.python.org
- [41] See www.mysql.com
- [42] T.M. Wonnacott, R.J. Wonnacott, *Introductory Statistics*, fifth ed., Wiley, New York, 1990.
- [43] Re-referenced Protein Chemical Shift Database of Wishart's group, see <http://redpoll.pharmacy.ualberta.ca/RefDB/>
- [44] B.R. Seavey, E.A. Farr, W.M. Westler, J.L. Markley, *J. Biomol. NMR* 1 (1991) 217.
- [45] See www.bmrb.wisc.edu