NMR chemical shift data and ab initio shielding calculations

Mulder, Frans A. A.; Filatov, Michael

Published in:
Chemical Society Reviews

DOI:
10.1039/b811366c

Publication date:
2010

NMR chemical shift data and *ab initio* shielding calculations: emerging tools for protein structure determination

Frans A. A. Mulder[a] and Michael Filatov[b]

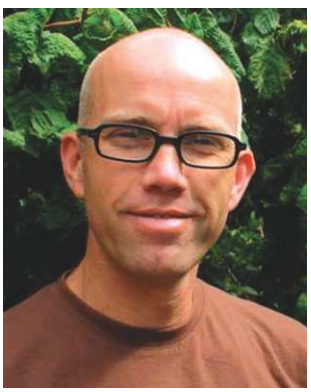
Received 29th June 2009
First published as an Advance Article on the web 4th November 2009
DOI: 10.1039/b811366c

In this *tutorial review*, we discuss the utilization of chemical shift information as well as *ab initio* calculations of nuclear shieldings for protein structure determination. Both the empirical and computational aspects of the chemical shift are reviewed and the role of molecular dynamics and the accuracy of different computational methods are discussed. It is anticipated that incorporating theoretical information on chemical shifts will increase the accuracy of protein structures, in the solid and liquid state alike, and extend the applicability of NMR spectroscopy to ever larger systems.

I. Introduction

Over the past decades nuclear magnetic resonance (NMR) spectroscopy has established itself as a powerful technique to determine the three-dimensional (3D) structures of biological macromolecules at atomic resolution, and a valuable complement to X-ray crystallography. Notably, it is the only method that can accurately define atomic structures in solution. To date (June 2009) over 5000 NMR structures of proteins (here taken to contain 50 or more residues) have been deposited at the Protein Data Bank (PDB), whereas the first deposited structures date back only to 1989. With few exceptions these structures have been determined in solution, similar to their natural milieu. The continued success of NMR spectroscopy is largely due to the constant development of NMR spectroscopy as a technique. Important advances include improvements in instrumentation hardware, pulse sequence development, access to residual anisotropic interactions, (bio)synthetic isotope enrichment schemes, and *in vivo*, in-cell detection. It is also evident that solid-state NMR spectroscopy is making headway, extending the applicability of NMR spectroscopy to study protein atomic structure in fibers, crystals, glasses and amorphous formulations. For the immediate future, the inclusion of chemical shift information promises to become a very powerful addition to structure determination, making use of semiempirical relationships as well as *ab initio* calculations. Recent progress in quantum chemical methods makes it possible to derive accurate relationships between calculated shielding parameters and protein 3D conformation and increases significantly the range of proteins that can be characterized structurally with NMR spectroscopy. Although

[a] *Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands*
[b] *Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands*



Frans A. A. Mulder

Frans A. A. Mulder (MSc and PhD cum laude from Utrecht University) was a post-doctoral fellow with Lewis E. Kay at the University of Toronto, and Mikael Akke at Lund University. In 2004 he moved to the University of Groningen, and in 2006 he was awarded a young investigator VIDI grant from the Netherlands Research Organization. In 2005, he spent a sabbatical with H. Vogel at the University of Calgary. Current research interests in his group include protein electrostatics, natively unfolded proteins, NMR pulse sequence development, protein molecular dynamics, and NMR structure determination of large proteins using bio-synthetic isotope labeling and computational methods.

Michael Filatov

M. Filatov (MSc University of Novosibirsk (Russia), PhD Institute of Catalysis (Novosibirsk, Russia)) was a postdoctoral fellow with Walter Thiel at Zürich University and with Sason Shaik at the Hebrew University of Jerusalem before working as an Assistant Professor at the University of Göteborg, Sweden from 2000 till 2005. Since 2005 he has been a Professor of Theoretical Chemistry at the University of Groningen (Netherlands). His research interests focus on the development of accurate quantum chemical computational schemes within the domain of density functional theory and on modeling ground and excited state electronic properties of organic, bio-inorganic and inorganic compounds.

chemical shieldings are very rich in information, owing to their tensorial character[1] we focus primarily on isotropic chemical shifts, as these are more relevant for comparison with experimental data.

II. Experimental determination of three-dimensional protein structures by NMR spectroscopy

All the parameters that can be measured by NMR spectroscopy are sensitive to molecular structure and dynamics and can be employed as restraints to construct models of the three-dimensional structures of proteins. The structures of small proteins (up to $\sim$100 amino acids) have traditionally been derived from two-dimensional, homonuclear ^{1}H NMR spectroscopy, and were primarily based on extensive lists of pairwise distances derived from nuclear Overhauser effects (NOEs), and from dihedral angle restraints derived from scalar spin–spin couplings (J).[2] The obvious advantage of this approach is that protein material can be isolated from a natural source, and, after purification, subjected to investigation. However, even a small protein of 100 residues will contain approximately 800 protons, hence the NMR spectra are highly congested. With increasing size, the number of signals that will appear in the same spectral region will become overwhelming, and this overlap problem becomes prohibitive for spectral assignment and structure determination. Multi-dimensional spectroscopy mitigates this problem somewhat, but many restraints remain ambiguous.

This problem can be overcome by making use of additional spins in proteins, such as ^{13}C and ^{15}N. However, because of the low natural abundance of the favorable spin-1/2 isotopes—1.1% for ^{13}C and 0.36% for ^{15}N—this approach only became feasible through the use of heterologous protein expression and isotope enrichment. The large spread in ^{13}C and ^{15}N chemical shifts, coupled with the use of [^{1}H–^{15}N] and [^{1}H–^{13}C] correlation spectroscopy, largely overcame the overlap problem. The widespread practice of uniform isotopic enrichment since the 1990s has pushed the molecular weight limit for routine structure determination towards larger systems (up to $\sim$250 amino acids), and provided additional probes such as ^{13}C–^{13}C and ^{1}H–^{13}C coupling constants, as well as access to protein dynamics from ^{15}N and ^{13}C spin relaxation.

Nonetheless, many proteins and complexes of biological interest are significantly larger than 30 kDa, and still remain out of reach. This is because the local magnetic fields from magnetic dipoles and anisotropic chemical shielding become increasingly effective in damping the NMR signal for larger molecules; the more rapid decrease of signal leads to broader lines, which reduces the resolution and sensitivity of the NMR experiment. A way around this issue has been to eliminate or dilute the abundant and strongly polarized proton spins by protein deuteration, effectively removing the strongest relaxation source. While a high degree of deuteration is beneficial to the spectral quality, the information for defining protein structure is now largely lost, as most aliphatic and aromatic protons have been removed. There is therefore clear need to use alternative sources of information on 3D protein structure. The most easily accessible alternative probes of molecular conformation in proteins are the chemical shifts of ^{13}C and ^{15}N nuclei. There are two major routes of using the information from chemical shifts: (i) an empirical route based on the observation that similar structures produce similar chemical shifts, and (ii) a first principles modeling of chemical shifts in specific protein secondary structures.

A. Empirical relationships between chemical shift and structure

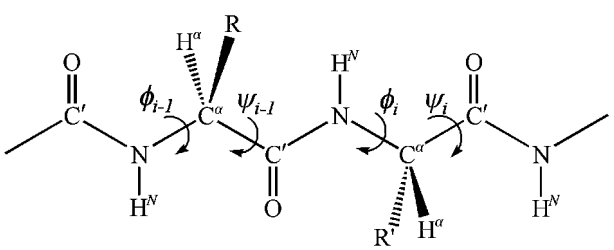
Chemical shifts are readily obtained for large proteins, and are also among the more easily accessible data for proteins in the solid state. Therefore, by using this information it should be possible to further expand the reach of NMR spectroscopy in structural biology. However, since many factors potentially contribute to protein chemical shifts it is important to establish their relative contributions, and assess their independence.[3] As an important first step, one needs to define a reference state from which all factors are to be counted. In practice it is common to define ''random coil'' chemical shifts for short peptides, which are assumed to be disordered, and devoid of persistent structure. Random coil chemical shifts rely on the specific identity of the side chain, and conformational sampling, and have proven very valuable when investigating the effects of amino acid sequence on chemical shifts.[4] In a second, alternative, approach the chemical shifts of a peptide fragment are obtained from so-called ''coil libraries''.[5] In this case, proteins of known structure are used to filter chemical shift data belonging to regions of regular secondary structure (α-helices, β-sheets and turns) from those of ''coil''. The resultant library is assumed to represent a generic reference state for disordered peptides, and the formation of structure will induce chemical shift changes through, for example, changes in backbone dihedral angles, hydrogen bonds and ring current shifts of nearby aromatic residues. In a third approach, a random coil chemical shift data set can be determined under strongly denaturing conditions. In an extensive investigation of this kind, Schwarzinger *et al.*[6] investigated the effects of neighboring amino acids in the peptides Ac-GGXGG-NH2, where X equals any one of the twenty naturally occurring amino acid types, and used this to compare with the protein *apo* myoglobin, under identical environmental circumstances.

^{1}H chemical shifts. The sensitivity of proton chemical shifts to structural and conformational effects in proteins have long been known, but it is still difficult to accurately decipher the multiple contributing factors. These primarily include sensitivity to conformation, hydrogen bonding, electric fields, and ring currents. Their relative contributions also vary for different chemical groups. For example, several studies have demonstrated that clear correlations exist between the $^{1}H^{\alpha}$ chemical shift and secondary structure, with upfield shift in α-helical conformations of about -0.3 ppm, and downfield shifts for β-sheets of approximately 0.5 ppm.[5,7] A relationship between amide proton shifts and backbone conformation also exists, contributing to the observed downfield shifts for β-sheets, but having an almost negligible influence on the shifts for α-helices. In addition, the distributions of the

secondary chemical shifts for amide protons are larger than those observed for $^{1}H^{\alpha}$,[8] indicating that their correlation with secondary structure is weaker. A second contribution arises from hydrogen bonds, and most strongly influences $^{1}H^{N}$ chemical shifts. Hydrogen bonds often are responsible for the most strongly shifted signals, and display a sensitive dependence on hydrogen bond acceptor–donor distance. A third, and sometimes large, contribution arises from the magnetic susceptibility anisotropies due to aromatic groups, and, to a lesser extent, peptide groups. The ring currents of nearby aromatic groups most significantly affect $^{1}H^{\alpha}$ chemical shifts, thereby blurring their relationship with conformation. Taking these effects into account prior to chemical shift analyses further explicates the structural correlation of ^{1}H chemical shifts. However, calculating ring current contributions is challenging, as they are very sensitive to the precise structure, which may not be available with sufficient accuracy, and subject to dynamics.

^{13}C chemical shifts. Backbone $^{13}C^{\alpha}$ and side chain $^{13}C^{\beta}$ carbon chemical shifts are first of all sensitive to amino acid side chain identity, due to the presence of electron-withdrawing substituents or charges. When these contributions are accounted for by subtracting random coil values for these amino acids, the remaining secondary chemical shifts, $\Delta\delta(^{13}C)$, display a high sensitivity to backbone ϕ/ψ torsion angles (see Scheme 1). Excluding glycines, $^{13}C^{\alpha}$ chemical shifts are shifted downfield by 2.2 (Ala) to 4.5 ppm (Thr) in helices, and upfield by −0.4 (Thr) to −1.8 (Arg) ppm in β-sheets.[5] Since the standard deviations for the secondary chemical shifts are only about 1.3 ppm this explains their power for the detection of secondary structure. $^{13}C^{\beta}$ secondary chemical shifts are somewhat smaller, and of opposite sign, such that the secondary chemical shift difference $\Delta_{\alpha\beta} = \Delta\delta(^{13}C^{\alpha}) - \Delta\delta(^{13}C^{\beta})$ is an even better predictor. Finally, backbone carbonyl $^{13}C'$ chemical shifts are sensitive to the amino acid side chains on both sides of the peptide plane, but the correction factors are small, with the exception of Pro (up to −2.5 ppm, depending on conformation) and aromatic amino acids (up to −0.5 ppm) at position +1 in the sequence. The $^{13}C'$ chemical shift is also quite sensitive to secondary structure, showing sizable (3 ppm) differences between α-helix and β-sheet.

A further contribution may arise from conformational preferences of substituents that are three bonds apart: a carbon nucleus in the *gauche* position about a subtending dihedral angle experiences increased shielding by up to 5 ppm, relative to the same group in the *trans* position. Depending on the probe nucleus considered, these gamma-*gauche* effects depend on backbone, and/or side chain rotameric states.



Scheme 1 Backbone dihedral angles in peptide chain.

Although these conformational effects have hitherto been largely neglected, two recent studies demonstrate that the effects can be sizable,[9,10] and may in part explain the remaining discrepancy between observed and calculated chemical shifts.[3] Of note, in contrast to the other backbone shifts, $^{13}C^{\alpha}$ shieldings are not sensitive to this effect, as the intervening dihedral angle is the near-planar peptide bond, which is always *trans* with the rare exception of *cis* Xxx-Pro (where Xxx denotes any amino acid).

Since carbon chemical shifts are so dominantly dependent on backbone geometry, they are the most accurate predictors of secondary structure.[3] Based on this premise, algorithms have been developed to score the presence of secondary structure from chemical shift data alone. The most well-known program is likely the Chemical Shift Index (CSI), due to Wishart and Sykes,[11] which originally was based on scoring $^{1}H^{\alpha}$ secondary chemical shifts, and later included $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$. This approach was subsequently further refined. For example, Wang and Jardetzky[8] used a joint probability calculation, including $^{1}H^{N}$, $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, and ^{15}N chemical shifts. They found that the reliability to discern α-helix from coil decreased in the order $^{13}C^{\alpha} > {}^{13}C' > {}^{1}H^{\alpha} > {}^{13}C^{\beta} > {}^{15}N > {}^{1}H^{N}$, whereas this order is $^{1}H^{\alpha} > {}^{13}C^{\beta} > {}^{1}H^{N} \sim {}^{15}N \sim {}^{13}C^{\alpha} \sim {}^{13}C'$ to detect β-sheet structures. These results indicate that $^{1}H^{N}$ and ^{15}N chemical shifts can also be useful indicators. Further improvement can be obtained by prior correction for neighboring effects on the random coil chemical shifts.[5–7]

^{15}N chemical shifts. Peptide amide nitrogen chemical shifts display a large variation, covering about 30 ppm. For example, the random coil chemical shifts of Gly are around 110 ppm, and well-separated from these of other residue types. ^{15}N chemical shifts for individual residue types have standard deviations of approximately 4 ppm. Much of the variation is due to the large influence of the preceding amino acid side chain; the ^{15}N chemical shift will be about 4 ppm higher for an amino acid preceded by the β-branched residues Val, Thr or Ile as compared with Ala or Gly. After taking these neighboring effects into account there remains less than one ppm variation unexplained for flexible peptides, such that the remaining variations in folded proteins must be explained by other contributions. Besides occasional ring current effects other significant variations result from hydrogen bonding, deviations from peptide bond planarity, and the presence of nearby charges.

Semiempirical chemical shift calculations. We consider here two recent, accurate programs to calculate chemical shifts from known 3D structures, which are freely distributed. These are SHIFTX by Neal, Wishart and co-workers,[12] and SPARTA by Shen and Bax.[7] The performance of these programs is summarized in Table 1.

The program SHIFTX was developed to calculate ^{1}H, ^{13}C and ^{15}N chemical shifts from atomic coordinates. It makes use of chemical shift hypersurfaces, derived from chemical shift and structural data, combined with semi-classical equations that relate the chemical shifts to ring currents, hydrogen bonding, solvent effects and electric field contributions. Many

Table 1 Correlation coefficient (r) between predicted and experimental chemical shifts and rms error for the semi-empirical programs SHIFTX[12] and SPARTA[7]

	$^1H^{\alpha}$	$^1H^N$	^{15}N	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C'$
SHIFTX[a]						
Correlation (r)	0.91	0.74	0.91	0.98	1.00	0.86
Rms error (ppm)	0.23	0.49	2.43	0.98	1.10	1.16
SPARTA						
Correlation (r)	0.88	0.72	0.89	0.98	0.99	0.86
Rms error (ppm)	0.27	0.51	2.52	0.98	1.07	1.09

[a] An analysis on the same set of proteins by both procedures performed by Shen and Bax[7] indicates that the SHIFTX performance on that data set deviates slightly from the numbers in the table. In that comparison, the rms errors using SHIFTX were 0.29, 0.54, 2.87, 1.12, 1.25 and 1.28 ppm for $^1H^{\alpha}$, $^1H^N$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$, respectively.

of the necessary parametrizations used in the calculation were derived from the extensive body of literature on these effects, in addition to a large body of three-dimensional protein structures and chemical shift data. The program is able to obtain very good correlations between calculated and observed chemical shifts.

The program SPARTA uses a database searching method, which utilizes both structural homology and protein sequence to predict the $^1H^N$, $^1H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, and ^{15}N chemical shifts, considering backbone ϕ/ψ torsion angles, side chain χ_1, hydrogen bonding, and ring current shifts. The predictive power of the method was optimized by iterative adjustment of weighting factors for the various contributions to yield the closest agreement between prediction and experiment. As can be seen from Table 1, the improvements are small, but as Shen and Bax argue,[7] in molecular fragment replacement methods the search for matching peptide fragments on the basis of multiple chemical shifts strongly reduces the search through the combined effect of a large number of independent, small improvements. To demonstrate the facility of the method, a plot of the correlations of predicted *versus* observed chemical shifts is shown in Fig. 1. Similar results were obtained using SHIFTX.[12]

B. Using chemical shifts to aid protein structure determination

With the advent of increased understanding of the various contributions to the chemical shift and as their quantitative prediction emerges, there is hope that they will be intelligible enough to enable *de novo* structure prediction. One early approach in this direction is the TALOS program,[13] which is based on the notion that similar structures are expected to yield similar chemical shifts. The TALOS database searching program contains extensive chemical shift assignments and high-resolution structures for several proteins, and can be used to search for contiguous segments of three amino acid residues which most closely agree with a segment of known structure. The output of the program can subsequently be used to restrain the backbone ϕ/ψ torsion angles in the process of structure calculation.

Recently two approaches were developed for 3D protein structure determination from chemical shift data alone that

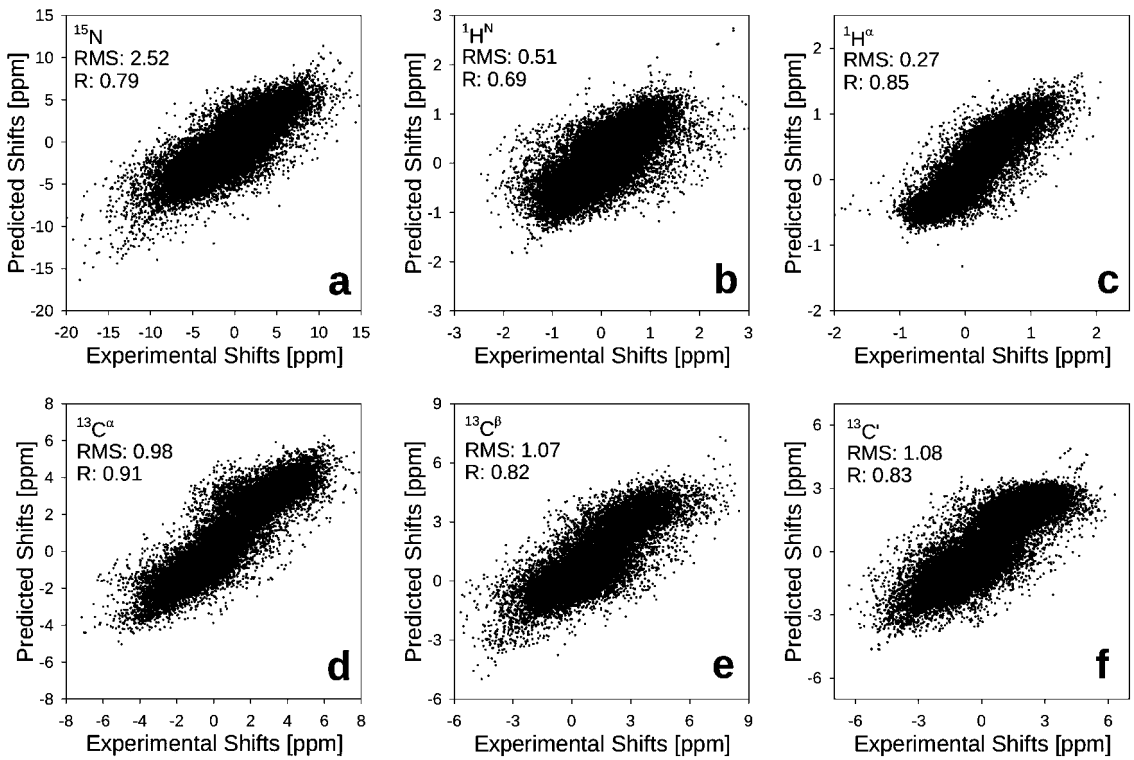


Fig. 1 Scatter plots comparing experimental and SPARTA-predicted secondary chemical shifts for backbone ^{15}N, $^1H^N$, $^1H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$ nuclei. The RMS deviations (in ppm) and Pearson correlation coefficients (R) between experimental and SPARTA-predicted shifts are indicated. Reprinted with kind permission from Springer Science + Business Media: *J. Biomol. NMR*,[7] copyright (2007).

agree rather closely with the target structures.[14,15] Here, the TALOS strategy for finding fragments is extended to contiguous segments, which are subsequently combined to build a complete 3D protein model. In the CHESHIRE algorithm[15] secondary chemical shift information is first used to determine secondary structure and backbone ϕ/ψ torsion angles, from which low-resolution structures without side chains beyond the C^{β} atom are assembled through simulated annealing. In the next step all atoms are represented explicitly, and refined by allowing movement about backbone and side chain torsion angles in an energy refinement protocol, augmented with information about side chain rotamer statistics. Using this approach, 11 proteins of 46–123 residues were determined to 2 Å or better rmsd for the backbone, whereas the agreement for side chains was poorer.

The program CS-ROSETTA was assembled by Shen, Bax, Baker and co-workers[14] using similar concepts. The ROSETTA machinery for structure prediction has been supplemented with the empirical relationships derived from SPARTA.[7] For 16 proteins, ranging in size from 56 to 129 residues, full-atom models were obtained that are within 0.7–1.8 Å backbone rmsd of the experimentally determined X-ray or NMR structures. Here, as with the CHESHIRE protocol, the current lack of accurate relationships between the chemical shifts of side chain nuclei and amino acid conformation in the models results in limited agreement for the side chains. Since quantum chemical calculations can yield this information, they can provide the missing insight into the relationship between the observed shifts and protein structure. Recent methodological advances enable one to accurately predict chemical shifts in large protein fragments from first principles, with the inclusion of important environmental and conformational factors that influence the chemical shifts.[16]

III. Theoretical methods of NMR chemical shift calculation

The resonance frequency $\nu_A = \gamma_A B_{0,z}/2\pi$ of a given magnetic nucleus A in external magnetic field $\mathbf{B}_0$ is determined by its magnetic moment $\vec{\mu}_A = \gamma_A \mathbf{I}_A$, which is related to the nuclear spin $\mathbf{I}_A$ and the magnetogyric ratio γ_A. Here and below we define that the magnetic field $\mathbf{B}_0$ as aligned with the z-axis and employ the atomic system of units in which the Planck constant $\hbar = h/2\pi$, Bohr radius a_0, electron charge e and electron mass m_e are set to unity. For practically affordable magnitudes of the magnetic field $\mathbf{B}_0$, the frequency ν_A lies within the radio frequency domain. Because different magnetic nuclei have different magnetogyric ratios (*e.g.*, $\gamma_{^1H} = 267.513 \times 10^6$ rad s^{-1} T^{-1}, $\gamma_{^{13}C} = 67.261 \times 10^6$ rad s^{-1} T^{-1}, *etc.*), they resonate at different frequencies.

This simple picture however is valid for a bare nucleus and, in atoms or molecules, it should be corrected for the presence of atomic or molecular electrons. Application of an external magnetic field induces electronic currents which, in turn, generate an additional local magnetic field. In a spherically symmetric atom, these currents are purely diamagnetic in origin and the resulting local magnetic field opposes the external field thus leading to shielding of the atomic nucleus (lowering of the resonance frequency). Because the induced diamagnetic current is proportional to the magnitude of the applied external field, this results in a local magnetic field $\mathbf{B}_{\mathrm{loc}}$ at the nucleus A given by

$$\mathbf{B}_{\mathrm{loc}} = (1 - \sigma^{d,A})\mathbf{B}_0 \tag{1}$$

where $\sigma^{d,A}$ is a small (usually in the range of 10^{-6}) diamagnetic shielding constant.

In molecules, the circulation of electronic currents around the target nucleus is hindered due to the presence of other nuclei and electrons revolving about them and this deviation from the spherical symmetry leads to emergence of an additional contribution to the total nuclear shielding,

$$\sigma^{\mathrm{tot},A} = \sigma^{d,A} + \sigma^{p,A} \tag{2}$$

which is known as the paramagnetic term, $\sigma^{p,A}$.[17] Because the paramagnetic contribution opposes the diamagnetic shielding (in the cases of interest in this review), this results in deshielding (an increase in the resonance frequency) of the target nucleus as compared to an isolated atom. The paramagnetic term involves the mixing between ground and excited states of the molecule due to the magnetic field, and it is rather sensitive to the molecular electronic structure.

In practice, nuclear shielding constants are not measured relative to a bare nucleus, as eqn (1) implies, but with respect to the target nucleus embedded in a reference compound.[18] Thus, the chemical shift δ^A is now defined as

$$\begin{aligned}\delta^A &= 10^6 \frac{\nu_A - \nu_{\mathrm{ref}}}{\nu_{\mathrm{ref}}} = 10^6 \frac{\gamma_A B_{0,z}(\sigma^{\mathrm{ref}} - \sigma^A)}{\gamma_A B_{0,z}(1 - \sigma^{\mathrm{ref}})} \\ &\approx 10^6(\sigma^{\mathrm{ref}} - \sigma^A)\end{aligned} \tag{3}$$

where it is used that the magnitude of nuclear shielding is of the order of 10^{-6}–10^{-4} for nuclei of light elements such as ^{1}H or ^{13}C. The so-defined chemical shift is expressed in parts per million (ppm) units.

The values of absolute shielding constants eqn (2) can be obtained from the measured chemical shifts provided that $\sigma^{\mathrm{tot},A}$ is known for at least one compound containing the target nucleus. The paramagnetic contribution $\sigma^{p,A}$ is related to the so-called spin-rotation constants which can be independently measured from molecular beam experiments.[19] The so-obtained $\sigma^{p,A}$ values are combined with the diamagnetic terms $\sigma^{d,A}$ obtained from *ab initio* calculations. This technique enables one to convert the entire set of experimental chemical shifts for a given nucleus into the absolute scale of shieldings. Currently, the experimental values of absolute shieldings are known with high accuracy for many nuclei, *e.g.* for ^{1}H and ^{13}C.[19]

A. Theory of chemical shielding tensor: diamagnetic and paramagnetic contributions

A detailed theory of chemical shielding in molecules can be derived from the energy expression for a molecule containing a magnetic nucleus placed in a magnetic field.[17] When a magnetic field of strength $\mathbf{B}_0$ is applied to a closed-shell molecule containing a magnetic nucleus A with magnetic moment $\vec{\mu}_A$ its ground state energy changes due to the interaction of the induced electronic currents with the nuclear magnetic moment and with the applied field. For weak static

perturbations, such as $\mathbf{B}_0$ and $\vec{\mu}_A$, the molecular electronic energy $E(\vec{\mu}_A, \mathbf{B}_0)$ can be expanded in a Taylor series around the unperturbed energy value. In this expansion, the term bilinear in $\mathbf{B}_0$ and $\vec{\mu}_A$ will be exactly equal to the change in the nuclear Zeeman energy $-\vec{\mu}_A \cdot (1 - \sigma^A)\mathbf{B}_0$ resulting from interaction with the surrounding electrons. Therefore the Taylor coefficient given by a second derivative of the molecular electronic energy $E(\vec{\mu}_A, \mathbf{B}_0)$ with respect to the field strength and the nuclear magnetic moment can be identified with the nuclear shielding tensor.

$$\sigma^A_{\alpha\beta} = \left.\frac{\partial^2 E(\vec{\mu}_A, \mathbf{B}_0)}{\partial \mu_\alpha \partial B_\beta}\right|_{\substack{\vec{\mu}_n \to 0 \\ \mathbf{B}_0 \to 0}} \alpha, \beta = x, y, z \qquad (4)$$

From eqn (4), the isotropic shielding constant σ^A_{iso} and the anisotropy parameter $\Delta\sigma^A$ can be obtained as[18]

$$\sigma^A_{\mathrm{iso}} = \tfrac{1}{3}(\sigma^A_{xx} + \sigma^A_{yy} + \sigma^A_{zz}) = \tfrac{1}{3}(\sigma^A_{11} + \sigma^A_{22} + \sigma^A_{33}) \qquad (5)$$

and

$$\Delta\sigma^A = \sigma_{33} - \tfrac{1}{2}(\sigma_{11} + \sigma_{22}). \qquad (6)$$

In eqn (5) and (6), $\sigma_{\alpha\beta}$ ($\alpha,\beta = x,y,z$) are the Cartesian components of the shielding tensor (a non-symmetric tensor of rank 2) and σ_{ij} ($i,j = 1,2,3$) are the components of the shielding tensor in the principal axes system, a coordinate system in which the symmetric part of the tensor is diagonal.

The magnetic field is incorporated into the quantum mechanical equations for electrons *via* the minimal coupling prescription within which the canonical momentum $\hat{\mathbf{p}}$ is replaced with the mechanical momentum $\hat{\mathbf{p}} + \mathbf{A}/c$, where $c \approx 137$ a.u. is the velocity of light. The vector potential $\mathbf{A}$ is due to both the external magnetic field $\mathbf{B}_0$ and the field of the magnetic nucleus A with the moment $\vec{\mu}_A$

$$\mathbf{A} = \mathbf{A}_0 + \mathbf{A}_{\mu_A} = \frac{1}{2}\mathbf{B}_0 \times (\mathbf{r}_{iA} - \mathbf{R}_0) + \frac{\vec{\mu}_A \times \mathbf{r}_{iA}}{r^3_{iA}}. \qquad (7)$$

Using mechanical momentum with eqn (7) in the molecular Schrödinger equation, Ramsey[17] derived the following expressions for the components of the shielding tensor:

$$\sigma^{d,A}_{\alpha\beta} = \frac{1}{2c^2}\langle\psi_0| \sum_i [\hat{\mathbf{e}}_\beta \times (\mathbf{r}_{iA} - \mathbf{R}_0)] \cdot [\hat{\mathbf{e}}_\alpha \times \mathbf{r}_{iA}]/r^{-3}_{iA}|\psi_0\rangle \qquad (8)$$

for the diamagnetic component and

$$\sigma^{p,A}_{\alpha\beta} = -\frac{1}{c^2}\sum_k (E_k - E_0)^{-1}\left[\hat{\mathbf{e}}_\alpha \cdot \langle\psi_0| \sum_i \hat{\mathbf{l}}_i/r^{-3}_{iA}|\psi_k\rangle\right] \times \left[\hat{\mathbf{e}}_\beta \cdot \langle\psi_k| \sum_i \hat{\mathbf{l}}'_i|\psi_0\rangle\right] \qquad (9)$$

for the paramagnetic component. In eqn (7)–(9), $\mathbf{R}_0$ is an arbitrary gauge origin (origin of the vector potential $\mathbf{A}_0$), $\hat{\mathbf{e}}_\alpha$ and $\hat{\mathbf{e}}_\beta$ are unit vectors along the respective Cartesian axes, $\mathbf{r}_{iA}$ is the position of electron i with respect to the nucleus A, $\hat{\mathbf{l}}_i = \mathbf{r}_{iA} \times \hat{\mathbf{p}}_i$ is its angular momentum relative to the nucleus A, and $\hat{\mathbf{l}}'_i = (\mathbf{r}_{iA} \times \mathbf{R}_0) \times \hat{\mathbf{p}}_i$ is its angular momentum with respect to the gauge origin $\mathbf{R}_0$.

In practical applications, eqn (8) and (9) are often replaced with a computationally more convenient expression for the shielding tensor[20]

$$\sigma^A_{\alpha\beta} = \sum_{\mu,\nu} D_{\mu\nu}\frac{\partial^2 h_{\mu\nu}}{\partial(\mathbf{B}_0)_\beta \partial(\vec{\mu}_A)_\alpha} + \sum_{\mu,\nu}\frac{\partial D_{\mu\nu}}{\partial(\mathbf{B}_0)_\beta}\frac{\partial h_{\mu\nu}}{\partial(\vec{\mu}_A)_\alpha} \qquad (10)$$

which is derived from eqn (4) with the help of the interchange theorem of perturbation theory. In eqn (10), $h_{\mu\nu}$ are the elements of the one-electron Hamiltonian matrix and $D_{\mu\nu}$ are the elements of the density matrix in the atomic orbitals (AO) representation. The first term in eqn (10) represents the diamagnetic part of the shielding tensor and the second term represents the paramagnetic part.

For large molecules, σ^d and σ^p are rather large in absolute magnitude and are of opposite sign thus nearly compensating each other in the total shielding constant. This requires the use of accurate computational approaches in connection with eqn (8), (9) or (10). However, simple qualitative arguments can be used to analyse different contributions to σ^d and σ^p. Thus, it is obvious from eqn (8) and (9) that non-vanishing contributions are made by the products of molecular orbitals (occupied–occupied orbital pairs for σ^d and occupied–virtual orbital pairs for σ^p) which comprise a rotation. Using these arguments a contribution of diamagnetic ring currents (that is nearly free circulation of electrons in the occupied π-orbitals) in aromatic π-systems to *e.g.*σ^d of ^{1}H can be identified.[21] Analogously, σ–π* and π–σ* excitations should play an important role for *e.g.*σ^p of ^{13}C in double bonds. Although partitioning of the overall shielding into contributions from ring currents, specific bonds, substituent groups, *etc*. can be useful for analysis of trends and for a rapid evaluation of chemical shifts (for example, they have been implemented in the SHIFTX program[12]), it should be realised that such a partitioning is somewhat artificial, because, in large molecules, there is no unambiguous way of separating different parts (for example, σ–π separation), and can not serve as a basis for accurate calculation of nuclear shieldings.[21]

B. Gauge invariance

In eqn (8) and (9), the operator kernels depend on the origin $\mathbf{R}_0$ of the vector potential $\mathbf{A}_0$ of the external uniform magnetic field $\mathbf{B}_0$, see eqn (7). Because the external field $\mathbf{B}_0(\mathbf{B}_0 = \vec{\nabla} \times \mathbf{A}_0)$ is independent of the gauge origin $\mathbf{R}_0$, so should be the shielding tensor. Although this property is not at all obvious from expressions (8) and (9), with the use of the hypervirial theorem $\langle\psi_k \sum_i \vec{\nabla}_i\psi_0\rangle = (E_0 - E_k)\langle\psi_k| \sum_i \mathbf{r}_i|\psi_0\rangle$ and the completeness relation $\sum_k |\psi_k\rangle\langle\psi_k| = 1$, the vector algebra relationships can demonstrate that the total nuclear shielding tensor is indeed independent of the choice of $\mathbf{R}_0$, that is

$$\sigma^A_{\alpha\beta}(\mathbf{R}_0) - \sigma^A_{\alpha\beta}(0) = 0 \qquad (11)$$

for any $\mathbf{R}_0$.[20] The fulfilment of eqn (11) is guaranteed when the completeness relation is satisfied and the unperturbed wave functions ψ_k are obtained variationally. Although the latter condition is easily satisfied in practice, the former condition requires the use of a complete basis set (CBS) which is not practically attainable. Therefore, in practical applications of

eqn (8), (9), or (10), there remains a dependence on the choice of the gauge origin $\mathbf{R}_0$ unless special care is taken to remove it.

C. Practical methods of NMR shielding tensor calculation

Perhaps the simplest starting point for the shielding calculation is in the use of the Hartree–Fock approximation, within which the ground state wave function is given by a single Slater determinant $\psi_0 = (N!)^{-1/2}\|\phi_1\bar{\phi}_1 \ldots \phi_j\bar{\phi}_j \ldots \phi_{N/2}\bar{\phi}_{N/2}\|$, where ϕ_j are the occupied molecular orbitals (MO). Commonly the MOs are represented as linear combinations of the atomic orbitals, the MO-LCAO approximation

$$\phi_j = \sum_{\mu} C_{\mu j}\chi_{\mu} \tag{12}$$

and the density matrix of a closed shell molecule is given by

$$D_{\mu\nu} = 2\sum_{i=1}^{N/2} C^*_{\mu i}C_{\nu i}. \tag{13}$$

The derivative of the density matrix with respect to the applied magnetic field $\partial D_{\mu\nu}/\partial(\mathbf{B}_0)_{\beta}$ can be obtained from the solution of the coupled-perturbed Hartree–Fock (CPHF) equations in which perturbation due to the external magnetic field (first term on the right hand side of eqn (7)) is included into the Fock operator.[22]

Common gauge origin methods. With the use of a common gauge origin $\mathbf{R}_0$ the perturbed Fock operator becomes gauge origin dependent and this dependence is transferred into the calculated nuclear shieldings unless a complete basis set is employed to expand the MOs (12).[22] In the theory with common gauge origin, both σ^d and σ^p are rather large and nearly cancel each other in a relatively small value σ^{tot}. Because σ^d is a ground state property and can be calculated with relatively high accuracy, even small inconsistencies in the description of the paramagnetic contribution, *e.g.* due to the use of insufficiently flexible basis sets, may translate to large errors in the total shielding. Although the shieldings become gauge origin independent in the CBS limit, the convergence with respect to basis set size is rather slow and requires the use of very large basis sets which severely restricts the applicability of such an approach.

Gauge-including atomic orbitals (GIAO) method. An approach to the calculation of shielding tensors in which the dependence on the gauge origin is removed is based on the use of so-called gauge-including atomic orbitals (GIAO). The first *ab initio* shielding calculations employing GIAOs were carried out by Ditchfield[20] and the developed computational scheme was revised by Pulay *et al.*[23] who formulated it in a computationally convenient form.

In the GIAO approach, complex basis functions which depend on the external magnetic field $\mathbf{B}_0$

$$\chi^{\mathrm{GIAO}}_{\mu}(\mathbf{B}_0) = e^{-i\mathbf{B}_0\times(\mathbf{R}_{\mu}-\mathbf{R}_0)\cdot\mathbf{r}/2c}\chi_{\mu} \tag{14}$$

are employed in eqn (12). In eqn (14), χ_{μ} is a field-independent real atomic orbital centered at the position $\mathbf{R}_{\mu}$. The property which makes the new functions $\chi^{\mathrm{GIAO}}_{\mu}$ especially convenient in shielding calculations is that, under the action of the mechanical momentum operator, the gauge origin $\mathbf{R}_0$ is shifted to $\mathbf{R}_{\mu}$, a point at which the orbital χ_{μ} is centered. Hence the use of GIAOs in the calculation of matrix elements of the molecular Hamiltonian leads to cancellation of the dependence on the gauge origin $\mathbf{R}_0$.[20]

In the MO-LCAO approximation using GIAOs, the coefficients $C_{\mu j}$ in eqn (12) become dependent on external magnetic field. Because the basis functions and the expansion coefficients become complex in the presence of an external magnetic field, the resulting GIAO-HF equations become more complicated than in the case of the common origin method.[20,23] An additional set of overlap integrals and two-electron integrals need to be evaluated within the GIAO-HF method, which makes this approach more time consuming. However, this is compensated by a much faster convergence of the calculated shielding parameters with respect to the basis set expansion.[23] In practical calculations with the GIAO approach, it is possible to obtain reasonably converged theoretical values of shielding parameters with the use of basis sets of modest size.

Local gauge origin methods: IGLO and LORG. A number of computational schemes were developed which avoid the necessity to compute a large number of additional two-electron integrals employed in the GIAO formalism. Probably the most economic gauge-independent approach to the calculation of nuclear shielding parameters is realized in the individual gauge for localized orbitals (IGLO) method of Schindler and Kutzelnigg.[24] In the IGLO method, the phase factors similar to (14) are attached to the molecular orbitals rather than to the AOs as in the GIAO approach. The new molecular orbitals φ_j are introduced *via*

$$\phi_j = e^{-i\mathbf{B}_0\times(\mathbf{R}_j-\mathbf{R}_0)\cdot\mathbf{r}/2}\varphi_j \tag{15}$$

where $\mathbf{R}_j$ is the position vector of the individual gauge origin of the orbital ϕ_j (usually chosen as the centroid of the charge of the orbital). Similar to the GIAO approach, this choice of MOs leads to cancellation of the dependence of the matrix elements in eqn (10) on the choice of the gauge origin $\mathbf{R}_0$.[24] Because the canonical molecular orbitals are delocalized over the entire molecule, it is more convenient to work with localized MOs. The use of localized (*e.g.* by the Foster–Boyd criterion) MOs offers a possibility to analyze the chemical shielding parameters in terms of individual contributions from atomic core electrons, chemical bonds, and lone pairs thus allowing for a transparent interpretation of the shielding parameters.[24]

Yet another theoretical method which yields gauge-origin independent shielding constants was developed by Hansen and Bouman[25] based on the Hartree–Fock wave function and is referred to as the localized orbital/local origin (LORG) method. In this method, individual gauge origins $\mathbf{R}_j$ are introduced for the occupied orbitals and the diamagnetic (8) and paramagnetic (9) contributions are split into two parts of which only one is dependent on the global gauge origin $\mathbf{R}_0$.[25] The dependence on $\mathbf{R}_0$ is then eliminated under the assumption of a complete basis set. Similar to the IGLO method, the local gauge origins $\mathbf{R}_j$ are selected as the charge centroids of the localized molecular orbitals.[25] With this choice of MOs, the LORG shieldings can be decomposed into the individual bond contributions. With the use of the complete basis set, the shielding tensors obtained with both methods, LORG and

IGLO, should converge to the values obtained with a common origin CPHF approach.[25]

Comparison between the GIAO, IGLO, and LORG approaches. Although the methods based on localized MOs, such as IGLO and LORG, are computationally less demanding than the GIAO approach, it is the latter approach that is currently more popular in practical calculations of NMR shielding parameters. More than 1400 scientific articles published since 1980 refer to GIAO as the method of shielding calculation, whereas 411 articles refer to IGLO and only 60 to LORG. Probably the turning point was the work of Pulay *et al.*[23] who achieved a very efficient implementation of the GIAO-HF method for NMR chemical shieldings with the use of efficient techniques for the two-electron integrals calculation in combination with fast algorithms for solving the CPHF equations. Although the IGLO method still remains computationally the fastest method for the calculation of chemical shieldings (it can be faster than GIAO[23,24] by a factor close to 2), the GIAO method shows much lower sensitivity with respect to the choice of basis set.[23] Besides being much less sensitive to basis set choice, the GIAO approach represents a very convenient starting point for the development of computational schemes which include electron correlation effects on calculated shieldings.

Electron correlation methods for nuclear shieldings. The first practical applications of the computational schemes described above have been done at the Hartree–Fock level of approximation in which it is assumed that any particular electron in the molecule moves in the average field of all other electrons. Within this approximation the correlation in the motion of electrons is lost, and, as a consequence, the Hartree–Fock method often yields rather poor results for many molecular properties, such as the molecular atomization energies, vibrational frequencies, dipole moments, *etc*. An accurate description of the nuclear shielding parameters requires inclusion of the electron correlation. This can be achieved, for example, *via* the use of many-body perturbation theory, either in the form of a finite perturbational expansion, as in the Møller–Plesset (MP) perturbation theory, or in the form of the coupled cluster (CC) approximation, where partial summation of the perturbation series is carried out up to infinite order.[26]

The importance of electron correlation for the nuclear shieldings was established already in early calculations which employed the common origin approach based on eqn (8) and (9). Electron correlation has the greatest effect on the paramagnetic contribution for which the Hartree–Fock approximation often underestimates the stability of the ground state relative to the excited states (see ref. 26 and references cited therein). Thus, the correlation may account for up to *ca*. 30% reduction in the absolute magnitude of the paramagnetic contribution in the compounds of light elements, especially those with lone electron pairs and multiple bonds.

Inclusion of the electron correlation into the local gauge origin methods, LORG or IGLO, is hampered by the necessity to use localized molecular orbitals.[26] Since most electron correlation theories are based on canonical Hartree–Fock orbitals, it appears most convenient to formulate the correlated theory of nuclear shieldings in terms of these

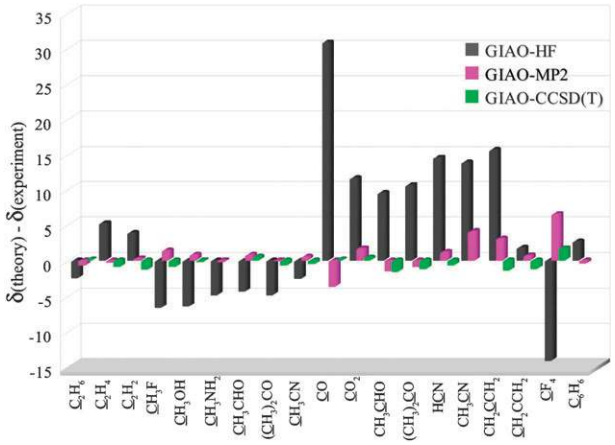


Fig. 2 Deviations from experiment of ^{13}C chemical shifts (with respect to CH_4) calculated with different *ab initio* theoretical methods. Data taken from ref. 29.

delocalized orbitals. Most straightforwardly this can be done with the use of the GIAO method, because it does not depend on a specific representation of MOs. Using the derivative approach to molecular properties, eqn (10), Gauss[27] has extended the GIAO-HF method of Pulay to second-order Møller-Plesset perturbation theory, GIAO-MP2. Although the GIAO-MP2 method is more time-consuming than GIAO-HF, its use leads to a marked improvement in the calculated nuclear shieldings in comparison with experiment.[27] This is illustrated in Fig. 2, where the deviations from experimental values in ^{13}C NMR chemical shifts are shown for a series of hydrocarbons and heterosubstituted compounds.

Even more accurate inclusion of the electron correlation effects has been achieved in the GIAO-CCSD and GIAO-CCSD(T) methods,[28] which employ the coupled-cluster Ansatz with single and double excitations (CCSD) and with perturbative treatment of triple excitations (CCSD(T)). Using the latter method, Gauss *et al.*[28,29] have achieved an accuracy of *ca*. 1 ppm (see Fig. 2 and Table 2) in describing the absolute ^{13}C isotropic shieldings in a series of molecules for which accurate gas-phase NMR data are available. To achieve this accuracy one needs to employ very large basis sets and incorporate vibrational corrections to the isotropic shieldings.[29] For the latter purpose, the shielding tensor is expanded around the molecular equilibrium geometry in a Taylor series in terms of the normal coordinates Q_r. Then, one obtains the vibrationally averaged shielding tensor σ_0[29] using the average values of displacements along the normal modes obtained from the anharmonic force fields, as in

$$\sigma_0 = \sigma_e + \sum_r \left(\frac{\partial\sigma}{\partial Q_r}\right)_{Q_r=0} \langle Q_r \rangle + \frac{1}{2}\sum_{r,s} \left(\frac{\partial^2\sigma}{\partial Q_r \partial Q_s}\right)_{Q_r,Q_s=0} \langle Q_r Q_s \rangle + \cdots. \quad (16)$$

Obviously, such an approach is very time consuming and can currently only be applied to relatively small molecules. However, the results of these calculations are indispensable for setting up proper scales of the nuclear shielding constants in the gas phase.

Table 2 Mean absolute deviations (MAD) and standard deviations from the experiment in the theoretical ^{13}C shieldings obtained with different methods. Data taken from refs. 29 and 36

Method	MAD	Standard deviation
GIAO-HF	8.8	10.7
GIAO-MP2	1.5	2.2
GIAO-CCSD(T)	0.7	0.8
GIAO-BP86	5.9	4.0
GIAO-B3LYP	6.8	5.3
IGLO-BPW91	3.4	5.4

Density functional calculations of NMR shielding tensors. A cost-effective alternative to correlated wave function *ab initio* methods is provided by density functional theory (DFT). DFT represents an ingenious reformulation of the many-body problem in quantum mechanics whereby the problem of solving the Schrödinger equation for interacting electrons is cast in the form of a much simpler set of one-electron equations, the so-called Kohn–Sham equations. The cost of a DFT calculation is roughly the same as the cost of a Hartree–Fock calculation. Modern density functional methods have been successful with regard to the calculation of molecular thermochemical and electronic properties such as equilibrium geometries, atomization energies, vibrational frequencies, dipole moments, excitation energies, *etc*. However, the commonly employed density functionals are less successful in the calculation of nuclear shieldings.[30,31]

Generally, the magnitude of the paramagnetic contribution is overestimated with the common density functionals, thus leading to much too deshielded nuclei.[30] This overestimation was explained as a consequence of too low orbital energy differences $\varepsilon_a - \varepsilon_i$ between the virtual and occupied MOs as produced by the approximate density functionals. This deficiency could be partially corrected with the use of the so-called hybrid functionals which mix in a fraction of the Hartree–Fock exchange energy, such as the B3LYP functional.[31] Alternatively, semi-empirical corrections were introduced into the expression for the paramagnetic term (9) which increase the magnitude of the orbital energy differences thus decreasing the magnitude of the paramagnetic term.[32]

Another approach to improve performance of the approximate density functionals in the calculation of nuclear shieldings was taken by Lee *et al.*[30] who used current density dependent functionals suggested earlier by Vignale, Rasolt, and Geldart.[33] However, the initial results were rather disappointing—the semi-empirical corrections[32] yielded more accurate shieldings[30]—and it has been suggested that further improvement of the functionals depending on the electronic current density *via* the vorticity $\vec{\nabla} \times (\vec{j}(\vec{r})/\rho(\vec{r}))$ is necessary. Recently, a noticeable improvement of the calculated shielding constants was obtained[34] with the use of the local multiplicative potentials obtained within the context of DFT with the help of the optimized effective potential technique. However, certain modification of numeric and implementational aspects of this promising technique is necessary before it can be routinely applied for the calculation of nuclear shielding parameters.[34]

Currently, density functional calculations of nuclear shieldings are available in combination with all approaches for the elimination of gauge dependence, IGLO-DFT, LORG-DFT, and GIAO-DFT.[35] Although a large number of approximate density functionals are currently available, none of the commonly used functionals has a clear advantage over the others. This can be seen from Table 2 and Fig. 3, which show deviations from the experimental values of the theoretical ^{13}C shieldings obtained with three popular density functionals. Even though these functionals perform better than the GIAO-HF method on average, all of them have conspicuous failures for certain molecules, especially those with multiple bonds and heteroatoms. However, for a particular class of molecules, the results of density functional calculations can reach acceptable accuracy in the calculation of chemical shifts. Because the cost of a density functional calculation is much lower than that of correlated *ab initio* wave function calculation, they represent an attractive alternative for large scale calculations.

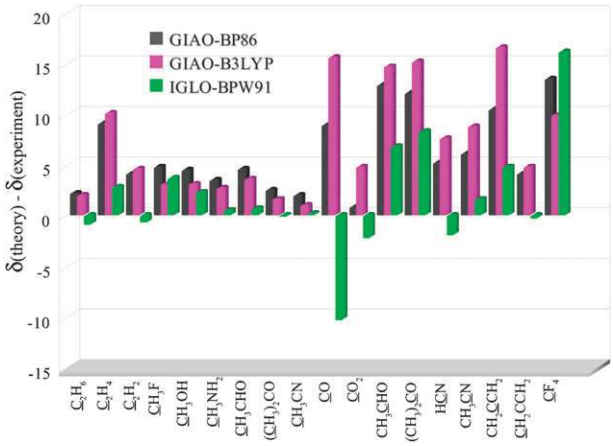


Fig. 3 Deviations from experiment of ^{13}C chemical shifts (with respect to CH_4) calculated with different DFT methods. Data taken from ref. 29 and 36.

IV. Theoretical simulations of NMR chemical shifts in peptides and proteins

The observation by Spera and Bax[37] of an empirical correlation between the secondary chemical shifts and the peptide backbone conformation has stimulated theoretical research of the factors determining the chemical shifts in natural conformations of proteins and peptides. In the pioneering work of de Dios *et al.*[38] and in a number of subsequent studies (see ref. 39 and references cited therein), a theoretical analysis of the structural and environmental dependencies of ^{1}H, ^{13}C, ^{15}N, and ^{19}F chemical shifts in peptides has revealed that the chemical shifts of $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{15}N are governed by changes in the electronic structure due to variation in the backbone torsion angles ϕ and ψ and also (primarily for ^{13}C) in χ_1 and χ_2 angles in side chains (see Scheme 1). Environmental effects due to the electric field inside the protein were found to be responsible for the secondary shifts of ^{19}F in fluorinated proteins. Besides these factors, hydrogen bonds may strongly affect shifts on ^{15}N, $^{1}H^{N}$, $^{13}C'$, and ^{17}O atoms. Because of the diversity of factors influencing the secondary shifts in proteins and peptides, it became clear that only accurate quantum chemical calculations could furnish a tool for analyzing relationship between the observed shifts and protein structure.[38]

The most important issues which need to be addressed when setting up computational models for the calculation of NMR shieldings in peptides concern: (i) selection of the appropriate basis sets, (ii) selection of the theoretical methods and (iii) modeling the effects of environment and nuclear (vibrational) motion. From the early works on chemical shieldings in peptides and proteins,[38,39] the concept of a locally dense basis set was used in combination with GIAO-DFT to calculate the shielding tensor. Because using even a medium sized basis set (*e.g.* 6-31G*) on all atoms in a peptide may result in as many as several thousand basis functions, it was suggested to employ the high quality basis sets only on a few atoms (usually the resonating atom and nearest neighbors) in the computational model. On the remaining atoms, a small basis set was used which led to considerable computational savings in the shielding calculations.

For a reliable interpretation of the experimental shifts, one needs to achieve rather high accuracy (<2–3 ppm for ^{13}C) in the calculated shieldings. The first applications of DFT methods to the calculation of ^{13}C absolute shieldings in a set of small organic molecules were rather disappointing because the average deviations from the experimental values were on the order of 10 ppm (see Table 2).[31] Despite the insufficient accuracy of DFT methods, the theoretically obtained absolute shieldings in a number of peptides and amino acids showed a linear correlation with the experimentally observed chemical shifts.[38,39] The approach based on cross-correlation between the observed chemical shifts and the calculated absolute shieldings is therefore widely employed in the study of peptides and proteins.[38] However the quality of the correlation as revealed by linear regression analysis can be rather poor leading to strong deviations of the slope of linear fit from the ideal value (1.0). Typically, the correlation coefficient r^2 of linear regression varies between 0.8 to 0.9 indicating reasonably good but non-perfect correlation.

The chemical shift is a relative quantity (see eqn (3)), and one can hope that some degree of error compensation may be achieved when taking the difference between the absolute shielding of a reference and target nucleus. This is however not guaranteed and errors made in the calculated shielding of reference compound may be amplified in the calculated chemical shifts. To improve the accuracy of the chemical shifts calculated with density functional methods an alternative procedure has been suggested by Forsyth and Seabag.[40] It was suggested to scale the theoretical shielding values using the slope a and intercept b parameters obtained in linear regression analysis of the experimental chemical shifts *versus* theoretical shieldings. The predicted shieldings are then obtained as

$$\delta^{\mathrm{pred}} = (b - \sigma^{\mathrm{calc}})/a \qquad (17)$$

where the parameters a and b are generally method and basis set dependent (*e.g.* for a typical computational setup B3LYP/6-31G*//B3LYP/6-31G* $a = -1.084$ and $b = 203.1$).[40] Although this is a purely empirical procedure, the mean absolute error in the so-obtained shifts was reduced to less than 2 ppm.

The empirical approach of Forsyth and Seabag was successfully used by Barone *et al.*[41] for structure validation of a number of naturally occurring organic compounds on the basis of the GIAO-HF ^{13}C chemical shifts. For all the compounds studied, the scaled chemical shifts for the correct conformations of target compounds were in much better correlation ($r^2 \approx 0.998$–0.999) with the experimental $\delta(^{13}\mathrm{C})$ values than for the wrong conformations. In a subsequent work,[42] this approach was extended to study conformations of flexible organic compounds. Prior to scaling, the GIAO-HF ^{13}C chemical shifts were averaged over the most popular conformations with the use of the respective Boltzmann factors. Again, a very satisfactory correlation between the theoretically obtained Boltzmann-averaged scaled chemical shifts and the experimental values was observed for the right stereoisomers of the studied compounds.

In a recent study of $^{13}C^{\gamma}$ chemical shifts in histidine dipeptides and in phenylalanine and tyrosine residues in dipeptides and proteins, Oldfield *et al.*[43,44] have used the scaled theoretical shifts (17) to achieve a rather good statistical correlation ($r^2 \approx 0.92$–0.94) with the observed values. On average, an accuracy of $\sim$1.3–1.6 ppm in the so-obtained theoretical shifts was achieved. In the latter work,[44] it has been also found that, for $^{13}C^{\gamma}$ chemical shifts, the protein environment plays a major role and a more or less reliable description of these effects could be achieved with the use of a self-consistent reaction field (SCRF) approach in the GIAO-HF and GIAO-DFT calculations.

The use of empirically corrected chemical shifts improves the statistical correlation between theoretical and experimental values considerably, thus improving the reliability of the theoretical predictions of protein structures. Another factor influencing the accuracy of the theoretical shieldings is the basis set dependence of the results.[43–45] The possibility of obtaining basis set independent shieldings by extrapolating to the CBS limit was explored in recent studies of chemical shifts and chemical shift anisotropies by Markwick and Sattler[46] and by Moon and Case.[47]

Moon and Case[47] have undertaken theoretical calculations of ^{13}C, ^{15}N, and ^{1}H absolute shieldings in *trans N*-methylacetamide and a model Gly–Gly dipeptide at both the wave function *ab initio* level of theory (GIAO-HF and GIAO-MP2 methods) and at the DFT level with a range of functionals varying from the oldest gradient-corrected functionals (such as BP86) to the most recent hybrid HF/DFT functionals. In comparison with the accurate experimental shieldings, the CBS extrapolated GIAO-MP2 results were clearly superior to the CBS extrapolated density functional shieldings. Interestingly, performance of (probably) the most popular and widely used B3LYP hybrid density functional was found to be inferior to other density functionals, such as PBE1 or B3PW91. It was suggested that basis set independent values of nuclear shieldings can be obtained from a two-point CBS extrapolation procedure

$$\sigma^{A}_{\mathrm{CBS}} = -0.73\sigma^{A}_{\text{cc-pVTZ}} + 1.73\sigma^{A}_{\text{cc-pVQZ}} \qquad (18)$$

which employs absolute shieldings obtained with the use of the correlation-consistent triple-zeta (cc-pVTZ) basis set and the quadruple-zeta (cc-pVQZ) basis set. However, the shieldings obtained at the density functional level with small basis sets may fortuitously be in better agreement with the experiment than the CBS extrapolated values obtained with more accurate methods.

It was also found by Moon and Case[47] that the use of locally dense basis sets yields more consistent and more reliable results than the use of partitioning techniques, such as the ONIOM method employed by Markwick and Sattler.[46] Besides being more accurate the use of locally dense basis sets is more economic than the use of the ONIOM method and considerable time savings can be achieved with the former method, especially if an extrapolation to CBS is required. It may therefore be conjectured that the use of locally dense basis sets combined with the accurate *ab initio* methods of wave function theory and CBS extrapolation techniques may furnish a very reliable tool for modeling shielding parameters in amino acids and short peptides.[47]

For modeling chemical shifts in long peptide chains, proteins or crystalline samples, even the use of locally dense basis sets may become prohibitively costly, if one attempts to include the interactions with all partner residues at the quantum mechanical level. Fortunately, nuclear shielding is a fairly local property and the strongest effect of distant atoms and residues on the shielding parameters at the target nucleus originates due to electric fields generated by charged centers.[38] In quantum chemical calculations, the electric field can be modeled by a field of point charges placed at the atomic positions of amino acid residues surrounding the residue of interest in the protein or crystal. This approach, named charge-field perturbation (CFP) GIAO method, was successfully applied by Oldfield *et al.*[38,39,43,44,48,49] for the calculation of ^{13}C and ^{19}F chemical shifts. Thus, the use of the electric field generated by neighboring amino acid molecules in an L-tyrosine crystal helped to considerably improve the statistical correlation of the ^{13}C shieldings obtained in the GIAO-HF calculation with experimental shifts; the r^2 value increased from 0.987 for the GIAO-HF shieldings obtained for a single molecule to 0.996 for CFP-GIAO-HF shieldings.[39,49]

Although the CFP approach was highly useful for the prediction of ^{19}F and most ^{13}C shieldings in proteins, it was less successful for modeling chemical shifts on $^{13}C^{\gamma}$ atoms in phenylalanine and tyrosine dipeptides, for which rather poor statistical correlation between the CFP-GIAO shieldings and experimental shifts obtained from solid state NMR was found by Mukkamala *et al.*[44] It was argued that the SCRF approach (using the polarizable continuum model (PCM)) can provide a better account of electrostatic interactions in these systems. This conclusion was supported by a considerably better agreement of the $^{13}C^{\gamma}$ shieldings obtained in the PCM-GIAO-HF and PCM-GIAO-B3LYP calculations with experiment.[44] However, the slope of linear regression varied between 1.24 and 1.85, thus strongly deviating from the ideal value of 1.0.[44] This indicates that, despite a good statistical correlation, there still remained some inconsistencies in the calculated shieldings.

Modeling of short peptide chains in solution may require explicit inclusion of the solvent molecules (for example, water molecules) in the calculation. Indeed, while the bulk solvent effects can be reliably modeled within the PCM or CFP model, the specific solvent–solute interactions, such as hydrogen bonding, can not be accurately treated within these models. The importance of explicit inclusion of water molecules in the quantum mechanical calculations along with using PCM for modeling the bulk of the solvent has been demonstrated by Han *et al.*[50] who studied theoretically Raman, Raman optical activity and vibrational circular dichroism spectra of alanyl dipeptide in aqueous solution. The molecular geometry obtained in this study was later verified in the liquid crystal NMR experiments which unambiguously shown that coordination with water molecules dictates preference for a specific conformation of alanyl dipeptide in aqueous solution.[51]

NMR shielding calculations are typically carried out using static geometries which can be obtained from X-ray data on crystalline samples or from the geometry optimizations using molecular mechanics force fields or using *ab initio* optimized geometries. Although the so-obtained nuclear shieldings seem to be sufficiently accurate for protein structure refinement, the effect of nuclear motion on the calculated shieldings deserves a more careful consideration. For small organic molecules in the gas phase, the inclusion of vibrational averaging of ^{13}C shieldings *via*eqn (16) improved the mean absolute accuracy by *ca.* 1.5 ppm.[29] A similar improvement in accuracy of the calculated ^{13}C shieldings was recently observed by Dumez and Pickard,[52] who studied nuclear shielding tensors of crystalline L-alanine and dipeptide β-L-aspartyl-L-alanine with the inclusion of vibrational and thermal averaging effects.

Dumez and Pickard[52] used two different approaches for obtaining vibrationally averaged shieldings: a Monte Carlo averaging of the shielding tensor and molecular dynamics (MD) simulations. In the former approach, a number of configurations $\mathbf{X}^{(n)} = \mathbf{X}^{(0)} + \sum_r s_r^{(n)}\mathbf{Q}_r$ was generated from the equilibrium structure $\mathbf{X}^{(0)}$ using random displacements along the normal modes (phonons) $\mathbf{Q}_r$ with the amplitudes $s_r^{(n)}$ randomly generated from the Gaussian probability distribution with variance $\sigma_r^2 = (\hbar/2\omega_r)\coth(\hbar\omega_r/2kT)$, valid for a harmonic oscillator with frequency ω_r at temperature T. Then shieldings calculated at each of the configurations were averaged, yielding the vibrationally averaged shielding $\langle\sigma\rangle$. In another approach, an MD simulation of the crystalline sample was used to generate a number of snapshots (molecular structures), and the nuclear shieldings were calculated at the snapshot geometries and averaged to $\langle\sigma\rangle$.[52]

This study has revealed that vibrational averaging (primarily zero-point motion) may result in variations of *ca.* 3–5 ppm for ^{13}C chemical shifts.[52] The Monte Carlo averaging yielded results that are in better agreement with experiment than the MD simulations. This was attributed to insufficient sampling of configurations around the equilibrium structure due to a limited duration of the MD simulation.[52] It is noteworthy that the calculated motional effects on the chemical shifts were of the same order of magnitude as the effect of replacing one density functional (*e.g.* LDA) with another (*e.g.* PBE gradient-corrected functional). This implies that certain improvements of the current approximations in DFT are also needed to improve agreement between the calculated and the experimental shifts.

V. Outlook and future perspectives

In the last decades, NMR spectroscopy has established itself as a reliable and accurate tool for the determination of secondary structure of proteins. However, if one looks at the distribution of the size of proteins found in living organisms (shown in the

upper panel of Fig. 4) and compares it with the molecular weight distribution of protein structures solved with the help of NMR and deposited in the Brookhaven Protein Data Bank (see lower panel of Fig. 4), then it becomes obvious that the use of NMR spectroscopy is limited to a very small fraction of all the proteins present in nature, and that it is strongly biased towards low molecular weight. This situation is understandable, because line broadening limits solution state NMR spectroscopy of fully protonated proteins to those with molecular weight less than approximately 25–30 kDa. Selective partial deuteration increases the protein size amenable to NMR considerably, but at the expense of removing the primary source of distance information, the interproton NOE. NMR chemical shifts may fill this gap; folding of a protein into its natural conformation leads to a large variation of the chemical shifts of the NMR signal with respect to the "random coil" protein chain. Although this fact has been known for several decades, the determination of protein 3D structure based solely on chemical shifts has been achieved only recently. Profiles of chemical shifts of characteristic atoms in amino acid residues as a function of backbone torsion angles[39,45] provide a large number of restraints used in the determination of protein 3D structure. In the works of Cavalli *et al.*[15] and Shen *et al.*,[14] complete 3D structures of a number of proteins were resolved (with rmsd of better than 2 Å) based on the combination of conventional molecular mechanics force fields with ^{13}C, ^{1}H and ^{15}N chemical shifts used as restraints. The quality of the resulting structures

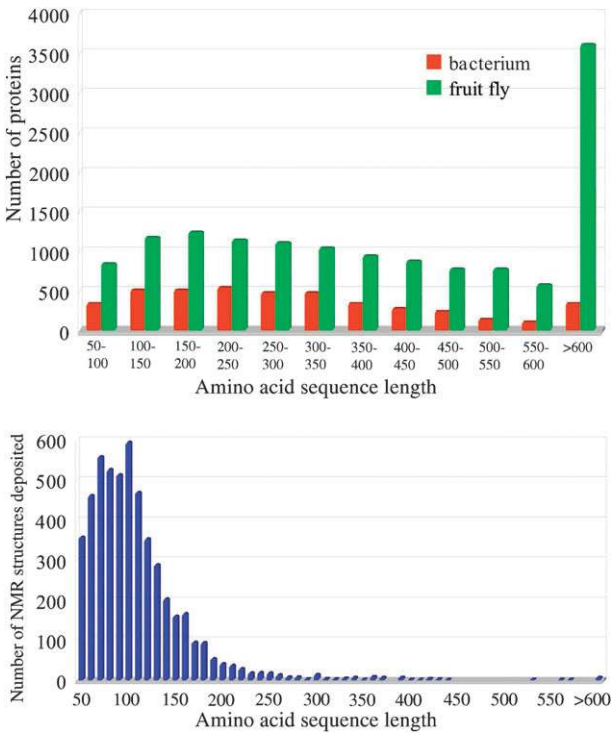


Fig. 4 Upper panel: distribution of predicted amino acid sequence lengths in the genomes of two organisms (a bacterium and a fruit fly). Picture is constructed using data from ref. 53. Lower panel: molecular weight distribution of the NMR structures deposited in the Brookhaven Protein Data Bank (June 2009).

suggests that this approach is a viable one, but certainly stipulates further methodological advances.

The use of modern computers for obtaining relationships between chemical shifts and 3D conformation of peptide chains from quantum chemical calculations is a valuable source of information which can increase the applicability of NMR spectroscopy for structure determination of proteins.[39] However, to reach this goal the accuracy of theoretically obtained nuclear shieldings needs to be increased. Although the currently used approach, in which shieldings obtained in DFT calculations are cross-correlated against experimental chemical shifts, was rather successful in the past, it has obvious limitations. In this perspective, it appears important to select those density functionals which are capable of yielding the most accurate shielding constants in comparison with accurate theoretical data generated in high-level *ab initio* calculations. Combined with extrapolation to the CBS limit[46,47] these density functionals should lead to improved predictions for variations of nuclear shieldings with protein structure. Besides the currently most popular GIAO approach, the use of local gauge-origin techniques, such as IGLO and LORG, can be beneficial for the shielding calculations, because the latter methods are computationally simpler and they allow for a transparent interpretation of nuclear shieldings in terms of contributions of neighboring groups and individual bonds.[24,25]

Another issue which needs to be resolved to improve the accuracy of theoretically obtained shieldings is to accurately take the protein environment into account. The use of point charges commonly adopted in general molecular mechanics force fields for modeling electric fields inside proteins does not always lead to satisfactory results for the calculated NMR shieldings.[44] The use of a polarizable continuum model was recommended to improve the correlation of calculated shieldings with experimental shifts,[44] however within such an approach all information about the structure of the surrounding protein is lost. For an accurate account of specific interactions, such as hydrogen bonds, it may be necessary to explicitly include the solvent molecules in the quantum mechanically calculated models.[50,51]

Very encouraging results were recently obtained by He *et al.*[16] who used the combined quantum mechanics/molecular mechanics (QM/MM) approach (which, for chemical shifts, is similar to the CFP method) combined with automated fragmentation (AF) to study ^{1}H, ^{13}C, and ^{15}N chemical shifts in the mini-protein Trp cage. Within the AF-QM/MM approach, the entire protein is split into a core region, for which chemical shifts are calculated quantum mechanically, and a buffer region which includes residues adjacent to the core region, which is also treated quantum mechanically. The environment is modeled with the use of molecular mechanics. Using this approach very good statistical correlation (with $R^2 > 0.95$) was obtained between the calculated and experimental ^{1}H chemical shifts. Moreover, experimental and calculated backbone ^{1}H, ^{13}C, and ^{15}N chemical shifts agreed with rms errors of 0.09, 0.32, and 0.78 ppm, respectively. It was found that the shielding constants are sensitive to the choice of partial atomic charges employed in the MM environment. In this respect, it seems beneficial to employ

polarizable force fields, such as the direct reaction field approach (DRF),[54] within which the electric field is generated by a set of polarizable charge distributions capable of self-consistently adapting themselves to the electrostatic potential and charge distribution of the quantum mechanical subsystem.

Currently, the state-of-the-art in using chemical shifts is limited to the polypeptide backbone, where dihedral angle predictions are based on the simple observation that similar amino acid sequences with similar chemical shifts for $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{15}N atoms have similar backbone geometry. Side chain chemical shifts have not been extensively employed, and the interior of the protein is built from educated guesses using database-derived conformational preferences and Monte Carlo sampling methods.[15] Clearly, packing of the amino acid side chains in the interior of proteins poses severe restrictions on the possible side chain conformations.[10] Because the side chains may adopt a number of conformations within very narrow energy ranges, the theoretical calculation of secondary chemical shifts requires proper averaging over the most popular conformations.[10] This necessitates the use of molecular dynamics methods for finding the most popular conformations of amino acid side chains in proteins. In this regard, the reliability of the results that can be obtained through use of common force fields needs to be carefully tested because even small energy differences may have a dramatic effect on the conformational ensemble.[55]

The use of ^{13}C and ^{15}N chemical shifts as a source of structural information can extend the applicability of NMR as a tool for 3D structure determination of proteins in solution. The range of proteins that can be characterized structurally with NMR spectroscopy can be markedly increased with the use of structural restraints obtained from a comparison of the theoretically calculated chemical shifts for specific spatial conformations of amino acid backbone and side chains, bringing within reach many systems of biological significance.

References

1 D. Sitkoff and D. A. Case, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1998, **32**, 165–190 and references cited therein.
2 K. Wüthrich, *Angew. Chem., Int. Ed.*, 2003, **42**, 3340–3363 and references cited therein.
3 D. S. Wishart and D. A. Case, in *Methods in enzymology, vol. 338: nuclear magnetic resonance of biological macromolecules, Part A*, ed. T. L. James, V. Dotsch and U. Schmitz, Academic Press, San Diego, 2001, pp. 3–34.
4 D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges and B. D. Sykes, *J. Biomol. NMR*, 1995, **5**, 67–81.
5 Y. J. Wang and O. Jardetzky, *J. Am. Chem. Soc.*, 2002, **124**, 14075–14084.
6 S. Schwarzinger, G. J. A. Kroon, T. R. Foss, J. Chung, P. E. Wright and H. J. Dyson, *J. Am. Chem. Soc.*, 2001, **123**, 2970–2978.
7 Y. Shen and A. Bax, *J. Biomol. NMR*, 2007, **38**, 289–302.
8 Y. J. Wang and O. Jardetzky, *Protein Sci.*, 2002, **11**, 852–861.
9 R. E. London, B. D. Wingard and G. A. Mueller, *J. Am. Chem. Soc.*, 2008, **130**, 11097–11105.
10 F. A. A. Mulder, *ChemBioChem*, 2009, **10**, 1477.
11 D. S. Wishart and B. D. Sykes, *J. Biomol. NMR*, 1994, **4**, 171–180.
12 S. Neal, A. M. Nip, H. Y. Zhang and D. S. Wishart, *J. Biomol. NMR*, 2003, **26**, 215–240.
13 G. Cornilescu, F. Delaglio and A. Bax, *J. Biomol. NMR*, 1999, **13**, 289–302.
14 Y. Shen, O. Lange, F. Delaglio, P. Rossi, J. M. Aramini, G. H. Liu, A. Eletsky, Y. B. Wu, K. K. Singarapu, A. Lemak, A. Ignatchenko, C. H. Arrowsmith, T. Szyperski, G. T. Montelione, D. Baker and A. Bax, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 4685–4690.
15 A. Cavalli, X. Salvatella, C. M. Dobson and M. Vendruscolo, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 9615.
16 X. He, B. Wang and K. M. Merz, Jr., *J. Phys. Chem. B*, 2009, **113**, 10380–10388.
17 N. F. Ramsey, *Phys. Rev.*, 1950, **78**, 699; N. F. Ramsey, *Phys. Rev.*, 1952, **86**, 243.
18 S. A. Smith, W. E. Palke and J. T. Gerig, *Concepts Magn. Reson.*, 1992, **4**, 107; J. Mason, *Solid State Nucl. Magn. Reson.*, 1993, **2**, 285.
19 C. J. Jameson, *Chem. Rev.*, 1991, **91**, 1375.
20 R. Ditchfield, *Mol. Phys.*, 1974, **27**, 789.
21 C. J. Jameson, in *Nuclear Magnetic Resonance vol. 10*, ed. G. A. Webb, RSC Publishing, Cambridge, 1981, pp. 1–17 and references cited therein.
22 R. M. Stevens, R. M. Pitzer and W. N. Lipscomb, *J. Chem. Phys.*, 1963, **38**, 550.
23 K. Wolinski, J. F. Hinton and P. Pulay, *J. Am. Chem. Soc.*, 1990, **112**, 8251.
24 W. Kutzelnigg, U. Fleischer and M. Schindler, in *NMR, Basic Principles and Progress*, ed. P. Diehl, E. Fluck, H. Günther, R. Kosfeld and J. Seelig, Springer, Berlin, 1990, vol. 23, p. 165.
25 A.E. Hansen and T. D. Bouman, *J. Chem. Phys.*, 1985, **82**, 5035.
26 J. Gauss and J. F. Stanton, in *Calculation of NMR and EPR Parameters*, ed. M. Kaupp, M. Bühl and V. G. Malkin, Wiley-VCH, Weinheim, 2004, p. 123.
27 J. Gauss, *J. Chem. Phys.*, 1993, **99**, 3629.
28 J. Gauss and J. F. Stanton, *J. Chem. Phys.*, 1996, **104**, 2574.
29 A. A. Auer, J. Gauss and J. F. Stanton, *J. Chem. Phys.*, 2003, **118**, 10407.
30 A. M. Lee, N. C. Handy and S. M. Colwell, *J. Chem. Phys.*, 1995, **103**, 10095.
31 J. R. Cheeseman, G. W. Trucks, T. A. Keith and M. J. Frisch, *J. Chem. Phys.*, 1996, **104**, 5497.
32 V. G. Malkin, O. L. Malkina, M. E. Casida and D. R. Salahub, *J. Am. Chem. Soc.*, 1994, **116**, 5898.
33 G. Vignale, M. Rasolt and D. J. W. Geldart, *Phys. Rev. B*, 1988, **37**, 2502.
34 A. V. Arbuznikov and M. Kaupp, *Int. J. Quantum Chem.*, 2005, **104**, 261–271 and references cited therein.
35 C. van Wüllen, in *Calculation of NMR and EPR Parameters*, ed. M. Kaupp, M. Bühl and V. G. Malkin, Wiley-VCH, Weinheim, 2004, p. 85.
36 L. Olsson and D. Cremer, *J. Chem. Phys.*, 1996, **105**, 8995.
37 S. Spera and A. Bax, *J. Am. Chem. Soc.*, 1991, **113**, 5490.
38 A. C. de Dios, J. G. Pearson and E. Oldfield, *Science*, 1993, **260**, 1491.
39 E. Oldfield, *Annu. Rev. Phys. Chem.*, 2002, **53**, 349.
40 D. A. Forsyth and A. B. Seabag, *J. Am. Chem. Soc.*, 1997, **119**, 9483.
41 G. Barone, L. Gomez-Paloma, D. Duca, A. Silvestri, R. Riccio and G. Bifulco, *Chem.–Eur. J.*, 2002, **8**, 3233.
42 G. Barone, D. Duca, A. Silvestri, L. Gomez-Paloma, R. Riccio and G. Bifulco, *Chem.–Eur. J.*, 2002, **8**, 3240.
43 F. Cheng, H. Sun, Y. Zhang, D. Mukkamala and E. Oldfield, *J. Am. Chem. Soc.*, 2005, **127**, 12544.
44 D. Mukkamala, Y. Zhang and E. Oldfield, *J. Am. Chem. Soc.*, 2007, **129**, 7385.
45 H. Sun, L. K. Sanders and E. Oldfield, *J. Am. Chem. Soc.*, 2002, **124**, 5486.
46 P. R. L. Markwick and M. Sattler, *J. Am. Chem. Soc.*, 2004, **126**, 11424.
47 S. Moon and D. A. Case, *J. Comput. Chem.*, 2006, **27**, 825.
48 H. Sun and E. Oldfield, *J. Am. Chem. Soc.*, 2004, **126**, 4726.
49 A. C. de Dios, D. D. Laws and E. Oldfield, *J. Am. Chem. Soc.*, 1994, **116**, 7784.
50 W.-G. Han, K. J. Jalkanen, M. Elstner and S. Suhai, *J. Phys. Chem. B*, 1998, **102**, 2587–2602.
51 C.-D. Poon, E. T. Samulski, C. F. Weise and J. C. Weishaar, *J. Am. Chem. Soc.*, 2000, **122**, 5642–5643.
52 J.-N. Dumez and C. J. Pickard, *J. Chem. Phys.*, 2009, **130**, 104701.
53 G. T. Montelione, D. Y. Zheng, Y. P. J. Huang, K. C. Gunsalus and T. Szyperski, *Nat. Struct. Biol.*, 2000, **7**, 982–985.
54 P. Th. van Duijnen, A. H. de Vries, M. Swart and F. Grozema, *J. Chem. Phys.*, 2002, **117**, 8442.
55 R. B. Best, N.-V. Buchete and G. Hummer, *Biophys. J.*, 2008, **95**, L07.