Rapid Computation of Protein NMR Properties – An Optimal Way to Chemical Shift Driven Protein Structure Refinement

IGOR JAKOVKIN, ULRICH STERNBERG, ANNE S. ULRICH
Institute for Biological Interfaces
Karlsruhe Institute of Technology
POB 3640, D-76021 Karlsruhe
GERMANY
igor.jakovkin@kit.edu

Abstract: Ab initio methods for chemical shift calculation are much too demanding for biopolymers and protein structure refinement. No ab initio chemical shift gradients have been developed up to now. Bond polarization theory (BPT) is a rapid and reliable semi-empirical method for computation of NMR parameters. We adapted BPT to the calculation of protein NMR chemical shifts. BPT routines are included into the COSMOS-NMR hybrid QM/MM force field enabling chemical shift driven geometry optimization. In this paper we describe the development of the protocol for chemical shift driven protein structure refinement and prove its robustness by applying it to seven randomly chosen structures from the RCSB protein data bank.

Key-Words: protein, NMR, chemical shift, geometry optimization, structure refinement, bond polarization theory, BPT

1 Introduction

Nuclear magnetic resonance spectroscopy (NMR) is a widely used method for protein structure elucidation. NMR chemical shifts (CS) of proteins contain a variety of information on their three-dimensional structure [1]. Empirical relations between torsion angles and chemical shift can be included into protein structure calculation [2] and there are several attempts of protein structure determination using solely chemical shifts [3, 4]. However empirical chemical shift prediction raises difficulties if applied to unusual protein folds, proteins containing non-standard amino acids, metalloproteins and posttranslationally modified proteins.

Ab initio methods provide an alternative approach to chemical shift calculation in biological systems. Unfortunately ab initio methods are computationally demanding and their application is practically restricted to polypeptides. At the first glance it seems to be possible to overcome this problem by applying hybrid ab initio quantum-mechanical/molecular mechanics methods (QM/MM) [5] and/or decomposition of protein structure into fragments. Vila et al. computed protein $^{13}\text{C}_\alpha$ chemical shift at the DFT level by constructing Ac-GXG-Me tripeptides with X being an amino acid from the protein sequence in experimental conformation [6]. Gao et al. [7] performed chemical shift calculations using fragment molecular orbital (FMO) method at the Hartree-Fock level. He et al. [8] proposed a more efficient approach with automated fragmentation QM/MM calculation considering many-body effects in the QM region and representing the atoms outside of the QM region by their force field point charges while in the FMO method only two-body interactions are considered.

In marked contrast to empirical methods ab initio and QM/MM methods for chemical shift calculation require accurate geometry. NMR derived protein structures (i.e. structures obtained from distance constraints with force field refined geometry) can be used for $^{13}\text{C}_\alpha$ chemical shift but not for $^{15}\text{N}$ chemical shift computation. Furthermore $^{15}\text{N}$ chemical shifts are extremely sensitive to intramolecular environment which cannot be sufficiently described by small clusters such as tripeptides or if many-body effects in the QM region are excluded [9]. In addition conventional force field point charges are not precise enough for ab initio calculations.

Thus protein chemical shift calculation with ab initio and QM/MM methods still remains a challenge as ab initio methods allow neither geometry optimization nor atomic charge calculation of a whole protein. It is also not possible to solve this problem by an ab initio chemical shift driven geometry optimization since ab initio chemical shift gradients have not been developed up to now.

Recently a semi-empirical approach was introduced in order to surmount this problem [10]. It is based on bond polarization theory (BPT) [11, 12] and enables rapid $^{13}\text{C}$ and $^{15}\text{N}$ chemical shift tensor prediction for well characterized structures (e.g. X-ray derived tripeptide crystals). $^{13}\text{C}_\alpha$ chemical shift prediction using NMR derived protein geometry is available within a few seconds on a normal desktop PC without protein fragmentation (the results are comparable to those of Vila et al. [6]). A significant improvement of chemical shift calculation for NMR derived protein structures is achieved by performing chemical shift driven protein structure refinement using experimental chemical shifts as a target function and computing the chemical shifts of the optimized structure. Geometry optimization
procedure ends in the local minimum of the difference between calculated and experimental chemical shifts with respect to the atomic coordinates and does not alter the structure significantly. Therefore the comparison between calculated and experimental chemical shifts allows to evaluate structural quality and to choose the structures optimally representing experimental chemical shifts from an NMR derived ensemble of structures.

In analogy to standard methods for protein structure determination from inter-proton distance restraints [13] the influence of experimental chemical shifts is represented by an additional energy term in COSMOS-NMR hybrid QM/MM force field [14]. The magnitude of this term is scaled by a force constant. The width of the chemical shift potential is adjustable. The initial NMR derived structure for chemical shift driven geometry optimization usually does not match experimental chemical shifts exactly and can be envisaged as a very high temperature conformation of the system. This is why a very thoroughful choice of the potential width and force constant is required; otherwise the geometry optimization will not converge. In this publication we develop a protocol for chemical shift driven geometry optimization with COSMOS-NMR force field by an empirical search of optimal potential width and force constant values. We demonstrate its robustness and applicability by optimizing seven randomly chosen proteins from the RCSB protein data bank reaching chemical shift convergence criteria in all cases.

2 Theory

A detailed description of bond polarization theory and its application to chemical shift calculation and chemical shift driven geometry optimization is presented in [11], [12] and [15]. Only a brief outline is given within this section.

2.1 Bond Polarization Theory

Bond polarization theory is designed for calculation of localized molecular properties such as NMR chemical shifts and atomic partial charges. Localized Slater type bond orbitals are constructed for every bond in the molecule. They form the ground state wave function. Anti-bonds are introduced to account for polarizations and delocalizations. For calculation of localized properties delocalizations can be mostly neglected. Polarization can be described by excited configurations where bonds are replaced by their own anti-bonds. In this way valence electrons are confined to the bonds and orbital localization is achieved. This is why molecular system can be divided into subunits. Approximating the Hamiltonian of the system by interacting atomic point charges the electronic structure in one subunit (e.g. bonds belonging to the nucleus of interest) depends on the localized wave function and the charge distribution of the molecular system.

2.2 Bond Polarization and Chemical Shift

BPT approach to the chemical shift calculation considers chemical shift expectation values of bonds in the ground state and investigates the change of this expectation values with the change of the molecular surrounding. A general formula is derived for the expectation value of one-electron operator applied to the perturbed wave function. It leads to a linear relationship between the expectation value of the NMR chemical shift and the bond polarization by atomic point charges of the molecular system:

\[
\delta = \delta^0 + AV
\]  

with \( \delta \) – bond contribution to the chemical shift of a nucleus, \( \delta^0 \) – CS contribution of an unpolarized bond, \( A \) – change of CS bond contribution with polarization and \( V \) – polarization energy of the bond. Parameters \( \delta^0 \) and \( A \) are defined by fitting Eq. 1 against known expectation values. Further calculations do not require explicit chemical shift operator evaluation, leading to significant speed-up of the computational process.

2.3 CS Driven Geometry Optimization

BPT is the only method for chemical shift calculation with analytically derived chemical shift gradients. Chemical shift gradients provide pseudo-forces for molecular mechanics:

\[
F^\text{CS}_\alpha = -f^\text{CS}_\alpha k^\text{CS} \left( \delta^\text{Exp} - \delta^\text{Calc} \right) \frac{\partial \delta^\text{Calc}}{\partial \alpha}
\]  

with \( F^\text{CS} \) – CS pseudo-force acting on a nucleus, \( \delta^\text{Exp} \) – experimental CS, \( \delta^\text{Calc} \) – calculated CS, \( \alpha \) – atomic coordinates, \( k^\text{CS} \) – force constant adjusting the magnitude of chemical shift pseudo-force and \( f^\text{CS} \) – scaling function controlling the width \( \Delta \) of the chemical shift potential:

\[
f^\text{CS}_\alpha = \frac{g - g^{-1}}{g + g^{-1}} (\delta^\text{Exp} - \delta^\text{Calc})
\]

\[
g = e^{\frac{\delta^\text{Exp} - \delta^\text{Calc}}{\Delta}}
\]

This function ensures strong chemical shift pseudo-forces if the differences between measured and calculated chemical shifts are in the order of \( \Delta \). For large deviations the forces become constant. This is necessary since there are local CS pseudo-energy minima for every nucleus and linearly rising pseudo-forces could easily deform the whole structure. Furthermore there is no single structure exactly matching all CS constraints because experimental CS values are structure ensemble and time averages.
3 Methods

A series of chemical shift driven geometry optimizations was performed on the first structure of ubiquitin structure ensemble (PDB ID 1D3Z). Initial CS force constants were chosen randomly. Initial CS potential width values reflected mean CS error in single point calculations. Separate force constants and potential widths were applied for $^{15}$N and $^{13}$C to account for different CS sensitivity upon molecular geometry errors. Force constants and potential widths were varied through a series of CS driven geometry optimizations to find optimal settings minimizing the root mean square deviation between calculated and experimental chemical shifts (cf. Table 1). To evaluate the robustness of optimal settings (cf. Section 4) the first structures of seven structure ensembles (PDB ID 1UEM, 2JUO, 2K14, 2K19, 2K53, 2K75 and 1Q02) were subjected to CS driven geometry optimization. Conjugate gradients method with an initial slope 0.2 (this corresponds to the step width of $0.2 \times 10^{-3}$Å) and gradient limit $1 \times 10^{-6}$ for the termination of the calculation was used for all CS driven geometry optimizations. The COSMOS software package was applied for all calculations [16].

### Table 1: Settings applied in a series of CS driven geometry optimizations of ubiquitin 1D3Z ($k_{\text{CS}}$– CS pseudo-force weighting factor, $\Delta$ – CS potential width, R – correlation coefficient between experimental and calculated CS, SD – root mean square deviation of experimental and calculated CS).

<table>
<thead>
<tr>
<th>Run</th>
<th>$k_{\text{CS}}$</th>
<th>$\Delta$</th>
<th>$k_{\text{CS}}$</th>
<th>$\Delta$</th>
<th>R</th>
<th>SD</th>
<th>R</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>20</td>
<td>0.15</td>
<td>2.5</td>
<td>0.94</td>
<td>2.38</td>
<td>0.63</td>
<td>4.04</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>15</td>
<td>5.00</td>
<td>2.5</td>
<td>0.29</td>
<td>53.7</td>
<td>0.66</td>
<td>3.76</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>15</td>
<td>1.5</td>
<td>2.5</td>
<td>0.85</td>
<td>3.43</td>
<td>0.81</td>
<td>2.78</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>15</td>
<td>2.00</td>
<td>2.5</td>
<td>0.83</td>
<td>3.51</td>
<td>0.85</td>
<td>2.48</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>15</td>
<td>1.25</td>
<td>2.5</td>
<td>0.85</td>
<td>3.38</td>
<td>0.80</td>
<td>2.81</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
<td>15</td>
<td>2.00</td>
<td>2.5</td>
<td>0.89</td>
<td>2.83</td>
<td>0.86</td>
<td>2.44</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>15</td>
<td>2.00</td>
<td>2.5</td>
<td>0.90</td>
<td>2.69</td>
<td>0.87</td>
<td>2.29</td>
</tr>
<tr>
<td>8</td>
<td>25</td>
<td>15</td>
<td>1.25</td>
<td>2.5</td>
<td>0.90</td>
<td>2.74</td>
<td>0.79</td>
<td>2.90</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>15</td>
<td>1.25</td>
<td>2.5</td>
<td>0.91</td>
<td>2.50</td>
<td>0.78</td>
<td>2.90</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
<td>20</td>
<td>0.15</td>
<td>2.5</td>
<td>0.90</td>
<td>2.70</td>
<td>0.81</td>
<td>2.78</td>
</tr>
<tr>
<td>11</td>
<td>30</td>
<td>20</td>
<td>0.15</td>
<td>2.5</td>
<td>0.91</td>
<td>2.64</td>
<td>0.82</td>
<td>2.63</td>
</tr>
<tr>
<td>12</td>
<td>30</td>
<td>15</td>
<td>2.25</td>
<td>2.25</td>
<td>0.91</td>
<td>2.56</td>
<td>0.86</td>
<td>2.35</td>
</tr>
<tr>
<td>13</td>
<td>30</td>
<td>15</td>
<td>2.50</td>
<td>2.25</td>
<td>0.91</td>
<td>2.50</td>
<td>0.87</td>
<td>2.24</td>
</tr>
<tr>
<td>14</td>
<td>30</td>
<td>15</td>
<td>2.75</td>
<td>2.25</td>
<td>0.90</td>
<td>2.56</td>
<td>0.89</td>
<td>2.18</td>
</tr>
<tr>
<td>15</td>
<td>30</td>
<td>15</td>
<td>3.00</td>
<td>2.25</td>
<td>0.90</td>
<td>2.64</td>
<td>0.90</td>
<td>2.03</td>
</tr>
<tr>
<td>16</td>
<td>30</td>
<td>15</td>
<td>2.00</td>
<td>2.00</td>
<td>0.91</td>
<td>2.57</td>
<td>0.86</td>
<td>2.37</td>
</tr>
<tr>
<td>17</td>
<td>30</td>
<td>15</td>
<td>2.25</td>
<td>2.00</td>
<td>0.91</td>
<td>2.52</td>
<td>0.86</td>
<td>2.37</td>
</tr>
<tr>
<td>18</td>
<td>30</td>
<td>15</td>
<td>2.50</td>
<td>2.00</td>
<td>0.89</td>
<td>2.78</td>
<td>0.93</td>
<td>1.76</td>
</tr>
<tr>
<td>19</td>
<td>30</td>
<td>15</td>
<td>2.75</td>
<td>2.00</td>
<td>0.90</td>
<td>2.61</td>
<td>0.87</td>
<td>2.23</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td>15</td>
<td>3.00</td>
<td>2.00</td>
<td>0.89</td>
<td>2.83</td>
<td>0.91</td>
<td>2.01</td>
</tr>
</tbody>
</table>

Fig. 1: Proteins used for the evaluation of the CS driven structure refinement protocol (from left to right; upper row: 2K19, 2K53, 1Q02, 2K14; lower row: 2K75, 1D3Z, 1UEM, 2JUO).

4 Results

A protocol for chemical shift driven protein structure refinement using BPT and COSMOS-NMR force field was successfully developed. Optimal settings for the force constant of the chemical shift pseudo-force and the width of the chemical shift potential are $k_{\text{CS}} = 30$ and $\Delta = 15$ ppm for $^{15}$N and $k_{\text{CS}} = 2.5$ and $\Delta = 2$ ppm for $^{13}$C. Deviations from these values increase the differences between the calculated and experimental chemical shifts after optimization. To evaluate the robustness of this protocol seven randomly chosen protein structures from the RCSB protein data bank were subjected to chemical shift driven protein geometry optimization. Correlation coefficient and root mean square deviation for all seven structures are 0.86 and 2.72 ppm for $^{15}$N and 0.85 and 2.57 ppm for $^{13}$C demonstrating the wide applicability of this protocol.

Chemical shift driven geometry optimization can be applied for structural quality evaluation. It allows to select ensembles of structures optimally representing experimental chemical shifts from larger sets of preselected structures. Furthermore the geometry of optimized structures is sufficient for $^{13}$N chemical shift tensor calculation. It opens a new way to the interpretation of NMR spectra of oriented samples and the study of protein dynamics. These investigations will be under the scope of a separate study.

References:


