Protein Dynamics

DOI: 10.1002/anie.201001898

NMR-Based Protein Potentials**

*Da-Wei Li and Rafael Brüschweiler**

The quality of molecular mechanics force fields is vital for the accurate in silico characterization of proteins. However, the development of better force fields has been a formidable challenge. Important improvements in force fields have been made recently; for example, CHARMM22/CMAP[1] and Amber's ff99SB[2] have been validated for several proteins by comparison of experimental NMR data, including spin relaxation data[1–3] and dipolar couplings,[4] with those predicted by molecular dynamics (MD) simulations. Another type of NMR observable is the chemical shift whose relationship to three-dimensional (3D) protein structures is increasingly well understood.[5] A recent comparison of calculated and experimental protein ^{13}C chemical shifts suggests that there is considerable room for additional improvements of the force field.[6]

In spite of their capacity to rigorously cross-validate MD trajectories, NMR parameters of proteins have not been used to actively guide the improvement of protein potentials. For each new combination of force-field parameters, weeks of computing time are required to generate new MD trajectories of whole proteins, thereby rendering a systematic exploration of force-field parameters prohibitively expensive. Therefore, past force-field developments have mostly relied upon quantum chemical calculations and spectroscopic data of small molecules and protein fragments.

Herein, we introduce a new approach for the optimization of force fields that is applicable to fully intact proteins for which NMR chemical shifts (or other NMR parameters) are available. To overcome the computational cost barrier, we reweight a parent trajectory performed with the original force field (V_{old}) for a new test force field (V_{new}) by using Boltzmann's relationship [Eq. (1)]:

$$p_{new}(j) = p_{old}(j)\ e^{-V_{new}(j)/kT} / e^{-V_{old}(j)/kT} \quad (1)$$

where $p_{old}(j)$ and $p_{new}(j)$ are the relative weights and $V_{old}(j)$ and $V_{new}(j)$ are the potential energies of a snapshot j for the old and new force field, respectively; k is Boltzmann's constant and T is the simulation temperature. Although reweighting is a common tool for enhancing conformational sampling,[7] it has not been used for force-field optimization directly applied to intact proteins.

Our approach starts with the calculation of chemical shifts of all carbon nuclei Cα, Cβ, and C′ for each snapshot of the parent MD trajectory, which are then stored for subsequent analysis. Time-averaged chemical shifts are calculated with equal weights, $p_{old}(j) = 1/N$, for all N snapshots and compared with the experimental chemical shifts by means of the root-mean-square deviation (RMSD) in ppm. The force field is then iteratively revised using the downhill simplex minimization algorithm, which in turn changes the weight of each snapshot according to Equation (1) and thereby allows a systematic improvement of the agreement between the experimental chemical shifts and the back-calculated average shifts from the new weights $p_{new}(j)$. In this way, a vast number of trial potentials can be screened for entire proteins through the reuse of the parent trajectories, thus providing an increase in analysis speed by a factor of 10^5 or more, depending on the trajectory lengths.

We demonstrate the method by deriving the new force field, ff99SBnmr1, from ff99SB[2] by modifying the potential of the backbone dihedral angles using the MD trajectories of four trial proteins (Figure 1). We then cross-validate the performance of ff99SBnmr1 for an additional 18 proteins. The proteins were selected based on the availability of 1) a relatively high-resolution X-ray crystal structure ($\leq$ 2.1 Å for 19 out of 22 proteins), and 2) the availability of NMR chemical shift assignments.

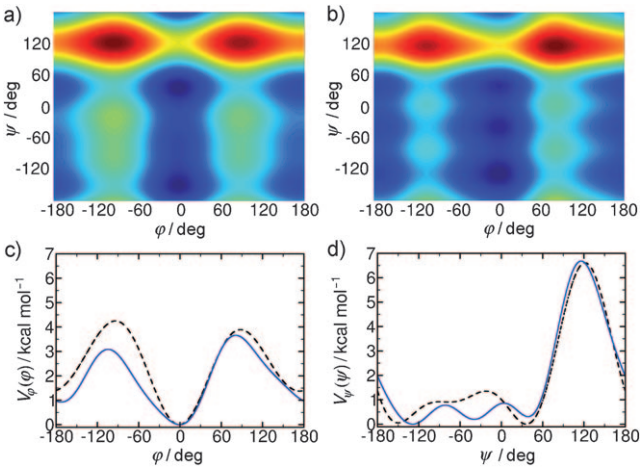

Figure 1. Comparison between the NMR-optimized potential for protein backbone dihedral angles defined by ff99SBnmr1 (b) and ff99SB[2] (a). Two-dimensional maps (a,b) and one-dimensional projections along ϕ (c) and ψ (d) dihedral angles for ff99SBnmr1 (solid line) and ff99SB (dashed line).

[*] Dr. D. W. Li, Prof. R. Brüschweiler
Department of Chemistry and Biochemistry
Florida State University and National High Magnetic Field Laboratory
Tallahassee, FL 32306 (USA)
Fax: (+1) 850-644-8281
E-mail: bruschweiler@magnet.fsu.edu
Homepage: http://spin.magnet.fsu.edu

[**] We thank Hugh Nymeyer for helpful discussions and Stephan Grzesiek for providing the NMR data of GB1. This work was supported by the National Science Foundation (grant MCB-0918362).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201001898.

The optimized dihedral angle potential of the ff99SBnmr1 force field is first tested by MD simulations of the four trial proteins. The sums of the RMSDs for the chemical shifts of the carbon atoms drop significantly from initial values of 3.30 (interleukin-4), 3.94 (engrailed homeodomain), 2.17 (GB1), and 3.11 ppm (ubiquitin), to 3.12, 3.64, 2.14, and 3.03 ppm, respectively. Figure 1 compares ff99SBnmr1 with ff99SB. Whereas the differences are small for positive ϕ,ψ angles, ff99SBnmr1 is mostly lower than ff99SB by up to 1.16 kcal mol^{-1} for the negative ϕ,ψ angles, which has a stabilizing effect on regular secondary structures.

Next, the new force field was cross-validated with 18 proteins of different topology and size (Figure 2). For all but one protein, ff99SBnmr1 yielded either improved or equal

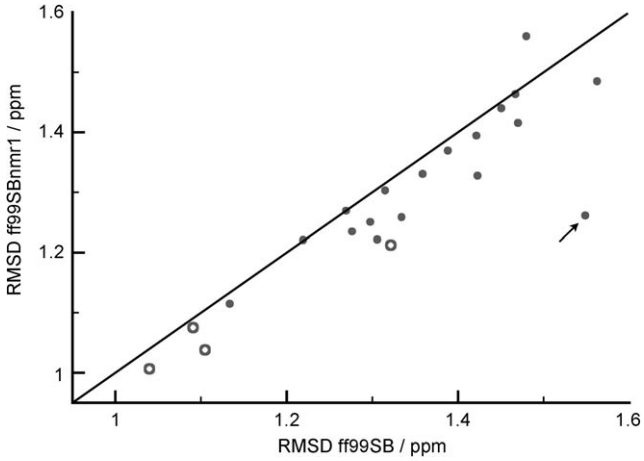

Figure 2. Cross-validation of the new force-field ff99SBnmr1 and comparison with the parent force-field ff99SB. Back-calculated RMSDs for the Cα, Cβ, and C′ chemical shifts for 18 proteins (filled circles) and the four trial proteins (open circles) were determined for trajectories using ff99SBnmr1 and ff99SB. Points on the diagonal indicate identical performance whereas points on the lower half indicate improved performance of ff99SBnmr1 over ff99SB. The DNA binding domain of *S. Aureus VraR* is indicated by an arrow.

results compared to those of ff99SB. The exception is the hypothetical protein JW2626 (2EA9), which starts out with one of the highest RMSDs for its chemical shifts. For all but one protein the initial 3D structure corresponds to the crystalline state, whereas for the remaining proteins all chemical shifts were determined in solution; therefore a large RMSD can be indicative of systematic differences between the structures in solution and in the crystalline environment. In such cases, a simulation length of 30 ns may be too short for adequate sampling of the solution-state ensemble. Larger changes can also occur for NMR structures, as is the case for the DNA binding domain (2RNJ) of *S. Aureus VraR*, which shows a major improvement in the RMSD of its chemical shifts using the new force field (Figure 2). The average improvement found for the 18 proteins used for cross-validation is similar to the one of the four trial proteins, thus providing evidence that the new potential is not the result of overfitting and that it is transferable between globular proteins of variable topologies and sizes.

The availability of complete sets of chemical shifts for over 5000 proteins from the BioMagResBank (BMRB)[8] makes this class of NMR parameters particularly attractive for force-field validation and refinement. Other experimental data that are complementary to chemical shifts, such as residual dipolar couplings and spin relaxation parameters, can be used for the same purposes (see the Supporting Information). Experimental NMR parameters are frequently used as pseudo-energy constraints during computer simulations of proteins to obtain better conformational ensembles.[9] The approach presented herein employs a novel strategy by using NMR information only to optimize the MD force field, and subsequent simulations employ the optimized force field without requiring the inclusion of NMR constraints.

Our results for 22 proteins with a cumulative simulation length of 1.76 μs indicate that the potential for the protein backbone dihedral angles determined by ff99SBnmr1 represents a significant overall improvement over ff99SB. This improvement is made possible by the efficient screening of candidate force fields through the re-weighting technique in which chemical shifts are used as local probes throughout the proteins. Continued advances in protein force fields are vital to achieving a predictive understanding of protein function at the atomic level by computer simulations. The approach presented herein should be useful for enhancing other force fields for proteins and other biomolecules, such as nucleic acids and carbohydrates.

Experimental Section

All simulations were performed with Amber 9[10] at 300 K with explicit solvent (SPC/E water) under PME periodic boundary conditions using the protocol described previously[3b,11] (see also the Supporting Information).

The experimental NMR chemical shifts data are either taken from the BMRB (entries 6475, 15536, 4094) or, in the case of GB1, obtained from S. Grzesiek. Cα, Cβ, and C' chemical shifts were predicted using the program SHIFTS.[5a]

The four trial proteins were interleukin-4 (PDB structure 1HIK as initial structure, and 40 ns trajectory lengths), engrailed homeodomain (1ENH, 100 ns), B1 immunoglobulin-binding domain of streptococcal protein G, GB1, (1PGA, 100 ns), and ubiquitin (1UBQ, 100 ns). The first two proteins are α helical and the other two have α + β folds. For each of the 18 test proteins (see the Supporting Information) two 30 ns MD trajectories were performed, one with ff99SB and one with ff99SBnmr1, using the same MD simulation protocol as described for the trial proteins.

Received: March 30, 2010
Revised: June 24, 2010
Published online: August 16, 2010

Keywords: conformation analysis · molecular dynamics · NMR spectroscopy · proteins · structural dynamics

[1] M. Buck, S. Bouguet-Bonnet, R. W. Pastor, A. D. MacKerell, *Biophys. J.* **2006**, *90*, L36.
[2] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins* **2006**, *65*, 712.
[3] a) A. J. Nederveen, A. M. J. J. Bonvin, *J. Chem. Theory Comput.* **2005**, *1*, 363; b) S. A. Showalter, R. Brüschweiler, *J. Chem.*

Theory Comput. **2007**, *3*, 961; c) A. N. Koller, H. Schwalbe, H. Gohlke, *Biophys. J.* **2008**, *95*, L4; d) N. Trbovic, B. Kim, R. A. Friesner, A. G. Palmer, *Proteins* **2008**, *71*, 684.
[4] S. A. Showalter, R. Brüschweiler, *J. Am. Chem. Soc.* **2007**, *129*, 4158.
[5] a) X. P. Xu, D. A. Case, *Biopolymers* **2002**, *65*, 408; b) S. Neal, A. M. Nip, H. Y. Zhang, D. S. Wishart, *J. Biomol. NMR* **2003**, *26*, 215; c) Y. Shen, A. Bax, *J. Biomol. NMR* **2007**, *38*, 289; d) K. J. Kohlhoff, P. Robustelli, A. Cavalli, X. Salvatella, M. Vendruscolo, *J. Am. Chem. Soc.* **2009**, *131*, 13894.
[6] D. W. Li, R. Brüschweiler, *J. Phys. Chem. Lett.* **2010**, *1*, 246.
[7] G. M. Torrie, J. P. Valleau, *J. Comput. Phys.* **1977**, *23*, 187.
[8] E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. Kent Wenger, H. Yao, J. L. Markley, *Nucleic Acids Res.* **2008**, *36*, D402.
[9] P. R. L. Markwick, T. Malliavin, M. Nilges, *PLoS Comput. Biol.* **2008**, *4*, e1000168.
[10] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, *J. Comput. Chem.* **2005**, *26*, 1668.
[11] D. W. Li, D. Meng, R. Brüschweiler, *J. Am. Chem. Soc.* **2009**, *131*, 14610.

