Importance of the CMAP Correction to the CHARMM22 Protein Force Field: Dynamics of Hen Lysozyme

Matthias Buck,* Sabine Bouguet-Bonnet,* Richard W. Pastor,[†] and Alexander D. MacKerell Jr.[‡]

*Department of Physiology and Biophysics, Case Western Reserve University, Medical School, Cleveland, Ohio 44106; [†]Laboratory of Biophysics, Center for Biologics Evaluation and Research, Food and Drug Administration, Rockville, Maryland 20852; and [‡]School of Pharmacy, University of Maryland, Baltimore, Maryland 21201

ABSTRACT   The recently developed CMAP correction to the CHARMM22 force field (C22) is evaluated from 25 ns molecular dynamics simulations on hen lysozyme. Substantial deviations from experimental backbone root mean-square fluctuations and N-H NMR order parameters obtained in the C22 trajectories (especially in the loops) are eliminated by the CMAP correction. Thus, the C22/CMAP force field yields improved dynamical and structural properties of proteins in molecular dynamics simulations.

As the timescale of molecular dynamics (MD) simulations of proteins is extended, the accuracy of the underlying force field becomes ever more important. Recently a grid-based correction, called CMAP (1), for the ϕ-, ψ-angular dependence of the energy has been introduced into the CHARMM22 force field (C22). This correction yields significant improvements in the residue-location specific distribution of the dihedral angles in protein crystal and solution simulations. For example, C22 yields a π-helix for certain model peptides, whereas C22/CMAP yields the experimentally observed α-helix (2). C22/CMAP has also been shown to improve the structural accuracy in extended simulations of globular proteins (1), including cases were the solvent is treated implicitly. In addition to the inaccuracies in the location of minima noted above, previous C22 simulations have revealed discrepancies with experimental measurements for the dynamic fluctuations in proteins (3). Such findings motivated us to examine the effect of C22/CMAP on protein internal dynamics.

Hen lysozyme has become a standard protein for comparisons between experimental relaxation and simulation derived N-H order parameters (S^2) (4,5). In this letter, we report a comparison with S^2 values derived from 25 ns MD simulations of hen lysozyme using C22 and C22/CMAP. Our findings strongly suggest that the C22/CMAP force field leads to an improved treatment of dynamical as well as structural properties of proteins in MD simulations.

The simulations were performed with the program CHARMM using the C22 all-atom protein force field alone and with the CMAP extension. Hen lysozyme (Protein Data Bank identifier 6LYT) was immersed in a 60 Å side-length water box with 11 chloride ions and was briefly equilibrated to 310 K employing particle-mesh Ewald with periodic boundary conditions. Separate 25 ns MD simulations were run in the NPT ensemble; the last 20 ns were used for analysis.

As a geometric measure, C_α atom positional root mean-square (RMS) differences from the crystal structure were evaluated; the average values are 1.8 ± 0.2 Å and 0.9 ± 0.1 Å for the C22 and C22/CMAP force fields, respectively. Thus, the new force field better reproduces the crystal structure of the protein, consistent with previous studies of the backbone ϕ-, ψ-dihedrals and RMS positional differences (1–3).

Our study of the dynamical properties of the hen lysozyme using the two force fields focused on a comparison with x-ray-derived B-factors and NMR-derived S^2 for the protein internal motion of main-chain N-H bonds. Analysis of the experimental B-factor and simulation-derived RMS fluctuations (Fig. 1 a) shows the C22/CMAP simulation to be in near quantitative agreement for the majority of the residues, whereas the C22 simulation considerably overestimates many RMS fluctuation values. In particular, using uncorrected C22, the main chain is too easily distorted in some of the regions with residues that have small side chains (Gly, Ala, and Asn). In the C22/CMAP simulation, discrepancies with experimental values remain in the vicinity of residue 47 and for residues 107–122. However, in these regions, there are close distance interactions (<3.5 Å) with surrounding protein molecules in the crystal lattice that are not present in solution and could account for the increased mobility in the simulation.

Comparisons between simulation and experimentally derived main-chain S^2 are shown in Fig. 1 b. Overall, it is clear that C22/CMAP yields improved agreement with S^2 derived from experiment. The RMS deviation from the experimentally derived S^2 is 0.179 for C22 and 0.067 for C22/CMAP. Hyperflexibility that is seen in the loop and several of the turn regions in the C22 simulation involves a spectrum of
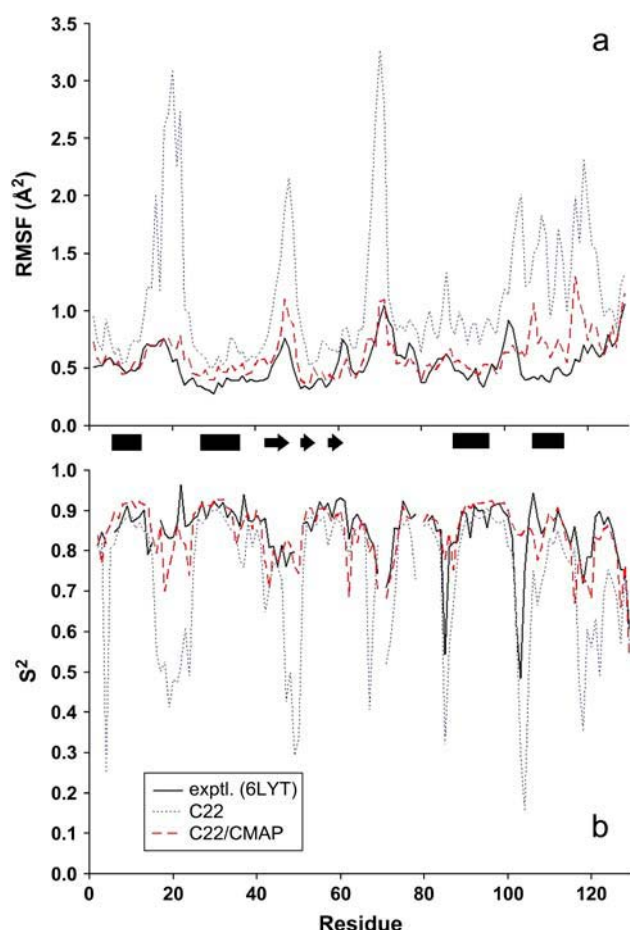
FIGURE 1 (*a*) Comparison of RMS C$_\alpha$ fluctuations over the last 20 ns of the C22 and C22/CMAP trajectories with those estimated from crystallographic B-factors of 6LYT as a function of protein sequence. $\langle\Delta x\rangle^2 = 3B/8\pi^2$ (not corrected for lattice disorder). Global motions are removed by superposition on the C$_\alpha$ frame of 6LYT in all analyses. (*b*) Comparison of simulation and experimentally derived order parameter S^2 for main-chain N-H. S^2 is estimated as the 6 ns time point of the reorientiational correlation function of the normalized vectors with reference to the fixed coordinate frame. (See Supplementary Material for details of the calculation, tables of values, and a list of residues that do not converge.) The experimental relaxation data (6) was input into the ModelFree suite of programs (7), deriving a correlation time of 5.78 ns for a global axially symmetric motion ($D_{//}/D\perp$ of 1.20), and using the Lipari-Szabo formalism for fitting S^2.

conformational transitions, some rapid and others much slower. In fact, using C22, 31 main-chain N-H sites included fluctuations that took place with a time constant in the range of or beyond the global tumbling of the protein ($\tau_R = 5.78$ ns). Some of the slower transitions would be difficult to detect in the experimental relaxation measurements and would thereby not fully contribute to the model-free derived S^2. However, it is not plausible that all such transitions would escape experimental detection. Hence, in a result consistent with that shown in Fig. 1 *a*, the low S^2 indicates that the backbone dihedrals in C22 are too flexible. In general,

residues with the best agreement to experiment undergo fast, small amplitude librational motions (4) in C22/CMAP (but also in C22), and are found in regions including the A-, B-, and C-helices (residues 6–13, 24–35, and 87–97) and part of the β-sheet (residues 50–60). Discrepancies with experiment in the fine variation of S^2 along the polypeptide in these regions may, in part, be associated with the assumption that the N-H bond length and ^{15}N chemical shift anisotropy values are invariant for all residues in the analysis of the experimental data (8). However, at certain residues, the C22/CMAP-calculated order parameters are significantly higher than experimental values (e.g., residues 47, 85, and 102–104—if these five residues are excluded from a comparison with the experimentally derived S^2, the correlation coefficient increases from 0.57 to 0.75 and the RMS deviation improves to 0.051). For these sites, it is possible that there are regions of conformational space that give additional flexibility but have not been accessed in the simulation. Are such regions not populated because the barriers introduced by CMAP are now too high? Or are the fluctuations that are responsible for the experimental relaxation data possibly more complex than can be described by the model free Lipari-Szabo formalism, fitting up to three residue-specific parameters here? These issues are beyond the scope of this report and will be addressed in a future publication.

Another parameter that can, in principle, be derived from both MD simulations and NMR relaxation data is the effective correlation time, τ_e. For C22 simulations, long correlation times (>1 ns) are seen in regions with low S^2, but uncertainties in both parameters are large, showing that the correlation functions are poorly converged (Supplementary Material). Only a few residues experience motions with correlation times >200 ps in the C22/CMAP simulation. Motions appear 2–5-fold faster than in the uncorrected C22 simulation. However, quantitative comparisons with τ_e derived from experiments are still poor, as seen in other studies (3,9). Recently it has been suggested that inaccurate model selection, as well as a lack of sensitivity in fitting experimental τ_e, are likely to be responsible for this lack of correspondence (10). Except for residues that show considerable disagreement in S^2, the motions in the C22/CMAP simulation occur over a range of timescales that are overall close to that of the experimental data.

Hen lysozyme has been used as a model system to evaluate a modification of the C22 protein force field that improved treatment of the ϕ-, ψ-energy surface via a grid correction energy map (CMAP). Here we have shown that the CMAP extension to the force field yields more accurate dynamic properties for this well-studied protein. Agreement with RMS fluctuation data derived from x-ray crystallography and with S^2 and τ_e derived from NMR spectrometry is improved. Such improvements are not unexpected, as the CMAP correction of C22 allows for better reproduction of quantum mechanical conformational energies for the entire ϕ-, ψ-surface as compared to C22 (11), or to AMBER and

OPLS. Our result suggests that other protein force fields may be improved by a similar CMAP correction.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

ACKNOWLEDGMENTS

REFERENCES and FOOTNOTES

1. MacKerell, A. D. 2004. Empirical force fields for biological macro-molecules: overview and issues. *J. Comput. Chem.* 25:1584–1606.

2. Freedberg, D. I., R. M. Venable, A. Rossi, T. E. Bull, and R. W. Pastor. 2004. Discriminating the helical forms of peptides by NMR and molecular dynamics simulation. *J. Am. Chem. Soc.* 126:10478–10484.

3. Philippopoulos, M., A. M. Mandel, A. G. Palmer, and C. Lim. 1997. Accuracy and precision of NMR relaxation experiments and MD simulations for characterizing protein dynamics. *Proteins:* 28:481–493.

4. Buck, M., and M. Karplus. 1999. Internal and overall peptide group motion in proteins. Molecular dynamics simulations for lysozyme compared with results from x-ray and NMR spectroscopy. *J. Am. Chem. Soc.* 121:9645–9658.

5. Soares, T. A., X. Daura, C. Oostenbrink, L. J. Smith, and W. F. van Gunsteren. 2004. Validation of the GROMOS force-field parameter set 45A3 against nuclear magnetic resonance data of hen egg lysozyme. *J. Biomol. NMR.* 30:407–422.

6. Buck, M., J. Boyd, C. Redfield, D. A. MacKenzie, D. J. Jeenes, D. B. Archer, and C. M. Dobson. 1995. Structural determinants of protein dynamics: Analysis of ^{15}N relaxation measurements for mainchain and sidechain nuclei of hen egg-white lysozyme. *Biochemistry.* 34:4041–4055.

7. Cole, R., and J. P. Loria. 2003. FAST-ModelFree: a program for rapid automated analysis of solution NMR spin-relaxation data. *J. Biomol. NMR.* 26:203–213.

8. Case, D. A. 2002. Molecular dynamics and NMR spin relaxation in proteins. *Acc. Chem. Res.* 35:325–331.

9. Pfeiffer, S., D. Fushman, and D. Cowburn. 2001. Simulated and NMR derived backbone dynamics of a protein with significant flexibility: a comparison of spectral densities for the βARK1 PH domain. *J. Am. Chem. Soc.* 123:3021–3036.

10. Chen, J., C. L. Brooks, and P. E. Wright. 2004. Model-free analysis of protein dynamics: assessment of accuracy and model selection protocols based on molecular dynamics simulation. *J. Biomol. NMR.* 29:243–257.

11. MacKerell, A. D., M. Feig, and C. L. Brooks. 2004. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25:1400–1415.