J|A|C|S
COMMUNICATIONS

Published on Web 12/30/2003

Improved Treatment of the Protein Backbone in Empirical Force Fields

Alexander D. MacKerell, Jr.,*,† Michael Feig,‡ and Charles L. Brooks, III‡

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, 20 Penn Street, Baltimore, Maryland 21201, and Department of Molecular Biology, The Scripps Research Institute, 10550 Torrey Pines Road, La Jolla, California 92037

Received June 28, 2003; E-mail: amackere@rx.umaryland.edu

Exploring structure−function relationships in proteins, including the protein-folding problem, represents a significant challenge in the theoretical study of biological macromolecules for which the accurate treatment of molecular conformations and interactions is likely to be crucial.[1] The application of empirical force field-based simulations have made significant progress toward approaching these objectives.[2] Despite these advances, limitations in the treatment of protein backbone conformational energy differences (i.e., the relationship of energy to the ϕ/ψ dihedral angles) remain. These limitations lead to systematic deviations in the ϕ/ψ angles in MD simulations of peptide models (e.g., the blocked peptide model of alanine or the alanine dipeptide, Ace-Ala-Nme)[3] and in proteins.[4] Moreover, limitations in the treatment of high-energy regions of the protein backbone may lead to misinterpretation of protein-folding pathways. To overcome these limitations, efforts have been made to refine the treatment of the protein backbone in empirical force fields.[5,6] While improvements are evident, inherent limitations in proper treatment of the ϕ/ψ dihedral terms remain. In this communication, results are presented from empirical force field calculations on the alanine dipeptide in solution and for several proteins in their crystal environments and in solution, using a grid-based correction to the full ϕ/ψ two-dimensional (2D) conformational energy surface. This extension of the potential energy function is shown to significantly improve the treatment of protein backbone conformational properties at both the model compound and macromolecular levels.

The peptide backbone ϕ/ψ dihedral angle energy terms are generally treated with a Fourier series of cosine terms, with the individual contributions of ϕ and ψ combined additively.[7,8] Studies indicate that this limits the ability of force fields to reproduce high-level quantum mechanical data on model compounds[9] as well as in the treatment of protein structure.[3,4] We overcome this limitation by extending the potential energy function used in the program CHARMM,[10] which is similar to the energy function used in most biomolecular empirical force fields,[11] to more accurately treat peptide and protein backbone ϕ/ψ conformational energies (Feig, M., Brooks, C.L., III, MacKerell, A. D., Jr., work in progress). Initial attempts to extend this functionality involved the introduction of ϕ/ψ dihedral cross terms. While this approach led to significant improvement over the additive treatment, it was still not able to reproduce QM energy differences between the α- and π-helical conformations. This is a problem that leads to the overestimation of π helices in molecular dynamics (MD) studies of proteins and peptides.[4] The introduction of a grid-based energy correction to the ϕ/ψ 2D surface, however, allows for the reproduction of any target 2D energy surface (e.g., a QM surface of the alanine dipeptide). Presently, this correction (see Supporting Information) has been applied to all nonproline residues in the CHARMM force field based on the alanine or glycine dipeptide QM ϕ/ψ energy surfaces. In addition, empirical adjustments to the alanine-based surface were incorporated to account for systematic deviations between MD-based and experimentally observed ϕ/ψ distributions. Such empirical corrections are also necessary to account for the inability of current empirical force fields to accurately treat both gas- and condensed-phase properties;[5] they provide mean-field many-body effects not included in the QM data and address possible limitations in the applied QM level of theory. Here, we present results based on a ϕ/ψ grid correction surface empirically optimized to reproduce experimental ϕ/ψ distributions during MD simulations of eight proteins in their crystal environments. Efforts to produce an empirical correction surface that accounts for the balance between helical and extended conformations and explicitly treats proline residues are ongoing, results from which will be reported at a later date.

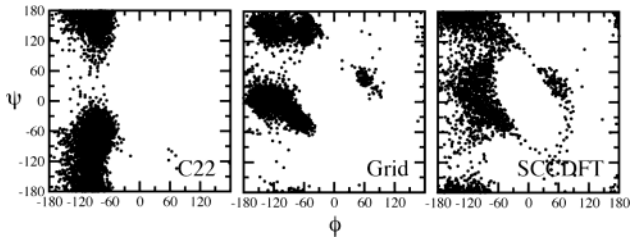

Figure 1. ϕ/ψ distribution from MD simulations of the alanine dipeptide (Ace-Ala-Nme) in solution using the CHARMM22[5] and CHARMM22 grid-corrected empirical force fields and previously published data from a QM/MM model (SCCDFT). Simulations were performed for 10 ns using the program CHARMM[10] with the TIP3P waters[15] with particle mesh Ewald[16] to treat the long-range electrostatic interactions. Details of the simulation protocol will be published elsewhere (Feig, M., Brooks, C. L., III, MacKerell, A. D., Jr., work in progress). QM/MM data are from a 3.6-ns MD simulation using a SCCDFT/AMBER model with the TIP3P water model.[3]

Parts a and b of Figure 1 show calculated ϕ/ψ distributions for the alanine dipeptide in solution using the CHARMM22 and grid-corrected force fields, respectively. In addition, recently published data from a QM/MM model[3] are presented in Figure 1c. As is evident, significant changes in the probability distributions occur because of the grid correction. In the grid-corrected surface, the distribution in the β-sheet region is wider, while in the helical region a more elongated, diagonal shape is evident where there is a correlated increase in ψ as ϕ increases from the α-helical conformation (−60°, −40°). In addition, sampling of the α_l conformation (60°, 30°) occurs in the new model. Importantly, the grid-corrected surface is in good agreement with the QM/MM-based probability distribution, indicating that the grid-corrected model is able to more accurately represent the conformational energy of the peptide backbone in the unfolded state.

† University of Maryland.
‡ The Scripps Research Institute.

10.1021/ja036959e CCC: $27.50 **© 2004 American Chemical Society**

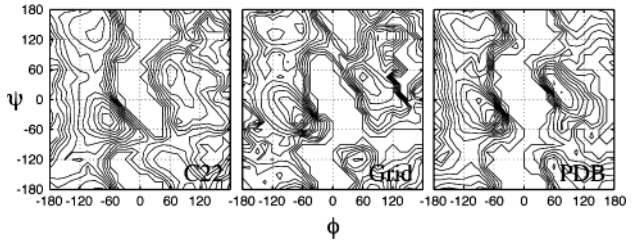

Figure 2. ϕ/ψ PMFs based on MD simulations using the CHARMM22 and CHARMM22 grid-corrected empirical force fields and from a survey of the PDB. Contours are in 0.5 kcal/mol increments up to 6 kcal/mol above the global minimum. MD PMFs based on 1.1-ns simulations of eight protein crystals (PDB identifiers 1AB1, 1I27, 5PTI, 1MBC, 1BZP, 135L, 3EBX, and 1BYI) in their explicit crystal environment using a protocol similar to that in Figure 1. PDB data included only those residues with B factors less than 30 Å.[3] PMFs were obtained from the respective probability distributions based on a Boltzmann distribution[17] and includes all residues in the proteins.

Evaluation of the grid-corrected model for treatment of ϕ/ψ distributions in proteins was performed by comparing ϕ/ψ potentials of mean force (PMF) from MD simulations of eight proteins in their crystal environment using both the CHARMM22 and grid-corrected force fields (Figure 2) with a PMF based on a survey[12] of the PDB.[13] Comparison of the three surfaces reveals the strong similarities of the grid-corrected and PDB-based PMFs. For the extended helical and α_l regions, the overall shapes of these two surfaces are quite similar, while significant discrepancies between the CHARMM22 and PDB PMFs are evident. Clearly, the grid-corrected force field leads to significant improvements in the treatment of the ϕ/ψ conformational energy in proteins as well. In addition, application of the grid correction to simulations of three proteins in solution used by Price and Brooks[14] yielded improvements in the agreement between simulation and crystallographic ϕ/ψ distributions for all three proteins (PDB identifiers: 1GPR, 1CLB, and 1HIJ, not shown).

The grid correction to the CHARMM22 peptide backbone force field represents a significant advance in the accurate treatment of protein backbone conformational energies. Considering that limitations in the treatment of the protein backbone as seen in CHARMM22 have also been reported for other commonly used force fields,[3] it is anticipated that the extension of current empirical force fields to include grid-based energy corrections similar to those used here will yield significant improvement in our ability to treat proteins and peptides via molecular simulations.

Acknowledgment. A.D.M., Jr. thanks C.L.B., III and his group members for the enjoyable sabbatical at The Scripps Research Institute. Financial support from the NIH (GM51505 to A.D.M., Jr. and GM37554, GM48807, and RR12255 to C.L.B., III) and computational support from the DoD High Performance Computing and the National Partnership of Advanced Computational Infrastructure are acknowledged. We thank Drs. Hu and Hermans for supplying the SCCDFT data in Figure 1.

Supporting Information Available: A description of the algorithm used in the grid-based energy correction and grid data required for conversion of the CHARMM22 ϕ/ψ energy surfaces to the corrected surfaces used in the present study (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Berendsen, H. J. C. *Science* **1998**, *282*, 642−643.
(2) Becker, O. M.; MacKerell, A. D., Jr.; Roux, B.; Watanabe, M. *Computational Biochemistry and Biophysics*; Marcel-Dekker: New York, 2001; p 512.
(3) Hu, H.; Elstner, M.; Hermans, J. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 451−463.
(4) Feig, M.; MacKerell, A. D., Jr.; Brooks, C. L., III. *J. Phys. Chem. B* **2002**, *107*, 2831−2836.
(5) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(6) Okur, A.; Strockbine, B.; Hornak, V.; Simmerling, C. *J. Comput. Chem.* **2003**, *24*, 21−31.
(7) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, U.K., 1987.
(8) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*; Wiley-Interscience: New York, 1988; Vol. LXXI.
(9) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908−5920.
(10) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(11) MacKerell, A. D., Jr. In *Atomistic Models and Force Fields*; Becker, O. M., MacKerell, A. D., Jr., Roux, B., Watanabe, M., Eds.; Marcel Dekker: New York, 2001; pp 7−38.
(12) Dunbrack, R. L., Jr.; Cohen, F. E. *Protein Sci.* **1997**, *6*, 1661−1681.
(13) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J. D.; Zardecki, C. *Acta Crystallogr., Sect. D* **2002**, *58*, 899−907.
(14) Price, D. J.; Brooks, C. L., III. *J. Comput. Chem.* **2002**, *23*, 1045−1057.
(15) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(16) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(17) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1976.

JA036959E