Comment on the Communication "The Key to Solving the Protein-Folding Problem Lies in an Accurate Description of the Denatured State" by van Gunsteren et al.**

Aaron R. Dinner and Martin Karplus*

In a recent communication,[1] van Gunsteren et al. proposed that the folding of proteins is much easier than has been believed[2] because the number of accessible conformers in the denatured state is only on the order of 10^9 for a 100-residue protein, instead of on the order of 10^{24} to 10^{30} used in previous analyses.[3, 4] This would make it possible for a polypeptide chain to fold to the native state in about 1 s by a random search since the transition from one conformer to another requires about 1 ns.[5] Van Gunsteren et al.[1] based their argument on simulations ranging from 32 to 200 ns of the folding and unfolding of six β and α amino-acid peptides in methanol and water. For these peptides, the number of conformers found by them corresponds to 1.22 to 1.74 conformers per torsional degree of freedom, in the same range as the value 1.33 we found[6] in Monte Carlo simulations with an implicit solvent model of a denatured 16-residue peptide that forms a β-hairpin in the folded state; the simple average of the number of conformers per torsion angle calculated by van Gunsteren et al. for the different peptides is 1.40.

If the number of conformers scales exponentially with the number of amino acids, an estimate of 1.33 (1.40) conformers per backbone torsion angle leads to a denatured state for a 100-residue protein with 1.33^{200} (1.40^{200}) $= 6 \times 10^{24}$ (1.7×10^{29}) accessible conformers,[3, 4] which is many orders of magnitude larger than the value (10^9) suggested by van Gunsteren et al.[1] The essential question concerns not the number of conformers observed in the peptide studies, but the statement that the number of conformers "does not grow exponentially with the length of the peptide" [p. 364/352 of ref. [1]]. It is the latter that leads to their optimistic conclusion concerning the simplicity of the protein folding problem. There is no evidence for or against this claim in their simulations. The peptides studied vary by only a factor of two in size; in fact, there is one short peptide (9 torsion angles) and the five others vary in length only between 16 and 21 torsion angles. Since the number of conformers shows essentially no correlation with the number of torsion angles (the Pearson linear correlation coefficient between the calculated number of conformers and the number of torsion angles is 0.184), it is not possible to draw any conclusion from the data concerning the relation between the length of a peptide chain and the number of conformers. In fact, the only correlation evident in the data is that between the number of conformers and the length of the simulation, independent of the size of the peptide (the Pearson linear correlation coefficient for these data is 0.94). This might suggest that convergence has not been achieved in the reported simulations.

If the length dependence of the number of conformers were nonexponential, this would be very important. Because there are no detailed all-atom simulations for a range of polypeptide sizes that can be used to resolve this question, we consider the results available from simpler models and experimental data. Enumerations of self-avoiding random walks of simple polymers on lattices[7, 8] and analytical models[9] indicate an exponential dependence on chain length. Moreover, experimental estimates[10] for the configurational entropies of proteins obtained by subtracting the entropies of hydration for polar and nonpolar groups from measured entropies of unfolding vary essentially linearly with the length of the sequence (see Table X of ref. [10]), which implies exponential scaling of the number of conformers via the standard Boltzmann definition of the (microcanonical) entropy.

If there are 10^{24} conformers for a 100-residue protein and the interconversion between backbone conformations occurs on a ns time scale (see above), it would take about 2×10^7 years to sample all of them in a random search for the native

---

[*] Prof. M. Karplus
Department of Chemistry & Chemical Biology
Harvard University
Cambridge, MA 02138 (USA)
Fax: (+1) 617 496-3204
E-mail: marci@tammy.harvard.edu
and
Laboratoire de Chimie Biophysique, ISIS
Université Louis Pasteur
4, Rue Blaise Pascal
67000 Strasbourg (France)
Fax: (+33) 3-90-24-15-62

Dr. A. R. Dinner
Department of Chemistry, University of California
Berkeley, CA 94720 (USA)

[**] We thank Eugene Shakhnovich (Harvard University) for pointing out the references on lattice polymer simulations and very helpful discussions. We also thank Wilfred van Gunsteren for comments on the manuscript. A.R.D. is a Burroughs Wellcome Fund Hitchings–Elion Postdoctoral Fellow. The work done at Harvard was supported in part by the National Institutes of Health.

state. This argument is the essence of the Levinthal paradox. Its resolution is based on the existence of a large enough bias in the effective energy surface to guide the system to the native state, such that only a relatively small fraction of the accessible conformers is sampled along any given trajectory [see ref. [4], and references therein]. For example, in a 27-residue heteropolymer lattice model for a protein, there are a total of 10^{16} possible conformers (taking account of excluded volume). Of these, on the order of 10^{10} conformers (about 2.2 per dihedral angle) are accessible at the fastest folding temperature,[3, 4, 11] while less than 10^6 are sampled in finding the native state. For a 125-residue lattice model, there are about 10^{54} accessible conformers (about 2.7 per dihedral angle),[12] but the folding time for optimized sequences increases by only a factor of 10^2–10^3 over that found for the 27-residue model. These results indicate that in the lattice simulations it is not the number of accessible conformers, but rather the number that are visited during a folding trajectory that determines the folding time. Only a limited search appears to be necessary due to the large number of transition states.

While we agree with van Gunsteren et al.[1] that the denatured ensemble needs to be represented correctly to obtain physically meaningful simulations of folding, the native state has to be represented accurately as well. The stability and the structure of the native state strongly influence the shape of the free energy surface, as evidenced by statistical correlations between these properties and experimentally measured folding rates[13, 14] and by the ability to increase the folding rate for lattice polypeptides by optimizing the native state stability[15, 16] and structure.[17] Whether a molecular dynamics simulation with an accurate force field and a fast supercomputer can fold a protein in the available time, in spite of the 10^{24} or more accessible conformers, depends on the nature of the transition state ensemble obtained, which appears to represent a smaller portion of the configuration space than that in the lattice model studies.[18] With all the effort concentrated on the *two* protein folding problems ("Prediction of Structure from Sequence" on the one hand and "Determination of How a Polypeptide Folds to the Native State" on the other), we look forward to a continuation of the rapid progress made in this area during the recent past.

---

[1] W. F. van Gunsteren, R. Bürgi, C. Peter, X. Daura, *Angew. Chem.* **2001**, *113*, 363–367; *Angew. Chem. Int. Ed.* **2001**, *40*, 351–355.

[2] M. Karplus, *Folding Des.* **1997**, *2*, 569–576.

[3] C. M. Dobson, A. Šali, M. Karplus, *Angew. Chem.* **1998**, *110*, 908–935; *Angew. Chem. Int. Ed.* **1998**, *37*, 868–893.

[4] A. R. Dinner, A. Šali, L. J. Smith, C. M. Dobson, M. Karplus, *Trends Biochem. Sci.* **2000**, *25*, 331–339.

[5] M. Gruebele, *Annu. Rev. Phys. Chem.* **1999**, *50*, 485–516.

[6] A. R. Dinner, T. Lazaridis, M. Karplus, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9068–9073.

[7] D. S. McKenzie, *Phys. Rep.* **1976**, *27*, 35–88.

[8] V. S. Pande, C. Joerg, A. Yu. Grosberg, T. Tanaka, *J. Phys. A* **1994**, *27*, 6231–6236; Corrigendum: *J. Phys. A* **1996**, *29*, 4753.

[9] K. A. Dill, *Biochemistry* **1985**, *24*, 1501–1509.

[10] G. I. Makhatadze, P. L. Privalov, *Adv. Protein Chem.* **1995**, *47*, 307–425.

[11] A. Šali, E. Shakhnovich, M. Karplus, *Nature* **1994**, *369*, 248–251.

[12] A. R. Dinner, M. Karplus, *J. Phys. Chem. B* **1999**, *103*, 7976–7994.

[13] K. W. Plaxco, K. T. Simons, I. Ruczinski, D. Baker, *Biochemistry* **2000**, *39*, 11177–11183.

[14] A. R. Dinner, M. Karplus, *Nat. Struct. Biol.* **2001**, *8*, 21–22.

[15] E. I. Shakhnovich, A. M. Gutin, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 7195–7199.

[16] E. I. Shakhnovich, *Phys. Rev. Lett.* **1994**, *72*, 3907–3910.

[17] A. R. Dinner, E. Verosub, M. Karplus, *Protein Eng.* **1999**, *12*, 909–917.

[18] M. Vendruscolo, E. Paci, C. M. Dobson, M. Karplus, *Nature* **2001**, *409*, 641–645.

Reply

Wilfred F. van Gunsteren,* Roland Bürgi, Christine Peter, and Xavier Daura

In their comment,[1] Dinner and Karplus express disagreement with our recent communication to *Angewandte Chemie*[2] based on their interpretation of the latter. They stress that there are indications from lattice model simulations as well as from experiment that the number of conformers of a protein increases exponentially with the number of residues, and they claim that we conclude, based on molecular dynamics simulations of small peptides, that the number of conformations of a protein scales nonexponentially with the number of residues. In this reply, we want to comment on the arguments given by Dinner and Karplus and to repeat our arguments such that it is clear what we observe, what we conclude, and what we suggest.

Dinner and Karplus quote as one source of indications for an exponential dependence of the number of conformers on chain length two studies of "self-avoiding random walks of simple polymers on lattices" (see refs. [7, 8] in their comment) and an analytical model based on lattice statistical mechanics (see ref. [9] in their comment). It is well known that the structural properties of a protein are not only the result of short-range (covalent and repulsive van der Waals) interactions, but to a large extent due to long-range (attractive van

---

[*] Prof. Dr. W. F. van Gunsteren, R. Bürgi, C. Peter, Dr. X. Daura
Laboratory of Physical Chemistry
Swiss Federal Institute of Technology Zurich
ETH Hönggerberg, 8093 Zurich (Switzerland)
Fax: (+41) 1-632-1039
E-mail: wfvgn@igc.phys.chem.ethz.ch

0044-8249/01/11324-4752 $ 17.50+.50/0