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

  1433-7851/01/4024-4616 $ 17.50+.50/0

der Waals and electrostatic) interactions. Self-avoiding random walks model short-range interactions only. Introducing additional nonbonded forces alters the behavior of polymers significantly, as is discussed in ref. [7] of the comment.[1] On the experimental side, the entropies estimated in ref. [10] of the comment[1] include both backbone and side-chain contributions and are therefore very difficult to relate to the number of relevant (low free energy, accessed in "equilibrium folding/unfolding pathways") backbone conformers of a protein. Consequently, we do not consider refs. [7–10] of the comment[1] as proofs of an exponential scaling of the number of equilibrium conformers with respect to chain length.

In their comment and in references therein, Dinner and Karplus calculate the number of conformers per backbone torsional degree of freedom and extrapolate to the number of conformers of an *n*-residue protein by assuming in both cases an exponential scaling. Since we do not want to make assumptions on the nature of the scaling, we did not attempt in our communication to interpolate to the number of conformers per backbone torsional degree of freedom from the peptides studied, nor extrapolate to the number of conformers of an *n*-residue protein.

The exact number of conformers (clusters in our communication[2]) depends not only on the number of residues of the peptide, but also on the intrinsic structural properties of the peptide (hydrogen bonding capacity, steric properties, etc.). It makes no sense to calculate the correlation between the number of clusters and the number of backbone torsional angles for peptides with very similar numbers of degrees of freedom, since the noise will be necessarily big. The only cases where the calculation of this correlation seems appropriate is between the smaller peptide (**A** in our communication) and all the other ones (**B** to **F**), where the number of degrees of freedom duplicates. Considering this, we agree that "the number of conformers shows essentially no correlation with the number of torsion angles", especially not an exponential one.

The exact number of conformers depends also on the clustering algorithm and the clustering criterion. The clustering algorithm used in refs. [3–8] allows for the existence of clusters populated by a single member only, even when they are in structural vicinity of other clusters. These single-member clusters were also counted in Table 1 of our communication. In all simulations we observe a low number of clusters which are significantly populated.[3–8] Increasing the simulation time beyond a given number of folding/unfolding events causes an increase in the number of sparsely populated clusters. The significantly populated clusters, however, remain essentially the same.

The point of our communication is that it is necessary to sample the correct ensemble of low-free-energy equilibrium conformers in order to draw conclusions on polypeptide folding. It goes without saying that "the native state has to be represented accurately as well". Proteins are inhomogeneous molecules. It is precisely this inhomogeneity and the very specific properties of each of the amino acids that makes them biologically important. Therefore, it is extremely important to model the representation of the protein and the interaction functions in such a way that the properties of interest are reproduced. In contrast to the studies cited in the comment,[1] we have shown that the ensembles of structures taken from our simulations accurately reproduce the available NMR data,[3–8] including NMR spectra for one peptide,[9] rather than just reproducing the predominant structure or justifying the simulation results with diffuse qualitative arguments. For the six peptides studied the number of relevant conformers is of the order $10^2 - 10^3$. Considering that these peptides fold in the simulations on a time scale of 10^{-9} s, we *suggest* that a protein that folds in a time scale of 10^{-3} s and shows similar equilibrium properties would access a space of around 10^9 relevant conformers. This estimate is based on the assumption that the number of relevant conformers scales linearly with the folding time, and *not*, as suggested by Dinner and Karplus,[1] by assuming a particular scaling of the number of conformers with respect to the number of backbone torsional angles.

We agree with Dinner and Karplus that it is the number of conformers "that are visited during a folding trajectory that determines the folding time". In contrast to the simulations they quote, our simulations sample an equilibrium of folded and unfolded states. All the conformations we observe are, therefore, part of an "*equilibrium* folding/unfolding pathway". If the reader would like to put it in terms of the protein folding funnel picture[10] this would mean that all conformers sampled belong to the bottom part of the funnel, that is are conformers with a low free energy. Since we do not randomize the conformations during the simulation, they do never leave this region of low free energy. We agree with ref. [11] in the comment[1] that the first stage of folding from a *random* high-free-energy conformation to low-free-energy conformations is fast. However, we consider it plausible that, once this region of low free energy is reached, the conformations are sampled according to their Boltzmann weight. This would imply that the folding time is roughly proportional to the number of accessible equilibrium conformers.

The argument of Dinner and Karplus against the dependence of the folding time on the number of accessible conformers is based on Monte Carlo simulations on a lattice of the folding of a 27-bead and a 125-bead model from random conformations to the "native" state (see refs. [11, 12] in their comment). Interactions are only calculated for nonbonded nearest neighbors. The number of accessible conformers for the smaller model is estimated using the number of native contacts as a coordinate to identify the accessible states; in the case of the larger model two subsets of native contacts are used. As mentioned above (second paragraph), we do not agree with the simplifications used in these models. Furthermore, the use of (subsets of) native contacts as reaction (folding) coordinate leads to a projection which we question. Therefore, we do not consider these calculations valid estimates of the number of equilibrium conformers of a protein of 27 and 125 residues, respectively.

In summary, we do not believe that a definitive scaling law for the number of equilibrium conformers as a function of the number of backbone torsional degrees of freedom can be derived from currently available theoretical or experimental data. Our simulations indicate rather a nonexponential scaling. We want to stress again that, in our opinion, it is

  1433-7851/01/4024-4617 $ 17.50+.50/0

fundamental to describe not only the native state but also the denatured state accurately in order to draw conclusions on the nature and mechanisms of protein folding. Only if the denatured state is well represented, conclusions can be made with respect to folding pathways, time scales, and reaction coordinates.

[1] A. R. Dinner, M. Karplus, *Angew. Chem.* **2001**, *113*, 4751 – 4752; *Angew. Chem. Int. Ed.* **2001**, *40*, 4615 – 4616.

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

[3] X. Daura, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark, *J. Mol. Biol.* **1998**, *280*, 925 – 932.

[4] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark, *Angew. Chem.* **1999**, *111*, 249 – 253; *Angew. Chem. Int. Ed.* **1999**, *38*, 236 – 240.

[5] C. Peter, X. Daura, W. F. van Gunsteren, *J. Am Chem. Soc.* **2000**, *122*, 7461 – 7466.

[6] R. Bürgi, X. Daura, A. Mark, M. Bellanda, S. Mammi, E. Peggion, W. F. van Gunsteren, *J. Peptide Res.* **2001**, *57*, 107 – 118.

[7] X. Daura, K. Gademann, H. Schäfer, B. Jaun, D. Seebach, W. F. van Gunsteren, *J. Am. Chem. Soc.* **2001**, *123*, 2393 – 2404.

[8] C. Santiveri, M. A. Jiménez, M. Rico, W. F. van Gunsteren, X. Daura, personal communication.

[9] C. Peter, X. Daura, W. F. van Gunsteren, *J. Biomol. NMR* **2001**, *20*, 297 – 310.

[10] P. E. Leopold, M. Montal, J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 8721 – 8725.

  1433-7851/01/4024-4618 $ 17.50+.50/0