SWARM-MD: Searching Conformational Space by Cooperative Molecular Dynamics

Thomas Huber[†] and Wilfred F. van Gunsteren*

Physical Chemistry, ETH Zentrum, CH-8092 Zürich, Switzerland

Received: January 5, 1998; In Final Form: April 20, 1998

A simulation algorithm is introduced, which uses a swarm of molecules to explore conformational space. The method uses multiple, different starting conformations and propagates them in time by integration of Newton's equation of motion. In contrast to conventional molecular dynamics simulation of a set of independent molecules, in this method each molecule of the swarm is in addition subject to an artificial field that keeps the trajectory of individual molecules tied to the average trajectory of the swarm. In this manner, a search for the global energy minima of many molecules is transformed into a cooperative search. It is shown that such a cooperative search is less attracted by local minima in the potential energy surface and that the total system is more likely to follow an overall potential energy gradient toward the global energy minima.

Introduction

The yearly increase of computer power brings the idea of simulating protein folding on a computer closer and closer to reality. If one day a sufficiently powerful supercomputer would be built that could simulate the time evolution of a protein over a period of a few seconds, the protein folding problem would be nearly solved. This would certainly be true if simulation would exactly reproduce reality, since it is a well-known fact that proteins are able to fold into a unique structure on a subsecond to minute time scale.

To explain the amazingly high success rate of proteins folding into their native conformation, funnel-like potential energy surfaces have been postulated.[1−3] Following this hypothesis, one should be able to simply follow the funnel by simulation of a protein using molecular dynamics techniques, i.e., by solving Newton's equation of motion. In practice, this idea breaks down for several reasons. First, the potential energy function is usually very complex with many barriers, and all but the smallest are surmounted very infrequently. Second, the energy hypersurface is of high dimensionality. This not only means that the conformational space to be searched is vast but also that many, kinetically equally likely, branching pathways may exist which lead to local minima on the energy surface. Keeping in mind that molecular dynamics simulation is a heuristic method to explore conformational space, kinetically even less likely pathways may be followed during a simulation.

One way toward the goal of making successful protein folding simulation more realistic is to develop new simulation methods in which the trajectory of a molecule is less affected by inaccuracies and details of the energy function and which thus make the trajectory of the molecule more likely to be a pathway of folding. Several methods have been proposed that rely on smoothing of the molecular potential energy surface. The probably most popular approach is simulated annealing (SA)[4] which takes advantage of smoothing an energy surface by entropic contributions. One might attempt to smooth a surface by filling in the energy hypervalleys.[5−7] Several other methods temporarily add artificial degrees of freedom[8−10] which can be seen as smoothing the energy surface with respect to the initial, physically real degrees of freedom. Some approaches are based on the smoothing of a mean field potential energy surface[11−18] or on a fuzzy description of the system[19−22] and, thus, a smoother energy surface. The strategy is more obvious in methods that apply smoothing procedures directly such as the deflation method[23] and the diffusion equation method (DEM).[24−29] In the last method, the diffusion equation is solved analytically for potential energy surfaces, and the original potential energy surface is restored by a time reversal process.

Our method here uses a slightly different approach to make a simulation less dependent on the detailed structure of an energy surface. The basic idea comes from model simulations of social insects.[30−32] These simulations were motivated by the astonishing insight that intelligent and efficient behavior of a whole swarm of insects can be achieved even in the absence of any particular intelligence or forethought of the individuals. Our method, in the following called SWARM-MD, transfers this kind of simulation to molecular modeling, and we investigated the effect of cooperative behavior of a swarm of molecules for conformational search.

A closely related, but different, method is the use of any additional term in the potential energy function that energetically penalizes the pairwise conformational differences between all pairs of a set of identical homologeous polymers.[33,34] It was inspired by the idea to use the similarity in structure of heteropolymers for guiding the search for the global minimum in Monte Carlo lattice simulations.

Theory

The concept of the method is based on the idea to combine a swarm of molecules with molecular trajectories into a cooperative system that searches conformational space. To build such a cooperative system, each molecule is, in addition to physical forces, subject to (artificial) forces that drive the trajectory of each molecule toward an average of the trajectories of the swarm of molecules.

Swarm Interaction Function. The interaction between molecules within the swarm is modeled by a function that exponentially decays with increasing distance measure. A

* To whom correspondence should be addressed.
[†] Present address: Supercomputer Facility, Australian National University, Canberra, ACT 0200, Australia.

S1089-5639(98)00625-2 CCC: $15.00
Published on Web 06/20/1998

widely used distance measure between two structures is the so-called root-mean-square positional distance (RMSD). This metric, however, requires to optimally superimpose molecules to remove translational and rotational shift, and it is, in this context, therefore less convenient to work with. The distance measure used in this work is the root-mean-square dihedral angle difference (DHAD) of N (selected) dihedral angles ϕ_i^j in molecule j and their corresponding swarm averages $\langle\phi_i\rangle$ over M molecules:

$$\langle\phi_i\rangle = \frac{1}{M}\sum_{j=1}^{M}\phi_i^j \tag{1}$$

Thus, the additional potential energy function that forces the swarm of molecules toward behaving as an organized system is defined by

$$V(\{\phi_i^j\})_{\text{swarm}} = \sum_{j=1}^{M} A\exp(-B\,\text{DHAD}) = \sum_{j=1}^{M} A\exp\left[-B\left(\frac{1}{N}\sum_{i=1}^{N}\left(\phi_i^j - \frac{1}{M}\sum_{j'=1}^{M}\phi_i^{j'}\right)^2\right)^{1/2}\right] \tag{2}$$

where the parameters A and B are used to define intercept and slope of this function.

As can easily be seen from this equation, the additional potential energy term V_{swarm} is dependent on the relative positions of molecules with respect to the swarm average, and the energy of a member of the swarm is not conserved.

The corresponding additional force acting on an atom k in molecule j is given by the opposite of the first derivative of the function V_{swarm},

$$\vec{f}_k = -\frac{\partial V_{\text{swarm}}}{\partial \vec{r}_k} = A\exp\left[-B\left(\frac{1}{N}\sum_{i=1}^{N}(\phi_i^j - \langle\phi_i\rangle)^2\right)^{1/2}\right]\times B\sum_{i=1}^{N}\left(\frac{(1/N)(\phi_i^j - \langle\phi_i\rangle)}{[(1/N)\sum_{i'=1}^{N}(\phi_{i'}^j - \langle\phi_{i'}\rangle)^2]^{1/2}}\right)\left(1 - \frac{1}{M}\right)\frac{\partial\phi_i^j}{\partial\vec{r}_k} \tag{3}$$

Methods

As a test, three distinct model molecules were used which only differ in the location of their global energy minima on the potential energy surface. The models are alkane-like chains with 50 methylene groups in a united atom representation. Nonbonded interactions were set to zero in these models. The location of the global energy minimum is therefore determined by a dihedral angle energy potential function of the form

$$V_\phi = k_\phi^1(1 + \cos(n_\phi^1\phi - \delta^1)) + k_\phi^2(1 + \cos(n_\phi^2\phi - \delta^2)) \tag{4}$$

For each dihedral angle in the chain the same potential energy function was used. The parameters and the dihedral angle value of lowest energy are given in Table 1 Figure 1 shows the potential energy profiles.

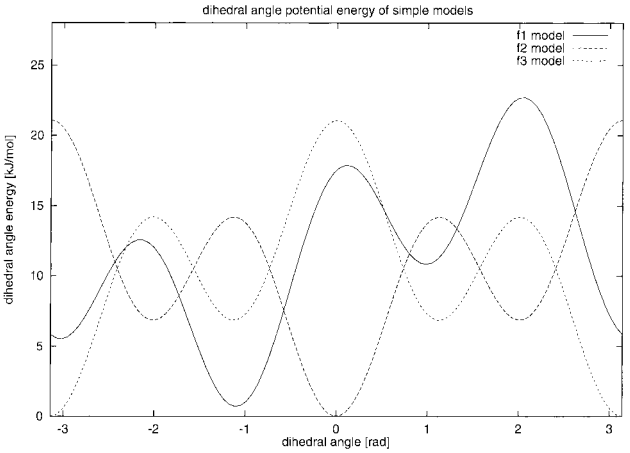

Figure 1. Dihedral angle energy functions (4) used in three simple molecular models. The parameters are listed in Table 1.

TABLE 1: Parameters of the Two Dihedral Angle Potential Energy Function Terms and the Dihedral Angle Value of Lowest Energy for Three Molecular Models

	model f1	model f2	model f3
k_ϕ^1 [kJ mol^{-1}]	5.85760	5.85760	5.85760
n_ϕ^1	3	3	3
δ^1 [deg]	180.0	180.0	180.0
k_ϕ^2 [kJ mol^{-1}]	5.85760	4.68608	4.68608
n_ϕ^2	1	1	1
δ^2 [deg]	90.0	180.0	0.0
ϕ_0 [deg]	−62.91	0.0	180.0

All molecular dynamics simulations were conducted in vacuo using the GROMOS87 37D4 vacuum force field parameters[35] and a leapfrog integration scheme with a time step of 2 fs. The SHAKE algorithm[36] was applied to constrain bond length with a relative accuracy of 10^{-4}. In the model systems bond angles were dynamically restrained using the GROMOS bond angle potential energy function and parameters for hydrocarbons. Berendsen's weak coupling method[37] was individually applied to each molecule to control the temperature. Expecting a high flux of energy during the folding process and in order to keep the temperature close to that of the simulated annealing protocol, when applied, a small temperature coupling time constant $\tau_T = 0.05$ ps was used. In the protein simulations, nonbonded interactions were evaluated with truncation of forces beyond a cutoff radius of 1 nm, while updating an interaction pair list every five simulation steps.

In simulations of model molecules, simulated annealing was performed by exponentially lowering the temperature from 400 to 100 K within the simulation time of 100 ps. In SWARM-MD simulation the parameters A and B in eq 2 were chosen to be $A = -200$ kJ mol^{-1} and $B = 0.8$ rad^{-1}, and the set of all $N = 47$ dihedral angles ϕ along the linear chain were used to calculate the additional potential energy contribution defined by eq 2.

The protein analyzed was the structured domain of chymotrypsin inhibitor 2 (CI-2, PDB acquisition code 3ci2), a 64-residue protein without disulfide bridges. Starting from a minimized structure of refined CI-2,[38,39] 50 conventional MD simulations of 1 ns at 500 K and 500 ps at 600 and 1000 K were performed to sample conformational pathways of unfolding of the protein. The runs were identical in setup, only differing in the initial random velocity assignment corresponding to the given temperature. RMSD values were calculated for C^α atoms only, and the calculation of DHAD was performed using backbone ϕ and ψ dihedral angles.

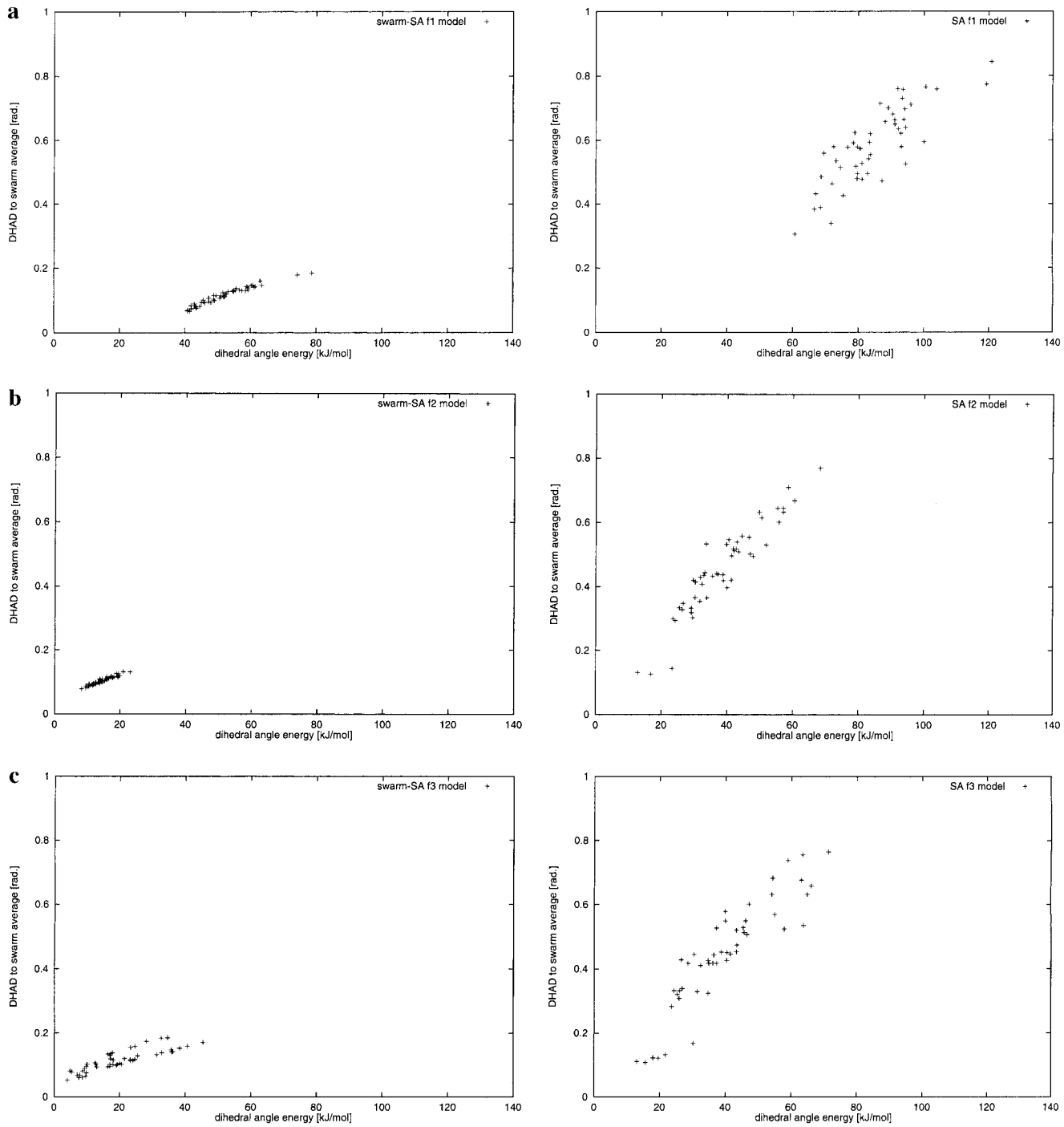

Figure 2. Comparison of the geometric spread of structures produced by the two optimization methods: Markers show the root-mean-square dihedral angle difference (DHAD) of all angles in the final structure compared to the lowest energy conformation. Left: 50 simulated annealing MD simulations of one molecule. Right: one simulated annealing SWARM-MD simulation with 50 molecules. (a) Model f1, (b) model f2, (c) model f3.

Results and Discussion

1. Model Systems. To test new search methods in molecular modeling with respect to their performance, it is essential to choose a molecular(- like) system for which knowledge about the shape of its potential energy surface is available. In many cases molecular systems are used to which several search and simulation methods were already applied, and thus, knowledge of their potential energy surface has been obtained. This approach, however, sets a limit to the size of a test system, since only the energy surface of relatively small molecules can be explored sufficiently by simulation. An alternative approach, which does not have an intrinsic limit with respect to size or complexity of the system, is to construct a model system for which the potential energy function is a sum of terms that are independent of each other. Test systems of this kind are used here. Nonbonded interactions are not considered, and the dihedral angle energy contributions to the total potential energy of the individual dihedral angles are independent of each other. Three different models of this type are used, and two features are common to all: (1) an overall energy gradient toward the global minimum (funnel) and (2) a high number of local minima ($3^{47} - 1$) separated by energy barriers of similar height. The three models only differ in the positions of the global energy

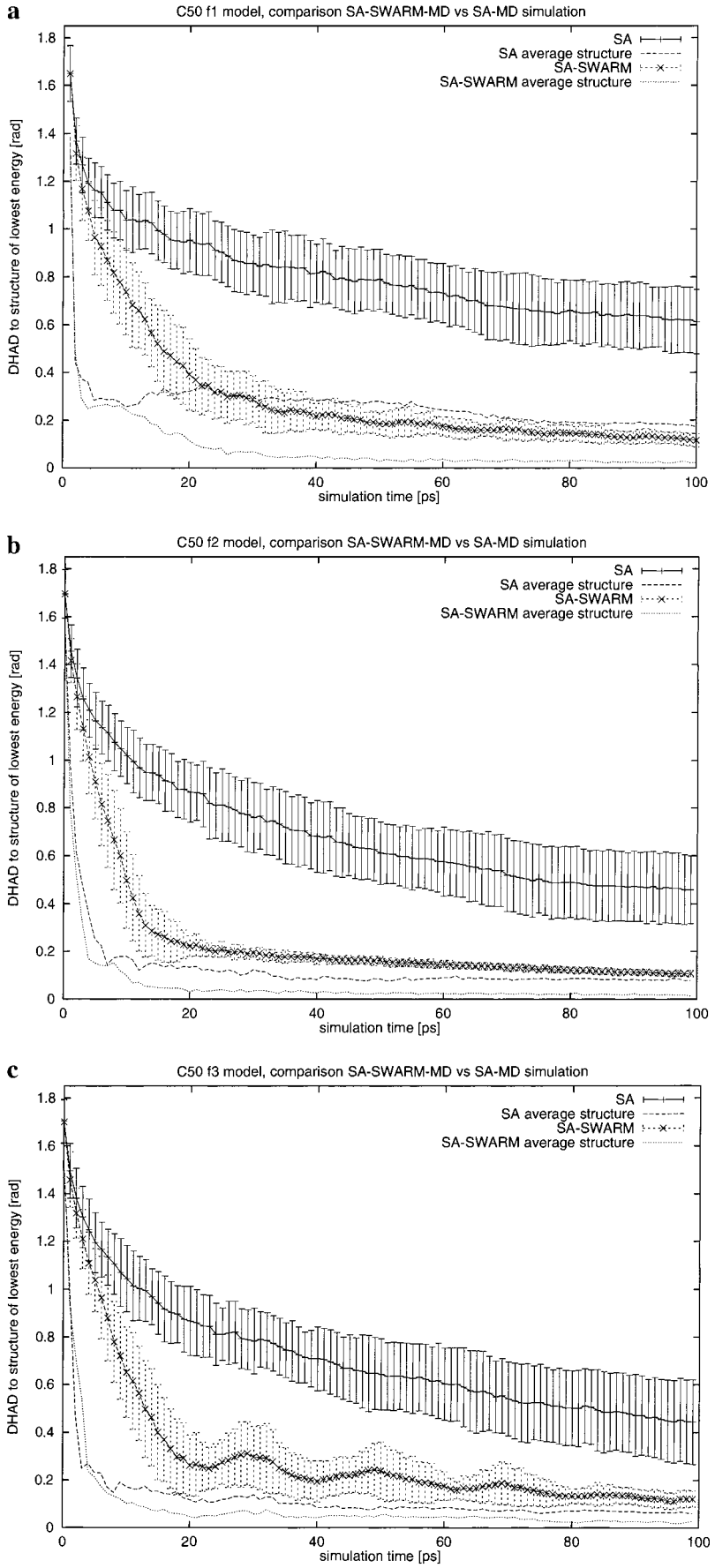

Figure 3. Comparison of 50 independent simulated annealing MD simulations of one molecule with one simulated annealing SWARM-MD simulation of 50 molecules. Lines with error bars indicate averages of individual root-mean-square dihedral angle differences (DHAD) to the lowest energy structure and standard deviation as a function of simulation time. Single lines indicate DHAD of the swarm dihedral angle averaged structure to the lowest energy structure. (a) Model f1, (b) model f2, (c) model f3.

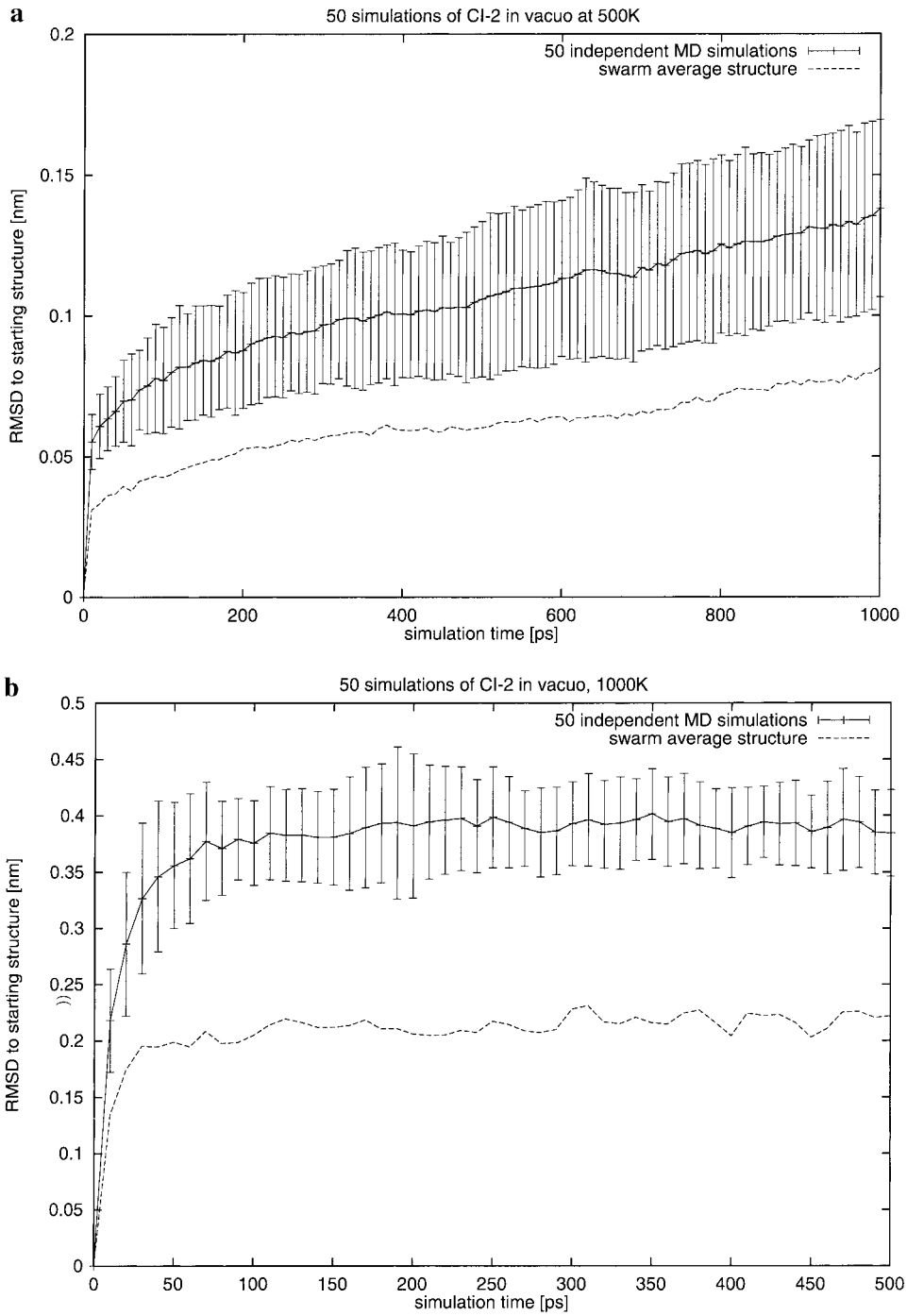

Figure 4. Comparison of the geometric evolution of structures in the course of 50 independent parallel unfolding simulations of CI-2 from the native starting structure. The continuous line with error bars indicates the average root-mean-square positional difference (RMSD) of C^{α} atoms and standard deviation. The dashed line shows the RMSD of the swarm positional coordinates averaged structure from all 50 independent parallel unfolding simulations. (a) At 500 K simulation temperature, (b) at 1000 K simulation temperature

minimum on the energy surfaces which were chosen to be distinctly different from each other.

There certainly are optimization methods that perform particularly well for these model systems. However, although these models are not suitable to judge the performance of all types of optimization methods, in our opinion, they are well suited for testing optimization methods based on molecular dynamics, such as analyzed here.

2. Results on Model Systems. Since in our model system the position of the global minimum is exactly known, one can analyze optimized structures relative to the structure of lowest energy. In Figure 2 the root-mean-square dihedral angle difference (DHAD) versus the dihedral angle potential energy is plotted for final structures of 50 simulated annealing MD simulations of one molecule and one simulated annealing SWARM-MD simulation using 50 molecules. Since simulated annealing is a heuristic method, calculations were repeated many times using starting structures with randomized dihedral angles and different initial velocity assignments. The same starting structures and initial conditions were used in the SWARM-MD simulation. Final structures from the SWARM-MD simulation show a smaller spread in energy and DHAD value and are closer to the global energy minimum. The smaller spread is partly due to the additional (SWARM) potential energy contribution, which reduces the spread.

Convergence to structures closer to the global energy

minimum illustrates that with cooperative search a molecule is less likely to get trapped in a local minimum, respectively, and is more easily able to surmount energy barriers with the help of forces driving it toward the dihedral angle averaged structure.

In Figure 3 the DHAD trajectory of the dihedral angle averaged structure and a statistical description from 50 individual structures are shown for both methods. All simulations start close from a DHAD value of random dihedral angle deviation ($DHAD_{random} = 1.81$ rad) and rapidly converge in the course of simulation. In all calculations the average structure approaches the structure of lowest energy faster than individuals do. This need not necessarily be the case but is based on the fact that individual dynamic systems distribute themself around the attracting global minimum. Thus, it is expedient to encourage individual structures to follow the average.

3. Analysis of Protein Unfolding Simulations. The success of the SWARM-MD method relies on the fact that the average structure of a swarm of molecules converges faster to the structure with lowest energy than individual molecules do. Whether this is true for more realistic models of molecules has not been shown so far. To test this hypothesis, we performed a great number of high-temperature unfolding simulations of the protein CI-2 and monitored the divergence of structures. A rather basic model was used. Solvent is neglected in the calculations, and it is not likely that the native conformation corresponds to the global energy minimum of the potential energy surface of this model. Furthermore, the process of high-temperature unfolding of a protein is not necessarily comparable to that of folding a protein. The aim of these simulations is not to give a detailed description of how this protein unfolds, but only to monitor the extent and possible direction of divergence of conformations from a well distinct energy minimum on the potential energy surface. In Figure 4 the time evolution of the RMSD value with respect to the starting structure is shown for 50 simulations starting with different initial velocities and at different temperatures. In addition, the RMSD between the starting structure and the average positional coordinates of structures from 50 parallel simulations is indicated by a dashed line. Similar to the observation with the simple model molecules, in all cases the distance from the native protein conformation to the average conformation is significantly smaller than the distance between any individual structure of the simulations and the starting structure. This holds, even when in the progress of a high-temperature simulation structures become quite dissimilar to each other and to the starting conformation. A similar trend is observed when the deviation measure in terms of internal coordinates (DHAD) is analyzed instead of the deviation in terms of positional coordinates (RMSD) (data not shown).

Therefore, in a refolding simulation using a more appropriate model to describe the system, one would expect that the convergence to the native (and global minimum) conformation is sped up when the system experiences an additional force driving the molecules toward their average structure.

Conclusion

It is shown that in a dynamics simulation average coordinates of a swarm of structures usually converge faster to the conformation with lowest energy than any individual within the collection of structures. On the basis of this observation, we proposed a method that biases the trajectory of motion of molecules in a swarm toward the swarm's average structure by the use of artificial forces. In this way searching properties of the whole system are enhanced. It is shown that the use of such a cooperatively acting system of many molecules is more likely to converge to the global minimum of the energy surface than a single molecule. One must, however, note that the dynamics of the system is, of course, artificial. The method changes the energy of molecules each time the swarm-averaged structure changes. Clearly this kind of searching can only be used when some kind of temperature regulation for the individual molecules is applied.

A clear disadvantage of the method is that the simultaneous use of numerous explicit molecules is computationally expensive. The concept of the method, however, is highly suited for parallel computer architectures for which a nearly optimal balance of processor load can be achieved. Furthermore, the amount of information exchanged between molecules or processes at each step is small, which makes a near linear scaling of performance with the number of processors is attainable.

References and Notes

(1) Betancourt, M.; Onuchic, J. *J. Chem. Phys.* **1995**, *103*, 773−787.
(2) Onuchic, J.; Wolynes, P.; Luthey-Schulten, Z.; Socci, N. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3626−3630.
(3) Bryngelson, J.; Onuchic, J.; Socci, N.; Wolynes, P. *Proteins* **1995**, *21*, 167−195.
(4) Kirkpatrick, S.; Gelatt Jr., C.; Vecchi, M. *Science* **1983**, *220*, 671−680.
(5) Glover, F. *ORSA J. Comput.* **1989**, *1*, 190−206.
(6) Glover, F. *ORSA J. Comput.* **1989**, *2*, 4−31.
(7) Huber, T.; Torda, A.; van Gunsteren, W. *J. Comput. Aided Mol. Design* **1994**, *8*, 695−708.
(8) Crippen, G. *J. Comput. Chem.* **1982**, *3*, 471−476.
(9) Purisima, E.; Scheraga, H. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 2782−2786.
(10) van Schaik, R.; Berendsen, H.; Torda, A.; van Gunsteren, W. *J. Mol. Biol.* **1993**, *234*, 751−762.
(11) Gerber, R.; Buch, V.; Ratner, M. *J. Chem. Phys.* **1982**, *77*, 3022−3030.
(12) Straub, J.; Karplus, M. *J. Chem. Phys.* **1990**, *94*, 6737−6739.
(13) Roitberg, A.; Elber, R. *J. Chem. Phys.* **1991**, *95*, 9277−9287.
(14) Olszewski, K.; Piela, L.; Scheraga, H. *J. Phys. Chem.* **1992**, *96*, 4672−4676.
(15) Zheng, Q.; Rosenfeld, R.; Vajda, S.; DiLisi, C. *Protein Sci.* **1993**, *2*, 1242−1248.
(16) Koehl, P.; Delaure, M. *J. Mol. Biol.* **1994**, *239*, 249−275.
(17) Mierke, D.; Scheek, R.; Kessler, H. *Biopolymers* **1994**, *34*, 559−563.
(18) Huber, T.; Torda, A.; van Gunsteren, W. *Biopolymers* **1996**, *39*, 103−114.
(19) Amara, P.; Hsu, D.; Straub, J. *J. Phys. Chem.* **1993**, *97*, 6715−9721.
(20) Ma, J.; Hsu, D.; Straub, J. E. *J. Chem. Phys.* **1993**, *99*, 4024−4036.
(21) Ma, J.; Straub, J. E. *J. Chem. Phys.* **1994**, *101*, 533−541.
(22) Orešič, M.; Shalloway, D. *J. Chem. Phys.* **1994**, *101*, 9844−9857.
(23) Crippen, G.; Scheraga, H. *Proc. Natl. Acad. Sci. U.S.A.* **1969**, *64*, 42−49.
(24) Piela, L.; Kostrowicki, J.; Scheraga, H. *J. Phys. Chem.* **1989**, *93*, 3339−3346.
(25) Kostrowicki, J.; Piela, L. *J. Opt. Theory* **1989**, *69*, 269−284.
(26) Kostrowicki, J.; Piela, L.; Cherayil, B.; Scheraga, H. *J. Phys. Chem.* **1991**, *95*, 4113−4119.
(27) Wawak, R.; Wimmer, M.; Scheraga, H. *J. Phys. Chem.* **1992**, *96*, 5138−5145.
(28) Kostrowicki, J.; Scheraga, H. *J. Phys. Chem.* **1992**, *96*, 7442−7449.
(29) Huber, T.; Torda, A.; van Gunsteren, W. *J. Phys. Chem. A* **1997**, *101*, 5926−5930.
(30) Hiebeler, D. The swarm simulation system and individual-based modeling Technical report, The Santa Fe Institute, Santa Fe, 1994.
(31) Chialvo, D.; Millonas, M. How swarms build cognitive maps Technical report, The Santa Fe Institute, Santa Fe, 1995.
(32) Rauch, E.; Millionas, M.; Chialvo, D. *Phys. Lett. A* **1995**, *A207*, 185−193.
(33) Keasar, C.; Elber, R. *J. Phys. Chem.* **1995**, *99*, 11550−11556.
(34) Keasar, C.; Elber, R.; Skolnick, J. *Folding Design* **1997**, *2*, 247−259.
(35) Groningen molecular simulation (GROMOS) library manual. van Gunsteren, W.; Berendsen, H.; Biomos, Groningen, The Netherlands, 1987.

(36) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys.* **1977**, *23*, 327−341.
(37) Berendsen, H.; Postma, J.; van Gunsteren, W.; DiNola, A.; Haak, J. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(38) Nanzer, A.; Poulson, F.; van Gunsteren, W.; Torda, A. *Biochemistry* **1994**, *33*, 14503−14511.
(39) Nanzer, A.; van Gunsteren, W.; Torda, A. *J. Biomol. NMR* **1995**, *6*, 313−320.