Thomas Huber*
Andrew E. Torda*
Wilfred F. van Gunsteren
Physical Chemistry
ETH Zentrum
CH-8092 Zürich, Switzerland

Optimization Methods for Conformational Sampling Using a Boltzmann-Weighted Mean Field Approach

An optimization protocol is proposed that combines a mean field simulation approach with Boltzmann-weighted sampling. This is done by including Boltzmann probabilities of multiple conformations in the optimization procedure. The method is demonstrated on a simple model system and on the side-chain conformations of phenylalanines in a small hexapeptide. For comparison, calculations were performed using classical stochastic dynamics simulations [*M. Saunders, K. N. Houk, Y. Wu, C. Still, M. Lipton, G. Chang, and W. C. Guida (1990)*, Journal of the American Chemical Society, *Vol. 12, pp. 1419*], *iterative optimization of probabilities on a fixed set of basis conformations* [*P. Koehl and M. Delaure (1994)*, Journal of Molecular Biology, *Vol. 239, pp. 249–275*], *and simulations with locally enhanced sampling* [*A. Roitberg and R. Elber (1991)*, Journal of Chemical Physics, *Vol. 95, pp. 9277–9287*]. *Although approximations are used in our method, the results may be more physically meaningful than those of the other procedures discussed. Furthermore, the approximate Boltzmann distribution allows generalization of the results.* © *1996 John Wiley & Sons, Inc.*

INTRODUCTION

Computer simulations of biological molecules may provide information at the atomistic level and offer insights unavailable with other methods. Generally, one wants to find the ensemble of structures spanning the physically important low energy conformations. Unfortunately, standard heuristic conformational search methods like Metropolis Monte Carlo (MC) or molecular dynamics (MD) simulations are poor at crossing energetic barriers at reasonable temperatures and they sample only small regions of conformational space.

Many methods based on dynamic or nondynamic schemes have been developed to attempt to search broader regions efficiently for low energy conformations. The most commonly applied is probably simulated annealing,[1] which relies on temperature evolution in time. Potential energy annealing conformational search[2] attempts to take advantage of the shape of the potential energy hypersurface. The local elevation technique[3] includes a long-time memory and enhances sampling by local modifications to the potential surface. There are several other methods that rely on temporarily increasing the dimensionality by adding artificial degrees of freedom.[4–9] Gerber et al.[10] introduced the idea of using the classical time-dependent Hartree (TDH) approximation for faster sampling. This uses the algebraic average field of a "bundle" of trajectories and was modified, investigated, and successfully applied by several authors,[11–16] including

Received May 15, 1995; accepted October 30, 1995.

* Present address: Research School of Chemistry, Australian National University, Canberra, 0200, Australia.

Biopolymers, Vol. 39, 103–114 (1996)

CCC 0006-3525/96/010103-12

Elber and co-workers[17–21] with his variation called locally enhanced sampling (LES).

A drawback of the current methods is that they are pure search procedures, usually with an unknown distribution of energies among conformations. This means that physical properties may be incorrectly estimated and a statistical generalization of the results can be difficult. Furthermore, in some systems, the global minimum may not be a satisfactory solution by itself. Observed properties reflect the distribution of the ensemble across all accessible minima, not just a single, idealized minimum energy structure.

Our approach[22] is intended to improve sampling of conformational space while giving an approximate Boltzmann distribution. This was done using an extension of the classical time-dependent Hartree approximation. The method can be seen as a modification of the TDH approximation with multiple copies of a system or part of a system. Normally, each copy within a system would be of equal probability, independent of time. In contrast, our method treats the probabilities of conformations as additional variables with the only constraint that the probabilities satisfy the Boltzmann relation. With these extra degrees of freedom, a system may more easily adopt a low (mean field) energy since it can change the mean conformation through either a change in the probability distribution or through conformational changes.

THEORY

We wish to obtain a Boltzmann ensemble of conformations of some system. This might be either a whole set of molecules, like a drug in a receptor surrounded by solvent, a molecule from the set, like the drug in a receptor, or a part of a molecule. The n atoms of the molecular system have atomic coordinates indicated by x_i ($i = 1, 2, \ldots, n$). The conformation X_α for molecule α is specified in terms of a particular set of atomic coordinate values.

$$X_\alpha = \{x_1^\alpha, x_2^\alpha \cdots x_n^\alpha\} \tag{1}$$

The energy of the conformation is given by

$$E_\alpha = V(X_\alpha) \tag{2}$$

where $V(X_\alpha) = V(\{x_1^\alpha, x_2^\alpha \cdots x_n^\alpha\})$ is an atomic interaction function.

For a Boltzmann ensemble of molecular conformations, the relationship between the probability of occurrence, indicated by p_α, and energy E_α of a conformation is given by the Boltzmann relation

$$p_\alpha = \frac{\exp[-(E_\alpha/k_BT)]}{\sum_{\alpha=}^{N_A} \exp[-(E_\alpha/k_BT)]} \tag{3}$$

where T is the absolute temperature, k_B is Boltzmann's constant, and the summation runs over all N_A members of the ensemble of conformations.

Physical properties of the molecular system are obtained as averages over the ensemble of conformations. For a physical quantity (observable) Q, which is a function of atomic coordinates, the ensemble average $\langle Q \rangle$ is defined by

$$\begin{aligned} \langle Q \rangle &= \sum_\alpha p_\alpha Q(X_\alpha) \\ &= t_{MD}^{-1} \int_0^{t_{MD}} Q[X(t')]\, dt' \quad \text{molecular dynamics} \\ &= N_{MC}^{-1} \sum_{\alpha \in MC}^{N_{MC}} Q(X_\alpha) \quad \text{Metropolis Monte Carlo} \end{aligned} \tag{4}$$

In MD simulation, the ensemble average is obtained as a time average over the atomic trajectories of time length t_{MD}. In MC simulation, the ensemble average is obtained as an unweighted average over the sequence of N_{MC} conformations. For the average energy of the molecular system, one obtains

$$\langle E \rangle = \sum_\alpha p_\alpha E_\alpha = \sum_\alpha p_\alpha V(X_\alpha) \tag{5}$$

In the mean field description, the molecular system is separated in M parts ($A, B, \cdots$), each represented by N_A identical copies ($\alpha, \beta, \cdots$) of O atoms ($i, j, \cdots$). Each of these copies moves in the mean field $\langle E \rangle$ of all other parts, given by

$$\langle E \rangle = E_{\text{tot}} = \sum_A^M \sum_\alpha^{N_A} p(A, \alpha) E(A, \alpha) \tag{6}$$

with $p(A, \alpha)$ the relative Boltzmann probability of copy α in part A

$$p(A, \alpha) = \frac{\exp\{-[E(A, \alpha)/k_BT]\}}{\sum_\alpha^{N_A} \exp\{-[E(A, \alpha)/k_BT]\}} \tag{7}$$

and $E(A, \alpha)$ the energy of a copy α in part A. For example, one could explicitly write the nonbonded interactions as

$$E(A,\alpha) = \tfrac{1}{2}\sum_{i}^{O_{A,\alpha}}\sum_{B\neq A}^{M}\sum_{\beta}^{N_B}\sum_{j}^{O_{B,\beta}} p(B,\beta) \times V(A,\alpha,i;B,\beta,j) + \tfrac{1}{2}\sum_{i}^{O_{A,\alpha}}\sum_{j\neq i}^{O_{A,\alpha}} V(A,\alpha,i;A,\alpha,j) \quad (8)$$

where $V(A, \alpha, i; B, \beta, j)$ is the atomic interaction function of atom i in copy α of group A with atom j in copy β of group B. From the last formula one can see that copy α of a group does not interact with any other copy representing this group and that the full interaction between atoms of this copy contributes to its energy. Next, the energy $E(A, \alpha)$ of a copy α depends on the set of probabilities from all copies of all other groups, but not on its own probability. With Eq. (7) this gives a self consistent set of nonlinear equations for the energies and probabilities of all copies in the system.

The computational problem is now to minimize $\langle E \rangle$ with the Boltzmann relation (7) as constraint. That is, one must find a Boltzmann-distributed ensemble of configurations on a very high-dimensional and complex interaction function $V(X_\alpha)$ as is generally used in biomolecular modeling.

Since the average energy $\langle E \rangle$ in Eq. (6) depends on both conformational coordinates X_α and conformational probabilities p_α of all copies, we distinguish three types of optimization algorithms.

Variation of Conformational Coordinates with Fixed Probability Distribution

Probabilities can be treated as fixed. This requires the approximation that a variation of conformational coordinates does not affect the probability distribution.[11-21] Because of the possible interchange of conformations, this results in uniform probability distributions. This, however, would only be physically reasonable if the simulation were performed at infinite temperature or the method produced a Boltzmann ensemble in finite simulation time, despite the complex and high-dimensional energy surface. Because one simulates at finite temperature and because simulations are of finite length, one must be careful interpreting such simulations.

Variation of Conformational Probabilities with Fixed Basis Set of Molecular Conformations

With the dependent Eqs. (7) and (8), the probability distribution for a fixed basis set $\{X_\alpha\}$ of conformations is fully determined. Therefore, the probability distribution can be calculated by either iterative evaluation of these equations to self-consistency or by matrix inversion techniques.[23] However, the quality of the solution depends critically on the choice of the fixed starting conformations, which serve as the basis for the probability distribution. Even if this were possible for very small, well-defined systems, it is obviously not possible to identify all important conformations for a realistic molecule with many degrees of freedom and a multitude of contributing states.

Simultaneous Variation of Conformational Coordinates and Probabilities

With simultaneous, or quasi-simultaneous, variation of both coordinates and probabilities, one is able to combine the advantages and avoid the disadvantages of the previous methods. A correct weighting between the conformations can be obtained because the probabilities are allowed to vary and adapt to the conformations. Furthermore, the choice of a basis set of starting conformations is not critical. All copies representing a single part of the molecule are individually changed according to classical equations of motion and automatically adopt low energy conformations. To evaluate the force on particle i of copy α in group A, $\vec{F}_m(A, \alpha, i)$, it is straightforward to calculate the negative derivative of the total (mean) energy with respect to the Cartesian coordinates $\vec{r}(A, \alpha, i)$ of particle i. With the approximation that probabilities are constant during a simulation step, they can be treated as parameters, thus

$$\begin{aligned}\vec{F}_m(A,\alpha,i) &= -\frac{\partial}{\partial \vec{r}(A,\alpha,i)} E_{\text{tot}} \\ &= -p(A,\alpha)\frac{\partial}{\partial \vec{r}(A,\alpha,i)} E(A,\alpha) \\ &= \tfrac{1}{2}p(A,\alpha)\sum_{B\neq A}^{M}\sum_{\beta}^{N_B}\sum_{j}^{O_{B,\beta}} p(B,\beta) \times \vec{F}(A,\alpha,i;B,\beta,j) \\ &\quad + \tfrac{1}{2}p(A,\alpha)\sum_{j\neq i}^{O_{A,\alpha}} \vec{F}(A,\alpha,i;A,\alpha,j)\end{aligned} \quad (9)$$

If $\vec{F}(A, \alpha, i; B, \beta, j)$ denotes the force on atom i in copy α of group A exerted by atom j in copy β of group B, then

$$\vec{F}(A, \alpha, i; B, \beta, j) = -\frac{\partial}{\partial \vec{r}(A, \alpha, i)} V(A, \alpha, i; B, \beta, j) \quad (10)$$

Since copy α of a group only contributes to the mean with the weight of its probability, the original masses $m(A, \alpha, i)$ of each atom i in each copy α and group A must also be scaled so as to conserve the total mass of the group. One then takes

$$m^*(A, \alpha, i) = p(A, \alpha) m(A, \alpha, i) \quad (11)$$

When integrating the equation of motion,

$$\frac{d^2\vec{r}(A, \alpha, i)}{dt^2} = m^*(A, \alpha, i)^{-1} \vec{F}_m(A, \alpha, i) \quad (12)$$

for a system, $E(A, \alpha)$ and $p(A, \alpha)$ should satisfy the Eqs. (7) and (8). This means that at each simulation time step the nonlinear set of self-consistent equations has to be solved, e.g., by iteration or matrix inversion. However, a good guess of the set $\{E(A, \alpha), p(A, \alpha)\}$ will be available from the previous MD step t_{n-1}. This means one can determine $E(A, \alpha)$ and $p(A, \alpha)$ at time t_n [viz. $E(A, \alpha; t_n)$ and $p(A, \alpha; t_n)$] using only one iteration. Thus, one can write

$$\begin{aligned} E(A, \alpha; t_n) &= \tfrac{1}{2} \sum_{i}^{O_{A,\alpha}} \sum_{B \neq A}^{M} \sum_{\beta}^{N_B} \sum_{j}^{O_{B,\beta}} p(B, \beta; t_{n-1}) \\ &\quad \times V(A, \alpha, i; B, \beta, j; t_n) \\ &\quad + \tfrac{1}{2} \sum_{i}^{O_{A,\alpha}} \sum_{j \neq i}^{O_{A,\alpha}} V(A, \alpha, i; A, \alpha, j; t_n) \end{aligned} \quad (13)$$

and

$$\vec{F}_m(A, \alpha, i; t_n) = -p(A, \alpha; t_{n-1}) \frac{\partial}{\partial \vec{r}(A, \alpha, i)} E(A, \alpha; t_n) \quad (14)$$

and

$$p(A, \alpha; t_n) = \frac{\exp\{-[E(A, \alpha; t_n)/k_B T]\}}{\sum_{\alpha}^{N_A} \exp\{-[E(A, \alpha; t_n)/k_B T]\}} \quad (15)$$

and

$$E_{\text{tot}}(t_n) = \sum_{A}^{M} \sum_{\alpha}^{N_A} p(A, \alpha; t_n) E(A, \alpha; t_n) \quad (16)$$

METHODS

All simulations were carried out using modified programs from the GROMOS suite.[24]

Iterative optimization procedures were started with a uniform probability distribution and continued until self-consistency. Self-consistency was detected by a small change in the probabilities between iterations. The criterion was that the sum of squared differences of probabilities with respect to the previous iteration be smaller than some threshold

$$\sum_{A}^{M} \sum_{\alpha}^{N_A} [p(A, \alpha; t_n) - p(A, \alpha; t_{n-1})]^2 < 10^{-16} \quad (17)$$

The instantaneous mean field energy of a conformation may fluctuate substantially between initial iteration steps. Because of the exponential weighting in the probability calculation, this could lead to large fluctuations in the probability of a conformation. Since these fluctuations are of little real significance and impair the convergence, a memory function was used to calculate the average energy over the recent history. It was this effective energy that was actually used to calculate the probability of the copy. Referring to this effective energy at iteration step t_n as $E_{\text{eff}}(A, \alpha; t_n)$, an exponential weighting was used with a decay constant τ_E

$$\begin{aligned} E_{\text{eff}}(A, \alpha; t_n) &= \exp(-\tau_E) E(A, \alpha; t_n) \\ &\quad + [1 - \exp(-\tau_E)] E_{\text{eff}}(A, \alpha; t_{n-1}) \end{aligned} \quad (18)$$

τ_E was set to 0.25 (a dimensionless value). This led to convergence in all simulations and no attempt was made to optimize this parameter for speed of convergence.

For all the dynamics simulations, a stochastic dynamics integrator[25] was used with a friction coefficient of $\gamma = 19\ \text{ps}^{-1}$, a time step of 2 fs and the SHAKE algorithm[26] applied to constrain bond lengths. Each system was coupled to a temperature bath[27] at 300 K with a coupling constant $\tau_T = 0.1$ ps in the pentane simulations, or τ_T equal to the time step. All interactions were evaluated without truncation of forces beyond a cutoff radius. All simulations used the GROMOS 37D4 vacuum force field.[24]

Pentane

The first simulations were performed on a simple model system to highlight the effect of the mean field approxi-

mation and the dependence of the mean field on the number of copies. Pentane, in the united atom representation, is a linear chain of only five atoms and its conformation is well described by just two dihedral angles. The simulations were performed with both the first (C_1) and the last (C_5) atom ($M = 2$) of the chain represented by either $N_A = N_B = 20$ or $N_A = N_B = 360$ copies. Initial positions were generated by taking a single-copy pentane and energy minimizing it in the force field. The terminal atoms were then copied and rotated in steps of 18° or 1° as appropriate. Unlike a normal MD simulation, these calculations require initial probabilities for copies, as well as coordinates. These probabilities were calculated by an iterative evaluation of the energy and corresponding Boltzmann probability at 300 K via Eqs. (7) and (8) until self consistency was reached.

Cyclo-(-D-Pro-Phe-Phe-Pro-Phe-Phe-) Peptide

The phenylalanine side-chain conformations play an important role in the biological activity of this small cyclic hexapeptide.[28] The conformation of a phenylalanine side chain is well described by the torsion angles χ_1 and χ_2. Thus there are eight independent variables and the evaluation of the conformational probabilities is not an easy task. To demonstrate this problem, a free stochastic dynamics simulation of this peptide over 1 ns was performed.

As a reference point for testing the methodology, the first calculations on the hexapeptide consisted of optimizing probabilities while keeping coordinates fixed. The aromatic ring of each of the four phenylalanines was copied 100 times. Within each side chain, the aromatic ring copies were all bound to a single C_β carbon. The copies at each of the four phenylalanine residues were then distributed evenly within the allowed space by rotating through 10 steps (of 36°) at both χ_1 and χ_2 dihedral angles. This resulted in a crude discretization of space, but with two dihedral angles (χ_1, χ_2), the number of copies grew quadratically with the number of steps per torsion angle. With only 100 copies, the number of atoms in the system grew from 62 in the original peptide to 2438 in the multiple sidechain one.

In a subsequent test, only 20 copies were used for each phenylalanine side chain and simulations performed with simultaneous dynamics of both coordinates and probabilities. The initial conformation was generated by setting the coordinates of each copy to be identical. This collection was then equilibrated at 600 K for 50 ps, resulting in a well distributed set of conformations, without unreasonable steric properties or impossible conformations. Simulations of 250 ps were performed using both a Boltzmann-weighted mean field and the LES method[29] in which each conformation was equally weighted.

RESULTS AND DISCUSSION

Pentane

This 5-atom model was used to demonstrate the basic features of our mean field approach.

Mean Field Properties. Because of the mean field description of the C_1 and C_5 atoms in the pentane model, the potential energy profiles along the two internal dihedral angles ϕ_1 and ϕ_2 become the same as the number of copies increases. Therefore no energetic discrimination of the conformations $\phi_1 = 60°/-60°$, $\phi_2 = -60°/60°$ against the conformations $\phi_1 = 60°/-60°$, $\phi_2 = 60°/-60°$ in the mean field is possible.

Temperature Effect. Figure 1a shows the relative Boltzmann probability distribution along the dihedral angle ϕ after optimization at different temperatures. As expected, the probability distribution is uniform for infinite temperature and the probabilities of low energy conformations increase when the temperature is lowered. In the limit of zero temperature, this results in a single conformation (the one with lowest energy). As the probability distribution changes, the mean field undergoes corresponding changes. This is shown in Figure 1b, where the field experienced by a test particle at the C_1 methyl position, due to the mean C_5 particle, is plotted as function of the dihedral angle ϕ_1 for several temperatures. The C_5 atom was represented by 360 copies. One can view this diagram differently. The field experienced by the C_1 atom is due to an average C_5 atom and this average C_5 atom includes Boltzmann weighting. This averaging over possible conformations with the correct weighting is similar in spirit to a free energy calculation. It is interesting to note that the plot has some characteristics of a free energy profile for rotation about the dihedral angle.

The plot suggests that local minima are smoothed out, while the location of the global minimum remains undisturbed. Unfortunately, this is not generally the case. At high temperature, the mean field can cause local minima to be of lower energy than the true global minimum (of the conventional force field). As long as temperatures are not too high, this is not a practical problem since the mean field represents only a minor distortion of the true force field.

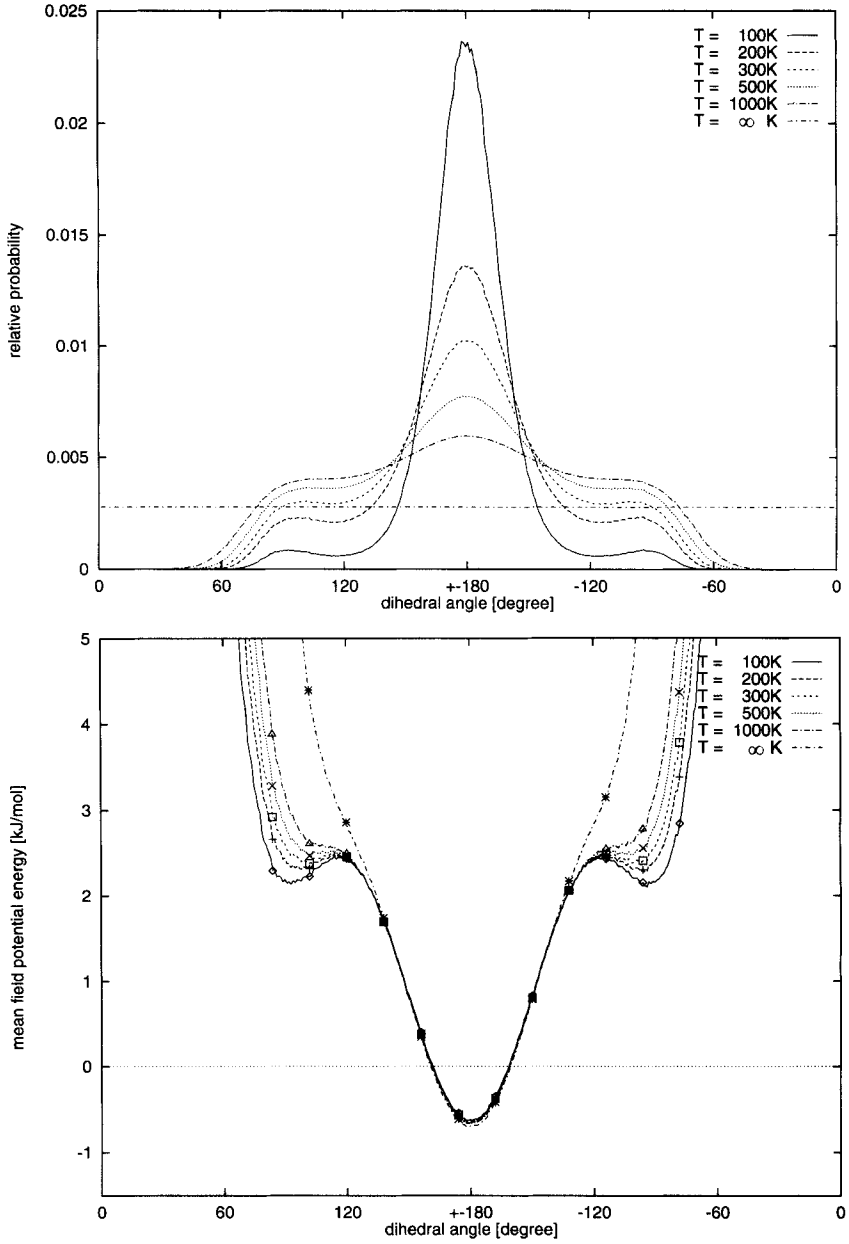

FIGURE 1 Pentane at different temperatures using 360 copies to represent C_1 and C_5 atoms. (a) Relative Boltzmann probability distribution vs internal dihedral angle ϕ. (b) Boltzmann weighted mean field potential vs internal dihedral angle ϕ. Points represent results from a 20-copy model.

Distribution Effect. A serious question for applications is how many copies are needed for a reasonable description of the mean field. To this end, the calculations with the pentane model were repeated, but with 20 instead of 360 copies. Figure 1b shows the comparison. Results from the 20-copy system were plotted as symbols on top of the continuous lines, which give the results of the 360-copy system. The good agreement shows that, in this very simple case, an initial grid spacing of 18° is adequate. In contrast, reducing the number of copies to five showed a deterioration of the mean field (data not shown). This deterioration is due to the relatively large signifi-

cance that interactions between individual copies take.

An intriguing problem that appears in the dynamic approach is that the copies tend to adopt a Boltzmann distribution due to the molecular dynamics technique. This can lead to double weighting of conformations when conformations rapidly exchange and sample a proper Boltzmann ensemble by themselves. This is the case in the simulation of the 20-copy pentane model. In most systems, however, simulation time is not sufficient to properly sample different conformations and the possibility of double weighting in our method leads, in our opinion, to a much smaller error than the use of equal weighting of each conformation.

Cyclo-(-D-Pro-Phe-Phe-Pro-Phe-Phe-) Peptide

This small hexapeptide was used for several reasons. First, the small number of atoms means that simulations are computationally cheap. Second, it has four bulky phenylalanines for which classical simulation methods like MD or SD fail to produce proper statistics for the different side chain conformations.

This system also served to give some idea of the computational expense of the method. In the case of 100 copies, the number of atoms increased from 62 to 2438. The computational effort, however, did not increase as would be expected for a system of more than 2×10^3 atoms. Copies of a single site do not interact with each other, so the number of interactions to be calculated does not grow as in a conventional system with the same number of particles.

Conformational Coordinate Optimization by Free Simulation. The difficulty in generating proper statistics for phenylalanine side-chain conformations in dynamics simulations is shown by Figure 2. This gives the internal angle χ_1^{Phe2} during a 1 ns SD simulation trajectory. Conformational transitions can be observed on a time scale of a few hundred picoseconds and only a few transitions occur during the whole simulation. The starting structure and the height of energy barriers completely determine whether or not the most probable conformations are sampled at all. For the simulation shown, an unlikely conformation was chosen to start with and no transition to the most favored conformation ($\chi_1 = 60°$) took place during the whole simulation. For this system, it is clear that side-chain statistics extracted from conventional molecular dynamics simulation, even if longer simulations are performed, have little physical meaning and are associated with huge errors.

Probability Optimization. The situation is different in the Boltzmann-weighted multicopy approach. Because each conformation can move to any other conformation along an artificial degree of freedom (the relative probability), normal physical transitions such as rotations about dihedral angles are not necessary. This is demonstrated on the model peptide where each phenylalanine is represented by 100 copies, uniformly distributed in χ_1-χ_2 space. While cartesian coordinates were kept fixed, the relative Boltzmann probability distributions of all four phenylalanines were optimized by iteration. In Figure 3 the mean field energy is plotted as a function of the iteration step. The energy converges to a final value of 254 kJ/mol. Due to the use of a memory term in the mean field potential energy [Eq. (18)], a plateau or even a slight increase of the mean field energy is observable. This shows why the change of probabilities was chosen as a criterion for convergence rather than an energy difference.

The converged probability distribution for the Phe^2 side chain is shown in Figure 4a. The maxima of probability are found for the conformation $\chi_1 \approx 60°$. Two structurally identical orientations of the aromatic ring, with a shift of 180°, can be seen along the angle χ_2. The probability maxima at $\chi_1 \approx 60°$ are markedly shifted from the usual $\chi_2 \approx 90°/-90°$. This could be a steric effect due to the other side chains, or an artefact due to the wide spacing of basis conformations. Although this plot gives a good impression of the probability distribution, it suffers from poor resolution. The spacing between neighboring conformations is 36° in the χ_1 or χ_2 dihedral angle and is not fine enough to pinpoint minima (see Figure 4a). Furthermore, the result may be biased since the conformations are generated by rotating the side chains of a minimized starting structure around the torsion angles χ_1 and χ_2. One conformation (the starting structure) is located in an energy minimum, whereas the other conformations are defined by relative dihedral angle differences to the starting structure and may not be located in minima of the potential energy. To overcome this problem, the next step was to allow the conformations to adapt

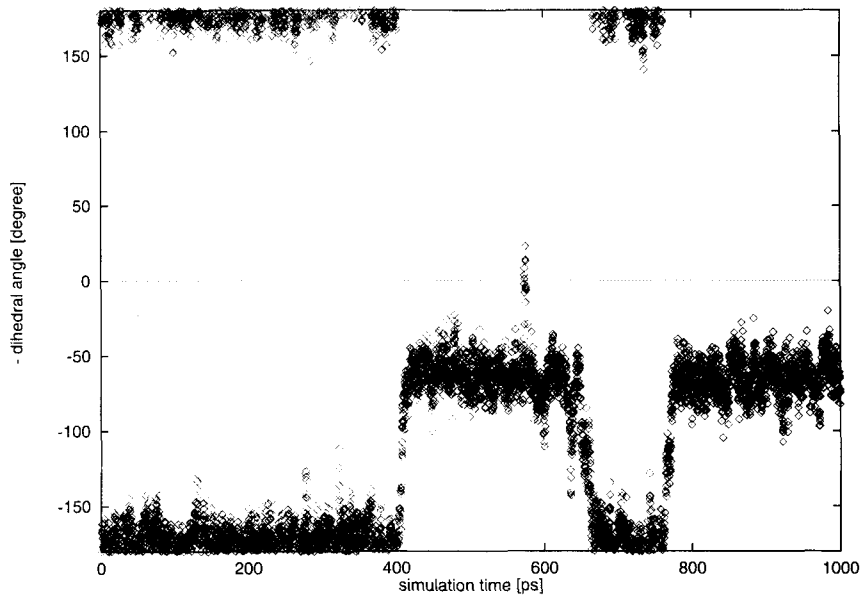

FIGURE 2 The χ_1 in Phe[2] during a 1 ns free stochastic dynamics simulation of cyclo-(-D-Pro-Phe-Phe-Pro-Phe-Phe-).

low energy conformations with a dynamics simulation technique.

Simultaneous Probability and Conformational Coordinate Optimization. For analyzing the results, we used two different relative probabilities of the conformations: First, the relative Boltzmann probability was used. This is the relative weight of a conformation in the simulation given by the Boltzmann relation. Second, the probability of occurrence was defined as the number of times the conformation occurs in a simulation trajectory as a

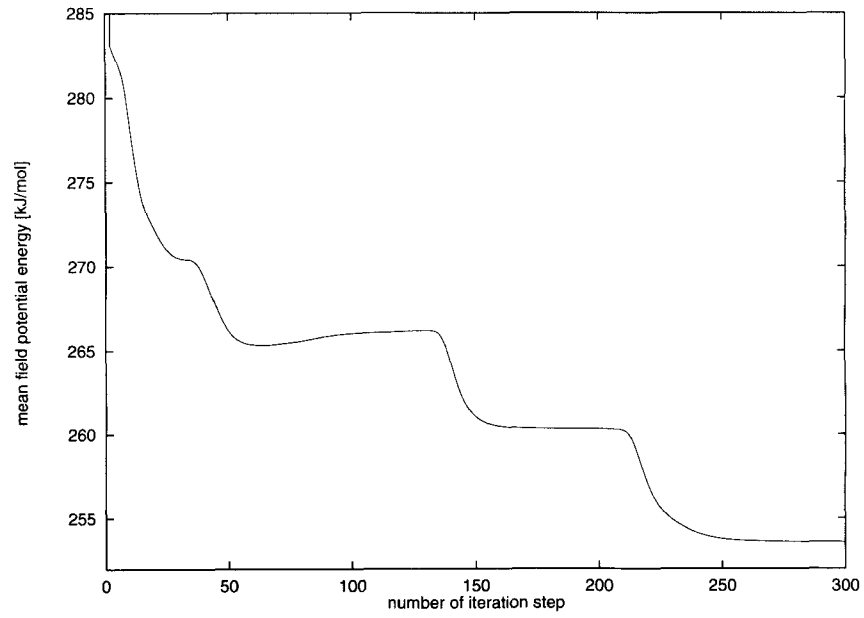

FIGURE 3 Mean field potential energy during iterative optimization of the probability distribution in cyclo-(-D-Pro-Phe-Phe-Pro-Phe-Phe-). Each phenylalanine was represented by 100 copies uniformly distributed in χ_1-χ_2 space.

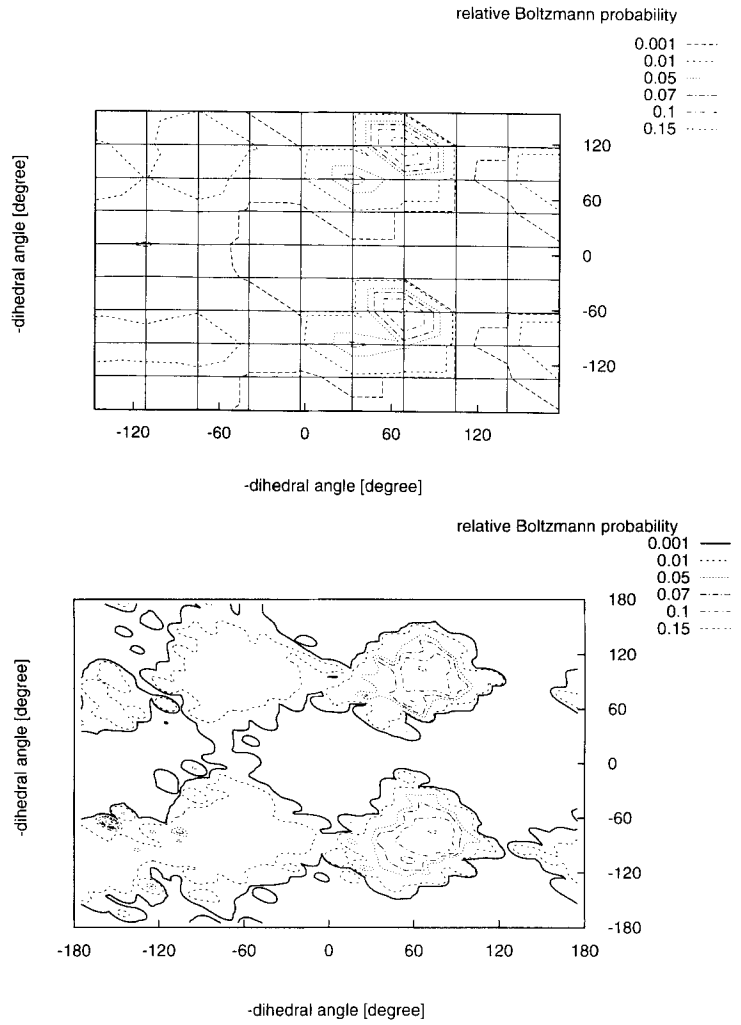

FIGURE 4 Relative Boltzmann probability distribution of the Phe^2 side chain in χ_1-χ_2 space after optimization. (a) Iterative optimization with a fixed set of 100 basis conformations. The grid shows the location of the basis conformations. (b) Stochastic dynamics simulation in a Boltzmann weighted mean field (averaged values).

fraction of all conformations. Since conformations are counted, each conformation gets the same weight.

In Figures 5a and 5b, both probabilities are shown for the χ_1 dihedral angle of the Phe^2 side chain during a Boltzmann-weighted mean field simulation. In Figure 5c the probability of occurrence of the same dihedral angle in a LES simulation is given. Since, in LES simulations, each conformation is equally likely or equally weighted, only the probability calculated by counting is plotted.

Although the probability of occurrence distributions from the LES simulation (Figure 5c) and the Boltzmann-weighted mean field simulation (Figure 5b) look similar, they differ significantly from the distributions when correct weighting (Figure 5a) is used. In the LES simulation, the individual copies of the side chain move in the arithmetically averaged field of the other side-chain copies, but in the Boltzmann weighted mean field simulation, they move in a completely different field. This different potential energy field is closer to the natural potential energy, so a different and more natural dynamics of the system is expected.

With the correct weighting, the conformations with χ_1 dihedral angle of 60° are much more fa-

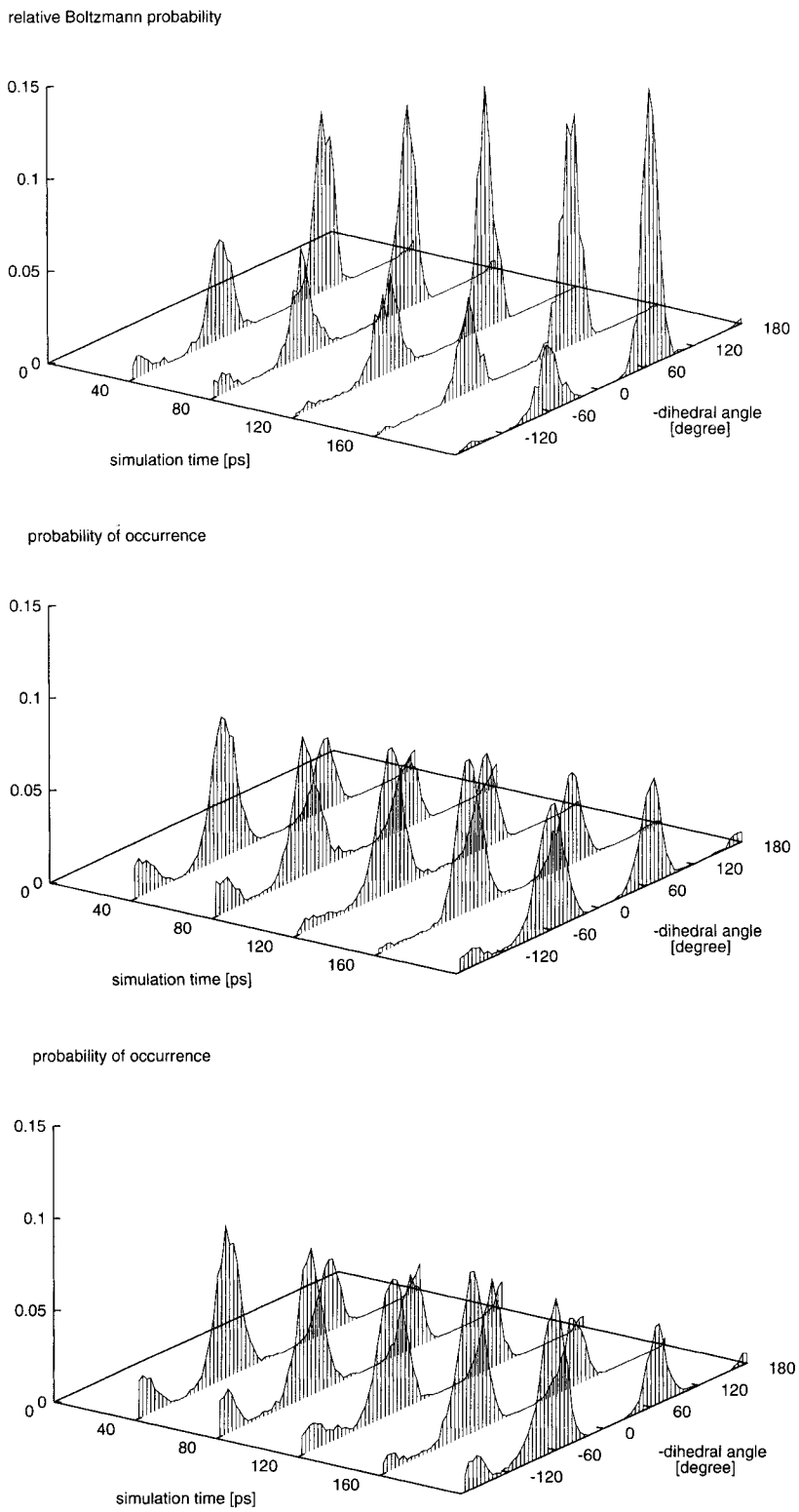

FIGURE 5 Time evolution of the probability distribution of χ_1 of the Phe[2] side chain during (a) Boltzmann-weighted mean field simulation, Boltzmann-weighted probabilities; (b) Boltzmann-weighted mean field simulation, probabilities of occurrence; (c) LES method, probabilities of occurrence.

vored due to lower energy, even if the conformation at $\chi_1 = -60°$ is sampled more often. This result is in good agreement with results in the previous section obtained from the iterative optimization of probabilities with a fixed basis set of conformations. However, in contrast to calculations with fixed discrete conformations (Figure 4a), a probability distribution in continuous conformational space is obtained (Figure 4b). Furthermore, one has a higher resolution with which to exactly locate probability maxima.

CONCLUSIONS

The iterative optimization of probabilities using a fixed basis set of conformations is an adequate method to obtain the relative Boltzmann probability distribution for a set of conformations. The results, however, are highly dependent on a proper choice of the basis conformations. Even for relatively simple situations such as the orientation of side chains in proteins, this is not easy. Small deviations of a conformation from the minimum-energy position can result in big changes of the potential energy. This results in even larger changes in the probability distribution due to the exponential dependence of the probability on the energy.

In contrast, it was shown that using a dynamic scheme to optimize both variables—probabilities and conformational coordinates—simultaneously results in much more accurate solutions. At the same time, one must be aware that the combination of molecular dynamics and an additional weighting of conformations can lead to an overweighting of low energy conformations. Unfortunately, it is not possible to state the severity of this problem in general. It depends on the barriers preventing the interchange of conformations and the problem vanishes in the limit of conformations which never interconvert. The hexapeptide example here is one case where conformations are relatively fixed (see Figure 2) and the Boltzmann weighting is essential. Despite these uncertainties, the dynamic optimization of Boltzmann-weighted conformations appears to be an attractive method to obtain proper probability distributions of conformations in situations where conventional methods fail.

Financial support from the Schweizerischen Nationalfonds (project 5003-034442) is gratefully acknowledged.

REFERENCES

1. Kirkpatrick, S., Gelatt, C. D., Jr. & Vecchi, M. P. (1983) *Science* **220,** 671–680.
2. van Schaik, R. C., van Gunsteren, W. F. & Berendsen H. J. C. (1992) *J. Comput.-Aided Mol. Design* **6,** 97–112.
3. Huber, T., Torda, A. E. & van Gunsteren, W. F. (1994) *J. Comput.-Aided Mol. Design* **8,** 695–708.
4. Weber, P. L., Morrison R. & Hare, D. (1988) *J. Mol. Biol.* **204,** 483–487.
5. Crippen, G. M. (1982) *J. Comput. Chem.* **3,** 471–476.
6. Crippen, G. M. (1989) *J. Comput. Chem.* **10,** 896–902.
7. Crippen, G. M. & Havel, T. F. (1990) *J. Chem. Inf. Comput. Sci.* **30,** 222–227.
8. Purisima, E. O. & Scheraga, H. A. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 2782–2786.
9. van Schaik, R. C., Berendsen, H. J. C., Torda, A. E. & van Gunsteren, W. F. (1993) *J. Mol. Biol.* **234,** 751–762.
10. Gerber, R. B., Buch, V. & Ratner, M. A. (1982) *J. Chem. Phys.* **77,** 3022–3030.
11. Straub, J. E. & Karplus, M. (1990) *J. Chem. Phys.* **94,** 6737–6739.
12. Hansmann, U. H. E. & Okamoto, Y. (1993) *J. Comp. Chem.* **14,** 1333–1338.
13. Rosenfeld, R., Zheng, Q., Vajda, S. & DiLisi, C. (1993) *J. Mol. Biol.* **234,** 515–521.
14. Rosenfeld, R., Zheng, Q., Vajda, S. & DiLisi, C. (1993) *J. Chem. Phys.* **99,** 8892–8896.
15. Zheng, Q. & Kyle, D. J. (1994) *Proteins* **19,** 324–329.
16. Mierke, D. F., Scheek, R. M. & Kessler, H. (1994) *Biopolymers* **34,** 559–563.
17. Elber, R. (1990) *J. Chem. Phys.* **93,** 4312–4321.
18. Verkhivker, G., Elber, R. & Gibson, Q. (1992) *J. Am. Chem. Soc.* **114,** 7866–7878.
19. Verkhivker, G., Elber, R. & Nowak, W. (1992) *J. Chem. Phys.* **97,** 7838–7841.
20. Ulitsky, A. & Elber, R. (1993) *J. Chem. Phys.* **98,** 3380–3388.
21. Ulitsky, A. & Elber, R. (1994) *J. Phys. Chem.* **98,** 1034–1043.
22. van Gunsteren, W. F., Huber, T. & Torda, A. E. (1995) in *European Conference on Computational Chemistry* (E.C.C.C. 1), Bernardi, F. & Rivail, Y.-L., Eds. American Institute of Physics Conference Proceedings, Vol. 330 pp. 253–268.
23. Koehl, P. & Delaure, M. (1994) *J. Mol. Biol.* **239,** 249–275.

24. van Gunsteren, W. F. & Berendsen, H. J. C. (1987) *Groningen Molecular Simulation (GROMOS) Library Manual,* Biomos, Groningen.
25. van Gunsteren, W. F. & Berendsen, H. J. C. (1988) *Mol. Simulat.* **1,** 173–185.
26. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. (1977) *J. Comput. Phys.* **23,** 327–341.
27. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984) *J. Chem. Phys.* **81,** 3684–3690.
28. Wagner, K. (1988) Ph.D. thesis, University of Frankfurt/Main, Germany.
29. Roitberg, A. & Elber, R. (1991) *J. Chem. Phys.* **95,** 9277–9287.