Chapter 3

Optimisation methods using Boltzmann weighted mean field approach

3.1 Summary

A Optimisation protocol is proposed which combines mean field approach with a Boltzmann weighted sampling. This is done by including Boltzmann probabilities of multiple conformations in the optimisation procedure. The method is demonstrated on a simple model system and on the side chain conformations of phenylalanines in a small hexapeptide. For comparison, classical stochastic dynamics (SD) simulations [1], iterative optimisation of probabilities on a fixed set of basis conformations [2] and simulations with locally enhanced sampling [3] were performed. Although approximations are used in our method, the results are more accurate and physically meaningful then the other procedures discussed. Furthermore the approximate Boltzmann distribution allows generalisation of the results.

3.2 Introduction

Computer simulations of biological molecules are a very powerful way getting deeper insights in molecular behaviour on an atomic level and therefore are becoming more and more important in molecular modeling. The particular aim in biomolecular modeling is to find an ensemble of lowest energy conformations of a molecule or parts of a molecule. Unfortunately, standard heuristic conformational search methods like Metropolis Monte Carlo (MC) or molecular dynamics (MD) show very poor ability to cross energy barriers at reasonable temperature and they sample only small regions in conformational space.

Many methods based on dynamic or non-dynamic schemes have been developed to attempt to search broader regions efficiently for low energy conformations. The most commonly applied is probably simulated annealing [4], which relies on temperature evolution in time. Potential energy annealing (PEACS) [5] attempts to take advantage of the shape of the potential hyper-surface. The local elevation technique [6] includes a long time memory and enhances sampling by local modifications of the potential surface. Several other methods exist, which rely on temporarily increasing the dimensionality by adding artificial degrees of freedom [7–12]. Gerber et. al. [13] 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 [14–19], including Elber and coworkers [3, 20–24] 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 amongst conformations. This means that physical properties may be incorrectly estimated and a statistical generalisation of the results can be difficult. Furthermore, in some systems, the global minimum may not be a satisfactory solution by itself. The observed properties reflect the distribution of the ensemble across all accessible minima.

Our approach [25] 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 by changing the mean conformation with a change in the probability distribution instead, or in addition, to conformational changes.

3.3 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 $(i = 1, 2, ...n)$. The conformation X_α for molecule α is specified in terms of a particular set of atomic coordinate values.

$$X_\alpha = \{x_1^\alpha, x_2^\alpha..x_n^\alpha\} \tag{3.1}$$

The energy of the conformation is given by

$$E_\alpha = V(X_\alpha) \tag{3.2}$$

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

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

$$p_\alpha = \frac{\exp^{-\frac{E_\alpha}{k_B T}}}{\sum_{\alpha=1}^{N_A} \exp^{-\frac{E_\alpha}{k_B T}}} \tag{3.3}$$

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

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

$$\begin{aligned}
\langle Q \rangle &= \sum_\alpha p_\alpha Q(X_\alpha) && (3.4)\\
&= t_{MD}^{-1} \int_0^{t_{MD}} Q(X\alpha(t'))\, dt' && \text{molecular dynamics}\\
&= N_{MC}^{-1} \sum_{\alpha \in MC}^{N_{MC}} Q(X_\alpha) && \text{Metropolis Monte Carlo}
\end{aligned}$$

$$(3.5)$$

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

$$\langle E \rangle = \sum_{\alpha} p_{\alpha} E_{\alpha} = \sum_{\alpha} p_{\alpha} V(X_{\alpha}) \tag{3.6}$$

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

$$\langle E \rangle = E_{tot} = \textstyle\sum_A^M \sum_\alpha^{N_A} \; p(A, \alpha) \; E(A, \alpha) \tag{3.7}$$

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

$$p(A, \alpha) = \frac{\exp^{-\frac{E(A,\alpha)}{k_B T}}}{\sum_\alpha^{N_A} \; \exp^{-\frac{E(A,\alpha)}{k_B T}}} \tag{3.8}$$

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

$$\begin{aligned} E(A, \alpha) &= \frac{1}{2} \sum_{i}^{O_{A,\alpha}} \sum_{B \neq A}^{M} \sum_{\beta}^{N_B} \sum_{j}^{O_{B,\beta}} \; p(B, \beta) \; V(A, \alpha, i; B, \beta, j) \\ &+ \frac{1}{2} \sum_{j}^{O_{A,\alpha}} \; V(A, \alpha, i; A, \alpha, j) \end{aligned} \tag{3.9}$$

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 the full interaction between atoms of this copy contribute 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 equation (3.8) this gives a self consistent set of non linear equations for the energies and probabilities of all used copies in the system.

The computational problem is now to minimise $\langle E \rangle$ with the Boltzmann relation (3.8) 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 (3.7) depends on both conformational coordinates X_α and conformational probabilities p_α of all copies, we distinguish three types of optimisation algorithms.

1. **Variation of conformational coordinates with fixed probability distribution**

   Probabilities can be treated as fixed with the approximation that a variation of conformational coordinates does not effect the probability distribution. Because of the possible interchange of two conformations, this results in uniform probability distributions. This, however, is only physically reasonable when the simulation is performed at infinite temperature or the method produces a Boltzmann ensemble, despite the complex and high–dimensional energy surfaces, in finite simulation time. Because the first case is very rare and no methods are known which fulfill the second criterion, one has to be rather careful when interpreting such simulations.

2. **Variation of conformational probabilities with fixed basis set of molecular conformations**

   With the dependent equations (3.8) and (3.9), the probability distribution for a fixed basis set 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. However, the physical quality of the solution depends critically on a proper choice of the fixed starting conformations, which are the basis of the probability distribution. Even if this were possible for very small well defined systems, it is obviously not possible for to identify all important conformations for a realistic molecule with many degrees of freedom a multitude of contributing states.

3. **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 start conformations is not critical. All copies representing a single part of the molecule are individually changed according to classical equations of motion and will therefore automatically adopt low energy conformations. To evaluate the partial force on particle i, $\vec{F}_m(A,\alpha,i)$, it is straight forward to calculate the negative derivative of the total (mean) energy. With the approximation that probabilities are constant during a simulation step, they can be treated as parameters

$$\begin{aligned}
\vec{F}_m(A,\alpha,i) &= -\ \frac{\partial}{\partial\vec{r}(A,\alpha,i)}\ E_{tot} \\
&= -\ p(A,\alpha)\ \frac{\partial}{\partial\vec{r}(A,\alpha,i)}\ E(A,\alpha) \\
&= \frac{1}{2}\ p(A,\alpha)\ \sum_{i}^{O_{A,\alpha}} \sum_{B\neq A}^{M} \sum_{\beta}^{N_B} \sum_{j}^{O_{B,\beta}}\ p(B,\beta)\ \vec{F}(A,\alpha,i;B,\beta,j) \\
&\quad + \frac{1}{2} \sum_{j}^{O_{A,\alpha}} \vec{F}(A,\alpha,i;A,\alpha,j) \qquad (3.10)
\end{aligned}$$

with $\vec{F}(A,\alpha,i;B,\beta,j)$ denoting the interaction force between atom i in copy α of group A with atom j in copy β of group B defined by

$$\vec{F}(A,\alpha,i;A,\alpha,j) = \frac{\partial}{\partial\vec{r}(A,\alpha,i)}\ V(A,\alpha,i;B,\beta,j) \qquad (3.11)$$

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

$$m^*(A,\alpha,i) = p(A,\alpha)\ m(A,\alpha,i) \qquad (3.12)$$

When integrating the equation of motion

$$\frac{d^2\vec{r}(A,\alpha,i)}{dt^2} = m^*(A,\alpha,i)^{-1}\vec{F}_m(A,\alpha,i) \tag{3.13}$$

for a system, $E(A,\alpha)$ and $p(A,\alpha)$ should satisfy the equations (3.8) and (3.9). This means that in each simulation time step the non linear set of self consistent equations have to be solved, e.g. by iteration or matrix inversion technique. However, if a good guess of the set $\{E(A,\alpha),\, p(A,\alpha)\}$ is available, for example from the previous MD step, t_{n-1}, one may determine $E(A,\alpha)$ and $p(A,\alpha)$ at time t_n (viz. $E(A,\alpha;t)$ and $p(A,\alpha;t)$) using only one iteration step, thus

$$\begin{aligned} E(A,\alpha;t_n) &= \frac{1}{2}\sum_{i}^{O_{A,\alpha}}\sum_{B\neq A}^{M}\sum_{\beta}^{N_B}\sum_{j}^{O_{B,\beta}} p(B,\beta;t_{n-1})\; V(A,\alpha,i;B,\beta,j;t_n) \\ &\quad +\frac{1}{2}\sum_{j}^{O_{A,\alpha}} V(A,\alpha,i;A,\alpha,j;t_n) \end{aligned} \tag{3.14}$$

and

$$\vec{F}_m(A,\alpha,i;t_n) = -\; p(A,\alpha;t_{n-1})\,\frac{\partial}{\partial\vec{r}(A,\alpha,i)}\,E(A,\alpha;t_n) \tag{3.15}$$

and

$$p(A,\alpha;t_n) = \frac{\exp^{-\frac{E(A,\alpha;t_n)}{k_B T}}}{\sum_{\alpha}^{N_A}\ \exp^{-\frac{E(A,\alpha;t_n)}{k_B T}}} \tag{3.16}$$

and

$$E_{tot} = \sum_{A}^{M}\sum_{\alpha}^{N_A} p(A,\alpha;t_n)\; E(A,\alpha;t_n) \tag{3.17}$$

3.4 Methods

All simulations were carried out using modified programs from the GROMOS suites [26].

Iterative optimisation procedures were started with an 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

$$\sum_{i}^{N} (p_i(t) - p_i(t-1))^2 \quad < \quad 10^{-16} \tag{3.18}$$

The instantaneous mean field energy of a conformation may fluctuate substantially between initial iteration steps. Because of the exponential term 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 worsen the convergence, a memory function was used to calculate the average energy over the recent history and this effective energy actually used to calculate the probability of the copy. Referring to this effective energy at iteration step t as $E_{eff}(t)$, an exponential weighting was used with a decay constant τ_E

$$E_{eff}(t) \quad = \quad exp(-\tau_E)\ E(t) \ + \ (1 - exp(-\tau_E))\ E(t-1) \tag{3.19}$$

τ_E was set to 0.25, which led to convergence in all simulations and no attempt was made to optimise this parameter for speed of convergence.

For all the dynamics simulations, a stochastic dynamics integrator [1] was used with a friction coefficient of $\gamma = 19$ ps^{-1} a time step of 2 fs and applying the shake algorithm [27] to constrain bond lengths. Each system was coupled to a temperature bath [28] 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 with no truncation of forces beyond a cutoff radius.

3.4.1 pentane

The first simulations were performed on a simple model system to highlight the effect of the mean field approximation 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 and the last atoms of the chain represented by either 20 or 360 copies. Initial positions were generated by taking a single–copy pentane and energy minimising 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 ground energy and corresponding Boltzmann probability until self consistency was reached.

A simulation was performed starting from the pentane structure with 20 copies representing both C_1 and C_5 and using their corresponding relative Boltzmann probabilities at T= 300 K. The probabilities were updated with a single iteration after each integration step.

3.4.2 cyclo-(-D-Pro-Phe-Phe-Pro-Phe-Phe-) peptide

The phenylalanine side chain conformations play an important role in the biological activity this small cyclic hexapeptide [29]. Even if a conformation of phenylalanine can in be described to a good approximation by the torsion angles χ_1 and χ_2, there are still eight independent variables left 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 optimising probabilities while keeping coordinates fixed. The aromatic ring of each of the two phenylalanines was copied 100 times. Within each side chain, the aromatic ring copies were all bound to a single C-beta carbon. The copies at each of the two phenylalanine residues were then distributed evenly within the allowed space by rotating through 10 steps (of 36°) at both chi-1 and chi-2 torsion angles. This resulted in only a crude discretisation of space, but with two torsion angles (χ_1, χ_2), the number of atoms 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 a next 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 well distributed conformations, but with no unreasonable steric properties or impossible conformations. Simulations of 250 ps were performed using both a Boltzmann weighted mean field and the LES method in which each conformation was equally weighted.

3.5 Results and Discussion

3.5.1 Mean field properties

Because of the mean field description of the $C1$ and $C5$ 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.

3.5.2 Temperature effect

Figure 3.1a shows the relative Boltzmann probability distribution after optimisation for 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 conformation with lowest energy. As the probability distribution changes, the mean field undergoes corresponding changes. This is shown in figure 3.1b where the field experienced by a test particle at the $C1$ methyl position, due to the mean $C5$ particle, is plotted as for several temperatures. The $C5$ atom was represented by 360 copies. Since the set of interacting conformations are weighted due to the Boltzmann relation this energy profile roughly reflects the free energy of pentane along an internal 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 raised too high, this is not a practical problem since the mean field is only a minor distortion of the true force field.

3.5.3 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 3.1b 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, the 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 significance that interactions between individual copies take.

An intriguing problem which 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 a possible 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 possible double weighting leads, in our opinion, to a much smaller error than the equal weighting of each conformation.

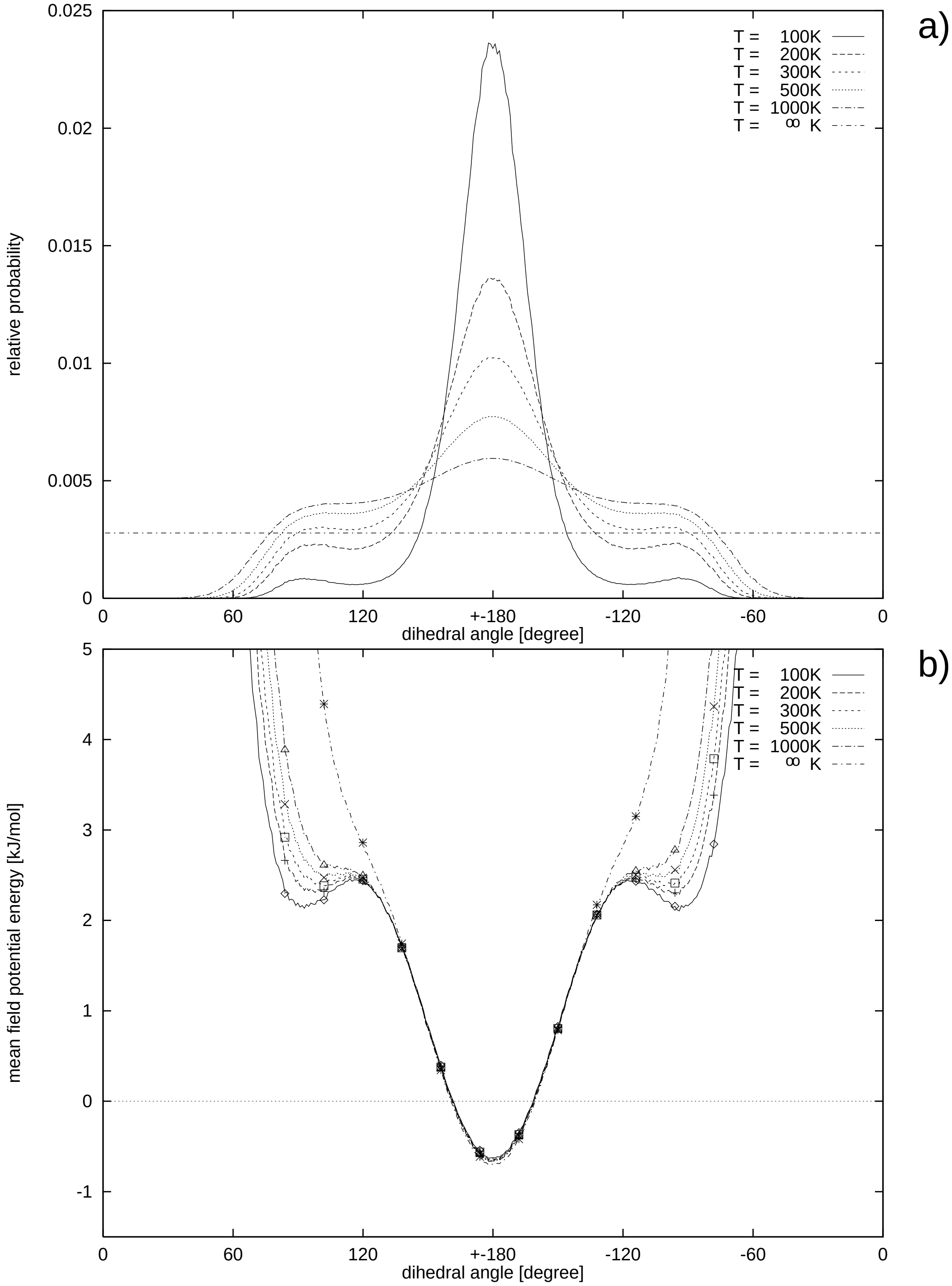


Figure 3.1: Pentane at different temperatures using 360 copies to represent $C1$ and $C5$ atoms. (a) relative Boltzmann probability distribution versus internal rotation angle ϕ. (b) Boltzmann weighted mean field potential versus internal rotation angle. Points present results from 20–copy model.

3.6 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 many bulky phenylalanines for which classical simulation methods like MD or SD fail to produce proper statistics for the different side chain conformations.

3.6.1 Free simulation

To illustrate the difficulty in generating proper statistics for phenylalanine side chain conformations in dynamics simulations, figure 3.2 shows 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 ps and only a few transitions occur over 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 and no transition to the most favoured conformation ($\chi_1 = 60°$) took place during the whole simulation. For this system, it is clear side chain statistics extracted from conventional molecular dynamics simulation, even if longer simulations are performed, have little physical meaning and be associated with huge errors.

This is different in the Boltzmann weighted multi–copy approach. Because each conformation can move to any other conformation along an artificial degree of freedom (the relative probability) transitions in the classical way 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. Whereas Cartesian coordinates were kept fixed, the relative Boltzmann probability distributions of all four phenylalanines were optimised by iteration. In figure 3.3 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, a plateau or even a slight increase of the mean field energy is observable. This is why the change of probabilities was chosen as a criterion for convergence instead of an energy criteria.

The converged probability distribution for Phe^2 side chain is shown in figure 3.4a. 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

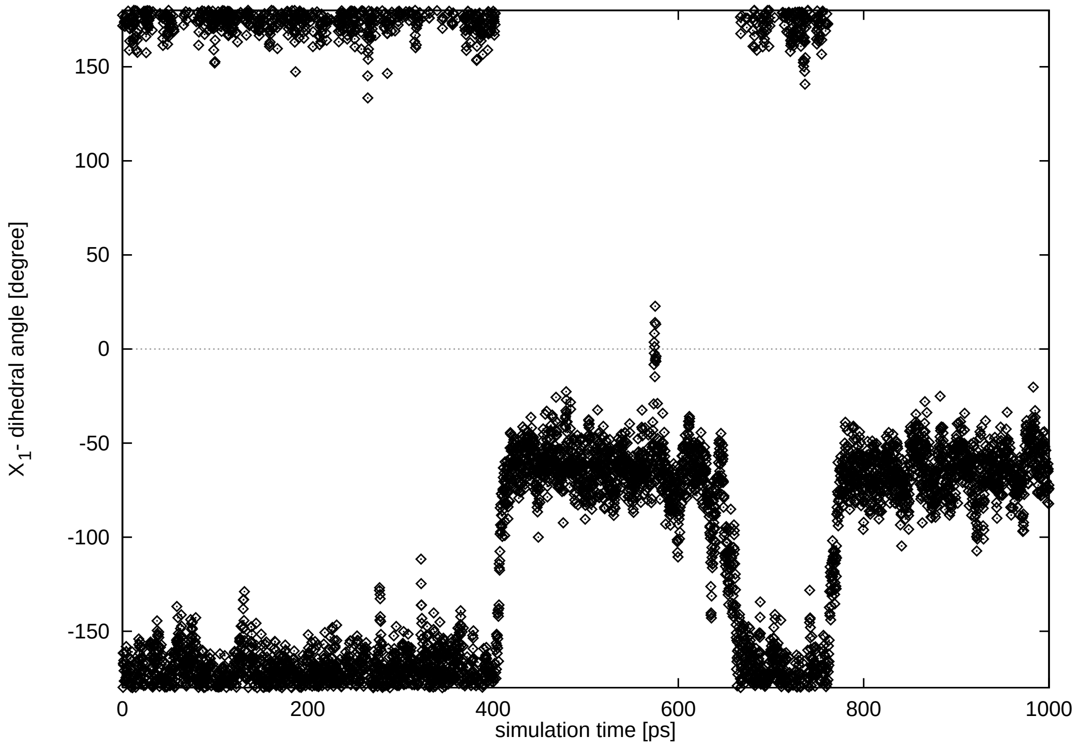


Figure 3.2: χ_1 in Phe^2 during a 1 ns free SD–simulation of cyclo-(-D-Pro-Phe-Phe-Pro-Phe-Phe-).

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 not fine enough to pinpoint minima (see figure 3.4a).

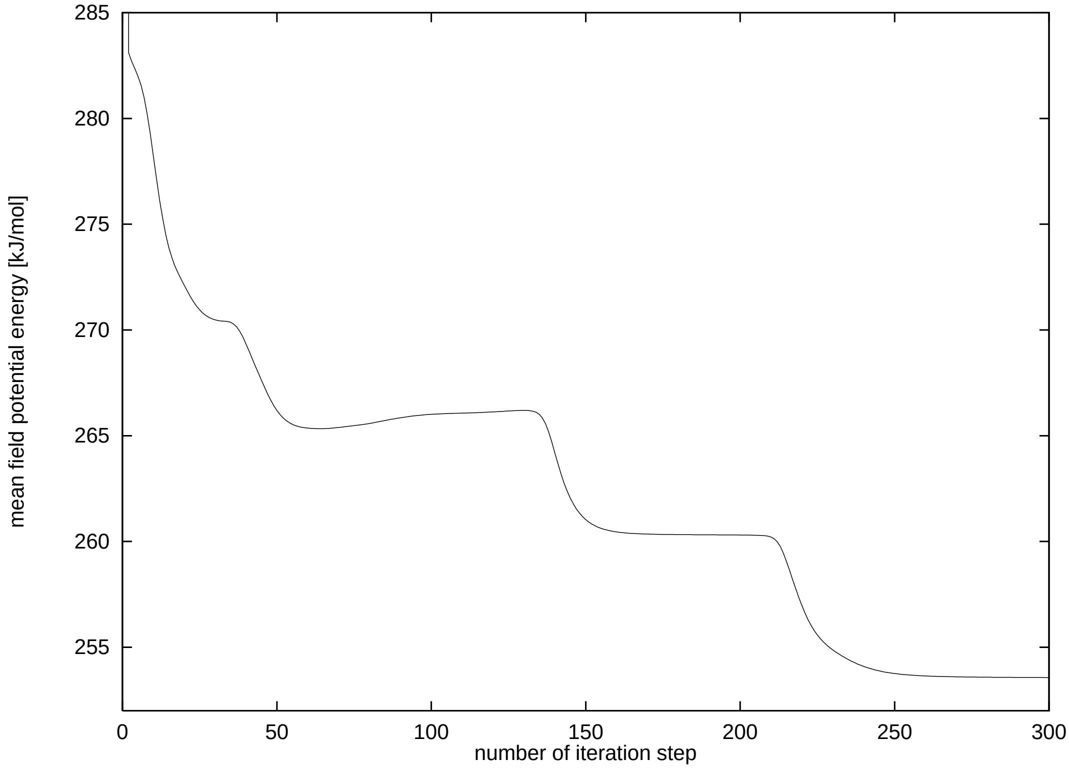
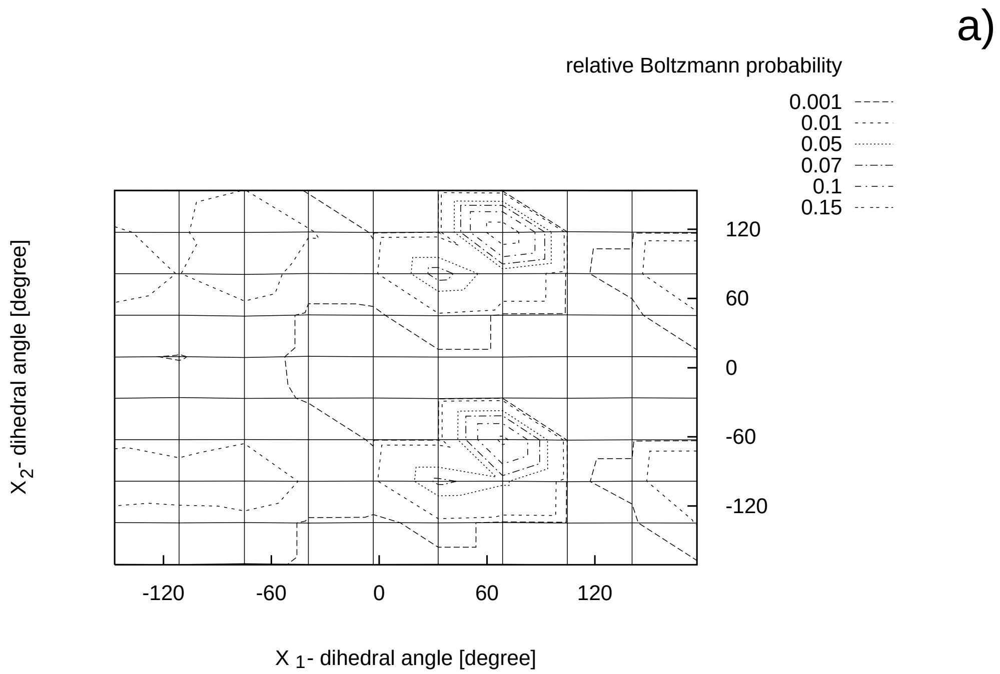


Figure 3.3: Potential energy of mean field during iterative optimisation of probability distribution in cyclo-(-D-Pro-Phe-Phe-Pro-Phe-Phe-). Each phenylalanine is represented by 100 copies uniform distributed in χ_1–χ_2 space.

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 start structure) is located in an energy minimum whereas the other conformations are defined by relative dihedral angle differences to the start 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 low energy conformations with a dynamics simulation technique.

3.6.2 Probability and cartesian coordinate optimization

For analysing the results, we used two different relative probabilities of the conformations: First, the probability of occurrence was defined as the number of times



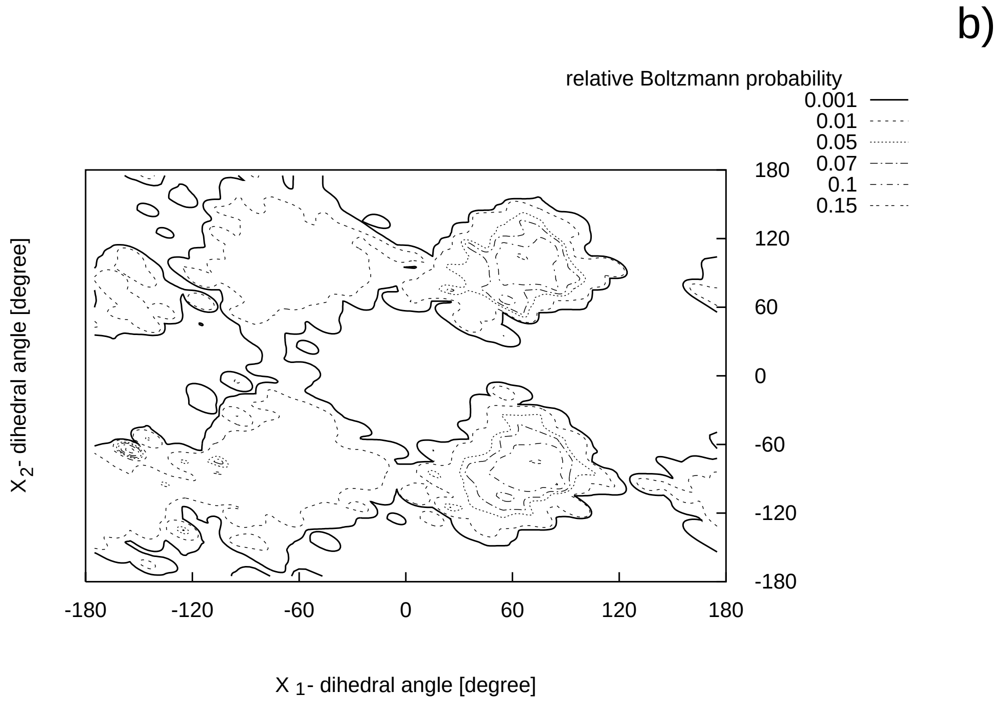


Figure 3.4: Relative Boltzmann probability distribution of Phe^2 side chain in χ_1–χ_2 space after optimization. (a) iterative optimisation with fixed set of 100 basis conformations. The grid shows the location of the basis conformations. (b) stochastic dynamics simulation in Boltzmann weighted mean field, (averaged values).

the conformation occurs in a simulation trajectory as a fraction of all conformations. Since conformations are counted, each conformation gets the same weight. Second, the relative Boltzmann probability was used. This is the relative weight of a conformation in the simulation given by the Boltzmann relation.

In figures 3.5a and 3.5b, 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 3.5c the probability of occurrence of the same torsion 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 probabilty distributions from LES simulation and Boltzmann weighted mean field simulation look similar, they differ significantly from the distributions when correct weighting is used. Whereas in the LES simulation the individual copies of the side chain are moving in the arithmetic averaged field of the other side chain copies, in the Boltzmann weighted mean field simulation, they are moving in a completely different field. This different potential 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 favoured 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 optimisation of probabilities with a fixed basis set of conformations. However, in contrast to calculations with fixed discrete conformations, a probability distribution in continuous conformational space is obtained (figure 3.4b). Furthermore, one has a higher resolution with which to exactly locate probability maxima.

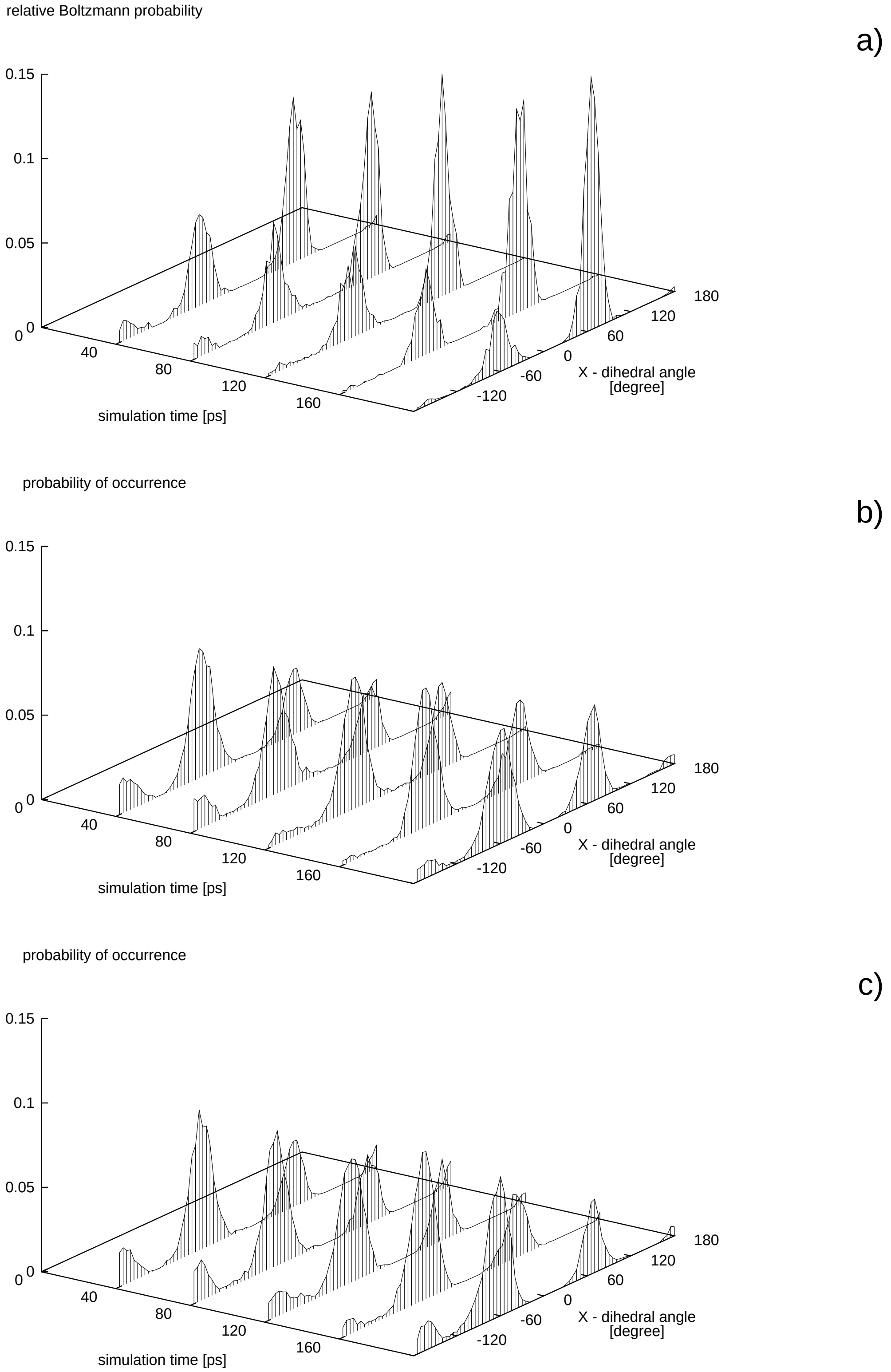


Figure 3.5: Time evolution of probability distributions of χ_1 of Phe^2 side chain during (a) Boltzmann weighted mean field simulation, boltzmann weighted probabilities. (b) Boltzmann weighted mean field simulation, probabilities of occurance. (c) LES method, probabilities of occurance.

3.7 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 well understood situations like the orientation of side chains in proteins, this is not a simple task, because small deviations of a conformation from the minimum–energy position can result in big changes of the energy. This results in even larger changes in the probability distribution.

In contrast, it was shown that using a dynamic scheme to optimize both variables – probabilities and conformational cordinates – simultaneously results in much more accurate solutions. Although we are aware of the fact that using the combination on a molecular dynamics method and performing an additional weighting of conformations can lead to an over–weighting of low energy conformations, this is probably a minor error in real systems where conformational transitions are rare events on the simulation time scale. Despite these approximations, 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.

References

[1] VAN GUNSTEREN, W.F. and BERENDSEN, H.J.C., A leap-frog algorithm for stochastic dynamics, *Mol. Simulation* **1**, 173–185 (1988).

[2] KOEHL, P. and DELAURE, M., Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy, *J. Mol. Biol.* **239**, 249–275 (1994).

[3] ROITBERG, A. and ELBER, R., Modelling side chains in peptides and proteins: Application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations, *J. Chem. Phys.* **95**, 9277–9287 (1991).

[4] KIRKPATRICK, S., GELATT, C.D.JR. and VECCHI, M.P., Optimization by simulated annealing, *Science* **220**, 671–680 (1983).

[5] VAN SCHAIK, R.C., VAN GUNSTEREN, W.F. and BERENDSEN, H.J.C, Conformational search by potential annealing algorithm and application to Cyclosporin A, *J. Computer Aided Mol. Design* **6**, 97–112 (1992).

[6] HUBER, T., TORDA, A.E. and VAN GUNSTEREN, W.F., Local elevation: A method for improving the searching properties of molecular dynamics simulation, *J. Computer Aided Mol. Design* **8**, 695–708 (1994).

[7] WEBER, P.L., MORRISON, R. and HARE, D., Determining stereo–specific ^{1}H nuclear magnetic resonance assignment from distance geometry calculations, *J. Mol. Biol.* **204**, 483–487 (1988).

[8] CRIPPEN, G.M., Conformational analysis by energy embedding, *J. Comput. Chem.* **3**, 471–476 (1982).

[9] CRIPPEN, G.M., Linearized embedding: A new metric matrix algorithm for calculating molecular conformations subject to geometric restraints, *J. Comput. Chem.* **10**, 896–902 (1989).

[10] CRIPPEN, G.M. and HAVEL, T.F., Global energy minimization by rotational energy embedding, *J. Chem. Inf. Comput. Sci.* **30**, 222–227 (1990).

[11] PURISIMA, E.O. and SCHERAGA, H.A., An approach to the multiple-minima problem by relaxing the dimensionality, *Proc. Natl. Acad. Sci. USA* **83**, 2782–2786 (1986).

[12] VAN SCHAIK, R.C., BERENDSEN, H.J.C., TORDA, A.E. and VAN GUNSTEREN, W.F., A structure refinement method based on molecular dynamics in four spatial dimensions, *J. Mol. Biol.* **234**, 751–762 (1993).

[13] GERBER, R.B., BUCH, V. and RATNER, M.A., Time dependent selfe consistent field approximation for intermolecular energy transfer. I. Formulation and application to dissociation of van der Waals molecules, *J. Chem. Phys.* **77**, 3022–3030 (1982).

[14] STRAUB, J.E. and KARPLUS, M., Energy equipartitioning in the classical time-dependent Hatree approximation, *J. Chem. Phys.* **94**, 6737–6739 (1990).

[15] HANSMANN, U.H.E. and OKAMOTO, Y., Prediction of peptide conformation by multicanonical algorithm: New approach to the multiple minima problem, *J. Comp. Chem.* **14**, 1333–1338 (1993).

[16] ROSENFELD, R., ZHENG, Q., VAJDA, S. and DILISI, C., Computing the structure of bound peptides: Application to antigen recognition by class I major histocompatiblility complex receptors, *J. Mol. Biol.* **234**, 515–521 (1993).

[17] ROSENFELD, R., ZHENG, Q., VAJDA, S. and DILISI, C., Theoretical ananlysis of the multicopy sampling method in molecular modeling, *J. Chem. Phys.* **99**, 8892–8896 (1993).

[18] ZHENG, Q. and KYLE, D.J., Multiple copy sampling: Rigid versus flexible protein, *Proteins* **19**, 324–329 (1994).

[19] MIERKE, D.F., SCHEEK, R.M. and KESSLER, H., Coupling Constants As Restraints In Ensemble Distance Driven Dynamics, *Biopolymers* **34**, 559–563 (1994).

[20] ELBER, R., Calculation of the potential of mean force using molecular dynamics with linear constraints: An application to a conformational transition in a solvated dipeptide, *J. Chem. Phys.* **93**, 4312–4321 (1990).

[21] VERKHIVKER, G., ELBER, R. and GIBSON, Q., Microscopic modelling of ligand diffusion through the Protein Leghemoglobin: Computer simulations and Experiments, *J. Am. Chem. Soc.* **114**, 7866–7878 (1992).

[22] VERKHIVKER, G., ELBER, R. and NOWAK, W., Locally enhanced sampling in free energy calculations: Application of mean field approximation to accurate calculation of free energy differences, *J. Chem. Phys.* **97**, 7838–7841 (1992).

[23] ULITSKY, A. and ELBER, R., The thermal equilibrium aspects of the time dependent Hartree and the locally enhanced sampling approximations: Formal properties, a correction, and computitional examples for rare gas clusters, *J. Chem. Phys.* **98**, 3380–3388 (1993).

[24] ULITSKY, A. and ELBER, R., Application of the locally enhanced sampling (LES) and a mean field with a binary collision correction (cLES) to the simulation of Ar diffusion and NO recombination in Myoglobin, *J. Chem. Phys.* **98**, 1034–1043 (1994).

[25] VAN GUNSTEREN, W.F., HUBER, T. and TORDA, A.E., Biomolecular modelling: Overview of types of methods to search and sample conformational space, in *Proceedings of the 1st European Conference on Computational Chemistry, American Institute of Physics Conf. Proc.*, edited by BERNARDI, F. and RIVALL, J.-L., pp. 253–268, New York, 1994.

[26] VAN GUNSTEREN, W.F. and BERENDSEN, H.J.C., *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos, Groningen, 1987.

[27] RYCKAERT, J.-P., CICCOTTI, G. and BERENDSEN, H.J.C., Numerical Integration of the cartesian equation of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Comput. Phys.* **23**, 327–341 (1977).

[28] BERENDSEN, H.J.C., POSTMA, J.P.M., VAN GUNSTEREN, W.F., DINOLA, A. and HAAK, J.R., Molecular dynamics with coupling to an external bath, *J. Chem. Phys.* **81**, 3684–3690 (1984).

[29] WAGNER, K., *Synthese und Konformationsanalyse von cytoprotektiv wirksamen cyclischen Hexapeptiden*, PhD thesis, University of Frankfurt/Main, Germany, 1988.