Theor Chem Acc (1999) 101:62–66
DOI 10.1007/s00214980m144

Theoretical
Chemistry Accounts
© Springer-Verlag 1999

Regular article

Adaptive umbrella sampling of the potential energy: modified updating procedure of the umbrella potential and application to peptide folding*

Christian Bartels[1], **Michael Schaefer**[1], **Martin Karplus**[1,2]

[1] Laboratoire de Chimie Biophysique, Institut Le Bel, Université Louis Pasteur, 4 rue Blaise Pascal, F-67000 Strasbourg, France
[2] Department of Chemistry, Harvard University, Cambridge, MA 02138, USA

Received: 11 July 1998 / Accepted: 22 September 1998 / Published online: 17 December 1998

Abstract. Adaptive umbrella sampling of the potential energy is used as a search method to determine the structures and thermodynamics of peptides in solution. It leads to uniform sampling of the potential energy, so as to combine sampling of low-energy conformations that dominate the properties of the system at room temperature with sampling of high-energy conformations that are important for transitions between different minima. A modification of the procedure for updating the umbrella potential is introduced to increase the number of transitions between folded and unfolded conformations. The method does not depend on assumptions about the geometry of the native state. Two peptides with 12 and 13 residues, respectively, are studied using the CHARMM polar-hydrogen energy function and the analytical continuum solvent potential for treatment of solvation. In the original adaptive umbrella sampling simulations of the two peptides, two and six transitions occur between folded and unfolded conformations, respectively, over a simulation time of 10 ns. The modification increases the number of transitions to 6 and 12, respectively, in the same simulation time. The precision of estimates of the average effective energy of the system as a function of temperature and of the contributions to the average effective energy of folded conformations obtained with the adaptive methods is discussed.

Key words: Adaptive umbrella sampling – Multicanonical sampling – Helical peptide – β-hairpin

1 Introduction

Molecular dynamics and Monte Carlo simulations are well-established methods to study systems of biological interest [1, 2]. However, with increasing complexity of the simulation system, it becomes more difficult to obtain reliable estimates of the equilibrium properties. This becomes particularly difficult if conformational transitions that are rare on the simulation time scale (e.g., between the folded and unfolded state) are important. To obtain estimates of the equilibrium properties of biomolecules, we recently introduced a molecular dynamics adaptive umbrella sampling technique [3] (see also Refs. [4, 5]) that leads to an increase in the number of transitions between the conformations that contribute to the equilibrium properties. The method introduces a biasing potential (umbrella potential) which is a function of the potential energy. The effect of the bias on estimates of the properties of the system can be calculated and quantitative estimates of the properties of the unbiased system can be obtained. For the penta-peptide met-enkephalin (sequence YGGFM) in vacuum and for the tetra-peptide RGDW in vacuum and in a 30 Å box of explicit water, it has been demonstrated that equilibrium properties can be estimated from simulations several nanoseconds in length since the umbrella potentials converge and several transitions between the different important conformations occur in these simulations [3, 6]. In the present work, we use the recently introduced analytical continuum solvent (ACS) model [7, 8] and investigate the convergence of the method for the peptide RN24, succinyl-AETAAAKFLRAHA-NH_2, which has been shown experimentally to have a high population of helix conformations [9] and the peptide BH8, NH_3^+-RGITVNGKTYGR-COO^-, which has a high population of β-hairpin conformations [10]. To increase the number of transitions between the unfolded and folded conformations of these peptides, we introduce a modification of the updating procedure for the umbrella potential. The effect of the modification on the precision of properties derived from the simulations is analyzed. Equilibrium properties of the two systems derived from these simulations such as estimates of the $^3J_{\mathrm{HNH}\alpha}$ coupling constants or the free energy of folding are

*Contribution to the Proceedings of Computational Chemistry and the Living World, April 20–24, 1998, Chambery, France

Correspondence to: M. Karplus

described in Ref. [8]. A comparison with simulations of the same peptides using a range of models for solvation effects is presented in a companion paper [11].

2 Methods

Conformations that determine the properties of a system at low (room) temperature generally have low potential energies. Transition-state conformations that are important for the equilibration of the system and for transitions between different important low-energy conformations have, in general, higher energies. Adaptive umbrella sampling [12–15] of the potential energy, referred to as "energy sampling" [3], leads to uniform sampling of the potential energy and provides a way of sampling conformations important at room temperature and the transition states connecting them. Thus, energy sampling should increase the number of transitions between the different conformations important at low temperatures (see later and Ref. [3]).

To achieve uniform sampling of the potential energy V, a number of simulations i with a modified Hamiltonian $H_i = H^\circ + U_i(V)$ are carried out in which the umbrella potential $U_i(V)$ is added to the Hamiltonian H° of the system. After each simulation the umbrella potential $U_i(V)$ is updated based on statistics of the sampling of the potential energy in the previous simulations, $j = 1, 2, \ldots, i$, such that more uniform sampling of the potential energy is expected. For this purpose, the potential energy is partitioned into bins $V \in (\xi_k, \xi_{k+1}]$ with indices k, where the ξ denote the boundaries between the bins, and the number of times $n_{j,k}$ in which the system is found in a particular bin k is determined for each of the simulations j. An estimate $\tilde{p}_k^\circ$ for the probability of finding the unperturbed system in a particular bin is calculated by determining a self-consistent solution of the weighted histogram analysis method (WHAM) Eqs. (1) and (2) [14, 16–18]

$$\tilde{p}_k^\circ = \frac{\sum_j n_{j,k}}{\sum_j N_j \tilde{f}_j c_{j,k}} \tag{1}$$

$$\tilde{f}_j = \frac{1}{\sum_k c_{j,k} \tilde{p}_k^\circ} \ , \tag{2}$$

with $c_{j,k} = \mathrm{e}^{-U_j(\bar{\xi}_k)/RT}$ and $N_j = \sum_k n_{j,k}$. In Eq. (1) the summations are over all simulations and in Eq. (2) over all bins. The symbol R is the gas constant, T is the temperature of the simulations, $\bar{\xi}_k = \frac{1}{2}\{\xi_k + \xi_{k+1}\}$ is the mid point of bin k, and $\tilde{f}_j$ is a scaling factor which arises from the condition that the probability of finding the system in any of the bins has to be 1. The umbrella potential $U_{i+1}(V)$ for the next simulation is obtained from the estimates $\tilde{p}_k^\circ$ as described in Ref. [14].

Energy sampling increases, in general, the number of transitions between the different important conformations. Nonetheless small residual free-energy barriers between the different important conformations may remain even after the umbrella potential has converged in the energy sampling runs [3]. The height of the residual free-energy barriers determines how often transitions occur. For met-enkephalin, the threonine dipeptide [3], the peptide RGDW [6], and the peptide BH8 (see later), several transitions occur during energy sampling runs of a few nanoseconds, indicating that the residual free-energy barriers are small. For the peptide RN24 only one folding and one unfolding event occurs during an energy sampling run of 10 ns (see later), indicating that the residual free-energy barriers are larger, so that convergence of structural and thermodynamic properties is not achieved.

To develop a scheme to increase the number of transitions, we consider a system with a local minimum from which the system has a very low probability of escaping in an unbiased simulation. If an energy sampling run starts in this local minimum, the statistics acquired in the first few simulations would faithfully represent the energy distribution (density of states) around this local minimum and the umbrella potential derived from these statistics would increase sampling of higher energy structures including transition states and thereby would enable the system to diffuse away from it. By contrast, if other regions of conformational space with a different energy distribution (e.g., a second local minimum with a somewhat higher potential energy) exist and if the system visits the first local minimum after it has been in the other regions of conformational space, then the acquired statistics would not represent the energy distribution around the local minimum anymore and the umbrella potential derived from the statistics would not necessarily increase the probability of sampling higher energy structures sufficiently for the system to diffuse away. However, if only the statistics acquired while the system was in the local minimum are retained and the earlier statistics from the other parts of the conformation space are suppressed, the umbrella potential would again reflect the region around the minimum and enable the system to diffuse away from it.

From the scenario of the previous paragraph it is to be expected that giving a high weight to the most recent statistics should lead to an umbrella potential that better represents the energy distribution around the current structure and enables the system to explore new structures. We implement such a weighting by replacing the statistics $n_{j,k}$ of the simulations $j = 1, 2, \ldots, i$ with

$$n'_{j,k} = n_{j,k} w^{i-j} \ , \tag{3}$$

where w is a positive weighting factor smaller than unity. The modified statistics $n'_{j,k}$ are then used in the WHAM Eqs. (1) and (2) to calculate the estimates $\tilde{p}_k^\circ$ and the umbrella potential $U_{i+1}(V)$ for the next simulation $i + 1$. Since estimates of properties calculated from the runs should be based on all regions of conformational space and not only on the region of conformational space sampled in the last few simulations, no weighting is used when calculating estimates in the final analysis of the run.

Energy sampling runs of the peptides BH8 and RN24 were performed at 1000 K using the program CHARMM [19] with the PARAM19 force field [19, 20] and the ACS potential [7, 8]. Nonbonded interactions were truncated with a switching function [19] from 8 to 12 Å. The time step of the integrator was set to 1 fs and the SHAKE algorithm [21] was used to fix the lengths of bonds involving hydrogens. Coordinates for the analysis were saved every 0.5 ps. The range of potential energies to be sampled was restricted to potential energies important in canonical ensembles with temperatures between 250 and 1100 K as described in Ref. [3] (see also Ref. [8]). Each run consisted of 1000 simulations $i = 1, 2, \ldots, 1000$ with 1000 molecular dynamics steps for equilibration followed by 9000 molecular dynamics steps for production. After each production simulation the umbrella potential was updated. Statistics on the sampling of the potential energies were acquired for the potentials energies relevant in canonical ensembles between 250 and 1100 K using 500 bins with a width of 2 kcal/mol per bin. For both peptides (BH8 and RN24) two energy sampling runs were carried out, one in which the statistics were not modified and one in which Eq. (3) was used. In the latter runs, the weighting factor w (Eq. 3) was set to $0.1^{\mathbf{1/100}} = 0.977$ which is a compromise between very small weighting factors which would introduce many transitions but would also lead to large fluctuations of the umbrella potential (see later) and a weighting factor of 1.

To analyze the content of secondary structures, we used the program DSSP [22] which classifies the residues into different secondary structure classes based on the hydrogen-bonding pattern. The helix and β-hairpin content is then calculated by dividing the number of residues having a helical or β-hairpin structure, respectively, by the total number of residues (see also Ref. [8]).

3 Results

Figure 1 illustrates the sampling of conformations during the four energy sampling runs. In the run of the β-hairpin peptide BH8 (Fig. 1A) without weighting of the statistics, hairpin conformations are sampled repeatedly with three instances in which the peptide is in a hairpin conformation. Here (and later), we count it as one instance of secondary structure formation if the

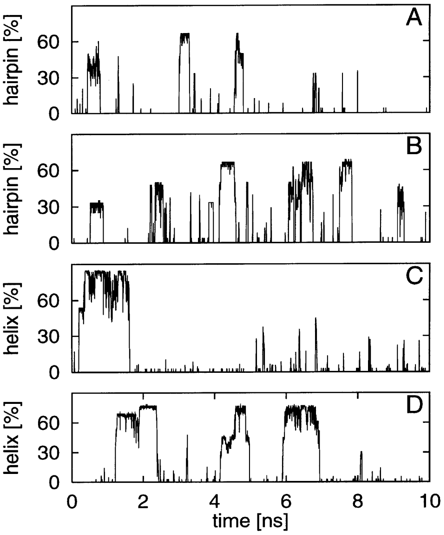

Fig. 1A–D. Simulation results for β-hairpin and helix content as a function of simulation time. **A** and **B** are from the energy sampling runs of the peptide BH8 without and with weighting of the statistics (Eq. 3, $w = 0.1^{1/100}$), respectively. **C** and **D** are from the energy sampling runs of the peptide RN24 without and with weighting of the statistics, respectively

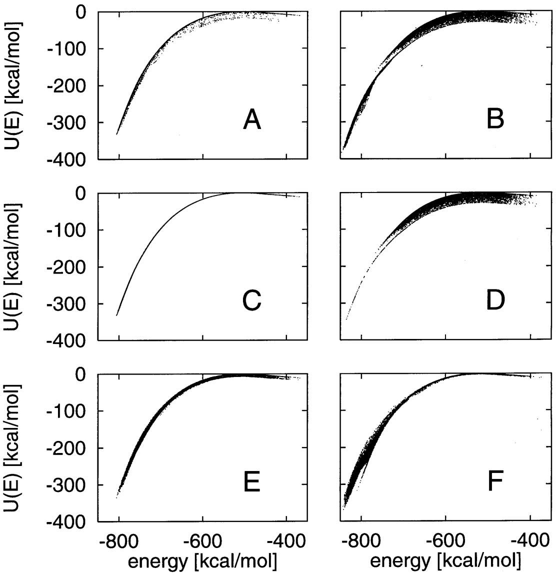

Fig. 2A–F. Umbrella potential during the energy sampling runs with and without weighting of the statistics. For each coordinate frame saved during the simulations, the umbrella potential is plotted versus the potential energy. **A** and **B** are from the entire runs (10 ns) without weighting of the statistics of the two peptides BH8 and RN24, respectively. **C** and **D** are from the last 8.5 ns of the two runs without weighting of the statistics, and **E** and **F** are from the last 8.5 ns of two runs with weighting of the statistics, respectively

peptide has more than 30% hairpin (or helix) content for more than 0.1 ns. In contrast, in the run of the α-helical peptide RN24 without weighting of the statistics (Fig. 1C), there is only a single instance during which a significant amount of helical conformation is present. This indicates that residual free-energy barriers are larger for the α-helical peptide RN24 than for the β-hairpin peptide BH8 and that much longer simulations would be required to obtain sufficient transitions for the peptide RN24. In the two runs in which Eq. (3) was used to emphasize the most recent statistics (Fig. 1B, D), more transitions occur: there are about six instances in which the peptide BH8 is in a hairpin conformation and three instances in which the peptide RN24 is in a helical conformation. Thus, the weighting scheme (Eq. 3) achieves the goal of increasing the number of transitions between folded and unfolded conformations.

Figure 2A, B shows the umbrella potentials in the two runs without weighting. The umbrella potential in the run of the β-hairpin peptide BH8 (Fig. 2A) fluctuates less than that in the run of the helical peptide RN24 (Fig. 2B). The fluctuations of the umbrella potentials are caused by the extrapolation used to obtain the umbrella potential for the bins for which no statistics have been acquired so far [3], and by random fluctuations. Once the entire potential energy range has been sampled and statistics have been acquired for all relevant bins, no extrapolation is used; in all the runs presented, this is the case after 1.5 ns. To study the random fluctuations only, the umbrella potentials from the last 8.5 ns of the runs without weighting are shown in Fig. 2C, D. There is virtually no variation in the umbrella potential of BH8 (Fig. 2C), for example, at a potential energy of −700 kcal mol^{-1} the umbrella potential varies only between −97.0 and −96.0 kcal mol^{-1}. In contrast, there are significant fluctuations in the umbrella potential of the run of RN24 without weighting (Fig. 2D), for example, at −700 kcal mol^{-1} the umbrella potential varies between about −103 and −78 kcal mol^{-1}. The umbrella potentials in the last 8.5 ns of the runs with weighting (Fig. 2E, F) fluctuate with a similar magnitude as the umbrella potential in the run of RN24 without weighting, for example, at −700 kcal mol^{-1} the umbrella potential varies between about −109 and −93 kcal mol-^{1} in the run of BH8 (Fig. 2E) and between about −83 and −73 kcal mol^{-1} in the run of RN24 (Fig. 2F).

To estimate observables of an unbiased system at a selected temperature, the conformations of an energy sampling run have to be weighted by appropriate factors [3, 14]. The determination of the weighting factors [14] does not make assumptions about the umbrella potential. From this point of view, estimates derived from an energy sampling run in which different umbrella potentials are used in the different simulations are as valid as estimates derived from simulations in which the umbrella potential is essentially the same in the entire run. However, it is important that the different simulations are in equilibrium, i.e., that the counts $n_{j,k}$ differ only by random fluctuations from their expectation values. This is the case for simulations that extend over a time period which is sufficiently long for all important transitions to occur a few times. In the run of BH8 without weighting, transitions between unfolded and folded conformations occur with a rate of about 0.3 ns^{-1}, which is slow compared to the duration of 10 ps of each individual simulation. However, since the umbrella potential has converged and so is essentially the same in the different simulations (Fig. 2C), the 1000 simulations of the run can be considered as a single simulation of 10 ns duration, during which several transitions between the different important conformations occur. Thus, the run of BH8 is in equilibrium and estimates derived from the run are affected by statistical errors only. In the remaining runs (BH8 with weighting, RN24 with and without weighting), important transitions have a rate which is slow compared to the length of individual simulations, and the umbrella potentials differ in the different simulations. Therefore, it cannot be decided, based on the above arguments, whether the runs are in equilibrium and whether estimates derived from the runs are affected by systematic errors due to the sampling procedure.

To assess the magnitude of possible errors introduced by the weighting of the statistics (Eq. 3), we compare estimates of the average effective energy of folded structures at 275 K and of the contributions of different terms in the energy function to the average effective energy derived from the different runs (Table 1) (note that the effective energy defined by the ACS potential is a free energy that contains contributions from the solvent entropy [7, 8]). The individual contributions to the average effective energy of folded structures differ by about 1–2 kcal mol^{-1} between the runs with and without weighting of the statistics (Table 1). This difference is small compared to the difference of up to 10 kcal mol^{-1} between the terms of the average effective energy of folded and unfolded conformations (not shown) and to the variation of the total average effective energy of more than 200 kcal mol^{-1} over the temperature range from 250 to 700 K. Differences in the estimates of the average effective energy of the system at different temperatures calculated from the runs with and without weighting (Fig. 3) are smaller than 5 kcal mol^{-1} for BH8 for all temperatures and for RN24 at low and at high temperatures. For RN24 at intermediate temperatures, the differences increase up to 24 kcal mol^{-1} at 350 K, which is large compared to the fluctuations of the average effective energy at a given temperature: these have values in the range 8–20 kcal mol^{-1} in both systems for temperatures between 250 and 700 K. At 275 K, the temperature of primary interest (when RN24 is stable), the error is only about 3 kcal mol^{-1}, which is sufficiently small to yield meaningful results.

These results suggest that for the peptide BH8, the weighting of the statistics (Eq. 3 with $w = 0.1^{1/100}$) has little effect on the estimates derived from the simulations. This is correlated with the facts that in the run without weighting a significant number of transitions occur and that the number of transitions is only slightly increased by the weighting of the statistics (Fig. 1). For the peptide RN24, the weighting leads to a significant increase in the number of transitions, which is correlated with the large differences between estimates derived from the runs of RN24 with and without weighting. These differences are pronounced only at potential energies in the intermediate potential energy range; at low and at high potential energies the differences are small (Fig. 3 and Table 1). The ensemble of structures at low potential energies is dominated by folded conformations, while the ensemble of structures at high potential energies is dominated by unfolded conformations. The ensemble at intermediate potential energies consists of both, folded and unfolded conformations, as well as transition states. Thus, it seems that properties related to transitions between the folded and unfolded conformation of RN24, such as the average effective energy at intermediate temperatures or the free-energy difference between folded and unfolded conformations, are affected by systematic errors due to the weighting of the statistics, whereas the average effective energy or structural properties of the folded conformations are not.

Comparison with experiments can give additional information on the quality of simulations. For the present systems, there is, in general, good agreement between experiments and simulations, although there are some minor differences [8]. For the system RN24, for example, both the experiments and the simulations

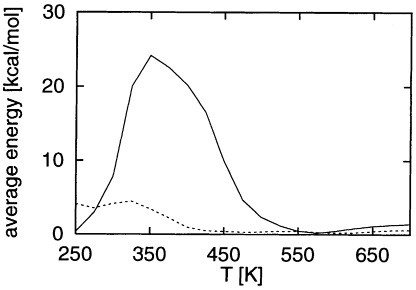

Fig. 3. Difference in the estimate of the average effective energy as a function of temperature calculated from the runs with and without weighting of the statistics. The *continuous line* is for the RN24 system, the *dashed line* for the BH8 system

Table 1. Contributions to the average effective energy of folded conformations at 275 K derived from the four energy sampling runs. The average effective energy and the contributions to the average effective energy from internal energy terms (internal), van der Waals energy terms (vdW), non-polar solvation free-energy terms (hydrophobic), and electrostatic free-energy terms (electrostatic) as defined by the analytical continuum solvent potential are listed. Values derived from the energy sampling runs with and without weighting of the statistics are given for β-hairpin conformations of the peptide BH8 and for helical conformations of the peptide RN24 (see Ref. [8] for definition)

Enthalpy (kcal mol^{-1})	BH8 (no weighting)	BH8 (weighting)	RN24 (no weighting)	RN24 (weighting)
Internal	101.2	102.3	102.3	102.7
vdW	−44.9	−46.5	−59.3	−59.7
Hydrophobic	29.3	29.3	30.9	30.8
Electrostatic	−857.1	−855.2	−884.5	−882.7
Total	−771.5	−770.1	−810.5	−808.8

clearly show that the peptide exists as a helix and an extended chain; however, the simulations provide no evidence for conformations with a distorted helix and a salt bridge between Glu 2 and Arg 10 which was proposed to be present as a third set of conformations based on the experiments. Drawing conclusions regarding the sampling procedure from these results is difficult, since comparisons of simulations with experiments are affected by a number of factors other than the sampling procedure, such as the quality of the force field and the quality of the experimental data.

4 Conclusions

Adaptive umbrella sampling of the potential energy (energy sampling) has been used to characterize the folding of peptides. For the β-hairpin peptide BH8, six transitions between folded and unfolded conformations occur in a energy sampling run of 10 ns. In the last 8.5 ns of the run, the umbrella potential varies by less than 1 kcal mol^{-1}, which provides strong evidence that the umbrella potential has converged and that the estimates derived from the run are affected only by statistical errors. The magnitude of the statistical errors of properties derived from energy sampling runs in which the umbrella potentials have converged were estimated in Refs. [3, 6] for the peptide RGDW and for met-enkephalin. Work to assess the magnitude of the statistical errors for the present systems is in progress. For the helical peptide RN24, only a single instance of helix formation is observed in a run of 10 ns and the umbrella potential varies significantly throughout the run. To increase the number of transitions, we introduced a scheme that gives higher weights to statistics from the most recent simulations (Eq. 3). This scheme enables the system to diffuse away from its current position and leads to more efficient exploration of conformational space. For the two peptide systems studied here, we found that the weighting with Eq. (3) increases the number of transitions between folded and unfolded conformations. In particular, there were three instances of helix formation in a 10 ns energy sampling run of the peptide RN24 with weighting. Since in the energy sampling runs with weighting of the statistics, the umbrella potentials differ in the different simulations, estimates derived from the runs may be affected by systematic errors due to the sampling procedure. For the peptide BH8, there were no significant differences in estimates derived from the runs with and without weighting, indicating that systematic errors due to the weighting of the statistics are not important in this case. For the peptide RN24, we found that properties of the folded and the unfolded conformations are not significantly affected by the weighting of the statistics; however, properties that depend on transitions between the folded and unfolded conformations are affected.

Increasing the number of transitions between different conformations leads to better sampling of conformational space and increases the probability of finding all the important folded conformations of the system. Since the estimates of the properties of the folded conformations seem not to be affected by the weighting, the proposed weighting of the statistics, appears to be a good method for identifying the folded conformations of the system and characterizing them. However, to obtain reliable estimates of properties related to transitions between folded and unfolded conformations, a scheme is needed that leads to a significant number of transitions between the conformations of interest and excludes systematic errors due to the sampling procedure. We are currently developing such a method that requires structural information on the conformations of interest. Such information can be obtained with the present method that does not depend on assumptions about the geometry of the various conformations, including the folded structure.

Acknowledgements. The Institut de Développment et des Ressources en Informatique Scientifique provided an allocation of CRAY T3E computer time which was used in part for this work. C.B. was supported by the EMBO long-term fellowship ALTF 448-1995 and the fellowship 823A-050424 of the Swiss National Science Foundation. The laboratory is supported in part by the CNRS (ISIS-UPRESA 7006), by the Ministère de l'Education Nationale, by a grant from the Association pour la Recherche contre le Cancer (France), and by a grant from the National Institutes of Health (USA).

References

1. Brooks CL III, Karplus M, Pettitt BM (1988) Proteins: a theoretical perspective of dynamics, structure, and thermodynamics. Wiley, New York
2. McCammon JA, Harvey S (1987) Dynamics of proteins and nucleic acids. Cambridge University Press, Cambridge, UK
3. Bartels C, Karplus M (1998) J Phys Chem B 102: 865
4. Nakajima N, Nakamura H, Kidera A (1997) J Phys Chem 101: 817
5. Hansmann UHE, Okamoto Y, Eisenmenger F (1996) Chem Phys Lett 259: 321
6. Bartels C, Stote RH, Karplus M (1998) J Mol Biol (in press)
7. Schaefer M, Karplus M (1996) J Phys Chem 100: 1578
8. Schaefer M, Bartels C, Karplus M (1998) J Mol Biol 284:835
9. Osterhout JJ, Baldwin RL, York EJ, Stewart JM, Dyson HJ, Wright PE (1989) Biochemistry 28: 7059
10. Ramírez-Alvarado M, Blanco FJ, Serrano L (1996) Nat Struct Biol 3: 604
11. Schaefer M, Bartels C, Karplus M (1998) Theor Chem Acc DOI 10.1007/s00214980m143
12. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1995) J Comput Chem 16: 1339
13. Hooft RWW (1992) J Chem Phys 97: 6690
14. Bartels C, Karplus M (1997) J Comput Chem 18: 1450
15. Mezei M (1987) J Comput Phys 68: 237
16. Boczko EM, Brooks CL III (1993) J Phys Chem 97: 4509
17. Ferrenberg AM (1989) Efficient use of Monte Carlo simulation data, Ph.D. Thesis, Carnegie-Mellan University
18. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM (1992) J Comput Chem 13: 1011
19. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comput Chem 4: 187
20. Neria E, Fischer S, Karplus M (1996) J Chem Phys 105: 1902
21. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) J Comput Phys 23: 327
22. Kabsch W, Sander C (1983) Biopolymers 22: 2577