ELSEVIER

Physica A 239 (1997) 244–254

PHYSICA A

Simulations of the thermodynamic properties of a short polyalanine peptide using potentials of mean force

Matteo Pellegrini[a], Niels Grønbech-Jensen[b,c,*], Sebastian Doniach[d]

[a] *Molecular Biology Institute, University of California, Los Angeles, CA 90095, USA*

[b] *Department of Physics, University of California, Los Angeles, CA 90024, USA*

[c] *Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

[d] *Departments of Applied Physics and Physics, Stanford University, Stanford, CA 94305-4090, USA*

Abstract

We report results of simulation studies of the equilibrium between helical and random coil states of dodeca-alanine by a method which systematically includes the effects of hydration. The statistical distribution of conformations at room temperature is determined and used to fit thermodynamic parameters for the helix to random coil equilibrium which may be directly compared with experimental data. The distribution of conformations is explored at atomic resolution by a constrained Langevin-dynamics simulated annealing technique which limits the effective number of degrees of freedom to two per residue. Water is included via a set of pair potentials of mean force. We show that the inclusion of hydration effects is essential for achieving a helical content comparable to that found experimentally. However, the cooperative nature of the helical state found in the simulations is low compared to the experimental value.

1. Introduction

During the past 10 years great progress has been made in the theoretical simulation of biomolecules. Molecular dynamics (MD) codes are at present able to simulate 10 ns trajectories of polyalanine peptides solvated in water molecules [1]. However, due to the limited number of trajectories one is able to compute in a reasonable amount of time, it has so far proved difficult to extract thermodynamic parameters from these simulations. Yet, the thermodynamic properties of this system are those that can be measured experimentally and there is, therefore, a great deal of interest in the

* Correspondence address: Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

PII S0378-4371(96)00483-9

simplification of atomic resolution simulations so that such parameters can be computed [2–5]. In this paper we report on a new approach that allows us to reduce the number of degrees of freedom used in the simulations, thus rendering the equilibrium properties of the system computationally accessible.

Since the simulations are conducted in an aqueous solvent, the first simplification we introduce lies in removing the explicit degrees of freedom associated with the water molecules while preserving their statistically averaged properties [6]. This is accomplished by generating pairwise potentials of mean force, between any two atoms in the peptide, that implicitly account for the average effect of water on the two solutes. We have previously shown that this approach adequately describes the solvated state of the alanine dipeptide [7].

The second approximation involves constraining most of the internal degrees of freedom of the peptide, by effectively freezing all bond lengths and angles other than the dihedral degrees of freedom. This is accomplished by using the constrained Langevin dynamics algorithm [8], which we have previously used to study the dynamics of simple polymers. In this algorithm, the bond lengths and angles, other than the dihedral angles, are maintained constant by subtracting the component of the force that perturbs them from the total force on each atom. By freezing the bond lengths and bond angles we are able to eliminate the contribution of fast vibrations to the dynamics, thus reducing the total degree for a 12 long polyalanine peptide from 216 to 24. These 24 are just the ϕ and ψ angles of the peptide backbone.

However, constraining the bond lengths and angles has the effect of raising the barriers between energy minima of the peptide. To reduce the barriers to the levels found in unconstrained simulations, we choose to describe the atoms as Gaussian distributions instead of point-like particles. This has the effect of smoothing the potentials used and, hence, of lowering the barriers between minima.

To sample the distribution of conformations for this system at room temperature, we apply the technique of simulated annealing. We demonstrate that the distributions of states obtained during the annealing runs are consistently sampled as a function of temperature, and that we are therefore not trapping the peptide into meta-stable high-energy minima. Using this procedure we are able to efficiently compute trajectories for about 1000 different randomly selected starting conformations of dodecaalanine.

From the final 1000 states of the annealing runs we obtain the room temperature distribution of states, from which we extract the thermodynamic parameters that describe the helical propensity of the peptide. It is known experimentally that similar peptides to the one we study undergo transitions from partly helical states to fully random coil states as a function of temperature. We will discuss how statistical mechanical models of helix–coil transitions may be used to extract the nucleation and propagation parameters that describe this transition from our distribution and will compare our results with the experimentally measured values.

2. Theory

2.1. Statistical mechanical models

In order to compare the results of our simulations with experimentally measured thermodynamic parameters that describe the helix to random coil transition, we use a statistical mechanical theory of this transition. Although the most popular model of the helix–coil transition is probably the one developed by Zimm and Bragg [9], we will present here a slightly different formulation, due to Lifson and Roig [10]. The two theories yield virtually identical results and the parameters of one can easily be inter-converted into the other [11].

The basic notion underlying the Lifson–Roig theory of the helix–coil transition is that the dihedral angle space of each amino acid in the polypeptide chain is divided into two regions: helical and coil. Each position along the chain is assigned one of three factors u, v or w, depending whether it is in the coiled state, the helical state with at least one coiled neighbor or the helical state with helical neighbors. A chain of length n may then be found in 2^n possible states, where the probability of occupying each state is given by the ratio of the product of the n factors to the partition function.

The significance of the w parameter is that when three consecutive residues $(i-1, i, i+1)$ are in the helical conformation, the carboxyl group of the peptide unit of the $i-2$ residue hydrogen bonds to the amide group of the $i+2$ peptide unit. Therefore, the power of w in the product of factors for one state of a peptide chain represents the number of hydrogen bonds.

The parameter that we will be concerned with the most here is the nucleation parameter v. It has generally been assumed that this parameter should be rather small. For instance, Baldwin's group usually fits their experimental data with a nucleation parameter $v=0.06$ [11,12]. The value of v is an indication of the cooperativity of the melting in the polymer. A small value of v and a large value of w (usually around 1.3 at room temperature), indicates that while the nucleation of a hydrogen bond is unlikely, once it has occurred the helix is very likely to propagate down the chain.

3. Methods

3.1. Potentials of mean force and constrained Langevin dynamics

The full atom model of polyalanine that we will use in our simulations consists of six atoms per amino acid since the CH and CH_3 groups are treated as united atoms. For simplicity, we have ignored terminal groups for the peptide, so that we have half a peptide bond at either end of the molecule. As explained above, all degrees of freedom, except for two dihedral angles per residue, are frozen in our simulation.

The potentials between the atoms are a modified version of those used in our previous work [7], to which we refer for a detailed discussion of their generation. Here we shall merely outline the procedure we use to compute them.

The bare potentials used in molecular dynamics simulations consist of bonded and non-bonded interactions. Since we freeze the bond lengths and angles, we are only concerned with the evaluation of the non-bonded term. In most formulations, the non-bonded term has two components: Lennard–Jones, to describe van der Waals interactions, and Coulomb terms.

In a typical simulation, water molecules are included around the solute molecule. The approximation we will employ here consists of incorporating the effect of water into the pairwise, intra-solute, non-bonded interactions, by computing potentials of mean force that average over water conformations. To compute the mean effect of water on Lennard–Jones interactions, we first calculate solute–solvent correlation functions using Monte Carlo sampling. These functions allow us to estimate the density around two solute molecules, and hence, compute the potential of mean force due to this density. The effect of water on the Coulombic term is approximated by incorporating a dielectric coefficient in this term. In this work we will use the value of 80 for the dielectric coefficient, the value for bulk water, which we have previously shown to yield reasonable values for the free-energy landscape of the alanine dipeptide.

These potentials are then incorporated into a novel technique that we have developed to integrate the equations of motion [8]. In typical molecular dynamics simulations, one must use time steps of 1 fs to obtain accurate estimates of the dynamics of a solute. In our approximation we ignore the fast vibrations that correspond to bond length and angle vibrations, and hence effectively generate longer timesteps. In the previous work we estimated, based on Stokes law of particles in bulk water, that these timesteps are of the order of $\sim 10^2$ fs. The fast vibrations are eliminated by introducing constraints into the dynamics of the solutes that keep the bond lengths and angles at their ideal values. In practice, this is accomplished by subtracting the component of the force on each atom that leads to bond angle and length perturbations, from the total force. The overdamped Langevin equations are then used to evaluate the dynamics of the constrained solute.

Since we have frozen all the bond angle and bond length degrees of freedom apart from the dihedral ones, if we were to rotate about the dihedral angles of a peptide we would find that we have introduced very large barriers between the minima. When we performed this exercise with the alanine dipeptide molecule, where the barrier heights have also been computed for the fully flexible model, we found that constraining all non-dihedral degrees of freedom produces barriers of the order of magnitude greater than in the unconstrained case. To correct this problem we must introduce some smoothing of the potentials that produce a softer repulsive component.

The smoothing technique we have adopted in this work lies in treating all atoms as Gaussian distributions rather than point-like particles. Thus, the potential that a particle 'feels' is the Gaussian average of the inter-atomic potential, where the Gaussian is centered at the particle's position and has a scale length of 0.3 Å (see Fig. 1).

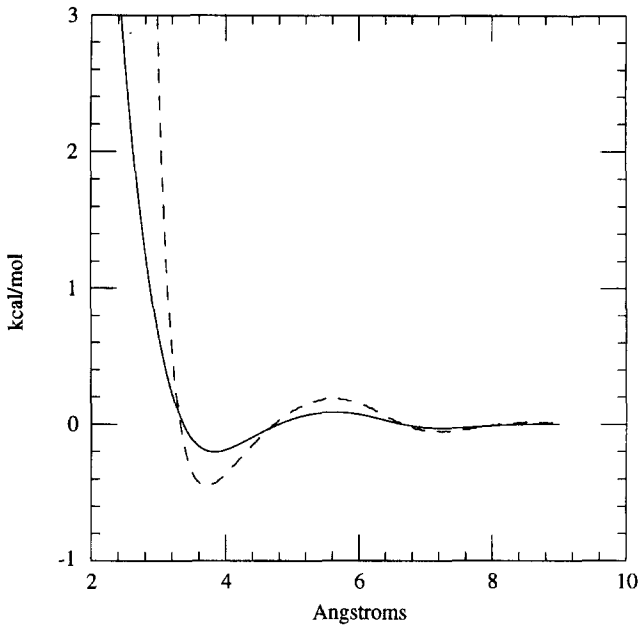

Fig. 1. The shape of the potential smoothed by convoluting it with a Gaussian distribution (solid line), superimposed on the original potential generated as described in our previous work (dashed line). In this case we are plotting the potential between two carbonyl carbons.

As explained in our previous work, we begin with the room temperature intra-solute two-body correlation function and perform a convolution with a Gaussian function. We then take the logarithm of the smoothed correlation function to obtain the intra-solute potential. We tested this approach on the alanine dipeptide and found that the barriers were now within a factor of two of those calculated with the unconstrained model.

3.2. Simulated annealing

To generate an equilibrium distribution of dodeca-alanine states we use the technique of simulated annealing [13]. That is, we ran approximately 1000 simulations, where we cooled the temperature from ten times the room-temperature to room temperature in 200 000 timesteps. We decreased the temperature linearly in T^{-1}. Each run takes approximately 30 min on a workstation. The room-temperature equilibrium distribution consists of the final state of the 1000 runs.

It is common during annealing procedures to trap the distribution of states in a local minima at higher temperatures. This may occur if the annealing is performed too rapidly or if the barriers between minima are too high, rendering difficult a thorough sampling of the phase space. To verify that we, in fact, obtained a room temperature and not higher-temperature distribution, we compared the density of states at each temperature. The temperature-dependent probability distribution of states is the product

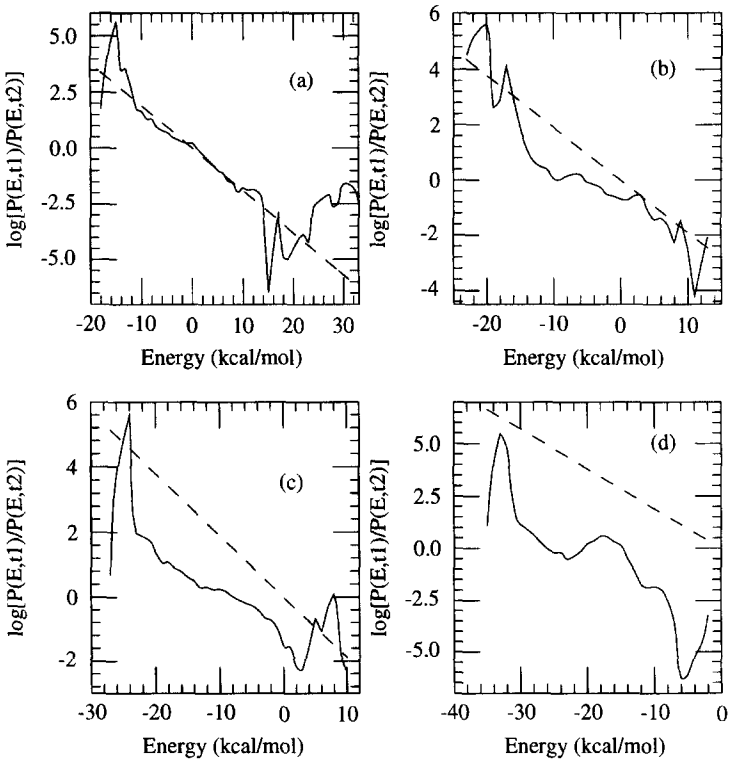

Fig. 2. As described in the text, we are plotting in solid lines the log of the ratio of the energy-probability distributions at different temperatures. The ratios are between the following temperature distributions: (a) 746 and 1042 K, (b) 476 and 588 K, (c) 403 and 476 K, (d) 280 and 310 K. The dashed lines represent the ratios that one obtains in the case of perfect Boltzman distributions. The slopes of our distributions are close to that of the ideal case in the middle, well-sampled, energy ranges. At very low and very high energies, the statistics are not large enough to obtain meaningful values of the ratios. Furthermore, due to our inability to correctly normalize the distributions, we observe a small offset between our ratios and the ideal ones. Nevertheless, the slopes of the ratios indicate that unlike the case of the un-smoothed potentials, we are not freezing the distributions at high temperatures.

of the Boltzman factor and the temperature-independent density of states:

$$P(E,T) = \exp\left(-\frac{E}{k_B T}\right)\Omega(E)\,. \tag{1}$$

If we, therefore, take the log of the ratio of the probability distributions,

$$\ln\frac{P(E,T_1)}{P(E,T_0)} = -E\left(\frac{1}{k_B T_1} - \frac{1}{k_B T_0}\right), \tag{2}$$

we expect to find a function that is linear in energy. This is, in fact, what we see for all distributions up to room temperature (see Fig. 2). We can, therefore, be fairly certain that we are not trapping the peptide into high-energy meta-stable states, but are obtaining a room-temperature equilibrium distribution.

4. Discussion

As discussed above, we produce a room temperature distribution of states for a full atom model of a 12 long polyalanine peptide. Along with the incorporation of the effect of water into a pairwise potential of mean force between any two atoms, we freeze all degrees of freedom of the peptide other than the principal dihedral angles, ϕ and ψ, for each residue. To compensate for the freezing of bond angle and length degrees of freedom, we treat each atom as a Gaussian-density distribution rather than a point-like particle. This allows us to lower the barriers between the minima in conformational space. We conduct 1000 simulated annealing runs, starting from a high temperature and slowly cooling down to room temperature. The distribution consists of the ensemble of the final structure that we collect at the end of each run.

The first thermodynamic parameter that we extract from this distribution is the fractional helix content of the peptide. We measure this by computing the number of helical hydrogen bonds in our structures. These bonds are defined as those between the i and $i+4$ residues, as in the standard alpha helix, or between the i and $i+3$ residues, as in the 3_{10} helix. For the formation of bond we use the following definition: the amide hydrogen and the carbonyl oxygen of the i and $i+3$, or the i and $i+4$ residues, must be within 3 Å of each other and the residues between these must lie within the alpha-helical region, defined by $\phi = 65° \pm 40°$ and $\psi = 50° \pm 45°$.

It has recently been argued that from the evidence of Fourier-transform infrared spectra and electron-spin resonance spectra that the more common state of a 17 long alanine-based peptide is, in fact, the 3_{10} helix and not the more common alpha helix [14]. We find that our helices form twice as many 3_{10} hydrogen bonds as i to $i+4$ ones. The total number of hydrogen bonds formed corresponds to a total helicity at room temperature of 16%. This value is in good agreement with that of 18%, reported by the Baldwin group [11] for a 14 long alanine-based peptide at room temperature, as measured by circular dichroism. By comparison, Okamoto and Hansmann [2], who do not include hydration effects, find helical contents of 80% at room temperature. Therefore, inclusion of hydration in the simulation is essential to recover the experimental values of helical content.

We next attempt to compute the thermodynamic parameters that describe the helix to coil transition of these peptides. This Lifson–Roig theory has been used by the Baldwin group to fit the melting curves of polyalanine-like peptides [11]. To solubilize the peptides they insert a lysine residue every fifth position. They measure the circular dichroism spectrum of solutions with peptides ranging from lengths of 14 to 100, as a function of temperature. The free parameters of the theory are the nucleation (v) and propagation (w) factors. Using the values of $v = 0.06$ and $w = 1.4$ at 0°C, the melting curves of the Lifson–Roig model are found to resemble closely the experimental data for all chain lengths.

The most direct procedure to measure these parameters with our simulations is to compute equilibrium distributions of the peptides as a function of temperature, thus computing the melting curve of the peptide. However, the potentials of mean force

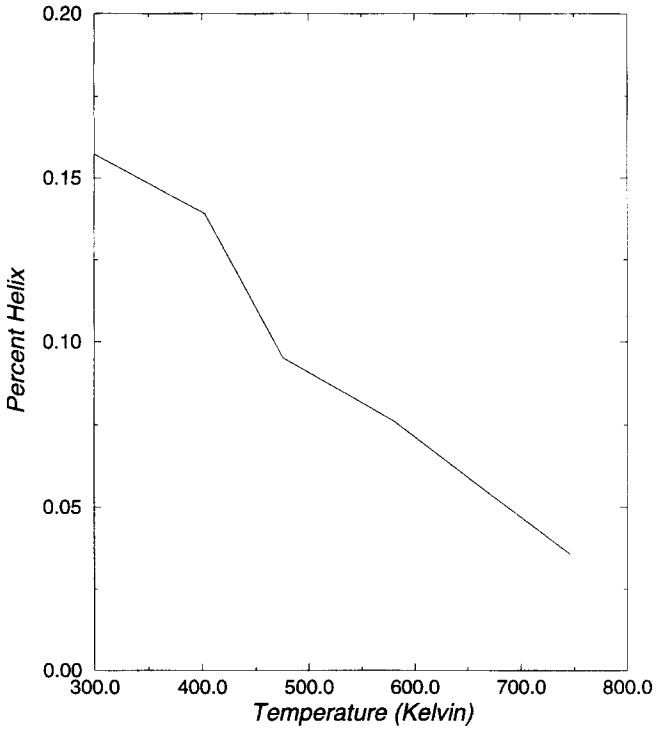

Fig. 3. The melting curve of the polyalanine peptide. We ran five sets of 1000 simulations in which we annealed the peptide from a starting temperature of 3000 K to the final temperatures of 300, 403, 476, 581 and 746 K. For each set of simulations we calculated the average fractional helical hydrogen-bond content from the 1000 final conformations.

that we use are calculated from room-temperature water distributions, and we would, therefore, not expect these to yield accurate distributions at higher temperatures. In fact, we see from Fig. 3 that the melting curve computed with these potentials is much too broad, and that there is significant helical content in our peptides at high temperatures where the experiments do not detect any.

Therefore, to extract the thermodynamic parameters that describe the melting curve, we fit the values of v and w so that the distribution of the number of hydrogen bonds per peptide predicted by the Lifson–Roig theory matched with that obtained from our simulation (see Fig. 4). This comparison is not completely satisfactory, since the Lifson–Roig theory assumes only i to $i+4$ hydrogen bonds, while both our simulation and experimental evidence suggest that the i to $i+3$ ones are more prevalent. However, since the experiments were analyzed with this theory, in order for us to produce parameters that may then be compared to the experimental ones, we are not free to modify it. Therefore, for the purposes of this comparison, we treat all hydrogen bonds as if they were of the i to $i+4$ type. Ideally, both the experiments and the simulations should be reanalyzed with a statistical mechanical model that includes the possibility of both types of hydrogen bond formation.

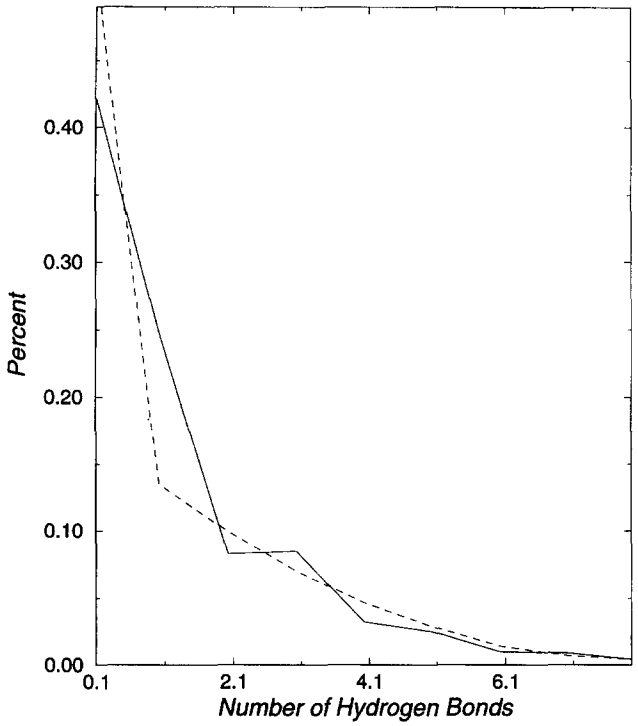

Fig. 4. A comparison of the hydrogen-bond distribution between our simulation (solid line) and that fitted to this using the Lifson–Roig statistical mechanical theory with the parameters $v = 0.3$ and $w = 1.2$ at 0°C (dashed line). The hydrogen-bond distribution from the simulation was computed from the final room-temperature conformation of our 1000 annealing runs. The distribution calculated from the Lifson–Roig theory was then fit to this using the above parameters.

We find that the best fit of the hydrogen bond distribution of the Lifson–Roig theory to our simulations is obtained with the values of $v = 0.3$ and $w = 1.2$ at 0°C. To obtain the value of w at room temperature we use the Arrhenius relationship:

$$\ln w = -\frac{\Delta H}{k_B T} + \frac{\Delta S}{k_B} . \tag{3}$$

These values can be converted to the parameters used in the Zimm–Bragg theory using the formulas [11]

$$\sigma = \frac{v^2}{1 + v^4} \tag{4}$$

and

$$s = \frac{w}{1 + v} , \tag{5}$$

and yield $\sigma = 0.03$ and $s = 0.92$. Therefore, we find that the nucleation value extracted from our simulation is larger than that measured by the Baldwin group, and the propa-

gation parameter is smaller. In other words, our system exhibits less cooperative behavior than the real peptide. However, recent measurements by other groups have found larger values for the nucleation parameter, $\sigma = 0.01$ [15], than that reported by Baldwin, suggesting that in fact the transition may be less cooperative than previously thought.

5. Conclusions

We have presented a formalism for simplifying atomic resolution molecular dynamics simulations in order to render the calculation of thermodynamic parameters computationally feasible. The simplifications in the potentials involve the replacement of explicit water molecules with potentials of mean force that represent some of the statistically averaged water properties. The dynamics have also been rendered more efficient by removing the fast degrees of motion from the trajectories by constraining the bond lengths and angles. Finally, we have compensated for the loss of bond flexibility by convoluting the potentials with a Gaussian-density distribution centered on each atom.

This procedure allows us to effectively sample the room temperature distribution of a 12 residue long polyalanine peptide. From this distribution we were able to extract the average helical content of the peptides, approximately 18%, which is in close agreement with experimentally measured values. We also find that by a ratio of 2 : 1, the helices are of the 3_{10} type rather than the α-helical type, a fact that is also supported by experimental measurements.

Finally, we demonstrated that by fitting the distribution of the number of hydrogen bonds per peptide predicted by the Lifson–Roig theory to our simulated distribution, we could extract the thermodynamic parameters that describe the helix to coil transition. In this case we found that the transition deduced from our simulations is less cooperative than that measured experimentally. In other words, the nucleation factor we measured, $v = 0.3$, was larger than the experimentally measured values of 0.06 to 0.14. Restating these results using the parameters of the Zimm–Bragg theory, we obtain $\sigma = 0.03$ instead of the experimentally measured values of 0.01 to 0.003.

We believe that the less cooperative nature of our simulated peptide may be partly due to the fact that we use two-body potentials of mean force, neglecting higher-order many-body contributions. Therefore, we are currently attempting to expand the accuracy of our method by including higher-order terms in the potentials. These potentials, together with the constrained Langevin dynamics algorithm, are beginning to provide a methodology to extract thermodynamic parameters from atomic scale simulations, a capability that is essential to test and improve molecular dynamics techniques.

Acknowledgements

We would like to thank Jim Hofrichter for useful conversations and permission to quote his unpublished data, and the Advanced Computing Laboratory (Los Alamos)

for access to their computing facilities. Parts of this work were conducted under the auspices of the US Department of Energy, supported (in part) by funds provided by the University of California for the conduct of discretionary research by Los Alamos National Laboratory. This work was also supported by the National Science Foundation.

References

[1] V. Daggett and M. Levitt, J. Mol. Biol. 223 (1992) 1121.
[2] Y. Okamoto and H.E. Hansmann, J. Phys. Chem. 99 (1995) 11276.
[3] L. Wang, T. O'Connell, A. Tropsha and J. Hermans, Proc. Nat. Acad. Sci. USA 92 (1995) 10924.
[4] A.S. Yangand and B. Honig, J. Mol. Biol. 252 (1995) 351.
[5] S-S Sung, J. Theor. Biol. 173 (1995) 389.
[6] G. Hummer, A.E. Garcia and D.M. Soumpasis, Biophys. J. 68 (1995) 1639.
[7] M. Pellegrini, N. Grønbech-Jensen and S. Doniach, J. Chem. Phys. 104 (1996) 8639.
[8] N. Grønbech-Jensen and S. Doniach, J. Comput. Chem. 15 (1994) 997.
[9] B.H. Zimm and J.K. Bragg, J. Chem. Phys. 31 (1959) 526.
[10] S. Lifson and A. Roig, J. Chem. Phys. 34 (1961) 1963.
[11] J.M. Scholtz, H. Qian, E.J. York, J.M. Stewart and R.L. Baldwin, Biopolymers 31 (1991) 1463.
[12] C.A. Rohl and R.L. Baldwin, Biochemistry 33 (1994) 7760.
[13] S. Kirkpatrick, C.D. Gelatt and M.P. Vecchi, Science 220 (1983) 671.
[14] S.M. Miick, G.V. Martinez, W.R. Fiori, A.P. Todd and G.L. Millhauser, 359 (1992) 653.
[15] P.A. Thompson, W.A. Eaton and J. Hofrichter, personal communication.