Advanced Series in Physical Chemistry – Vol. 7

RECENT DEVELOPMENTS IN THEORETICAL STUDIES OF PROTEINS

Editor

Ron Elber

Department of Physical Chemistry
Hebrew University
Israel

World Scientific
Singapore • New Jersey • London • Hong Kong

CONTENTS

CHAPTER 3

OPTIMIZATION TECHNIQUES WITH APPLICATIONS TO PROTEINS

JOHN E. STRAUB

Department of Chemistry, Boston University
Boston, Massachusetts 02215, USA

Contents

1. The Rugged Energy Landscape of Proteins

Proteins are complex macromolecules which exhibit a variety of interesting structural and dynamical properties. While they typically consist of thousands of conformational degrees of freedom, proteins fold into well-defined three-dimensional structures which allow them to carry out functions such as reaction catalysis with great precision. Studies on protein folding have demonstrated that the number of conformations available to a protein is far too large to allow for the folding to occur by a random search.[1] Still, proteins fold. Apparently, the sequence of a protein has been designed to fold rapidly into a stable native state reliably. Unraveling exactly how that is done has been recognized as one of the most interesting and important scientific challenges of our time.[2,3]

It is clear that proteins fold into compact low energy states with remarkably harmonious designs. This property has been expressed in the "consistency principle" of Gō,[4] the "principle of minimal frustration" of Bryngelson and Wolynes,[5] and in the studies of the importance of the "stability gap" between the lowest energy state and the manifold of higher energy native-like states[5,6] and the "energy gap" between the lowest energy and first excited conformations of model proteins.[7] If the native state of the protein is assumed to be the global free energy minimum, the problem of identifying the native state is reduced to the problem of finding the global minimum of a complicated many dimensional energy function. This view has been made popular principally through the work of Scheraga who has also contributed a fantastic number of creative approaches to the difficult problem of optimizing protein structure.[8] The difficulty of finding the global minimum is closely related to the character of the multidimensional energy function — the energy landscape — of the protein.

Experiment and computer simulations have had some success in providing the details of the energy landscape of proteins. For example, there

is some fuzziness to the definition of the native state of a protein. The dominant folded states of proteins are often thermally stable, but at times appear to be metastable states.[9] Moreover, even when a protein structure is well-defined, X-ray crystal structures as well as solution NMR studies have demonstrated that there is some variety in the ensemble of structures making-up the sample of the native protein. This disorder, often important to the function of the protein, is in the form of conformational substates — conformations of the protein which are similar in structure and energy but which represent distinct conformations of the protein. The seminal experiments of Frauenfelder and coworkers on the kinetics of ligand rebinding in myoglobin following photodissociation have shown that these substates can be biologically active but may carry out their function with varying ability.[10] Conformational substates have also been identified in the computer simulation study of Elber and Karplus who found that in a 300 ps trajectory of myoglobin the protein underwent many transitions between substates (or local energy minima).[11] Similar studies by Noguti and Gō of a trypsin inhibitor have drawn the similar conclusion that during an equilibrium simulation of less than 1 ns a variety of substates are sampled.[12] They classed the transitions between substates as (1) local conformational changes involving a few residues which lead to elastic deformations of the protein and (2) cooperative rearrangements which induce plastic deformations in the whole molecule.

Simulation studies of Straub and Thirumalai involving ribonuclease A have demonstrated the existence of conformational substates which are sampled in a dynamics trajectory of less than 100 ps at room temperature.[13] It was determined that these substates are separated by barriers of a few kcal/mol. Their studies have demonstrated the existence of a broad distribution of energy barriers ranging from a tenth to tens of kcal/mol.[13] It is likely that the distribution of barrier heights is intimately connected to the broad distribution of relaxation times found in proteins. More insights into the gross features of the energy landscape have been provided by studies of protein folding based on simple "minimal" models of the protein. Camacho and Thirumalai have studied the thermodynamic and kinetic properties of model heteropolymers.[14] They found that there exists a relatively small number of compact states of low energy and that these states are separated by significant energy barriers. Detailed analysis of the density of states has been carried out by Sali, Shakhnovich and Karplus for lattice Monte Carlo simulations of a simple model protein in three dimensions.[15]

These studies demonstrate that the thermodynamics and kinetics of proteins are best described in terms of the details of the potential energy hypersurface — the energy landscape.[16] The picture of the energy landscape of a protein which is consistent with the experimental and simulation evidence is a rugged one with a large number of minima and a distribution of energy scales. The general features of the landscape are dominated by a relatively small number of low lying energy minima. The lowest energy state in many cases is believed to represent the native state of a protein. This picture is in harmony with the frustration observed in low temperature kinetics experiments as well as the important features of protein folding observed in the simulation of folding of model heteropolymers. Superimposed on this gross structure is a substructure of many local minima. This substructure explains the minor disorder observed in X-ray crystal structures and solution NMR structures of proteins as well as the frequent transitions observed between substates on a picosecond time scale in room temperature computer simulations. Landscapes of this kind have been used to explain many properties of spin glasses and a variety of theories developed for their study have been profitably extended to the study of proteins, most notably by the group of Wolynes,[5] and by Shakhnovich and coworkers.[17]

The topological details of these rugged energy landscapes are currently being defined using computer simulations of minimal models of proteins as well as all atomic models in terms of the distribution of energy minima, the connectivity of the minima, and the distribution of energy barriers separating the minima. In the study of protein folding, more than the distribution of energy minima, it is important to identify the lowest lying energy states of the compact protein which are likely to include folding intermediates and the subset of native protein configurations. Therefore, it is important to develop reliable methods for finding the compact, lowest lying energy states of the protein. From this point of view, an important part of the protein folding problem is a global energy minimization problem — a search for the lowest energy compact states of the system. This chapter will review some of the more promising algorithms for finding low lying energy minima of complex biomolecules.

2. Strategies for Global Energy Minimization

Problems centering on the global optimization of proteins share many characteristics of a large class of "hard" problems known as $\mathcal{NP}$-

complete.[18,19] These problems can be solved by a straightforward, exhaustive search of the set of possible solutions, such as the set of all possible conformations of the protein. Unfortunately, for this class of hard problems, the complexity increases faster than any polynomial in N, where N is a measure of the system size such as the number of atoms. This makes the solution of a problem, such as finding the lowest energy conformation of a large protein and folding the protein *in machina*,[20] a currently intractable problem — one which cannot be solved in a realizable time using an exhaustive search of configuration space.

This is good and bad news. The bad news is that we will have to be very clever to make progress in finding the lowest energy configuration for a protein of even 50 residues. The good news is that because this is a common problem found throughout the physical sciences, there are many techniques being developed which have provided very promising results. In this chapter I try to provide a unified way of viewing what I believe to be the most promising approaches to the global optimization problem for proteins.

All problems of global optimization may be reduced to the same problem — how best to find the extremum of a "cost function" which is a measure of how well the problem has been solved. In the best known optimization problem, that of the "traveling salesman," the goal is to find the shortest path which connects a set of cities that the salesman must visit.[21,22] For a small number of cities, the number of paths is small and the solution is easily found using an exhaustive search. However, as the number of cities increases, the number of possible paths rapidly exceeds the number of paths one can consider, and an exhaustive search is rendered impotent. In the protein folding problem, one might assume that the native folded state of the protein is the thermodynamically most probable configuration. The cost function is the potential energy (or more accurately and ambitiously the free energy) of the protein. To find the folded configuration of the protein one must find the global minimum of the cost function — the potential or free energy of the protein.[23] Other examples include the problem of identifying the ground state configuration of a spin glass or very large-scale integrated circuits,[24] the optimum reaction coordinate in a many dimensional chemical reaction,[25–27] or the optimum configuration of a molecule subject to distance constraints provided by X-ray diffraction or NMR experiments.[28–31] Having abandoned the exhaustive search, we must

find more intelligent strategies which can simplify the problem without loss of essential details.

2.1. *Simulated Annealing (SA)*

A breakthrough was made in 1983 by Kirkpatrick, Gellat and Vecchi[24] who proposed what is now known as the Simulated Annealing (SA) algorithm. This method is based on an analogy with statistical mechanics where the cost function may be thought of as a rugged potential energy surface. The temperature of the system is at first very high, and barriers in the system are easily crossed. The temperature is then slowly reduced according to a "cooling schedule" until the temperature is small or zero.[18] The slow cooling of the system has the effect of annealing the system, reducing strains by finding lower and lower energy basins on the vast potential energy hypersurface. While the "optimal" cooling schedule[32] cannot be realized in most problems of interest, many realistic cooling schedules have provided fabulous results on a wide variety of problems ranging from the design of circuits to the global energy minimization of molecular clusters. However, we are interested in proteins, and the results of a number of careful studies indicate that for proteins SA is an ineffective strategy for finding the lowest energy configuration.[33,34] I believe this is due to the wide array of energy scales in the biopolymers which tends to frustrate SA by making the optimal cooling schedule particularly difficult to realize. This point is discussed in detail in Sec. 4.

2.2. *Potential Smoothing and Coarse Graining*

Hoare and McInnes[35] noted that softer potentials tend to favor a regular crystal configuration (as the global minimum) while shorter range potentials tend to encourage amorphous structures. In an elegant analysis of the nonlinear optimization problem, Stillinger and Weber[36] recognized that while the number of local energy minima is a strongly (exponentially[37]) increasing function of the system size, how fast the absolute number of minima increases has a great deal to do with the length scale of the interparticle interactions.[35] Systems with short range potentials dominated by nearest neighbor interactions have large numbers of local minima. When the range of interaction is increased, the number of minima can be drastically reduced. Several detailed studies have shown how dramatic this effect can be.[38,39] The program of smoothing the potential energy hypersurface to

remove high lying local minima, and to deepen and broaden the global energy minimum, has been was referred to as the "antlion strategy."[36,38] Berry has written a fine review of the related work done examining the character of the rugged energy landscapes of atomic and molecular clusters.[40]

The structure of this chapter rests on the foundation of these two important advances. I hope to convince you that the most promising methods for global optimization of proteins can be understood as variations on the simulated annealing and antlion paradigms. The next section is a selective review of some useful "tailored" optimization strategies which have been developed for specific applications. The most interesting of these algorithms are at heart potential smoothing algorithms. The rest of the chapter will concentrate on "general" algorithms for global optimization which may be

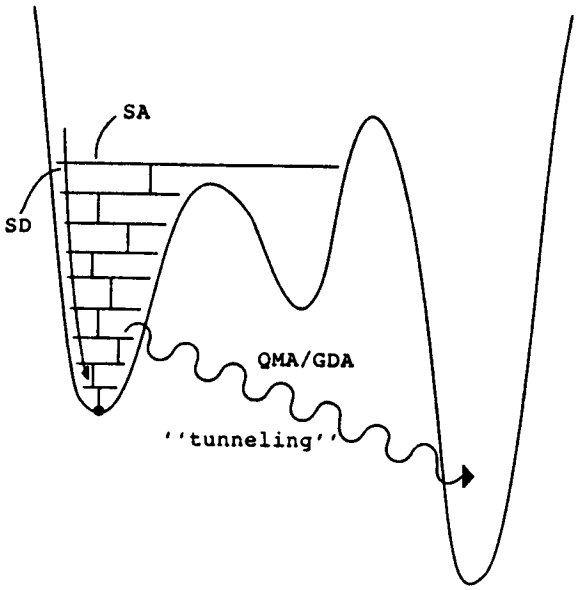

Fig. 1. A schematic of the paths taken by three optimization algorithms on a rugged potential energy surface. The local steepest descent minimizer (SD) locates the nearest local minimum. The simulated annealing (SA) algorithm searches a larger region of configuration space, but with rapid (more practical) cooling schedules, one may fail to clear important barriers and become trapped in a local minimum. The quantum mechanical annealing (QMA) method, as well as the GPP and GDA algorithms, has the advantage of working with a coarse grained potential energy surface where the wave packets can tunnel through potential barriers.

applied to any molecular structural optimization problem and to a variety of problems outside of molecular energy minimization. The third section describes simulated annealing in time using an approximate classical dynamics which results from an attempt to anneal simultaneously an ensemble of systems rather than a single point on the vast configuration space. In the third section, I discuss an unconventional annealing strategy (Gaussian Density Annealing, or GDA) where the system is followed in temperature rather than time, avoiding the problematic cooling schedule. Alongside this method, I mention two related techniques which are in the same spirit. The fourth section covers a variety of methods for performing "quantum mechanical annealing" (QMA) where the annealing process can benefit from tunneling (which dominates the system relaxation at low temperatures) as well as activated barrier crossing (effective only at high temperatures). A graphical overview of these techniques is provided in Fig. 1.

Several excellent overviews of local optimization techniques exist which are complimentary to the methods discussed in this chapter.[41-44] While many of the algorithms discussed here have been tested on proteins, some have not. The global energy minimization of clusters of Lennard-Jones atoms has emerged as a standard test case for algorithms designed for molecular optimization problems and many results presented here deal with this useful system with known results.[40,45-47]

3. Tailored Algorithms

In a broad sense, tailored algorithms refer to optimization methods which are designed *ad hoc* for a specific problem. In the cases I will discuss, what is tailored about an algorithm is that it only applies to potential functions of a certain symmetry or functional form. For example, an algorithm which exploits the functional form of a centrosymmetric potential is not easily generalized to treat covalently bonded systems, and vice versa. However, as will be clear, many of these algorithms, while tailored to specific problems, may be applied to the broad class of problems involving molecular energy minimization.[48] Moreover, they are recognizable as specific forms of the more general class of potential smoothing algorithms.

3.1. *Potential Shift and Bond Scaling*

Piela and coworkers have presented a method for global energy minimization designed for centrosymmetric potentials.[49] The method is based on a

potential deformation where the range of the attractive interaction is effectively increased relative to the length scale of the repulsive (excluded volume) interaction. In an application to atomic clusters, the Lennard-Jones potential was transformed as[49]

$$V_{LJ}(r_{ij}) \rightarrow V_{LJ}(r_{ij} + ar_{\min}), \tag{1}$$

where $0 \leq a \leq 1$ and $r_{\min}$ is the minimum of the initial potential energy function $r_{\min}$ (see Fig. 2). When $a = 1$ the potential is transformed such that the potential minimum is moved to the origin $r_{ij} = 0$ and the portion of the potential with a repulsive force is eliminated — the potential is purely attractive. As a decreases to zero the physical potential is restored.

In practice, the algorithm is implemented by (1) finding the minimum of the scaled potential with $a = 1$ (in the case of atomic clusters all atoms are

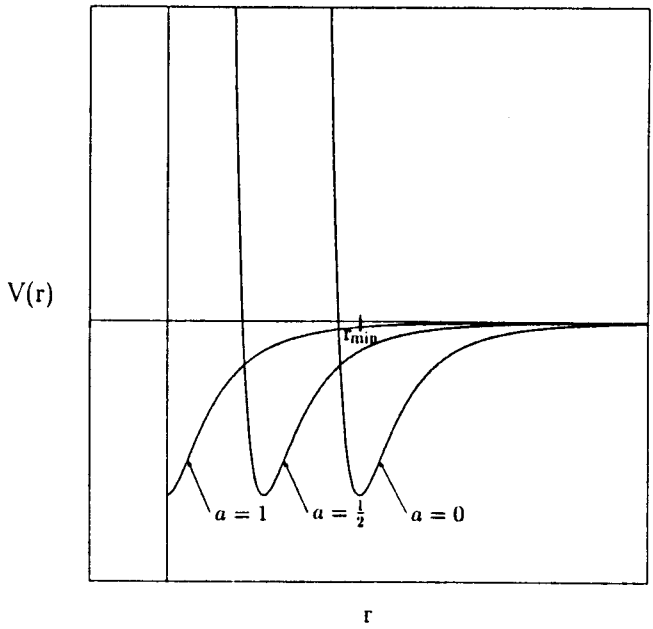

Fig. 2. A schematic of the potential energy surface for a Lennard-Jones interaction $V(r)$. The physical potential energy has a minimum at $r_{\min}$. When the potential is transformed through the shift of $r \rightarrow r + ar_{\min}$ the potential minimum moves towards the origin, increasing the relative importance of the attractive interaction.

superimposed at the origin), (2) changing $a \rightarrow a - \delta$ followed by (3) energy minimization to track the minimum on the new surface, and (4) a return to step (2) until the energy minimum is mapped back to the physical potential where $a = 0$. This method has been successful in finding the global energy minimum for a series of Lennard-Jones clusters ranging from 5 to 19 atoms but appears to be quite CPU intensive (reported to take 100 hours on an IBM 386 for the 19 atom cluster).

Loop closure is a fundamental problem in protein modeling.[50] DeLisi and coworkers[51] have introduced an effective method for loop closure which has been applied to model loop segments (of seven residues), including α-helical, β-strand and surface loops, in a variety of systems[52] as well as to bound peptides in class I major histocompatibility complex receptors.[53] The algorithms are based on a simple scaling of the bond lengths and van der Waals interactions coupled with energy minimization. Suppose the goal is to model a peptide segment into a site with a known end-to-end distance d. (1) The peptide segment is extended using random values of the backbone dihedrals leading to an initial end-to-end length of d_0. (2) The bond lengths are then uniformly scaled by adjusting the equilibrium distance as

$$r_0 \rightarrow r_0(d/d_0)^{(N-i)/N} \tag{2}$$

such that the end-to-end distance equals the known value (see Fig. 3). (3) The energy of the peptide is then minimized (if possible in the external field of the protein) where the repulsive van der Waals interactions are also scaled by a factor of $(d/d_0)^{2(N-i)/N}$. (4) Returning to step (2) the cycle is repeated N times until the peptide bonds are scaled back to their equilibrium lengths and the segment is energy minimized (hopefully) globally. This procedure should be repeated with a variety of initial conditions leading to a final ensemble of possible configurations. While this method has been employed with a local minimizer in step (3) it might be improved by substituting a more effective energy minimization method such as a form of simulated annealing discussed in the next section.

The scaling of the peptide bond is quite similar to the bond shift algorithms of Piela and coworkers.[49] Note that a primitive version of the bond shift algorithm of Eq. (1) is

$$V(r_{ij}) \rightarrow V(r_{ij} + ar_0)\,, \tag{3}$$

where $1 - d/d_0 \leq a \leq 1$ and r_0 is the minimum of the bond stretching potential energy function. When $a = 1 - d/d_0$ the potential is transformed

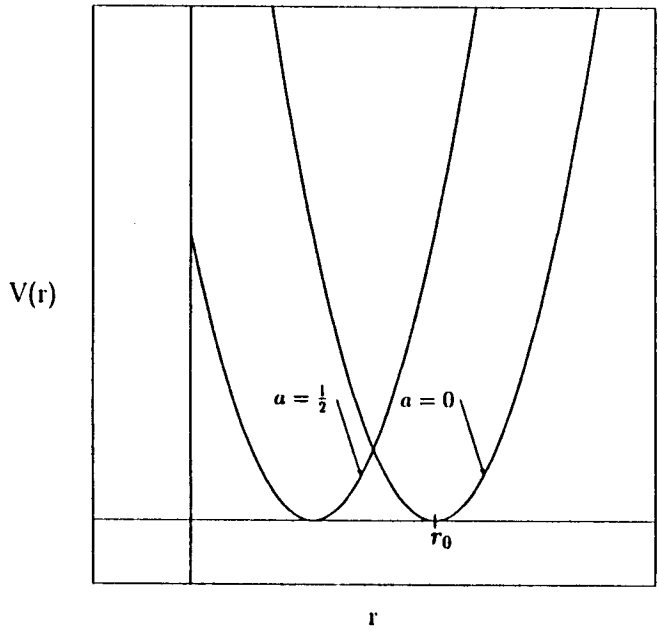

Fig. 3. A schematic of the potential energy surface for a bond stretching interaction $V(r)$. The physical potential energy has a minimum at r_0. When the potential is transformed through the shift of $r \rightarrow r + ar_0$ the potential minimum moves towards the origin, increasing the relative importance of the attractive interaction.

such that the potential minimum is moved toward the origin $r_{ij} = 0$ and the portion of the potential with a repulsive force is reduced (when $a = 1$ the potential is purely attractive). As a decreases to zero the physical potential is restored. A combination of the primitive bond shift algorithm with the nonbonded shift of Piela and coworkers discussed here is a variation of the successful loop closure algorithm of DeLisi and coworkers worth investigation.

3.2. *Sigma and Torsion Potential Fluctuation*

An inventive method for enhanced conformational space sampling has been presented by Liu and Berne.[54] This method rests on the very general idea of allowing constants which determine the length and energy scales of the

interaction potential to fluctuate during a simulation according to deterministic (or stochastic) equations of motion. Their method is in the spirit of the extended dynamical methods of Andersen,[55] and Nosé,[56] and Hoover[57] for applying external constraints on temperature or pressure. While their applications involve Monte Carlo studies of a binary Lennard-Jones fluid and chain folding in normal alkanes, they provide a clear prescription for applications involving molecular dynamics which may be more suitable for biomolecular systems.[54]

For example, consider a Lennard-Jones fluid of particles with physical diameter σ. We allow the particle diameters to fluctuate between the values of $\sigma_{\min}$ and σ_0. Writing the Lennard-Jones diameters σ as variables which depend on a parameter w we have

$$\sigma(w) = \sigma_0 + (\sigma_{\min} - \sigma_0)s(w)\,, \tag{4}$$

where $s(w)$ is a sigmoidal function which varies between 0 and 1 switching σ between σ_0 and $\sigma_{\min}$. By defining a "kinetic energy" for the parameters w_k of the kth atom with fictitious mass m_k the generalized Hamiltonian energy function becomes[54]

$$\mathcal{H} = \frac{1}{2}\sum_{k=1}^{N} M_k \dot{r}_k^2 + \frac{1}{2}\sum_{k=1}^{N} m_k \dot{w}_k^2 + V(\mathbf{r}, \mathbf{w}) + F(\mathbf{w}) \tag{5}$$

with corresponding equations of motion

$$M_k \ddot{r}_k = -\nabla_{r_k} V(\mathbf{r}, \mathbf{w}) \tag{6}$$

$$m_k \ddot{w}_k = -\nabla_{w_k} V(\mathbf{r}, \mathbf{w}) - \nabla_{w_k} F(\mathbf{w})\,. \tag{7}$$

$F(\mathbf{w})$ is a cost function which constrains the fluctuations of the variables $\mathbf{w}$ during an otherwise standard molecular dynamics simulation. The first set of equations is for molecular dynamics on a potential which depends on the fluctuating parameter $\mathbf{w}$. Fluctuations in $\mathbf{w}$ will deform the potential and occasionally lower barriers allowing infrequent events to occur with greater frequency. This leads to enhanced sampling of the configuration space (see Fig. 4).

Similarly, in a polymer the torsion angle potential can be written[54]

$$V(\phi, w) = V_0(\phi) + [V_m(\phi) - V_0(\phi)]s(w)\,, \tag{8}$$

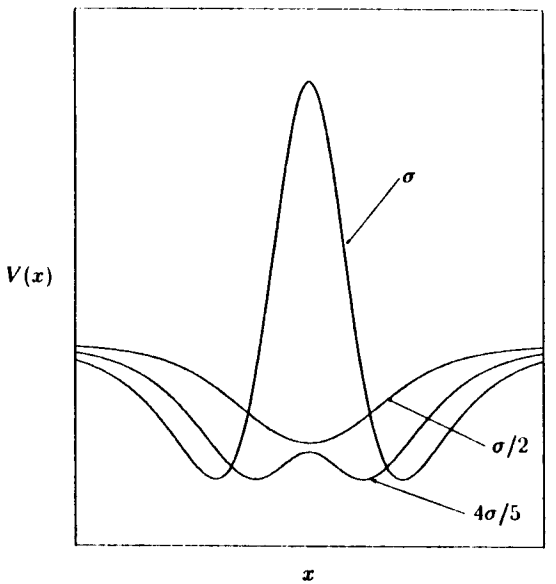

Fig. 4. The potential energy for the interaction of a single atom passing along the x-axis between two atoms fixed at $(x = 0, y = \pm\Delta)$. When the diameter σ of the interior atom is allowed to fluctuate to smaller values, the potential surface for moving between the two particles changes from a separated "double well" with a high barrier, to a double well with significantly lower potential barrier, to a minimum with no barrier. This mechanism allows for enhanced barrier crossing and configurational sampling in the sigma fluctuation algorithm.

where $V_0(\phi)$ is the physical potential and $V_m(\phi)$ is a potential with lower barriers than $V_0(\phi)$. The fluctuations of the Lennard-Jones atomic diameters and torsion angle potential have been shown to lead to greatly enhanced sampling in Monte Carlo simulations of n-pentane and C_{50} H_{102}.[54] For the Lennard-Jones simulations, the effect of the sigma fluctuations is to soften the interactions allowing for closer contacts and smaller amplitude oscillations in the structure (radial distribution function) indicating a lowering in the barriers of the potential of mean force which leads to enhanced sampling. The qualitative trends are similar to those seen in the approximate dynamical annealing methods discussed in the next section.[58] The sigma fluctuation method shares many of the features of the potential smoothing algorithms featured in this chapter. Moreover, it is also possible to compute averages in the ensemble of the physical potential (fixed σ) using

the enhanced sampling trajectories (fluctuating σ) via umbrella sampling techniques.[54] Similar results are expected in simulations of proteins where the use of molecular dynamics, rather than Monte Carlo, may be more attractive. This enhanced conformational space sampling should lead to more rapid folding as well as information on the distribution of low energy conformations of the molecule.

3.3. *The Antlion Method applied to Proteins*

A slightly different approach which combines potential smoothing with a novel implementation of secondary structure prediction information has been developed by Stillinger and coworkers in the form of the "antlion method." This method has been applied to a blocked alanine dipeptide and tetrapeptide[59] as well as the α-helical protein melittin.[60] In each case the potential surface was deformed by applying constraints on the (ϕ, ψ) angles which bias the configuration towards a given local secondary structure. In the case of melittin, the deformation included (1) a bias towards the *trans* peptide backbone configuration, (2) a bias towards the L conformer of each amino acid using an improper torsion potential around the C_α, (3) a penalty function to encourage a set of ideal (ϕ_0, ψ_0) angles of the form

$$V_{\phi,\psi} = k_\phi[1 - \cos(\phi - \phi_0)] + k_\psi[1 - \cos(\psi - \psi_0)], \tag{9}$$

where the values of the force constants k_ϕ and k_ψ are determined by a neural network trained on a learning set, (4) a penalty function in the form of a Coulombic potential which encourages hydrogen bonding between the backbone carbonyl O of residue i and the backbone amide H of residue $i+4$ where the charges were determined using a neural network, and (5) setting to zero the side chain atomic charges.

Following the determination of parameters for the penalty functions (to encourage secondary structure formation) the conformation of melittin was minimized from an initial extended conformation on the smoothed potential surface. (For the application to melittin, the ideal (ϕ_0, ψ_0) angles were those of an α-helix $(-57°, -47°)$.) This was followed by minimization on the physical potential surface until convergence. This method succeeded in finding a structure within an overall 2.45 Å root-mean-square difference from the minimized crystal structure.[60]

Some protein optimization algorithms, such as that of Friesner and coworkers which was applied to myoglobin,[61] rely on an initial step of secondary structure prediction. If it is possible to identify regions of secondary

structure, extended regions of protein substructure can be fixed, thereby greatly reducing the number of degrees of freedom which must be included in the search for the global energy minimum. With regions of secondary structure frozen in, folding is expected to occur as described in the diffusion-collision model of Karplus and Weaver.[62] However, secondary structure prediction methods are less than perfect (usually around 70% accuracy per residue) and any folding optimization algorithm which assumes this knowledge will have difficulties with the accumulating error. The antlion method overcomes this problem by using a penalty function which biases the residue towards a given set of idealized (ϕ, ψ) angles rather than rigidly fixing the backbone (ϕ, ψ) angles of a given residue. However, one must specify the ideal angles (ϕ_0, ψ_0) about which the penalty function will constrain the residue. In the case where the secondary structure is of one kind, such as melittin[60] or even the example of myoglobin,[61] this might succeed. However, in cases where there is a known mix of secondary structure types, the problem of best choosing (ϕ_0, ψ_0) must be addressed. Nevertheless, this tailored algorithm appears to be very promising.

The methods discussed in this section have been tailored to exploit detailed aspects of specific problems. In the remainder of this chapter, I will describe three classes of general algorithms for the optimization of molecular systems which have much in common with the successful tailored methods discussed here.

4. Classical Simulated Annealing in Time

The paradigm of global optimization of complex systems is simulated annealing.[18,24,63,64] This method makes use of a powerful analogy between the statistical mechanical process of annealing and the challenge of global optimization of a complex cost function. In global energy minimization, where our cost function is the potential energy surface, this requires no stretch of the imagination. The system is simulated at an initially high temperature using molecular dynamics or Monte Carlo. The temperature is slowly lowered according to a cooling schedule, annealing the system to a low or zero temperature. The final configuration of the system is the guess at the global energy minimum. If the cooling schedule is chosen properly, we can expect to find the global minimum with a high probability.

The difficulty comes in finding a practical cooling schedule which can guarantee a high probability of finding the global energy minimum.[65] A

typical cooling schedule is defined by decreasing the temperature by a factor $\xi < 1$ from one step to the next so that[18]

$$T_{k+1} = \xi T_k \,. \tag{10}$$

While this is convenient and faster for smaller values of ξ, it is arbitrary. In many complex systems, we know that there are a variety of energy scales on the rugged energy landscape.[10,12,13] It is reasonable to expect that the optimal cooling schedule will be intimately related to the disparity between the largest and smallest energy scales in the system. A simple heuristic analysis in one dimension demonstrates this by estimating the lower bound on the simulation time t_{sim} necessary for finding the global minimum with a reasonable probability.

Taking the global energy minimum to be the zero of energy, E^*_{max} is the highest barrier on the potential surface (see Fig. 5). If we choose the initial

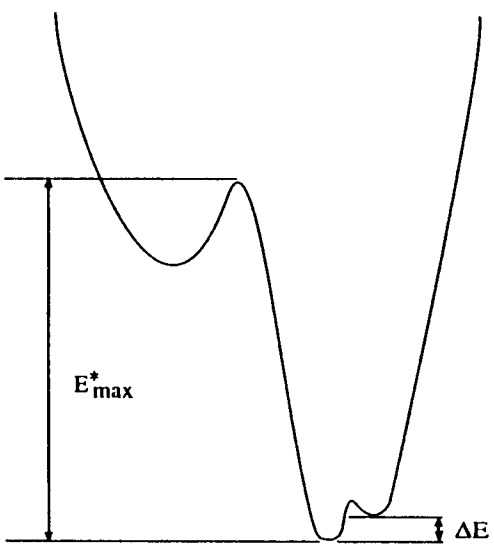

Fig. 5. A schematic of an energy landscape showing the important energy scales which determine the effectiveness of the simulated annealing algorithm. At the high end is the largest energy barrier which must be crossed E^*_{max} for a trajectory to explore all of the configuration space. At the low end is the energy gap ΔE separating the global and next lowest energy minimum. For many problems, the ratio of energy scales $E^*_{\text{max}}/\Delta E$ defines the difficulty of the optimization problem.

maximum value of the temperature such that

$$k_B T_{\max} = 1/\beta_{\min} \simeq E^*_{\max}$$

then

$$e^{\beta_{\min} E^*_{\max}} = \mathcal{O}(1). \tag{11}$$

At this high temperature, the trajectory will be free to cover the entire configuration space unhindered by the energy barriers in the system.

On the other hand, the lowest temperature we need to consider is one where $k_B T_{\min} = 1/\beta_{\max} < \Delta E$ where $\Delta E = E_1 - E_0$ is the "energy gap" between the global energy minimum and the next lowest energy minimum. This inequality should be satisfied so that the mole fraction of being found in the global energy minimum X_0 (compared to the mole fraction for all other minima $1 - X_0$) is large. For many problems, at low enough temperatures there is an effective two level system dominated by the global and next lowest energy minimum, so we simply demand that

$$e^{\beta_{\max} \Delta E} = \frac{X_0}{1 - X_0} \tag{12}$$

be reasonably large (meaning that X_0 is reasonably close to 1).

Ideally, at any point along the annealing trajectory, the system will be at thermal equilibrium. To effectively sample the system at a given β the trajectory should spend a time of at least $t = \tau_s \exp(\beta E^*_{\max})$ required to cross any barrier in the system; τ_s is a characteristic time scale for vibrational motion. For that to be true, the time spent (dt) in a given temperature interval $(\beta, \beta + d\beta)$ should be at least

$$dt = E^*_{\max} \tau_s e^{\beta E^*_{\max}} \, d\beta. \tag{13}$$

τ_s is a short time scale fixed by the highest frequency motion in our system which sets the upper bound on the time step we can take in the integration of our system.

The *lower bound* on the total simulation time t_{sim} can be found by integrating Eq. (13) between the initial $\beta_{\min}$ and final $\beta_{\max}$ leading to a number of simulation steps $N_s = t_{\text{sim}}/\tau_s$ which scales as

$$N_s \sim \left(\frac{X_0}{1 - X_0} \right) e^{E^*_{\max}/\Delta E}. \tag{14}$$

This result can be used to estimate the simulation time required in molecular dynamics to find the global energy minimum with a probability X_0. The exponential dependence of N_s on the ratio of energy scales $E_{max}/\Delta E$ shows that the most difficult optimization problems involve systems with a great disparity in energy scales. It is not possible to explain the difficulty of a given minimization problem solely in terms of a small energy gap separating the lowest and next lowest energy minima, or the presence of high barriers separating regions of configuration space. *It is the ratio in the coarsest and finest energy scales that dictates the simulation time needed.* Similar results hold for more commonly employed cooling schedules.

When the rate of lowering temperature in time is fixed, the simulation time required will be set by the limits of the initial and final temperatures. In the global optimization of biomolecular systems there is a range of energy scales, the ratio $E^*_{max}/\Delta E$ is large, and the necessary simulation time will be significantly longer than for simple systems. A rough estimate of this time, taking $X_0 = 0.01$ and $E^*_{max}/\Delta E = 10$, indicates that such a simulation is out of the question at the present time. In fact, reasonable estimates should be close to the real time required to simulate the dynamical folding of the protein *in machina* without the help of simulated annealing which now can only be performed for model systems.[66-68]

For simulated annealing to be an effective strategy for optimization of proteins, some changes are required. Think of applying the simulated annealing algorithm using an ensemble of systems in parallel. One may then explore a variety of mean field approximations for the approximate integration of an ensemble of systems. I will summarize a few of these techniques in this section. Unlike an *ad hoc* potential transform, these methods are rooted in the approximate integration of the classical Liouville equation which describes the time evolution of the classical density distribution (an ensemble of classical systems) just as Newton's equations employed in classical simulated annealing describe the time evolution of a single trajectory.

4.1. *Classical Dynamical Annealing using the Liouville Equation*

A method for global optimization which is becoming increasingly popular involves the use of straightforward ensemble integration on massively parallel computers. In the simplest application, an ensemble of initial configurations is generated (representing copies of the system) and each copy is simulated independently, for example, using simulated annealing. When

it is determined that a trajectory has reached its lowest energy, or has moved outside of the region of interest in configuration space, the system is removed and replaced by a new copy of the system. In this way an ensemble of systems — a density distribution — is simulated. However, this is computationally demanding and the time over which the system dynamics must be followed is very long (according to the argument given above).

In the development of new methods, it is often desirable to describe the time evolution of an ensemble of systems (to derive the benefits of enhanced sampling) while treating the dynamics approximately (to reduce the computational cost of simulating so many systems in parallel). One place to start is the classical Liouville equation which defines the exact dynamics of the classical density distribution $\rho(r,p,t)$.

In the time evolution of an ensemble of classical systems,[69,70] where each system follows the classical mechanics defined by Newton's equations of motion,[71] the time evolution of the distribution $\rho(r,p,t)$ is described by the Liouville equation[72,73]

$$\frac{\partial}{\partial t}\rho(r,p,t) = -\mathcal{L}_0\rho(r,p,t) \tag{15}$$

where $\mathcal{L}_0$ is the Liouville operator

$$\mathcal{L}_0 = \frac{p}{m}\cdot\frac{\partial}{\partial r} + F(r)\cdot\frac{\partial}{\partial p} \tag{16}$$

and $F(r)$ is the force and m is the mass. $F(r)$, r and p are d-dimensional vectors.

An exact description of the dynamics of $\rho(r,p,t)$ is provided by the equations of motion for the average position and momentum[74]

$$\frac{d\langle r\rangle}{dt} = \frac{\langle p\rangle}{m} \qquad \frac{d\langle p\rangle}{dt} = \langle F\rangle \tag{17}$$

and for the higher-order moments of position and momentum

$$\frac{dM_{n,k}}{dt} = \frac{n}{m}M_{n-1,k+1} + kW_{n,k-1}\,. \tag{18}$$

The moments of the distribution are defined as $M_{n,k} = \langle(r-r_0)^n(p-p_0)^k\rangle$ and $W_{n,k} = \langle(r-r_0)^n(p-p_0)^k(F-F_0)\rangle$ where $r_0 = \langle r\rangle$, $p_0 = \langle p\rangle$, and $F_0 = \langle F(r)\rangle$. The brackets $\langle\ldots\rangle$ imply an average over the density distribution[69]

$$\langle A\rangle = \int dr dp A(r,p)\rho(r,p,t)\,. \tag{19}$$

Integration of this hierarchy of equations provides an exact description of the dynamics of the ensemble. However, in practice this is intractable since for most anharmonic potentials the moments are coupled (up to infinite order moments). Therefore, one must either truncate the moment expansion (which can be numerically unstable) or approximate $\rho(r,p,t)$.

As a first approximation, Ma, Hsu and Straub[74,75] have taken $\rho(r,p,t)$ for a many body system to be a product of single particle distributions represented by d-dimensional spherically symmetric Gaussian phase packets (GPP).[76,77] Each GPP is completely defined by the first and second moments — the packet center (r_0, p_0) and widths $(M_{2,0}, M_{1,1}, M_{0,2})$ in phase space. It is straightforward to derive equations of motion for each Gaussian using Eq. (18).

The appeal of these simple equations of motion comes from their origin as an approximate solution of the Liouville equation for an ensemble of systems, while requiring little more computational investment than is required to integrate Newton's equations for a single representation of the system. The equations of motion are variationally optimized, meaning that the error between this approximate solution and the exact solution has been minimized in a least squares sense. While being approximate, the algorithm leads to significantly enhanced sampling in both applications to equilibrium averaging and global optimization.[74]

In general

$$m\ddot{r}_0 = -\nabla_{r_0}\langle V\rangle \tag{20}$$

which has the form of Newton's equation for the Gaussian center r_0 moving on a coarse-grained effective potential $\langle V\rangle$. When $M_{2,0} = 0$ the equations of motion reduce to the usual equations of classical molecular dynamics. The second moment equations depend on the Laplacian of the effective potential. It has been found[74] that the average value of the second moment in position $\bar{M}_{2,0}$ is related to the generalized Einstein frequency[78,79]

$$\Omega_E^2 = \frac{1}{dm}\langle\nabla^2 V\rangle \sim dk_B T/2m\bar{M}_{2,0}\,. \tag{21}$$

This provides some insights into the dynamics of the approximate Gaussian density distribution. The center of the distribution moves to minimize the force while the widths adjust to the curvature of the effective potential. Moreover, they adjust in a way that instantaneously provides the ensemble of states represented by a fluctuation about the average structure. In

this sense it is similar to an effective harmonic or quasiharmonic model of the protein where the configuration space is represented by a given conformation of the protein (the center of the distribution) and the harmonic fluctuations around this configuration. In the Gaussian phase packet representation, this effective harmonic representation follows the evolution of the centers. At each instantaneous configuration the density distribution represents a distribution of protein structures and no normal mode analysis or time averaging of fluctuations (quasiharmonic analysis) is required. Furthermore, as the centers evolve, the density distribution carves out a tube in phase space representing a volume of many single protein trajectories. It is this enhanced sampling which provides the ability to find the global energy minimum most effectively during simulated annealing.

4.1.1. *Evaluating the Effective Potential for Proteins*

Any difficulty associated with the application of the GPP dynamics comes in calculating the effective potential $\langle V \rangle$ for a protein using a general empirical energy function of the ECEPP[80,81] or CHARMM[82] form. Once $\langle V \rangle$ is known, the GPP dynamical trajectory is found by solving the first order differential equations in the fashion of molecular dynamics.[71,82,83] For covalently bonded systems such as polymers and proteins this is a simple matter while, in contrast, defining effective Monte Carlo moves can be quite challenging.

How do we calculate $\langle V \rangle$ for a protein? To start we need to calculate the effective potential between a pair of interacting atoms. For a many body system it is convenient to approximate the N body distribution function as a Hartree product of the single particle distribution functions[84]

$$\rho(r^N, p^N, t) = \prod_{k=1}^{N} \rho_k(r_k, p_k, t) . \tag{22}$$

This approximation makes it particularly easy to evaluate the pair potential

$$\begin{aligned} \langle V \rangle &= \sum_{i>j=1}^{N} \int dr^N \int dp^N \ \rho(r^N, p^N) V(|r_i - r_j|) \\ &= \sum_{i>j=1}^{N} \int dr_i \int dp_i \int dr_j \int dp_j \, \rho_i(r_i, p_i) \rho_j(r_j, p_j) V(|r_i - r_j|) . \end{aligned} \tag{23}$$

The final effective potential is the interaction energy of the Gaussian densities of the two atoms. That is, the probability of the two atoms being a certain distance apart is smeared out over a range of distances. The result is an averaged potential energy with the features of a coarse grained potential — raised and fewer minima and lower barriers.

The effective pair potential $\langle V \rangle$ is a function of the separation of the centers of the two interacting density packets (as well as the widths of those packets). It turns out that we can reduce the $2d$-dimensional integral above to a one-dimensional integral over the interpacket separation.[74] The result is that the potential is averaged as

$$\langle V \rangle = \int_0^\infty dr G(r) V(r) \tag{24}$$

over a distribution function

$$G(r) = (\zeta/\pi)^{1/2} \frac{r}{r_{ij}} \left[\exp(-\zeta(r - r_{ij})^2) - \exp(-\zeta(r + r_{ij})^2) \right] \tag{25}$$

where $r_{ij} = |\mathbf{r}_0^{(i)} - \mathbf{r}_0^{(j)}|$ is the interpacket center separation and $\zeta = d/[2(M_{2,0}^{(i)} + M_{2,0}^{(j)})]$ is the reciprocal sum of the density packet widths (the mean square spread in position). The effective potential is easily evaluated when $V(r)$ is a Gaussian, exponential, or a polynomial. As we will discuss below, it is most common to represent the nonbonded potential energy as a sum of Gaussians or exponentials, and the intermolecular potential in the form of polynomials.

The general form for the empirical potential energy function most popular in studies of proteins is[81,82,85]

$$V(r^N) = V_{\text{bond}} + V_{\text{angle}} + V_{\text{torsion}} + V_{\text{LJ}} + V_{\text{Coulomb}} . \tag{26}$$

In evaluating the effective potential for the density distribution we can employ a very useful trick — we will write the total potential as a sum of pair potential terms. This is a simple matter for the bond potential which is typically modeled as a harmonic Hook's law form $V_{\text{bond}} = \kappa |r_i - r_{i+1}|^2$. The nonbonded interaction is typically a pair potential consisting of the Lennard-Jones interaction of a hard core repulsion (modeled as a $1/r^{12}$ form or just as well as an exponential) and a longer range attractive London dispersion energy (which varies as $1/r^6$). In some cases it is important to include ion-induced dipole terms (varying as $1/r^4$ and proportional to

the atomic polarizability). Higher order effects such as self-consistent polarizability can be added but with a significant computational overhead. For all but the last interactions, the potential is a simple pairwise additive function.

For the Coulombic interaction (which varies as $1/r$), the average potential of Eq. (24) can be calculated exactly in a simple form. For the Lennard-Jones potential we find it easiest to fit the potential to a sum of Gaussians and/or exponentials for which the average potential is easily calculated.[74]

Other interactions are often treated using three or four atom interactions. The angle term is usually treated using internal angle coordinates defined by the three atoms forming the angle. However, this form may be replaced by a Urey-Bradley interaction of the form $V_{\text{angle}} = \kappa|r_i - r_{i+2}|^2$ (however, the force constants must be adjusted). For the case of rigid bonds, the usual angle term may be written as a harmonic interaction in $|r_i - r_{i+2}|^2$. The torsional potential is usually written in terms of four atom torsional angles but in the case where the bonds and angles are constant it may be rewritten using a simple polynomial as a sum over powers of $|r_i - r_{i+3}|^2$ making it a simple pair potential. The take home message is that the protein potential function can be readily written in terms of pairwise additive interactions. Once this is done, we can use Eq. (24) to evaluate the effective potential energy of interaction between the density of two atoms in the protein modeled using Gaussian phase packets.

An alternative is to treat the internal potential (bond, angle, torsions) as a force acting only on the packet centers. The effective potential is

$$\langle V \rangle = V_{\text{bond}} + V_{\text{angle}} + V_{\text{torsion}} + \langle V \rangle_{\text{LJ}} + \langle V \rangle_{\text{Coulomb}}, \tag{27}$$

where only the nonbonded potential is averaged over the density distribution. The centers of the packets define an underlying protein structure which is confined to a manifold of states of well defined bond lengths and angles.

4.1.2. *Temperature Control during Annealing*

In applying the GPP algorithm to an optimization problem we perform a simulated annealing of the continuous density distribution, which requires that we control the temperature during cooling. There are a variety of ways to do this. One method involves a rigid constraint on the temperature.

Other methods involve coupling the system to a heat bath such that the temperature of the system may fluctuate. We will discuss and compare these methods in the next two sections.

4.1.2.1. *A Rigid Constraint on the Temperature*

To apply the constant temperature constraint which truly holds the temperature constant, we appeal to Gauss's principle[86,87] and write the generalized Liouvillian[74]

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_c = \mathcal{L}_0 - \gamma \frac{\partial}{\partial p} \cdot p, \tag{28}$$

where $\mathcal{L}_0$ is the streaming operator of the system and $\mathcal{L}_c$ is the collision operator which couples the system to the bath. γ is determined by the external temperature constraint equation $dT(t)/dt = 0$ or equivalently $d\langle p^2\rangle/dt = 0$ which guarantees that the temperature remains constant at every instant in time.

Using the Liouville equation we find equations of motion for the average position and momentum[74]

$$\frac{d\langle r\rangle}{dt} = \frac{\langle p\rangle}{m}, \qquad \frac{d\langle p\rangle}{dt} = \langle F\rangle - \gamma\langle p\rangle \tag{29}$$

and for the higher-order moments of position and momentum

$$\frac{dM_{n,k}}{dt} = \frac{n}{m} M_{n-1,k+1} + k\left(W_{n,k-1} - \gamma M_{n,k}\right). \tag{30}$$

The above equations are exact and completely general. For the Gaussian phase packet, the time evolution of the first and second-order moments is given by[74]

$$\dot{r}_0 = \frac{p_0}{m}, \qquad \dot{p}_0 = -\nabla_{r_0}\langle V\rangle - \gamma p_0 \tag{31}$$

$$\dot{M}_{2,0} = \frac{2}{m} M_{1,1}$$

$$\dot{M}_{1,1} = \frac{1}{m} M_{0,2} - \frac{1}{d} M_{2,0} \nabla^2_{r_0}\langle V\rangle - \gamma M_{1,1}$$

$$\dot{M}_{0,2} = -\frac{2}{d} M_{1,1} \nabla^2_{r_0}\langle V\rangle - 2\gamma M_{0,2}.$$

The temperature constraint equation which specifies the cooling schedule $dT(t)/dt$ determines the value of γ. For an exponential cooling schedule $dT(t)/dt = -\gamma T(t)$. Note that $\langle V \rangle$ is the bare potential energy averaged over the phase space distribution

$$\langle V \rangle \equiv \int dr \int dp\ \rho(r,p) V(r)\,. \tag{32}$$

Averaging over the density distribution $\rho(r,p)$ effectively coarse-grains the potential surface. In Fig. 6 we show how the effective potential is smoothed as the width of the density distribution increases.

Note that in general

$$m\ddot{r}_0 = -\nabla_{r_0}\langle V \rangle - \gamma p_0 \tag{33}$$

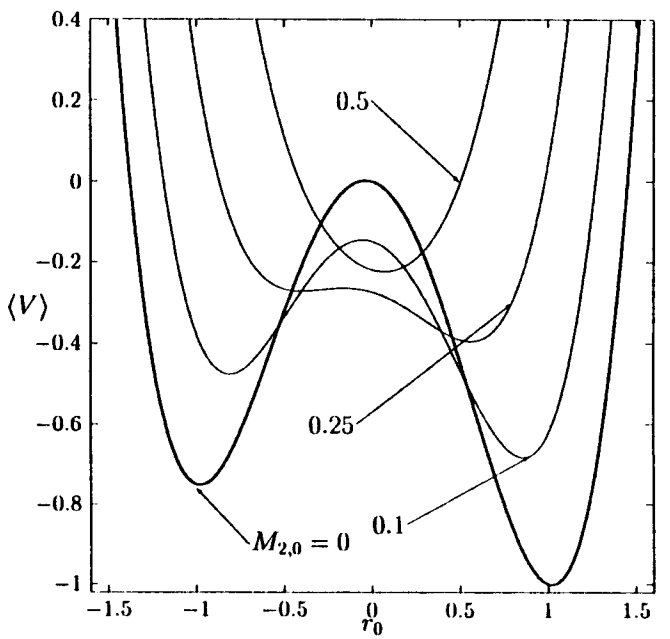

Fig. 6. The effective potential defined by Eq. (32) is plotted for an asymmetric double well for a series of squared widths $M_{2,0}$. The bare potential used in molecular dynamics corresponds to $M_{2,0} = 0$. The units are arbitrary. The potential is approximately what would be seen by a Lennard-Jones atom moving along r between two atoms (of different types) centered at $r = \pm 2$.

which has the form of a Langevin equation (minus the explicit random force) for the Gaussian center r_0 moving on a coarse-grained effective potential. In the limit that $M_{2,0} = 0$ the equations of motion reduce to the constant temperature algorithm of Evans and Hoover[87] for classical point particles.

4.1.2.2. *Coupling to a Bath through Weak and Strong Collisions*

The rigorous constraint on the temperature leads to a set of equations which are easily integrated and which in practice lead to a continually fluctuating $\rho(r,p,t)$. In a more realistic treatment of coupling a system to a heat bath we expect that the density distribution will relax to the static equilibrium density distribution where $\dot{\rho}_{eq} = -\mathcal{L}\rho_{eq} = 0$. This can be achieved, even in an approximate representation of $\rho(r,p,t)$, by coupling the system to the heat bath through a collision operator $\mathcal{L}_c$ of the Fokker–Planck[88] or BGK form.[89]

Consider a "weak collision" Fokker–Planck operator[88]

$$\mathcal{L}_c = -\gamma \frac{\partial}{\partial p} \cdot \left[p + mk_BT \frac{\partial}{\partial p} \right] . \tag{34}$$

The final equations of motion for the Gaussian density distribution of a particle coupled to a heat bath through the Fokker–Planck collision operator are identical to Eq. (31) with the exception that the equation of motion for the second moment of the momentum $M_{0,2}$ is

$$\dot{M}_{0,2} = -\frac{2}{d} M_{1,1} \nabla_{r_0}^2 \langle V \rangle - 2\gamma \left[M_{0,2} - dmk_BT \right] .$$

A related result has been presented where the momenta are adiabatically eliminated leading to a Gaussian packet solution of the Smoluchowski equation.[153,154]

Following the hybrid constant temperature algorithm of Andersen,[55] one might choose a BGK-like collision operator[89,90]

$$\mathcal{L}_c = \gamma \left[1 - \left(\frac{m}{2\pi k_BT} \right)^{d/2} \exp(-p^2/2mk_BT) \int dp \right] \tag{35}$$

which thermalizes the momentum with a rate γ according to the Maxwell distribution. In our GPP dynamics algorithm this corresponds to integrating a deterministic dynamics specified by Eq. (31) where the equation of

motion for $M_{0,2}$ is replaced by[74]

$$\dot{M}_{0,2} = -\frac{2}{d} M_{1,1} \nabla^2_{r_0} \langle V \rangle - \gamma \left[M_{0,2} - (p_0^2 + dmk_B T) \right] . \tag{36}$$

In each algorithm the value of the friction γ is fixed and the temperature is controlled by an external cooling schedule. If the equations of motion are integrated at a fixed T for a long enough time (long compared to $1/\gamma$) the distribution is damped and $\rho(r,p)$ approaches an estimate of the steady state distribution.

The simulated annealing using GPP dynamics and an exponential cooling schedule has been applied to the global energy minimization of atomic clusters ranging from 2 to 55 atoms with excellent results.[74,91] The results

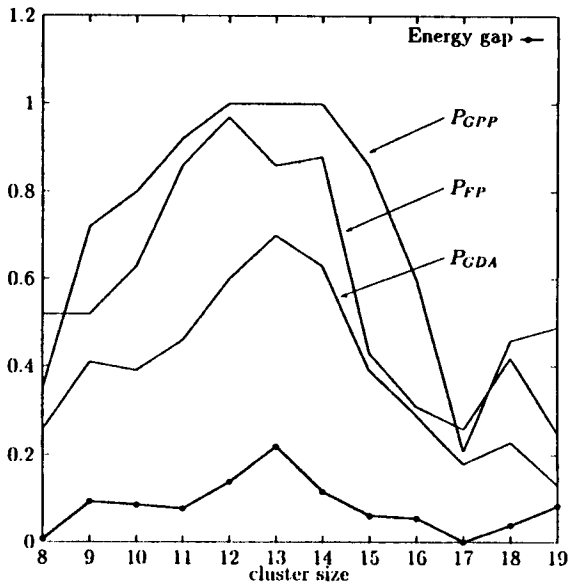

Fig. 7. The probability of locating the global energy minimum for a sample of one hundred independent configurations is plotted for a series of Lennard-Jones clusters for three algorithms: simulated annealing with Gaussian Phase Packets (GPP), simulated annealing using the Fokker–Planck GPP algorithm (FP) and simulated annealing in temperature using the Gaussian density annealing (GDA) algorithm. The energy gap ΔE between the global energy minimum and the next lowest energy state found is plotted (in Lennard-Jones reduced units per particle) to show the inverse correlation between the size of the energy gap and the probability for finding the global minimum.

for annealing using the (1) a rigid temperature control (GPP) and (2) coupling to a bath using the Fokker–Planck collision operator (FP) are summarized in Fig. 7. For comparison, the energy gap $\Delta E = E_1 - E_0$ between the global and next lowest lying energy minimum found in the study is plotted. There is a strong correlation between the probability of finding the global minimum and the magnitude of the energy gap. The hardest optimization problems are those for which the energy gap is small (for example, $n = 17$).

4.2. *Annealing using Locally Enhanced Sampling (LES)*

For conformational searches of peptides and small proteins, a powerful method which stays close to the spirit of sampling conformation space through molecular dynamics is the Locally Enhanced Sampling (LES) algorithm developed by Elber and coworkers.[11,84,92] This method relies on making a "mean-field" approximation where a set of atoms or amino acid residues are simulated by an averaged "field of copies" of those residues. To explain, assume we have a protein which binds small peptides. While the protein is large, we are mostly interested in the binding-site region and how the peptide binds to it. If we were to run Z simulations of the full system, for all these simulations much of the protein will be in the same conformation. That is, the time spent updating the unaffected portion of the system is wasted.

In the LES procedure we divide the atoms or residues into a "system," consisting of the peptide and those residues in the binding region, and a "bath" of residues which are not strongly affected by the binding process. Z trajectories of the "system" are run simultaneously feeling the full force of the single set of "bath" coordinates; atoms of the "bath" feel the averaged force of the Z systems. This is an approximate method which has been quite successful in exploring reaction pathways for carbon monoxide diffusing through the heme protein myoglobin where the ligand is treated as the "system" by Z copies (Z as large as 60).[11] For most problems of interest the "system" will be small compared to the "bath" and the LES algorithm provides an approximate method of dramatically increasing the sampling of peptide conformations (optimally by a factor of Z) while only marginally increasing the computer time required.

Using the formalism of the Liouville equation based algorithms discussed above, we express the density distribution of the LES algorithm as

$$\rho(r^N, p^N, t) = \rho_{\text{bath}}\rho_{\text{system}}$$
$$= \prod_{k=1}^{M} \delta(r - r_k)\delta(p - p_k) \prod_{k=M+1}^{N} \rho_{\text{system}}(r_k, p_k) \qquad (37)$$

and

$$\rho_{\text{system}}(r_k, p_k) = \frac{1}{Z}\sum_{i=1}^{Z} \delta(r - r_{ki})\delta(p - p_{ki}) \qquad (38)$$

where Z is the number of "copies" of the system particles. The density distribution of the peptide bath interacts with an approximate mean field distribution of the protein system.

While the LES method is a powerful technique, it is approximate. For example, as a result of the lack of equal-and-opposite forces the system does not equipartition energy as one would find in normal molecular dynamics — energy is collected in the "system" coordinates effectively heating the system to Z times the temperature of the bath.[93] However, there are approximate methods for correcting this problem[94] including holding the bath coordinates fixed when no sampling of the bath coordinates is desired.[95,96] An additional difficulty arises from the excluded volume of the system. In the limit that the system copies have little overlap, the excluded volume will be Z times that of a single system. Nevertheless, the LES algorithm is a very effective enhanced sampling algorithm which is easily applied using molecular dynamics or Monte Carlo.

Roitberg and Elber have applied the LES method to optimization problems using a dynamical simulated annealing strategy with impressive results for two tetrapeptides and aromatic side chains in BPTI.[97] Moreover, they provide a beautifully simple proof that the result of an ideal annealing run using the LES algorithm will result in the *exact* result for the global energy minimum which would be found using MD simulated annealing.

The strength of the LES algorithm is that is it very simple to implement. No effective potential is required, and the equations of motion are Newton's equations. For problems involving proteins, the empirical potential energy function is a combination of terms simply expressed in either Cartesian or internal coordinates, but not both. This makes the calculation of the effective potential required by the wave packet methods rather tedious. The LES algorithm maintains many of the desirable features of a

mean field approximation (lower barriers, fewer minima) while remaining simple.

4.3. *Selectively Optimized Sampling*

We can imagine a combination of the Gaussian phase packet (GPP)[74] and LES[11] algorithms which provides enhanced sampling for a selected subset of coordinates (as in LES). The bath coordinates are represented by a delta-function distribution of single point particles while a mobile basis of Gaussian packets is used to represent the system coordinates where enhanced sampling is desired.[98] For example, the density distribution of the peptide "system" coordinates can be smeared out as a product of single Gaussian phase packets while the protein "bath" is represented using point particles as in standard molecular dynamics. Such a choice for the density distribution allows for optimized sampling of a *selected set of coordinates*, $i \in (M+1, N)$, while the remaining coordinates, $i \in (1, M)$, are treated as point particles in the standard way using Newton's equations of motion

$$m\ddot{r}_k = \sum_{j=1}^{M}{}' F(|r_k - r_j|) + \sum_{j=M+1}^{N} \langle F(|r_k - r_j|)\rangle_j \,. \tag{39}$$

The primed sum indicates the restriction $j \neq k$. The bracket $\langle\rangle_j$ indicates an average over the density distribution of the jth phase packet. We expect to find the correct energy equipartitioning and avoid any excluded volume problems using this hybrid representation.

5. Classical Simulated Annealing in Temperature

The nut to crack in applying simulated annealing to proteins is the definition of a cooling schedule which leads to a reasonable probability of finding the global minimum in a computationally realizable time. An appealing alternative is to replace the cooling schedule necessary when annealing in real or Monte Carlo time with direct integration in temperature. As we shall see later, this is closely related to quantum dynamics in imaginary time. There is no real time and therefore no cooling schedule that must be specified as part of the annealing protocol. There is only a set of equations describing the evolution of the ensemble of systems in temperature. In this section I discuss a method which directly anneals an approximation to the

classical density distribution in temperature and then describe a few closely related algorithms.

5.1. *The Gaussian Density Annealing (GDA) Algorithm*

Ma and Straub[91] have recently explored the possibility of performing simulated annealing directly in temperature by approximately integrating the classical Bloch equation[99]

$$\frac{\partial \rho_{eq}}{\partial \beta} = -(\mathcal{H} - \langle \mathcal{H} \rangle)\rho_{eq} \tag{40}$$

to obtain the equilibrium classical density distribution for the canonical ensemble $\rho_{eq}(r, p, \beta), = \exp(-\beta\mathcal{H})/Q(\beta)$ at a given temperature. $\mathcal{H}(r, p)$ is the classical Hamiltonian energy function and $Q(\beta) = \int dr dp \exp(-\beta\mathcal{H})$ is the canonical partition function. In fact, Eq. (40) is the classical analog of the imaginary time Schrödinger equation (discussed in the next section). For an N body system it is convenient to make the Hartree approximation to the many body density distribution as a product of the single body density distributions (employed in the GPP integration[91] of the Liouville equation)

$$\rho(r, \beta) = (2\pi M_2/d)^{-d/2} \exp\left[-\frac{d}{2M_2}(r - r_0)^2 \right]. \tag{41}$$

The equations of motion in reciprocal *temperature* for the center r_0 and second moment M_2 of a single Gaussian packet in d-dimensions are

$$\begin{aligned} \frac{\partial r_0}{\partial \beta} &= -\frac{1}{d} M_2 \nabla_{r_0} \langle V \rangle \\ \frac{\partial M_2}{\partial \beta} &= -\frac{1}{d^2} M_2^2 \nabla_{r_0}^2 \langle V \rangle . \end{aligned} \tag{42}$$

$\langle V \rangle$ is the pair potential averaged over the density distribution

$$\langle V \rangle_{GDA}(r_0, \beta) = (2\pi M_2/d)^{-d/2} \int dr' V(r') e^{-d||r_0 - r'||^2/2M_2} . \tag{43}$$

This effective potential is of the form that appears in the approximate solution of the Liouville equation using Gaussian phase packets. The evaluation of this potential for a protein was discussed in the previous section.

The form of these equations is quite appealing. The centers move according to a steepest descent energy minimization equation on the effective

potential energy surface while the widths of the density distribution adjust themselves to the curvature of the effective potential surface. Therefore, the method incorporates the general properties of potential smoothing algorithms. Moreover, the annealing minimization evolves directly in temperature with no real time dynamics. While the equations define a simulated annealing protocol the algorithm is independent of cooling schedule — if the equations of motion are integrated accurately we should have the optimal annealing protocol for that representation of the density distribution.

The strength of the GDA method is that the problematic cooling schedule has been replaced by a set of deterministic equations of motion for the annealing of the density distribution. It is a simple matter to interpret the evolution of the density distribution in temperature. Starting with the known infinite temperature distribution

$$\rho_{\text{eq}}(\beta = 0) = \text{constant} \tag{44}$$

the equations are integrated in temperature to a final value β_{max} which must satisfy $1/\beta_{\text{max}} \ll \Delta E$ where ΔE is the energy gap between the global and next lowest energy minimum (see Fig. 7). For small β, the Gaussian representation of the density distribution should be quite reasonable. In the intermediate β regime, we expect that many minima contribute to the exact equilibrium density distribution of the protein and the Gaussian density approximation will be at best an incomplete representation of the equilibrium ensemble of configurations. However at large β (low enough temperatures), we expect that the equilibrium density distribution will be localized in a single potential minimum and the Gaussian density approximation should be a good representation of the exact equilibrium distribution.

This method has been applied to the standard case of Lennard-Jones clusters ranging from 2 to 55 atoms with excellent results[91] (see Fig. 7). It has also been shown that the evolution of the density distribution in temperature according to the GDA algorithm leads to physically reasonable results for the system density and distribution functions. Initially, at small β (high temperature), the density corresponds to that of a gas. As the temperature is lowered there is a weak phase change to a liquid-like density distribution, followed by a stronger phase change to a solid-like density distribution localized in the global (or a low lying) energy minimum. During the periods of phase change the equations of motion can become stiff;

however, as long as they are integrated accurately there is an adequate amount of time spent in the transition region. This is in contrast to the standard simulated annealing algorithm where one must guess how long to spend in a region of phase change where there are long length scale correlations and relaxation times.

Amara and Straub[100] have applied the GDA algorithm to a model protein.[68] The model consists of 22 residues where each residue is a sphere of neutral (N), hydrophobic (B) or hydrophilic (L) nature. A pair of hydrophobic sites interact with a favorable Lennard-Jones 12-6 potential. A neutral site interacts with any other site through a purely repulsive $1/r^{12}$ interaction. A hydrophilic site interacts with a hydrophilic or hydrophobic residue through a 12-6 type potential where the $1/r^6$ term, as well as the core potential, is purely repulsive. The torsion potential for dihedral angles

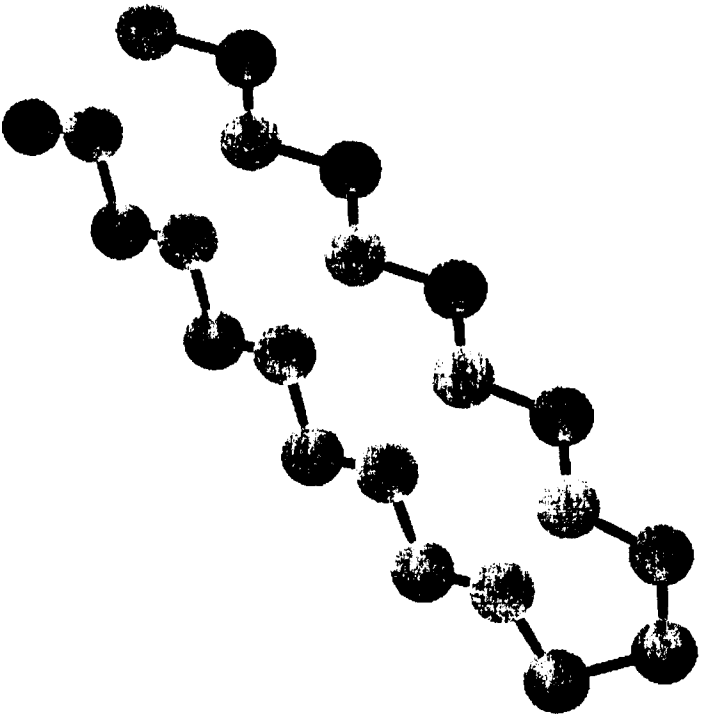

Fig. 8. The global energy minimum of the model protein studied by Amara and Straub using the GDA algorithm.

made-up of two or more neutral residues (in the center of the chain) are floppy, one fold sinusoidal functions while all other dihedrals are more rigid and made of a linear combination of one and three fold sinusoidal terms. The details of this potential can be found elsewhere.[68]

The sequence studied is $B(LB)_5\ N_2\ (LB)_5$ which leads to a global energy minimum configuration of a β-sheet (see Fig. 8). The role of the rigid dihedrals is to form the two linear chain regions while the more flexible central torsion angles allow for the bend to form a β-sheet. The global energy minimum configuration allows for a maximum number of energetically favorable hydrophobic pair contacts. This is a minimal model for protein folding

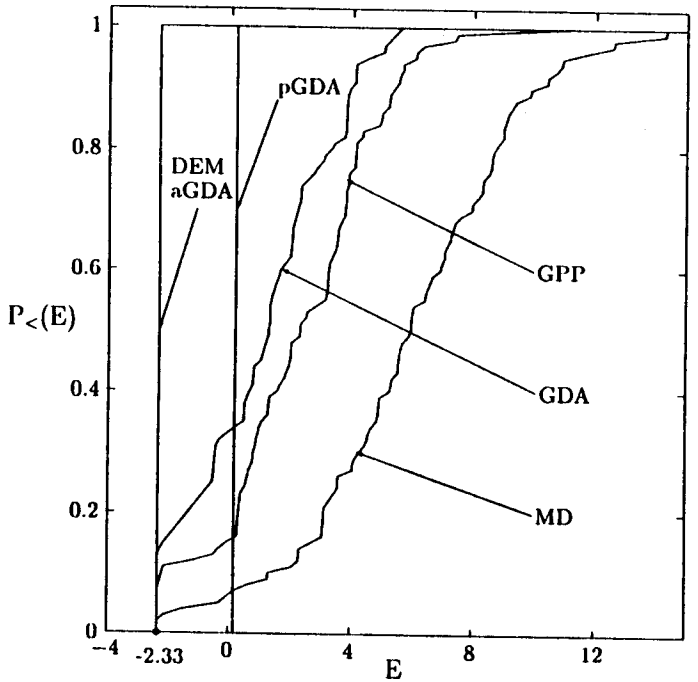

Fig. 9. The probability of locating a configuration with energy E or less from an initial distribution of one hundred independent configurations is plotted for a model protein for five algorithms: simulated annealing in time using molecular dynamics (MD) and Gaussian phase packet dynamics (GPP), simulated annealing in temperature using the Gaussian density annealing (GDA) algorithm, the "adiabatic" GDA algorithm (aGDA), the GDA algorithm using a single "preconditioning" step (pGDA) and the diffusion equation method (DEM) of Scheraga and coworkers.

and represents a serious test of the optimization algorithms. It is a short step from the potential function used for this model protein to the more general all atom empirical potential energy function such as ECEPP or CHARMM.

The results for the energy minimization of this model are presented in Fig. 9.[100] Simulated annealing using molecular dynamics (MD) was applied using an exponential cooling schedule. This schedule is not optimal and was chosen as a standard for the comparison of algorithms. The same cooling schedule was employed in a simulated annealing protocol using the Gaussian phase packet (GPP) dynamics with the rigid temperature constraint. The third method applied was the Gaussian Density Annealing (GDA) algorithm which is free of a cooling schedule. The initial mean squared width for the Gaussian density of each residue was $M_2 = 5$. In each case a weak pairwise harmonic boundary potential was used to encourage the molecule to explore compact configurations and the effective potential of Eq. (27) was used. The results are summarized in Fig. 9 in terms of the probability of finding a minimum energy configuration with energy E or less from a distribution of independent configurations chosen from a high temperature molecular dynamics trajectory. The data indicate that the GPP annealing algorithm performs significantly better than the standard MD simulated annealing for the same cooling schedule. The GPP results locate the global energy minimum more frequently and tend to find significantly lower local energy minima. The GDA algorithm outperforms both the GPP and MD simulated annealing algorithms by isolating the global minimum more frequently and tending to locate only very low energy local minima. However, while the GPP results can be improved by optimizing the cooling schedule (allowing for slower cooling), the GDA algorithm (which is free of any cooling schedule) can be improved only through a better choice of initial conditions and boundary potential or a more complete representation of the density distribution $\rho(r, \beta)$.

5.2. *The Diffusion Equation Method (DEM)*

I find it useful to think of the highly successful DEM of Piela, Kostrowicki and Scheraga[101] as a classical annealing algorithm. In the DEM, a steepest descent minimization is carried out on a transformed potential energy surface. The potential transformation is the convolution of the potential with a Gaussian function of a given width. This leads to a coarse graining of the

potential function over a length scale defined by the width of the Gaussian. Initially, the Gaussian width is taken to be large and it is hoped that the transformed potential has one or a few minima. (This can be guaranteed by a clever adjustment of the boundary potential which primarily changes the initial conditions.) Energy minimization is used to find the minimum of the transformed surface. The width is then gradually decreased and the minimum is mapped to an energy minimum of the physical potential surface (where the Gaussian widths are zero).

The transformed potential energy function used in the DEM can be written[102]

$$\langle V\rangle_{DEM}(r,t) = (4\pi t)^{-d/2} \int dr' V(r') e^{-||r-r'||^2/4t}, \qquad (45)$$

where t is the diffusion "time." The larger the value of t, the wider the Gaussian and the longer the length scale for the potential smoothing. This transformed interaction potential employed in the DEM is exactly the effective potential employed in the GDA algorithm where the squared width of the Gaussian $M_2 = 2td$ is proportional to the diffusion "time" of the DEM. In fact, the *ad hoc* potential deformation used in the DEM results from a fundamental treatment of the temperature evolution of the equilibrium density distribution for the special case of a Gaussian density approximation.

We can take this comparison an extra step. In the GDA algorithm one initially chooses large M_2's and then allows the centers to follow a steepest descent trajectory to the minimum potential energy conformation. In the DEM there is a single diffusion time meaning that all particles are constrained to have the same M_2, and as the diffusion time is reduced, corresponding to $\beta \to \infty$, the effective widths decrease monotonically. In the case of the GDA algorithm the widths are variationally optimized depending on the environment of each particle. The DEM can be thought of as a special case of the GDA algorithm where (1) the second-moments M_2 for all wave packets are identical, and (2) the magnitude of M_2 is initially set to something large and then monotonically decreased to zero while integrating the r_0 equation to converge at each value of M_2. For this reason, it is profitable to think of the DEM as a classical annealing algorithm.

The DEM has been applied to water clusters with mixed results.[103] As has been noted, a Gaussian coarse grained smoothing of the sort employed in the GDA and DEM might fill in a deep narrow well (the global energy

minimum) before it fills a shallow broad well. Perhaps this "pathology" is present in the water–water interactions which makes certain cluster sizes difficult optimization problems. Of course, it might be that the higher lying broad *potential* minimum is a lower *free energy* minimum.[104]

In an application of the DEM to the Met-enkephalin pentapeptide (see Fig. 10), the torsional angle potential was set to zero and the bond lengths and angles were fixed. The solution found by the diffusion equation method was one on the manifold of peptide conformational states where the bonds and angles are fixed and the nonbonded (Lennard-Jones and Coulombic) potential is favorable. In this study, Kostrowicki and Scheraga showed that the initial boundary conditions are extremely important in the solution of the DEM. Depending on the form of the boundary potential and the initial diffusion time it is possible to find one or a number of initial conformations and some or none of these can lead to the global energy minimum. For example, a two Gaussian representation of the nonbonded and hydrogen bonding functions of the ECEPP potential was used. After the initial minimization on the transformed potential there was one minimum corresponding to a rather extended conformation. This minimum mapped onto

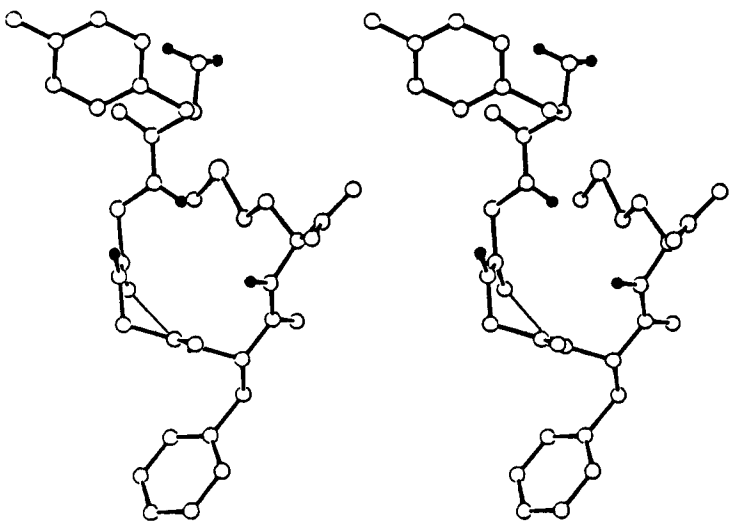

Fig. 10. View of the Met-enkephalin pentapeptide energy minimum isolated by the DEM taken from the study of Kostrowicki and Scheraga.[152]

a local (and relatively high lying) potential minimum for the pentapeptide. However, when a three Gaussian representation of the nonbonded and hydrogen bonding functions was used, two compact initial minima were found for the transformed potential and both of these minima mapped to a single structure which was very close to the form of the global energy minimum. The small differences between the DEM result and the global minimum are due to the absence of the torsional angle potential.

These results demonstrate the crucial importance of the initial conditions used for the DEM. Similar conclusions follow for the application of the GDA method. Their study also demonstrates that with a good choice of boundary conditions the DEM can be a very successful algorithm for energy minimization on the complex potential energy surface of peptides and proteins. This point has been discussed in detail by Scheraga and coworkers.[105] More research needs to be performed to understand the details of the influence of the initial (boundary) conditions for these methods. Another important point is that the DEM is a deterministic procedure, in that once the few minima are initially isolated on the transformed potential surface, there is an exact procedure for mapping these minima on to the physical potential surface. This is a positive aspect of the DEM and GDA methods which those methods built on simulated annealing in time lack.

To allow for comparison with other optimization methods discussed in this chapter, the DEM was applied to the model protein (discussed in the previous section) and the results are displayed in Fig. 9. (The comparison made here is used to demonstrate the general properties of the algorithms and not the absolute best results which can be obtained.) Note that the DEM isolates a single final configuration and that configuration is the global energy minimum. In contrast, the GDA algorithm finds a distribution of minima which includes the global minimum and all low lying "excited" states. In the DEM, the first step consisted of a local energy minimization on the transformed potential surface for a set of one hundred random configurations taken from a high temperature molecular dynamics run. With a large enough initial diffusion time (in this case 1000) and a proper choice of boundary conditions one can expect to find a few (in this case one) energy minima on the transformed potential surface. The DEM then consisted of a single trajectory which mapped the lone minimum energy configuration on the transformed surface on to the global energy minimum of the physical potential surface.

5.3. *Adiabatic Gaussian Density Annealing (aGDA)*

An important variation of the GDA algorithm is the "adiabatic GDA" (aGDA) method. As I noted above in the discussion of the simulated annealing algorithms, the hope is that at each temperature the density distribution simulated is close to the equilibrium distribution. In the primitive GDA algorithm there are errors that develop due to the fact that the basis set for $\rho(r)$ is incomplete. These errors can lead the density distribution away from the equilibrium value. Tsoo and Brooks recognized that such errors propagate and increase during integration.[106] They suggested that this problem could be minimized by occasionally fixing the values of the set of $\{M_2\}$ and running a steepest decent trajectory on the corresponding effective potential surface. This step allows the centers to adjust to the set of packet widths leading to a closer approximation to the "steady state" or hopefully equilibrium solution. The results for a series of Lennard-Jones clusters were excellent. They also applied the modified GDA algorithm to water clusters with interesting results.[106]

In earlier work on a related imaginary time optimization algorithm,[107] it was suggested that the equations of motion for r_0 and M_2 could be decoupled. For the primitive GDA, problems result when the M_2 values decrease too quickly. Therefore, a useful "adiabatic" approximation provides that for every set of $\{M_2\}$ the position of the density distribution center relaxes "instantaneously" to its equilibrium value. The algorithm is applied as follows. (1) The steepest descent equation

$$\dot{r}_o = -\nabla_{r_0}\langle V(r_0, M_2)\rangle \tag{46}$$

is solved with a fixed set of widths $\{M_2\}$ to find the minimum on the effective potential surface for a particular temperature $r_0^{ad}(\beta)$. This can be performed using a favorite local minimizer. (2) The widths of each packet are integrated out in β holding fixed the set of center positions $\{r_0^{ad}\}$ by solving

$$\frac{\partial M_2}{\partial \beta} = -\frac{1}{d^2} M_2^2 \nabla_{r0}^2 \langle V(r_0^{ad}, M_2)\rangle \,. \tag{47}$$

(3) Return to step (1) and repeat the cycle.

The resulting algorithm is very much like the DEM in that the positions of the centers follow changes in the widths adiabatically. We refer to this method as the "adiabatic" GDA algorithm (aGDA). An important difference between this algorithm and the DEM is that each width is allowed to

evolve in an optimal way which depends on the curvature of the effective potential surface; in the DEM each particle "density" has the same width. This difference should be particularly important in studies of inhomogeneous systems such as proteins which lack the symmetry of the solutions for the atomic clusters. The results for the aGDA algorithm are shown in Fig. 9. As we found in the case of the DEM, the distribution of initial configurations mapped on to a single energy minimum on the transformed potential surface where the particle widths were $M_2 = 5$. Following the initial minimization, there was a single trajectory which led to the global energy minimum.

It is also possible to perform a "preconditioning" step by solving Eq. (46) for the initial set of $\{M_2\}$ and subsequently running a trajectory according to the usual GDA equations of motion Eq. (42). This preconditioned GDA (pGDA) was applied to the model protein and the results are shown in Fig. 9. Like the aGDA and DEM algorithms, the pGDA algorithm tends to isolate very few minima. However, for this particular run the aGDA and DEM isolated the global energy minimum while the pGDA algorithm found a higher lying local minimum. Similar results were found by Tsoo and Brooks[106] in their optimization study of water clusters. It is important to allow the packet centers to "keep up" with any rapid variation in the packet widths. This seems to explain the success of the aGDA method in finding the global energy minimum.

5.4. *Shalloway's Packet Annealing (PA)*

The strength of both the Gaussian Density Annealing algorithm[91] and the Diffusion Equation Method[101] rests on the use of a similar transformed potential energy function. In the former case, the use of the transformed potential energy follows from a solution of the classical Bloch equation. However, while the lowest potential energy is desirable, thermodynamically, it is the basin of lowest free energy that will predominate at low temperatures. Shalloway has developed an intriguing method which resembles the GDA and DEM algorithms, with the important difference that the optimization problem is solved on a transformed *free energy* surface.[108-110]

The GDA and DEM methods employ a Gaussian transform of the potential energy function. In Shalloway's Packet Annealing (PA) method, the transformed potential is replaced by a Gaussian transform of the Boltzmann factor[108]

$$\langle F\rangle(r) = -\frac{1}{\beta}\ln\left[\left(\pi\Lambda^2\right)^{-d/2}\int dr' e^{-\beta V(r')}e^{-\|r-r'\|^2/\Lambda^2}\right]. \qquad (48)$$

The Boltzmann factor is coarse grained over a region of length scale Λ to provide an effective free energy $\langle F\rangle$. This is simple to write, but difficult to evaluate (except for quadratic potential functions). Nevertheless, Shalloway and coworkers have developed an efficient quadrature method which has been applied to the Lennard-Jones potential in the optimization of atomic clusters.[110]

The PA method begins with the specification of an initial value of the temperature and the smoothing length scale Λ. A schedule for the reduction of the temperature and Λ is then specified. At each temperature, a Monte Carlo search of the configuration space is carried out with a careful choice of step size. The distribution generated by the Monte Carlo search on the effective free energy surface is annealed according to the cooling schedule until a low temperature is reached. The resulting distribution provides a guess at the global energy minimum. The number of Monte Carlo steps taken at a given temperature was varied as $n\log(n)(1-\log T)$ as the temperature decreased, n being the number of atoms in the cluster. The cooling schedule was defined by decreasing the temperature by a factor of $\xi < 1$ from the temperature at the previous step.[110]

Reasonable results are obtained for the optimization of Lennard-Jones clusters using the strategy described above. Coleman, Shalloway and Wu[110] examined two variations: (1) where a local minimization algorithm was applied at the end of the packet annealing process described above, and (2) where local annealing was performed at the end of each cooling step. The latter algorithm which makes use of the information provided by the whole trajectory, rather than the endpoint above, was the more effective one in our finding the global minima of Lennard-Jones clusters from 3 to 27 atoms as well as larger clusters of 36, 54 and 100 atoms with reasonable probability.

The most appealing aspect of the PA method is that it employs an effective free energy function rather than the potential energy surface. Like the GDA and DEM, the PA method has had success for the optimization of atomic clusters. However, for the optimization of complex molecular systems, calculation of the effective free energy function is expected to be significantly more time consuming. As discussed above, the technical problems are easily solved for the calculation of $\langle V\rangle$, required by the GDA

and DEM algorithms, from an empirical energy function of a protein. One expects to have a more difficult time in the calculation of $\langle F \rangle$. Recent application of the PA method to met-enkephalin has provided very encouraging results.[155]

6. Quantum Mechanical Annealing

The dynamics of classical and quantum mechanical systems are qualitatively different. To a classical point particle, energy barriers are impregnable. Energy barriers must be crossed over, and if the total energy for the system is less than the barrier energy, the barrier cannot be crossed. However, a quantum mechanical particle may tunnel through a barrier whose height is greater than the total energy of the system. As a result, quantum mechanical annealing methods are particularly attractive as methods for global optimization (see Fig. 1).

I will mention several approaches to this problem. Each method is an extension of a well developed method for finding the ground state wave function of a quantum mechanical system. The hope is that for a system with a nondegenerate ground state, the density $||\phi(r)||^2$ associated with the wave function $\phi(r)$ will be localized in the region of the classical global energy minimum. The optimization process is then approached in two steps, (1) an estimate for the ground state wave function is found, and (2) the classical limit is taken (by reducing $\hbar$ to zero, or increasing the mass, or simply reducing the ratio $\hbar^2/2m$ to zero).

6.1. *Variational Calculations*

The most successful methods for finding an estimate of the ground state wave function of an atomic or molecular system rely on the time-indepen -dent variational method.[111] Given a system where the ground state energy is $\mathcal{E}_0$, any guess $\phi(r)$ at the wave function of the system will have an energy

$$E = \langle \phi | \mathcal{H} | \phi \rangle \geq \mathcal{E}_0 \,, \tag{49}$$

where $\mathcal{H}$ is the Hamiltonian operator. That is, the energy of the guess is always greater than or equal to the exact ground state energy.

Somorjai proposed using a distributed fixed Gaussian basis set.[112] While this suggestion is a reasonable approach for one-dimensional problems,[113] there are reasons to doubt the efficacy of such an approach for larger systems

of hundreds of atoms which can extend over large regions of configuration space. For a long chain molecule, extended states of the system may explore large volumes while occupying relatively little of it at any given time. The space fixed basis set is everywhere, all the time. As such, it can spend significant amounts of computational time monitoring empty space. This problem can be addressed by exploring mobile basis sets that are functions of atomic or internal molecular coordinates.

Olszewski, Piela and Scheraga[114] have developed a method for finding the global energy minimum of a protein by approximating the ground state wave function of the system. Their Self-Consistent Multi-Torsional Field (SCMTF) method is based on approximating the protein wave function by a Hartree product of normalized single torsion angle wave functions $\phi_k(\theta_k)$, leading to a set of coupled one-dimensional time-independent Schrödinger equations with the effective Hamiltonian

$$\mathcal{H}_k = -\frac{\hbar^2}{2I_k}\frac{\partial^2}{\partial\theta_k^2} + \langle V\rangle_k(\theta_k) \tag{50}$$

where I_k is an averaged moment of inertia for the kth torsion angle, and $\langle V\rangle_k(\theta_k)$ is the mean field potential energy for the kth torsion angle — the potential energy averaged over all degrees of freedom other than θ_k. The minimization procedure consists of (1) generating an initial configuration of the protein (a set of dihedral angles), (2) estimating the various

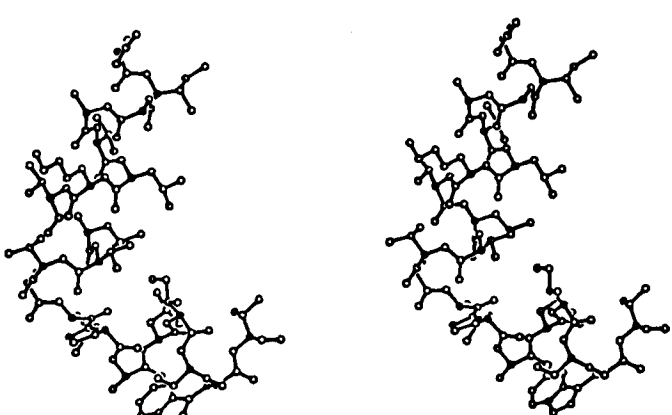

Fig. 11. View of the global energy minimum configuration of melittin isolated using the SCMTF method taken from the study of Olszewski, Piela and Scheraga.[116]

moments of inertia I_k from an ensemble of configurations, (3) calculation of the mean field effective potential $\langle V\rangle_k(\theta_k)$ where the required integrals are computed using a Monte Carlo procedure, and (4) solution of the set of coupled one-dimensional Schrödinger equations, iterating until the estimates of the $\phi_k(\theta_k)$ no longer change. The lowest energy configuration found in this procedure is the guess at the global minimum. The SCMTF method has been applied to Met-enkephalin,[114] decaglycine and icosalanine,[115] and a 20 residue portion of the protein melittin.[116] For melittin the SCMTF method effectively located the global potential energy minimum for the system from four independent starting configurations (see Fig. 11).

6.2. *Imaginary (Euclidean) Time Methods*

The time-dependent Schrödinger equation is[117]

$$i\hbar\frac{\partial}{\partial t}\phi(r,t) = -\frac{\hbar^2}{2m}\nabla^2\phi(r,t) + V(r)\phi(r,t) \tag{51}$$

where $\phi(r,t)$ is in general a complex function. $\hbar$ is Planck's constant, $V(r)$ is the potential energy, and m is the mass. The imaginary (Euclidean) time form of the time-dependent Schrödinger equation is defined by substituting the imaginary time variable $\tau = it/\hbar$ so that

$$\frac{\partial}{\partial\tau}\phi(r,\tau) = \frac{\hbar^2}{2m}\nabla^2\phi(r,\tau) - V(r)\phi(r,\tau)\,. \tag{52}$$

We arbitrarily take $\phi(r,\tau)$ to be real. The wonderful property of this equation is that the formal solution is of the form

$$\phi(r,\tau) = e^{-\mathcal{H}\tau}\phi(r,0)\,. \tag{53}$$

Suppose we know the eigenfunctions and energies for $\mathcal{H}$ which are $u_n(r)$ and $\mathcal{E}_n$ for the nth excited state. We can rewrite the wave function as

$$\phi(r,\tau) = \sum_n c_n u_n(r) e^{-\mathcal{E}_n\tau}\,. \tag{54}$$

At large enough values of τ the dominant contribution to the wave function comes from the ground state

$$\phi(r,\tau) \simeq u_0(r)e^{-\mathcal{E}_0\tau} \tag{55}$$

while the contribution from all excited states is exponentially smaller. Starting from a guess at $\phi(r,\tau)$, the imaginary time method consists of integrating Eq. (52) to find an estimate of the ground state wave function. This method is widely employed and can be used to find accurate ground and excited state energies of small molecular systems.

6.2.1. *Quantum Mechanical Annealing using Diffusion Monte Carlo (DMC)*

As Schrödinger originally noted, the imaginary time Schrödinger equation has the familiar form of the classical diffusion equation

$$\frac{\partial}{\partial\tau}C(r,\tau) = \mathcal{D}\nabla^2 C(r,\tau) - kC(r,\tau) \tag{56}$$

with concentration $C(r,\tau)$, diffusion constant $\mathcal{D}$, and rate constant k (for the first-order rate process).[118] As this analogy makes clear, one can simulate the imaginary time Schrödinger equation as a classical diffusion equation for a distribution of system replicas (a concentration profile or density distribution). The first-order rate process leads to the spawning and death of system replicas. In practice, one simulates an ensemble of replicas of the system. Each replica diffuses through a Monte Carlo walk in configuration space. At each step a decision is made whether to spawn or kill system replicas based on the rate constant $k = V(r)$.[119,118]

Doll and coworkers have recently adapted the DMC method[119,120] to treat the classical optimization problem. They apply the DMC to determine the ground state wave function and energy at a given value of $\hbar^2/2m$. The mass of the particles is then gradually increased, each time applying the DMC method, until the classical limit is reached. At this point, they arrive at an estimate of the global energy minimum of the system. Their method has been applied to Lennard-Jones atomic clusters with perfect results for clusters of up to 19 atoms.[121]

The use of DMC is particularly attractive since (1) no approximation to the wave function is required, and (2) no effective potential (or matrix elements) or derivatives are required, only evaluations of the potential energy function. It appears to be a very attractive method for applications to biomolecules where an empirical energy function is complicated by the mix of terms depending on Cartesian and internal coordinates (see Sec. 4.1.1).

There are two possible drawbacks. One is that DMC is exact and becomes more computationally intensive than an approximate solution. Moreover, there is the difficulty of applying Monte Carlo to proteins in an efficient manner. It is well-known that a simple Metropolis algorithm of random atomic moves distributed uniformly in a cube will typically lead to an increase in energy due to the stiff bond and angle constraints. However, more intelligent algorithms have been designed which make use of internal coordinates or normal modes[122,123] and these appear to be quite promising. Nevertheless, molecular dynamics remains the paradigm for the simulation of thermodynamic properties of proteins. For these reasons, it is important to develop an approximate method for solving the imaginary time Schrödinger equation based on the solution of deterministic equations of motion (akin to molecular dynamics). One such method is based on Gaussian wave packet dynamics in imaginary time.

6.2.2. *Quantum Mechanical Annealing using Gaussian Wave Packets*

The formal solution of the imaginary time Schrödinger equation is $\phi(r,\tau) = \exp(-\mathcal{H}\tau)\phi(r,0)$. The form of the equilibrium distribution function is $\exp(-\mathcal{H}/k_BT)$, where k_B is Boltzmann's constant and T is the temperature. Here, time plays the role of an inverse temperature. Lowering the temperature to anneal the system is equivalent to increasing the imaginary time. Amara, Hsu and Straub have exploited this idea in searching for the global energy minimum through approximate solution of the imaginary time Schrödinger equation for the ground state wave function using Gaussian wave packets.[107]

The starting point is an alternative normalized expression which is a variation of Eq. (52), namely

$$\frac{\partial}{\partial\tau}\phi(r,\tau) = -\left(\mathcal{H} - \langle\mathcal{H}\rangle\right)\phi(r,\tau)\,. \tag{57}$$

If we take the wave function of each particle to be a Gaussian

$$\phi(r,\tau) = (2\pi M_2/d)^{-d/4}\exp\Big[-\frac{d}{4M_2}(r-r_0)^2\Big] \tag{58}$$

we can fully define the Gaussian by its average position $r_0 = \langle r\rangle$, and its squared width or second moment, $M_2 = \langle(r-r_0)^2\rangle = d\sigma^2$. It is a simple matter to derive equations of motion for the evolution of the GWP

in imaginary time[107]:

$$\dot{r}_0 = -\frac{2}{d} M_2 \nabla_{r_0} \langle V \rangle , \tag{59}$$

$$\dot{M}_2 = \frac{d\hbar^2}{2m} - \frac{2}{d^2} M_2^2 \nabla_{r_0}^2 \langle V \rangle .$$

These are the optimal equations of motion for the Gaussian which minimizes the square of the error between the approximate and exact solutions. (1) As has been noted by Heller, in real time the equations of motion for a GWP have no explicit dependence on $\hbar$.[124] However, the imaginary time equations do depend explicitly on $\hbar$. (2) The evolution of the packet center resembles that of a steepest descent[44] down the effective potential $\langle V \rangle$

$$\dot{r}_0 \propto -\nabla_{r_0} \langle V \rangle . \tag{60}$$

The term $d\hbar^2/2m$ acts to expand the wave packet and is a measure of the quantum nature of the particle. For an effectively classical system we expect $\hbar^2/2m$ to be small. If we choose, we can increase this parameter to encourage the system to delocalize and tunnel. (3) The packet will evolve to minimize the effective *total* energy, a sum of the potential $\langle V(r_0, M_2) \rangle$ and kinetic $d\hbar^2/8mM_2$ energies (see Fig. 12). Larger values of $\hbar$ increase the importance of the kinetic energy and raise the zero point energies of the well minima. The effective total energy surface has many of the features of quantum effective potentials which may be used in the simulation of quantum systems.[58,125] (4) For a many-body system we approximate the total wave function by a Hartree product of N single particle wave functions

$$\psi(r^N, \tau) = \prod_{k=1}^{N} \phi_k(r_k, \tau) . \tag{61}$$

The effective potential energy can be written as a sum over the individual pair interaction energies

$$\langle V \rangle = \int dr \prod_k \rho_k(r_k) V(r) \tag{62}$$

where the density of a given particle is $\rho_i(r_i) \equiv \phi_i^*(r_i)\phi_i(r_i)$ and $V(r)$ is averaged over the positions of all the other particles — a mean field potential.[126] This method has been very successful in locating the global energy minimum for a series of Lennard-Jones clusters.[107]

The difficulty in applying this method to biomolecules is that the effective potential $\langle V \rangle$ must be calculated. When the GWP is written in Cartesian coordinates, the effective potential for the Coulombic interaction may be calculated exactly. The Lennard-Jones potential (or an ion induced dipole interaction, etc.) must be fit to a series of Gaussians. The dihedral potential may be expressed in terms of a polynomial in the distance $r_{i,i+4}$ to perform the integrals in Eq. (62). So, it is possible to evaluate $\langle V \rangle$ for the standard empirical energy function for a protein, but with some added complexity when compared with the DMC method which requires only an evaluation of the empirical energy function itself (see Sec. 4.1.1).

The appeal of the imaginary time method is that the equations of motion are deterministic and, due to the small number of parameters which must be followed, the integration is readily performed for a biomolecule or polymer. One is relieved of the burden of finding clever Monte Carlo moves which efficiently explore configuration space. A natural and promising extension of the methods discussed in this section involves quantum mechanical annealing using a path integral representation of the protein.[127,128]

6.3. *"Tunneling" in General*

In an early study of this kind of quantum mechanical annealing, Ruján adapted the Green Function Monte Carlo (GFMC) technique to create an algorithm called "simulated tunneling."[129] In the GFMC method, an ensemble of systems (or population of configurations) is generated, and each member of the ensemble is allowed to diffuse until an equilibrium fluctuating population of configurations is found. The end result is an estimate of the energy and wave function of the quantum mechanical ground state.

Since the goal is to find the classical global energy minimum, and not the quantum mechanical ground state energy, Ruján made some simplifications of the GFMC method inspired by classical simulated annealing which led to his simulated tunneling algorithm. (1) An ensemble of systems is generated and a cooling schedule is chosen. (2) Each system in the ensemble is allowed to diffuse following a Monte Carlo procedure as the temperature is lowered according to a cooling schedule. (3) If a new minimum is located by the ensemble, return to (1) and generate a new ensemble of states; if not, stop.

Ruján applied his method to the traveling salesman problem, with apparently excellent results. He observed that this algorithm, which employs a diffusing population of systems, resembles not only imaginary time

quantum mechanical evolution and classical reaction diffusion dynamics, but also evolutionary systems, and computational ecologies.

Reflecting on our discussion of the GDA method, we can recognize that while a single classical point particle cannot tunnel through an energy barrier, a distribution of classical particles can. This is evident in Fig. 12 where the barrier in the effective total energy is lopped off even in the classical $\hbar = 0$ limit. With a distribution of particles, even if the *average* energy of the distribution is less than the barrier energy, the distribution can cross the barrier. This is a form of "classical tunneling" which provides a mechanism for barrier crossing in simulated annealing optimization employing a density distribution (of the GPP or GDA sort) rather than a single point particle.

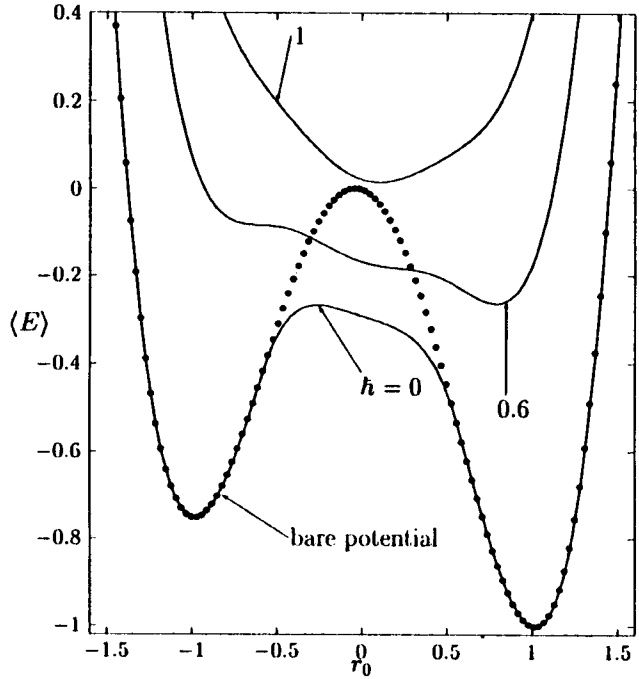

Fig. 12. The effective energy of the quartic double well potential is plotted for three values of $\hbar = 0$ (classical limit), 0.6 and 1 in Lennard-Jones reduced units. The effective energy is determined as the sum of the kinetic energy, $d\hbar^2/8mM_2$, and the effective potential, $\langle V(r_0, M_2^{ad})\rangle$, using the optimal value of M_2 at each value of r_0.

7. Sampling with Monte Carlo or Molecular Dynamics

In this review, I have tried to emphasize what a variety of successful and promising optimization strategies have in common. I have also stated what I think are the important differences. Occasionally, I have referred to the usefulness of having a deterministic set of equations of motion to integrate, as opposed to a function to sample using Monte Carlo. Also, some of the techniques presented involve some form of molecular dynamics or Monte Carlo calculation. As such, it seems worthwhile to mention some important developments in MC and MD methodology.

A great deal of work has been done since the first study of the effectiveness of Monte Carlo sampling for proteins.[130] Currently, there is some disagreement over the effectiveness of Brownian dynamics,[67] noisy molecular dynamics,[68] or molecular dynamics[131] when applied to large organic molecules and proteins.[132–135] Some evidence supports the use of improved Monte Carlo sampling using internal coordinate (torsion angle) moves[122,123] and Monte Carlo schemes coupled with energy minimization[136] which have led to significantly improved results over simulated annealing strategies[33] in applications to enkephalin.[136] However, most of this discussion has occurred before recent improvements in the numerical efficiency of MC and MD calculations on proteins.

A problem facing MC or MD simulations of complex systems is related to "broken ergodicity." The presence of high energy barriers separating regions of configuration space can cause a trajectory (Monte Carlo or molecular dynamics) to be locked into one phase of the system.[13] It may well be that there are phases of lower free energy meaning that the simulated phase is a supercooled state of the system.[137] A number of advances in Monte Carlo methodology which address the problem of broken ergodicity[138] have been reported in recent years. Cao and Berne have developed an "anti-force bias" method which helps to accelerate barrier crossing.[139] Ferrenberg and Swendsen,[140] and Frantz, Freeman and Doll[141] have proposed methods which use a high temperature MC run, where barriers may be crossed easily, overcoming problems of broken ergodicity, to compute averages at a lower temperature of interest. These latter methods have been combined by Tsai and Jordan to examine phase changes in small rare gas and water clusters.[142] Extended electrostatics techniques,[143,144] like potential smoothing methods, enhance both MD and MC calculation.

More recently, Andricioaei and Straub[156] have developed an enhanced sampling algorithm which uses a Monte Carlo protocol to explore a generalized statistical distribution of the form proposed by Tsallis

$$p(r) \propto [1 - (1 - q)\beta V(r)]^{\frac{1}{1-q}}$$

where $V(r)$ is the potential energy and "q" is a parameter. When $q \to 1$ the standard Gibbs–Boltzmann statistics are recovered. Otherwise, it has been shown that for $q > 1$ the sampling is greatly enhanced as is the probability of locating the global energy minimum.[156] The mechanism underlying the enhanced sampling is a dramatic increase in the probability of being in a barrier region relative to being in a minimum. The result is an increase in the rate of barrier crossing and sampling phase space. This is similar to the effect of tunneling in quantum annealing methods and barrier lowering due to potential smoothing.

There have also been significant advances in molecular dynamics methods. Multiple time step algorithms[145] can enhance molecular dynamics calculations on proteins by at least a factor of 4 to 5 over traditional integration methods.[146] This should be important not only for extending simulation times in the study of dynamics, but also in the use of MD for configurational space sampling.

Furthermore, molecular dynamics in higher dimensions, which provides a mechanism for lowering barriers, is akin to potential smoothing.[147] The transform from three to four dimensions appears as

$$r^2 \to r^2 + w^2 = x^2 + y^2 + z^2 + w^2 \tag{63}$$

and is easily implemented in a standard molecular dynamics calculation. A boundary potential may be placed on w to control its fluctuation. Fluctuations of the parameter w in the fourth dimension can help to lower barriers in the system by acting as an effective "random smoothing" of the potential in three dimensions. This method looks a great deal like the potential shift algorithm of Piela and coworkers mentioned previously, and has some conceptual similarity to distance geometry[28,148] and energy embedding procedures.[149,150] Similar effects have been achieved through coupling of the system to an "energy bath" which has a well-defined reference potential energy.[151]

Clearly, there are many possibilities involving both MD and MC techniques, combined with some form of potential smoothing and modern algo-

rithms for integration and updating of nonbonded interaction lists, which will lead to improved algorithms for enhanced sampling of proteins.

8. Future Directions

There are many important directions for future research into optimization algorithms for proteins. This chapter was written to give an overview, however incomplete, of a number of areas where progress has been made and where improvements are likely to follow. In addition to many areas in the field of optimization not discussed in this chapter, I believe there is no method discussed here which can be described as finished or complete. As such, there remain many interesting unanswered questions.

The work discussed in this chapter points to the importance of potential smoothing techniques in molecular optimization algorithms. While smoothing methods may be applied *ad hoc* to take advantage of the special properties of a given problem, it is also possible to develop algorithms which represent the system using continuous classical density distributions or quantum mechanical wave functions which display the positive properties of smoothing methods. The quantum mechanical annealing methods have the further advantage of being able to tunnel to a lower energy minimum, although algorithms based on the purely classical density distribution display generalized "tunneling" as well — an ensemble of states may pass over a barrier whose energy is higher than the mean energy of the ensemble.

The GDA algorithm, which is essentially an annealing algorithm but which evolves in temperature rather than time, has the desirable property of being free of a cooling schedule. This property is shared by the quantum mechanical algorithms which evolve in imaginary time. It is easy to imagine a variety of quantum mechanical annealing methods beyond the wave packet and DMC methods discussed here and these methods seem well worth exploring.

Some of the algorithms discussed are deterministic and their results are strongly dependent on the choice of boundary conditions. An important area for future work on methods such as the DEM and aGDA algorithms is in specifying improved boundary conditions and understanding in detail their link to the final minimum energy structures isolated. Other methods such as the GDA and LES algorithms consist of deterministic equations of motion which evolve structures from a random set of initial conditions. In this way these algorithms have a non-deterministic component — the

choice of initial conditions. This randomness can be seen as a complication, requiring more effort to solve the set of equations, but can also be of some value in providing a distribution of low energy structures.

An important question concerns the relative importance of new potentials for proteins and optimization algorithms.[15] Clearly, the development of coarse grained or minimal models of proteins is an important direction of research. These model potentials have the benefit of being considerably simplified when compared to all atom models. At the same time, the optimization problem remains, even when using model potentials. Developments in each area will represent important contributions to the effort to identify effective algorithms to explore protein folding.

Acknowledgments

I have enjoyed discussions with my students Patricia Amara and Jianpeng Ma who have made many contributions to the work summarized in this chapter. Special thanks to David Hsu for his numerous insightful observations and helpful comments. Thanks to Bruce Berne, Anders Wallqvist and Dave Thirumalai for comments on this manuscript. I thank Harold Scheraga for providing me with copies of Figs. 10 and 11 and the American Chemical Society for permission to use them. This work was supported in part by a grant from the donors of the Petroleum Research Fund, administered by the American Chemical Society, and a grant from the National Science Foundation (CHE-9306375).

References

1. C. Levinthal, in *Mossbauer Spectroscopy in Biological Systems* eds. P. Debrunner, J. C. M. Tsibris and E. Münck (Univ. of Illinois Press, Urbana-Champaign, 1969) p. 22.
2. G. D. Fasman, *Prediction of Protein Structure and the Principles of Protein Conformation* (Plenum, New York, 1989).
3. T. E. Creighton, *Protein Folding* (Freeman, New York, 1992).
4. N. Gō, *Ann. Rev. Biophys. Bioeng.* **12**, 183 (1983), N. Gō and H. Abe, *Adv. Biophys.* **18**, 149 (1984).
5. J. D. Bryngelson and P. G. Wolynes, *J. Phys. Chem.* **93**, 6902 (1989); J. D. Bryngelson and P. G. Wolynes, *Biopolymers* **30**, 177 (1990), *Proc. Natl. Acad. Sci. USA* **84**, 7524 (1987).
6. Z. Guo, D. Thirumalai and J. D. Honeycutt, *J. Chem. Phys.* **97**, 525 (1992).
7. M. Karplus and E. Shakhovich, in *Protein Folding*, ed. T. E. Creighton (Freeman, New York, 1992).

8. I will not attempt to provide a summary of Scheraga's many contributions to this problem. Fortunately, he has written a number of clear and recent reviews which I can recommend as a summary of the many interesting approaches to the global energy minimization of proteins which do not fit the scope of this chapter. See H. A. Scheraga, *Rev. Comp. Chem. III*, 73 (1992), eds. K. B. Lipkowitz and D. B. Boyd; *Theor. Biochem. Mol. Biophys.* **2**, 231 (1991); *Chemica Scripta* **29A**, 3 (1989); *Ann. N.Y. Acad. Sci.* **439**, 170 (1985).
9. M. B. Smith, *Aust. J. Biol. Sci.* **17**, 261 (1964); M. B. Smith and J. F. Back, *ibid.* **18**, 365 (1965); **21**, 539 (1968); **21**, 549 (1968).
10. H. Frauenfelder, S. G. Sligar and P. G. Wolynes, *Science* **254**, 1598 (1991); H. Frauenfelder, F. Parak and R. D. Young, *Ann. Rev. Biophys. Chem.* **17**, 451 (1988).
11. R. Elber and M. Karplus, *J. Am. Chem. Soc.* **112**, 9161 (1990); R. Czerminski and R. Elber, *Proteins* **10**, 70 (1991).
12. T. Noguti and N. Gō, *Proteins* **5**, 97 (1989); *ibid.* **5**, 104 (1989); *ibid.* **5**, 113 (1989); *ibid.* **5**, 125 (1989); *ibid.* **5**, 132 (1989).
13. J. E. Straub, A. Rashkin and D. Thirumalai, *J. Am. Chem. Soc.* **116**, 2049 (1994); J. E. Straub and D. Thirumalai, *Proc. Natl. Acad. Sci. USA* **90**, 809 (1993); J. E. Straub and D. Thirumalai, *Proteins* **15**, 360 (1993).
14. C. Camacho and D. Thirumalai, *Proc. Natl. Acad. Sci. USA* **90**, 6369 (1993).
15. A. Sali, E. Shakhnovich and M. Karplus, *Nature* **369**, 248 (1994).
16. H. Frauenfelder, S. G. Sligar and P. G. Wolynes, *Science* **254**, 1598 (1991).
17. E. I. Shakhnovich and A. M. Gutin, *Biophys. Chem.* **34**, 187 (1989).
18. E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines* (Wiley, New York, 1990).
19. The term $\mathcal{NP}$-complete stands for "non-deterministic polynomial time" referring to the fact that problems in this class do not have solutions which can be found using a deterministic search algorithm in a time which scales as a polynomial of N, where N is a measure of the size of the problem.
20. P. G. Wolynes, in *Biologically Inspired Physics*, ed. L. Peliti (Plenum Press, New York, 1991) p. 15.
21. E. L. Lawler, J. K. Leustra, A. H. G. Rinnooy Kan and K. B. Shmoys, *The Travelling Salesman Problem* (Wiley, New York, 1985).
22. G. Baskaran, Y. Fu and P. W. Anderson, *J. Stat. Phys.* **45**, 1 (1986).
23. Nevertheless, as Scheraga has stated "It is an article of faith that we are looking for the global minimum." See H. A. Scheraga, *Alfred Benzon Symp.* **28**, 343 (1989).
24. S. Kirkpatrick, Jr., C. D. Gelatt and M. P. Vecchi, *Science* **220**, 671 (1983).
25. R. Czerminski and R. Elber, *Int. J. Quant. Chem.* **24**, 167 (1990); R. Elber and M. Karplus, *Chem. Phys. Lett.* **139**, 375 (1987).
26. R. Czerminski and R. Elber, *Proc. Natl. Acad. Sci. USA* **86**, 6963 (1989); R. Czerminski and R. Elber, *J. Chem. Phys.* **92**, 5580 (1990).
27. S. Fischer and M. Karplus, *Chem. Phys. Lett.* **96**, 5272 (1992).

28. T. F. Havel and K. Wütrich, *Bull. Math. Biol.* **46**, 673 (1984); *J. Mol. Biol.* **182**, 281 (1985).
29. G. M. Clore, A. M. Gronenborn, A. T. Brünger and M. Karplus, *J. Mol. Biol.* **186**, 435 (1985).
30. L. Nilsson, G. M. Clore, A. M. Gronenborn, A. T. Brünger and M. Karplus, *J. Mol. Biol.* **188**, 455 (1986).
31. A. T. Brünger, G. M. Clore, A. M. Gronenborn and M. Karplus, *Proc. Natl. Acad. Sci. USA* **83**, 3801 (1986).
32. H. J. Kushner, *SIAM J. Appl. Math.* **47**, 169 (1987).
33. A. Nayeem, J. Vila and H. A. Scheraga, *J. Comput. Chem.* **12**, 594 (1991).
34. R. Brower, G. Vasmatzis, M.. Silevrman and C. Delisi, *Biopolymers* **33**, 329 (1993).
35. M. R. Hoare and J. A. McInnes, *Adv. in Phys.* **32**, 791 (1983).
36. F. H. Stillinger and T. Weber, *J. Stat. Phys.* **52**, 1429 (1988). The effect on the number of minima of coarse graining the potential hypersurface was discussed earlier in F. H. Stillinger, *Phys. Rev.* B**32**, 3134 (195).
37. A recent study of C. J. Camacho and D. Thirumalai, *Phys. Rev. Lett.* **71** 2505 (1993), suggests that while the number of compact states of a protein grows exponentially with the number of amino acids, the number of *minimum energy* compact states is far smaller. They suggest that these minimum energy states act as basins of attraction in the intermediate stages of protein folding, and that the native state is reached from one of these minimum energy compact states.
38. F. H. Stillinger and D. K. Stillinger, *J. Chem. Phys.* **93**, 6106 (1990).
39. P. A. Braier, R. S. Berry and D. J. Wales, *J. Chem. Phys.* **93**, 8745 (1990).
40. R. S. Berry, *Chem. Rev.* **93**, 2379 (1993).
41. P. E. Gill, W. Murray and M. H. Wright, *Practical Optimization* (Academic Press, San Diego, 1981).
42. D. A. Pierre, *Optimization Theory with Applications* (Dover, New York, 1986).
43. T. Schlick, in *Rev. Comp. Chem.*, Vol. III, eds. K. B. Lipkowitz and D. B. Boyd, (VCH, New York, 1992) p. 1.
44. S. A. Teukolsky, W. H. Press, B. P. Flannery and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge Univ. Press, Cambridge, 1986).
45. M. R. Hoare and P. Pal, *Adv. Phys.* **20**, 161 (1971).
46. J. A. Northby, *J. Chem. Phys.* **87**, 6166 (1987).
47. C. D. Maranas and C. A. Floudas, *J. Chem. Phys.* **97**, 7667 (1992).
48. The focus of this chapter, which is methods for global energy minimization of macromolecules, is a subset of the large class of hard optimization problems and as such the "general" methods discussed in the remaining sections may appear to many to be tailored to problems of molecular optimization. However, all of those methods are in fact easily generalized to problems such as that of the traveling salesman.

49. J. Pillardy, K. A. Olszewski and L. Piela, *J. Phys. Chem.* **96**, 4337 (1992).
50. M. Dygert. N. Gō and H. A. Scheraga, *Macromol.* **8**, 750 (1975); K. A. Palmer and H. A. Scheraga, *J. Comp. Chem.* **12**, 505 (1991); *ibid.* **13**, 329 (1992).
51. Q. Zheng, R. Rosenfeld, S. Vajda and C. DeLisi, *J. Comput. Chem.* **14**, 556 (1993).
52. Q. Zheng, R. Rosenfeld, S. Vajda and C. DeLisi, *Protein Science* **2**, 1242 (1993).
53. R. Rosenfeld, Q. Zheng, S. Vajda and C. DeLisi, *J. Mol. Biol.* **234**, 515 (1993).
54. Z. Liu and B. J. Berne, *J. Chem. Phys.* **99**, 6071 (1993).
55. H. C. Andersen, *J. Chem. Phys.* **72**, 2384 (1980).
56. S. Nosé, *J. Chem. Phys.* **81**, 511 (1984).
57. W. G. Hoover, *Phys. Rev.* **A31**, 1695 (1985).
58. Similar effects are seen in the difference in structure between classical and quantum mechanical representations of atomic fluids using path integral methods or quantum effective potentials of the Feynman-Hibbs or Wigner-Kirkwood form. See D. Thirumalai, R. Hall and B. J. Berne, *J. Chem. Phys.* **81**, 2523 (1984). As we will see in Sec. 5, the use of a quantum effective potential which allows for tunneling can lead to effective global optimization routines.
59. T. Head-Gordon, F. H. Stillinger and J. Arrecis, *Proc. Natl. Acad. Sci.* **88**, 11076 (1991).
60. T. Head-Gordon and F. H. Stillinger, *Biopolymers* **33**, 293 (1993).
61. J. R. Gunn, A. Monge, R. A. Friesner and C. H. Marshall, *J. Phys. Chem.* **98**, 702 (1994).
62. M. Karplus and D. Weaver, *Biopolymers* **18**, 1421 (1979); D. Bashford, D. L. Weaver and M. Karplus, *J. Biomol. Struct. Dyn.* **1**, 1243; D. Bashford, F. E. Cohen, M. Karplus, I. D. Kuntz and D. L. Weaver, *Proteins* **4**, 211 (1988).
63. D. Vanderbilt and S. G. Louie, *J. Comput. Phys.* **56**, 259 (1984).
64. L. T. Wille, *Letters to Nature* **324**, 46 (1986).
65. L. T. Wille, *Chem. Phys. Lett.* **133**, 405 (1987).
66. J. D. Honeycutt and D. Thirumalai, *Proc. Natl. Acad. Sci. USA* **87**, 3526 (1990).
67. J. Rey and J. Skolnick, *J. Chem. Phys.* **158**, 199 (1991).
68. J. D. Honeycutt and D. Thirumalai, *Biopolymers* **32**, 695 (1992).
69. R. C. Tolman, *The Principles of Statistical Mechanics* (Dover, New York, 1979).
70. D. A. McQuarrie, *Statistical Mechanics* (Harper and Row, New York, 1976).
71. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford, Bristol, 1990).
72. B. J. Berne and R. Pecora, *Dynamic Light Scattering* (Wiley-Interscience, New York, 1976).

73. R. W. Zwanzig, in *Lectures in Theoretical Physics*, Vol. III, eds. W. E. Britton, B. W. Downs and J. Downs (Wiley Interscience, New York, 1961) p. 106; see also R. Zwanzig, *J. Chem. Phys.* **33**, 1338 (1960).
74. J. Ma, D. Hsu and J. E. Straub, *J. Chem. Phys.* **99**, 4024 (1993).
75. J. Grad, J. Y. Yan and S. Mukamel, *Chem. Phys. Lett.* **134**, 219 (1987); J. Grad, Y. J. Yan, A. Haque and S. Mukamel, *J. Chem. Phys.* **86**, 3441 (1987).
76. S. Mukamel, *J. Phys. Chem.* **88**, 3185 (1984).
77. R. D. Coalson and M. Karplus, *J. Chem. Phys.* **79**, 6150 (1983); *ibid.* **81**, 2891 (1984); *ibid.* **93**, 3919 (1990). See also A. D. McLachlan, *Mol. Phys.* **8**, 39 (1964); E. J. Heller, *J. Chem. Phys.* **64**, 63 (1976).
78. J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Academic Press, London, 1986).
79. R. A. Laviolette and F. H. Stillinger, *J. Chem. Phys.* **83**, 4079 (1985).
80. G. Némethy, K. D. Gibson, K. A. Palmer, C. N. Yoon, G. Paterlini, A. Zagari, S. Rumsey and H. A. Scheraga, *J. Phys. Chem.* **96**, 6472 (1992); I. K. Roterman, M. H. Lambert, K. D. Gibson and H. A. Scheraga, *J. Biomol. Struct. Dyn.* **7**, 421 (1989).
81. F. A. Momany, R. F. McGuire, A. W. Burgess and H. A. Scheraga; *J. Phys. Chem.* **79**, 2361 (1975); G. Nementhy, M. S. Pottle and H. A. Scheraga, *ibid.* **87**, 1883 (1983); M. J. Sippl, G. Nementhly and H. A. Scheraga, *ibid.* **88**, 6231 (1984).
82. C. L. Brooks, M. Karplus and M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics* (John Wiley and Sons, New York, 1988).
83. J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids* (Cambridge Univ. Press, New York, 1987).
84. The classical trajectory-bundle TDSCF method provides greater flexibility than Gaussian wavepackets at the cost of introducing more degrees of freedom. See R. B. Gerber and M. A. Ratner, *Adv. Chem. Phys.* **70**, 97 (1988); G. C. Schatz, V. Buch, M. A. Ratner and R. B. Gerber, *J. Chem. Phys.* **79**, 1808 (1983); V. Buch, R. B. Gerber and M. A. Ratner, *Chem. Phys. Lett.* **101**, 44 (1983); R. B. Gerber, V. Buch and M. A. Ratner, *J. Chem. Phys.* **77**, 3022 (1982); *Chem. Phys. Lett.* **91**, 173 (1982).
85. H. A. Scheraga, *Adv. Phys. Org. Chem.* **6**, 103 (1968).
86. D. J. Evans, W. G. Hoover, B. H. Failor, B. Moran and A. J. C. Ladd, *Phys. Rev.* **A28**, 1016 (1983).
87. W. G. Hoover, A. J. C. Ladd and B. Moran, *Phys. Rev. Lett.* **48**, 1818 (1982); D. J. Evans, *J. Chem. Phys.* **78**, 3297 (1983).
88. H. Risken, *The Fokker-Planck Equation* (Springer-Verlag, Berlin, 1989).
89. J. L. Skinner and P. G. Wolynes, *J. Chem. Phys.* **69**, 2143 (1978); *J. Chem. Phys.* **72**, 4913 (1980).
90. B. J. Berne, J. L. Skinner and P. G. Wolynes, *J. Chem. Phys.* **73**, 4314 (1980).

91. J. Ma and J. E. Straub, *J. Chem. Phys.* **101**, 533 (1994): *ibid.* **103**, 9113 (1995).
92. G. Verkhiver, R. Elber and W. Nowak, *J. Chem. Phys.* **97**, 7838 (1992).
93. J. E. Straub and M. Karplus, *J. Chem. Phys.* **94**, 6737 (1991).
94. A. Ulitsky and R. Elber, *J. Chem. Phys.* **98**, 1034 (1994); *ibid.* **98**, 3380 (1994).
95. A. Miranker and M. Karplus, *Proteins: Structure, Function, and Genetics* **11**, 29 (1991).
96. A. Caflisch, A. Miranker and M. Karplus, *J. Med. Chem.* **36**, 2142 (1993).
97. A. Roitberg and R. Elber, *J. Chem. Phys.* **95**, 9277 (1991).
98. N. Corbin and K. Singer, *Mol. Phys.* **46**, 671 (1982); K. Singer and W. Smith, *Mol. Phys.* **57**, 761 (1986) for Gaussian wavepackets applied to the quantum simulation of neon atoms in real time.
99. The unnormalized equilibrium density distribution for the canonical ensemble $\hat{\rho}(r,p,\beta) = \exp(-\beta H)$ satisfies the Bloch equation

$$\frac{\partial \hat{\rho}}{\partial \beta} = -H\hat{\rho}. \tag{64}$$

[See R. W. Zwanzig, in *Lectures in Theoretical Physics*, Vol. III, eds. W. E. Britton, B. W. Downs and J. Downs (Wiley Interscience, New York, 1961) p. 106.] Integrating the normalized distribution avoids the numerical difficulties associated with integration of unnormalized quantities, such as $\hat{\rho}(r;\beta)$, which become vanishingly small for large values of β.
100. P. Amara and J. E. Straub, **99**, 14840 (1995).
101. L. Piela, J. Kostrowicki and H. A. Scheraga, *J. Phys. Chem.* **93**, 3339 (1989).
102. J. Kostrowicki, L. Piela, B. J. Cherayil and H. Scheraga, *J. Phys. Chem.* **95**, 4113 (1991).
103. R. J. Wawak, M. M. Wimmer and H. A. Scheraga, *J. Phys. Chem.* **96**, 5138 (1992).
104. H. A. Scheraga, *Rev. Comp. Chem*, **III**, 73 (1992).
105. J. Kostrowicki, L. Piela, B. J. Cherayil and H. A. Scheraga, *J. Phys. Chem.* **95**, 4113 (1991); J. Kostrowicki and H. A. Scheraga, *J. Phys. Chem.* **96**, 7442 (1992).
106. C. Tsoo and III C. L. Brooks, *J. Chem. Phys.* **101**, 6405 (1994).
107. P. Amara, D. Hsu and J. E. Straub, *J. Phys. Chem.* **97**, 6715 (1993).
108. D. Shalloway, in *Recent Advances in Global Optimization*, eds. C. A. Floudas and P. M. Pardalos, (Princeton Univ. Press, Princeton, 1992).
109. D. Shalloway, *Global Optimization* **2**, 281 (1992); M. Orešič and D. Shalloway, *J. Chem. Phys.* **101**, 6405 (1994).
110. T. Coleman, D. Shalloway and Z. Wu, *Technical Report: Cornell Theory Center* (November 1992).
111. L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1968).
112. R. L. Somorjai, *J. Phys. Chem.* **95**, 4141 (1991).
113. M. Sylvain and R. L. Somorjai, *J. Phys. Chem.* **95**, 4147 (1991).

114. K. A. Olszewski and H. A. Scheraga, *J. Phys. Chem.* **96**, 4672 (1992).
115. K. A. Olszewski, L. Piela and H. A. Scheraga, *J. Phys. Chem.* **97**, 260 (1993).
116. K. A. Olszewski, L. Piela and H. A. Scheraga, *J. Phys. Chem.* **97**, 267 (1993).
117. A. Messiah, *Quantum Mechanics* (Wiley and Sons, New York, 1976).
118. J. B. Anderson, *J. Chem. Phys.* **63**, 1499 (1975); *ibid.* **65**, 4121 (1976); *Int. J. Quantum Chem.* **15**, 109 (1979).
119. D. Ceperley and B. Alder, *Science* **231**, 555 (1986).
120. D. L. Freeman and J. D. Doll, *J. Chem. Phys.* **82**, 462 (1985).
121. A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson and J. D. Doll, *Chem. Phys. Lett.* **219**, 343 (1994).
122. T. Noguti and N. Gō, *Biopolymers* **24**, 527 (1985).
123. A. Kitao and N. Gō, *J. Comp. Chem.* **12**, 359 (1991).
124. E. J. Heller, *J. Chem. Phys.* **62**, 1544 (1975); E. J. Heller, *Acc. Chem. Res.* **14**, 368 (1981).
125. R. P. Feynman and H. Kleinert, *Phys. Rev.* A**34**, 5080 (1986).
126. A. D. McLachlan, *Mol. Phys.* **8**, 39 (1964).
127. R. P. Feynman, *Quantum Mechanics and Path Integrals* (Wiley, New York, 1965).
128. C. Zheng, C. F. Wong, J. A. McCammon and P. G. Wolynes, *Nature* **334**, 726 (1988).
129. P. Ruján, *Z. Phys. B - Condensed Matter* **73**, 391 (1988).
130. S. H. Northrup and J. A. McCammon, *Biopolymers* **19**, 1001 (1980).
131. Y. Sun and P. A. Kollman, *J. Comp. Chem.* **13**, 33 (1992).
132. M. Saunders, K. N. Houk, Y. -D. Wu, W. C. Still, M. Lipton, G. Chang and W. C. Guida, *J. Am. Chem. Soc.* **112**, 1419 (1990).
133. T. Schaumann, W. Braun and K. Wüthrich, *Biopolymers* **29**, 679 (1990).
134. B. Von Freyberg and W. Braun, *J. Comp. Chem.* **12**, 1065 (1991).
135. D. G. Garrett, K. Kastella and D. M. Ferguson, *J. Am. Chem. Soc.* **114**, 6555 (1992).
136. Z. Li and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **84**, 6611 (1987).
137. J. P. Valleau and S. G. Whittington, in *Statistical Mechanics, Part A*, ed. B. J. Berne, (Plenum Press, New York, 1976).
138. J. I. Siepmann and M. Sprik, *Chem. Phys. Lett.* **199**, 220 (1992).
139. J. Cao and B. J. Berne, *J. Chem. Phys.* **92**, 1980 (1990).
140. A. M. Ferremberg and R. H. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988).
141. D. D. Frantz, D. L. Freeman and J. D. Doll, *J. Chem. Phys.* **93**, 2769 (1990).
142. C. J. Tsai and K. D. Jordan, *J. Chem. Phys.* **99**, 6957 (1993).
143. R. H. Stote, D. J. States and M. Karplus, *J. Chim. Phys.* **88**, 2419 (1991).
144. H. -Q. Ding, N. Karasawa and W. A. Goddard III, *J. Chem. Phys.* **97**, 4309 (1992).
145. M. E. Tuckerman, G. J. Martyna and B. J. Berne, *J. Chem. Phys.* **93**, 1287 (1990); M. E. Tuckerman, B. J. Berne and A. Rossi, *J. Chem. Phys.* **94**,

1465 (1991); M. E. Tuckerman, B. J. Berne and G. J. Martyna, *J. Chem. Phys.* **94**, 6811 (1991).
146. D. D. Humphreys, R. A. Friesner and B. J. Berne, *J. Phys. Chem.* **98**, 6885 (1994).
147. R. C. van Schaik, H. J. C. Berendsen, A. E. Torda and W. F. van Gunsteren, *J. Mol. Biol.* **234**, 751 (1993).
148. L. M. Blumenthal, in *Theory and Applications of Distance Geometry* (Chelsea, New York, 1970) p. 97; G. M. Crippen, *Distance Geometry and Conformational Calculations* (Research Studies Press, England, 1981).
149. G. M. Crippen, *J. Comp. Chem.* **3**, 471 (1982); *J. Comp. Chem.* **5** 548 (1984).
150. E. O. Purisima and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **84**, 2782 (1986); *J. Mol. Biol.* **196**, 697 (1987).
151. R. C. van Schaik, W. F. van Gunsteren and H. J. C. Berendsen, *J. Computer-Aided Mol. Design* **6**, 97 (1992).
152. J. Kostrowicki and H. A. Scheraga, *J. Phys. Chem.* **96**, 7442 (1992).
153. J. E. Straub, J. Ma and P. Amara, *J. Chem. Phys.* **103**, 1574 (1995).
154. P. Amara, J. Ma and J. E. Straub, in *Global Minimization of Nonconvex Energy Functions: Molecular Conformation and Protein Folding*, eds. P. M. Pardalos, D. Shalloway and G. Xue (American Mathematical Society, 1996), p. 41.
155. B. Chorch, M. Orešič and D. Shalloway, in *Global Minimization of Nonconvex Energy Functions: Molecular Conformation and Protein Folding*, eds. P. M. Pardalos, D. Shalloway and G. Xue (American Mathematical Society, 1996), p. 41.
156. I. Andricioaei and J. E. Straub, *Phys. Rev.* **E53**, 3055 (1996).