JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 12, Number 10, 2005

Pp. 1275–1288

HOPE: A Homotopy Optimization Method for Protein Structure Prediction

DANIEL M. DUNLAVY,[1] DIANNE P. O'LEARY,[2] DMITRI KLIMOV,[3]
and D. THIRUMALAI[4]

ABSTRACT

We use a homotopy optimization method, HOPE, to minimize the potential energy associated with a protein model. The method uses the minimum energy conformation of one protein as a template to predict the lowest energy structure of a query sequence. This objective is achieved by following a path of conformations determined by a homotopy between the potential energy functions for the two proteins. Ensembles of solutions are produced by perturbing conformations along the path, increasing the likelihood of predicting correct structures. Successful results are presented for pairs of homologous proteins, where HOPE is compared to a variant of Newton's method and to simulated annealing.

Key words: protein structure prediction, energy minimization, global optimization, homotopy method, simulated annealing.

1. INTRODUCTION

GIVEN AN ENERGY FUNCTION FOR A POLYPEPTIDE CHAIN, the task is to determine the native structure that corresponds to the lowest energy. Potential energy functions for proteins typically have multiple local minima whose number increases exponentially with the number of degrees of freedom (Li and Scheraga, 1987; Wilson *et al.*, 1988). The NP-complete nature of the problem prompted Ngo and Marks (1992) to suggest, "function-minimization algorithms can be efficient for protein structure prediction only if they exploit protein-specific properties."

There are several approaches for determining the native conformation of a protein using energy minimization. Some of the more effective methods include the truncated Newton method (Xie and Schlick, 1999) as well as a combination of the limited memory BFGS quasi-Newton and Hessian-free Newton methods (Das *et al.*, 2003), genetic algorithms (Le Grand and Merz, 1993; Brodmeier and Pretsch, 1994),

[1]Applied Mathematics and Scientific Computation Program, University of Maryland, College Park, MD 20742.
[2]Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742.
[3]Program in Bioinformatics and Computational Biology, School of Computational Sciences, George Mason University, Manassas, VA 20110.
[4]Department of Chemistry and Biochemistry and Biophysics Program, Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742.

smoothing methods (Schelstraete *et al.*, 1998), and simulated annealing (Wilson *et al.*, 1988; Kawai *et al.*, 1989; Wilson and Cui, 1990; Shin and Jhon, 1991). However, few of these methods exploit similarities of sequence-related pairs of proteins.

In this paper, we apply a new method—*homotopy optimization using perturbations and ensembles* (HOPE) (Dunlavy and O'Leary, 2005)—to find the global minimizer of the potential energy function associated with a particular protein model. A *homotopy* function is defined that deforms the potential energy function of a *template* protein into that of the *target* protein whose native conformation is sought. We assume that the native structure of the template sequence is known. Starting with the known structure, HOPE finds local minimizers of the homotopy function as it is deformed into the target protein's potential energy function. To increase the likelihood of convergence to the global minimum, the local minimizers are perturbed at each step in the deformation and used to find nearby structures corresponding to local minima. This procedure creates ensembles of local minimizers of the homotopy function throughout the deformation process.

HOPE is similar to comparative modeling methods (Fiser and Sali, 2003) in that it uses the properties of a template protein to help predict the native conformation of a target protein. In contrast, HOPE in the current implementation uses a single template protein, whereas many comparative modeling methods use pieces of one or more template proteins to help predict the native conformation of a test sequence. HOPE is also related to smoothing methods for energy minimization. In such methods, the deformation starts with a smooth approximation of the template's potential energy function, and typically only a single local minimizer is found. Finally, HOPE can be viewed as a simulated annealing method on an evolving energy landscape defined by the homotopy function, using a constant temperature ($T = 0$) in its annealing schedule and a move class that includes only local minimizers as candidate conformations.

Homotopy methods (Allgower and Georg, 2003) have been used previously for exploring potential energy surfaces and computing stationary points of energy functions (Ackermann and Kliesch, 1998) and for computing optimal configurations of atomic and molecular clusters (Hunjan *et al.*, 2002; Coleman and Wu, 1994). In both instances, standard homotopy functions (convex, fixed point, etc.) were employed, and such functions do not exploit the protein-specific features of potential energy functions. Moreover, the homotopy functions used were smoothing functions for the potentials of a single cluster of atoms. In contrast, we use a homotopy function that deforms the potential energy function of a template protein into that of a target protein. Thus, HOPE takes advantage of the sequence-based and/or structural relationships between proteins in predicting the native structure of the target protein.

In this paper, we use HOPE to predict accurately the native conformations of model proteins. Simulated annealing, parameterized to use more than twice the amount of computational resources to solve the same problem, was unsuccessful. We show that HOPE without the use of perturbations is more successful at predicting the native conformations than a globally convergent variant of Newton's method. We also discuss the prospect of applying HOPE to attain native structures described by more realistic energy functions.

2. PROTEIN MODEL

We use a previously introduced coarse-grained model of proteins (Veitshans *et al.*, 1997) to test the efficacy of HOPE. In the coarse-grained protein model, each amino acid is represented by its α-carbon. Thus, a polypeptide chain is modeled as a chain of particles, where each particle corresponds to an α-carbon atom and models one of three types of residues: hydrophobic, hydrophilic, or neutral. The features of proteins that are most responsible for structural stability are included in the model—hydrophobic forces, van der Waals interactions, and torsional strain—and both bond lengths and bond angles are allowed to be variable. The diversity of hydrophobic species in real proteins is modeled in the interactions between the particles corresponding to hydrophobic residues.

2.1. Potential energy function

Let $X \in \mathbb{R}^{3n}$ denote the Cartesian coordinates of a chain of n particles in three dimensions, with $X_k \in \mathbb{R}^3$ containing the coordinates of the k^{th} particle in the chain. We also use the notation X_k to denote the k^{th}

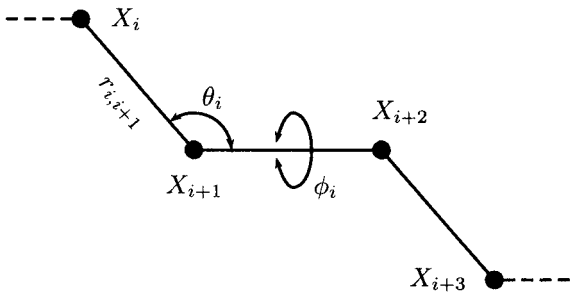


FIG. 1. Geometry of the model protein.

particle. The geometry of a chain is presented in Fig. 1. Four particles, $X_i, \ldots, X_{i+3}$, are depicted as nodes in the figure, with lines between nodes representing the (virtual) bonds between particles.

The distance between particles X_i and X_j is denoted by $r_{ij} = \|X_j - X_i\|_2$. The distance, $r_{i,i+1}$, between consecutive particles X_i and X_{i+1} is called the *bond length* for those particles. The angle $\theta_i \in [0, \pi]$, formed between three consecutive particles, X_i, X_{i+1}, and X_{i+2} is called the *bond angle*. The angle $\phi_i \in [-\pi, \pi]$, formed between the vectors normal to the plane defined by particles X_i, X_{i+1}, and X_{i+2} and that defined by particles X_{i+1}, X_{i+2}, and X_{i+3}, is called the *dihedral angle*. The choice of sign for the dihedral angles conforms to the rules set forth by the IUPAC-IUB (1970)

Each particle is assigned a particle type, p, depending on the type of residue to which it corresponds: hydrophobic (B), hydrophilic (L), or neutral (N). For example, $p_k = B$ if particle X_k represents a hydrophobic residue.

The total potential energy of a chain of particles, $E : \mathbb{R}^{3n} \to \mathbb{R}$, is

$$E(X) = E_{bl}(X) + E_{ba}(X) + E_{dih}(X) + E_{non}(X) \tag{1}$$

where E_{bl}, E_{ba}, E_{dih}, and E_{non} correspond to the bond length, bond angle, dihedral angle, and nonbonded potentials, respectively. The bond length and bond angle potentials are

$$E_{bl}(X) = \sum_{i=1}^{n-1} \frac{k_r}{2} \left(r_{i,i+1} - \bar{r}\right)^2 \tag{2}$$

$$E_{ba}(X) = \sum_{i=1}^{n-2} \frac{k_\theta}{2} \left(\theta_i - \bar{\theta}\right)^2 \tag{3}$$

where $k_r = 100$, $\bar{r} = 1$, $k_\theta = 20/(rad)^2$, and $\bar{\theta} = 105°$.

The dihedral angle potential is taken to be

$$E_{dih}(X) = \sum_{i=1}^{n-3} [A_i(1 + \cos\phi_i) + B_i(1 + \cos 3\phi_i)] \tag{4}$$

where A_i and B_i depend on P_i, the number of neutral particles in the subchain forming the dihedral angle ϕ_i. Denoting $\varepsilon_h = 1$ as the average strength of the hydrophobic interactions, the dihedral potential parameters take the values $A_i = B_i = 1.2\varepsilon_h$ when $P_i \leq 1$; otherwise $A_i = 0$ and $B_i = 0.2\varepsilon_h$. Thus, there is enhanced flexibility in the dihedral angles for chains containing two or more neutral particles.

The nonbonded potential is given by

$$E_{non}(X) = \sum_{i=1}^{n-3} \sum_{j=i+3}^{n} \gamma_{ij} \left\{ \alpha_{ij} \left(\frac{\bar{r}}{r_{ij}}\right)^{12} - \beta_{ij} \left(\frac{\bar{r}}{r_{ij}}\right)^{6} \right\}. \tag{5}$$

TABLE 1. PARAMETERS FOR THE NONBONDED POTENTIAL, $E_{non}(X)$

p_i	p_j	α_{ij}	β_{ij}	γ_{ij}
L	L,B	1	1	$4\varepsilon_L$
N	L,N,B	1	0	$4\varepsilon_L$
B	B	1	-1	$4\nu\varepsilon_h$

The parameters used in each pairwise interaction of particles i and j are given in Table 1, where $\varepsilon_L = \frac{2}{3}\varepsilon_h$ and ν is a dimensionless parameter that is assumed to have Gaussian distribution with a mean value of 1 and a standard deviation of σ. The diversity in the hydrophobic residues is controlled by σ, with $\sigma = 0$ leading to no diversity (i.e., all hydrophobic residues are identical). We assume that interactions between hydrophobic residues are attractive; thus, only positive values of ν are used (more details for the choice of ν are given by Veitshans *et al.* (1997)).

3. ALGORITHMS

HOPE attempts to find the lowest energy conformation of a target chain by producing a sequence of conformations that converges to a local minimizer of $E(X)$. The starting conformation is the lowest energy conformation of a template chain with known structure, whose sequence matches, to some extent, that of the target chain. The HOPE algorithm is derived from a homotopy optimization method.

3.1. Homotopy method for unconstrained minimization

Given the potential energy function of a target chain, denoted by $E^1(X)$, the goal is to solve the unconstrained minimization problem

$$\min_{X \in \mathbb{R}^{3n}} E^1(X) \tag{6}$$

whose solution is denoted by X^*.

The homotopy optimization method (HOM) (Watson and Haftka, 1989) is a continuation-like method (Allgower and Georg, 2003) for solving this problem. Let $\lambda \in [0, 1]$ and let $H(X, \lambda)$ be a function such that

$$H(X, \lambda) = \begin{cases} E^0(X), & \text{if } \lambda = 0, \text{ and} \\ E^1(X), & \text{if } \lambda = 1, \end{cases} \tag{7}$$

where $E^0(X)$ is the potential energy function of a template chain. To simplify the discussion, we assume that the first and second derivatives of H exist and are continuous. HOM, presented in Fig. 2, produces a sequence of points starting at $(X^0, 0)$ and ending at $(X^1, 1)$, where X^1 is an approximation of X^*.

1. Input: $X^{(0)} = X^0$, a local minimizer of $H(X, 0)$; $m \geq 1$
2. $\lambda^{(0)} = 0$; $\Delta\lambda = 1/m$
3. **for** $k = 1, \ldots, m$
4. $\quad \lambda^{(k)} = \lambda^{(k-1)} + \Delta\lambda$
5. $\quad$ minimize $H(X, \lambda^{(k)})$ w.r.t. X, starting at $X^{(k-1)}$, obtaining $X^{(k)}$
6. **end**
7. Output: $X^1 = X^{(m)}$

FIG. 2. HOM Algorithm.

HOM generates points satisfying

$$\nabla_X H(X, \lambda) = 0, \text{ and} \tag{8}$$

$$\nabla_X^2 H(X, \lambda) \text{ is positive semi-definite,} \tag{9}$$

where $\nabla_X H(X, \lambda)$ and $\nabla_X^2 H(X, \lambda)$ denote the first and second partial derivatives of $H(X, \lambda)$ with respect to X. Note that $\nabla_X^2 H(X^1, 1) = \nabla^2 E^1(X^1)$, so X^1 is a local minimizer of $E^1(X)$.

The main difference between HOM and a continuation method is in Step 5, where a continuation method would solve a system of equations. Also, HOM will "jump" over a turning point or a bifurcation point. Such jumps allow the convergence of HOM to a local minimizer of $E^1(X)$ without having to explicitly handle turning or bifurcation points.

Note that HOM is guaranteed to find only a local minimizer of $E^1(X)$ and we are interested in finding the *global minimizer* of $E^1(X)$.

3.2. HOPE

In order to increase the likelihood of finding the global minimizer of $E^1(X)$, we minimize $H(X, \lambda)$ in Step 5 of HOM using an ensemble of starting conformations instead of just a single starting conformation. This extension is the essential difference between HOPE and HOM.

In Step 5 of the HOM algorithm, the next local minimizer in the sequence, $X^{(k)}$, is found via local minimization starting at the previous conformation in the sequence, $X^{(k-1)}$. In the HOPE algorithm, the next ensemble of local minimizers is found via local minimization starting at the $c^{(k-1)}$ conformations in the previous ensemble along with $\hat{c}$ *perturbed versions* of each of those conformations. Since this leads to exponential ensemble growth, at most c_{max} distinct local minimizers are included in the next ensemble. In the end, HOPE produces an ensemble of local minimizers of E^1, from which we choose the one with the lowest function value as the best approximation to the true solution.

The HOPE algorithm is presented in Fig. 3, where the overall structure of HOM is retained. A perturbed version of X is denoted by $\xi(X)$, where $\xi : \mathbb{R}^N \to \mathbb{R}^N$ is a function that stochastically perturbs one or more particle coordinates in X. Note that for $c_{max} = 1$ and $\hat{c} = 0$ HOPE reduces to HOM.

In Steps 11–12 of HOPE, the ensemble of local minimizers to be used in the next iteration is determined. If the number of distinct local minimizers found at the current iteration is less than the maximum ensemble size, then all are used in the next iteration; otherwise, we must choose the "best" subset. What constitutes

1. Input: $X_1^{(0)} = X^0$, a local minimizer of E^0; $m \geq 1$; $c_{max} \geq 1$; $\hat{c} \geq 0$
2. Initialize: $\lambda^{(0)} = 0$; $\Delta\lambda = 1/m$; $c^{(0)} = 1$
3. **for** $k = 1, \ldots, m$
4. $\quad \lambda^{(k)} = \lambda^{(k-1)} + \Delta\lambda$
5. $\quad$ **for** $j = 1, \ldots, c^{(k-1)}$
6. $\quad\quad$ minimize $H(X, \lambda^{(k)})$ w.r.t. X, starting at $X_j^{(k-1)}$, obtaining $X_{j,0}^{(k)}$
7. $\quad\quad$ **for** $i = 1, \ldots, \hat{c}$
8. $\quad\quad\quad$ minimize $H(X, \lambda^{(k)})$ w.r.t. X, starting at $\xi(X_j^{(k-1)})$, obtaining $X_{j,i}^{(k)}$
9. $\quad\quad$ **end**
10. $\quad$ **end**
11. $\quad c^{(k)} = \min\{c^{(k-1)}(\hat{c}+1), c_{max}\}$
12. $\quad X_1^{(k)}, \ldots, X_{c^{(k)}}^{(k)} =$ the $c^{(k)}$ "best" (unique) local minimizers among $X_{j,i}^{(k)}$, $j = 1, \ldots, c^{(k-1)}$, $i = 0, \ldots, \hat{c}$
13. **end**
14. Output: $X^1 =$ the best conformation among $X_j^{(m)}$, $j = 1, \ldots, c^{(m)}$

FIG. 3. HOPE Algorithm.

the best subset may depend on the iteration number (k), the values of the algorithm parameters (m, c_{max}, and $\hat{c}$), or the choice of local minimization routine (along with *its* parameterization).

An obvious measure of what constitutes the best conformations, and the one used in the experiments presented in this paper, is the homotopy function value: conformations with the lowest function values are considered the best. However, there may be other suitable (or perhaps even better) measures depending on the homotopy function used. We have investigated using the amount of conformational change in three dimensions before and after minimization and using the number of minimization iterations required before convergence as alternate measures. However, these measures invariably led to a degradation in the success rate of HOPE for our homotopy functions.

3.3. Homotopy function

We now define the homotopy function, $H(X, \lambda)$, used in HOPE. First, the dihedral potentials of $E^0(X)$ and $E^1(X)$ are partitioned into two terms containing low frequency ($\cos\phi_i$) and high frequency ($\cos 3\phi_i$) terms:

$$E_{dih}(X) = E_{dih1}(X) + E_{dih2}(X) \tag{10}$$

with

$$E_{dih1} = \sum_{i=1}^{n-3} A_i(1 + \cos\phi_i), \text{ and} \tag{11}$$

$$E_{dih2} = \sum_{i=1}^{n-3} B_i(1 + \cos 3\phi_i). \tag{12}$$

Using these new dihedral terms, the homotopy function is

$$\begin{aligned} H(X, \lambda) = {} & E^0_{bl}(X) + E^0_{ba}(X) \\ & + (1 - \rho_1(\lambda))E^0_{dih1}(X) + \rho_4(\lambda)E^1_{dih1}(X) \\ & + (1 - \rho_2(\lambda))E^0_{dih2}(X) + \rho_5(\lambda)E^1_{dih2}(X) \\ & + (1 - \rho_3(\lambda))E^0_{non}(X) + \rho_6(\lambda)E^1_{non}(X) \end{aligned} \tag{13}$$

where $\rho_i(\lambda)$, $i = 1, \ldots, 6$, are continuous weighting functions dependent on the homotopy parameter, λ. To satisfy the conditions that $H(X, 0) = E^0(X)$ and $H(X, 1) = E^1(X)$, these functions must satisfy

$$\rho_i(\lambda) = \begin{cases} 0, & \text{if } \lambda = 0 \\ 1, & \text{if } \lambda = 1 \end{cases}, \; i = 1, \ldots, 6. \tag{14}$$

The convex homotopy, defined using $\rho_i(\lambda) = \lambda$, did not yield good results but was used as a starting point for developing a more useful homotopy. Specifically, we performed computations with HOM to identify modifications to the convex homotopy that increased the success rate of predicting the correct conformations for the target chains.

Figure 4 shows plots of the weighting functions, $\rho_i(\lambda)$, used in the present computations. A convex homotopy deforms E^0_{non} into E^1_{non} in the first half of the homotopy ($\lambda \in [0.0.5]$) so that in the second half of the homotopy ($\lambda \in [0.5, 1]$), E^1_{non} is the only nonbonded potential contribution in H. The template dihedral terms (E^0_{dih1} and E^0_{dih2}) are driven to zero during the first quarter of the homotopy, and the target dihedral terms are not included until the second half of the homotopy. Thus, during the second quarter of the homotopy ($\lambda \in [0.25, 0.5]$) there are no dihedral angle potential contributions in H. This allows the nonbonded interactions to determine all conformational stability. We found this necessary for

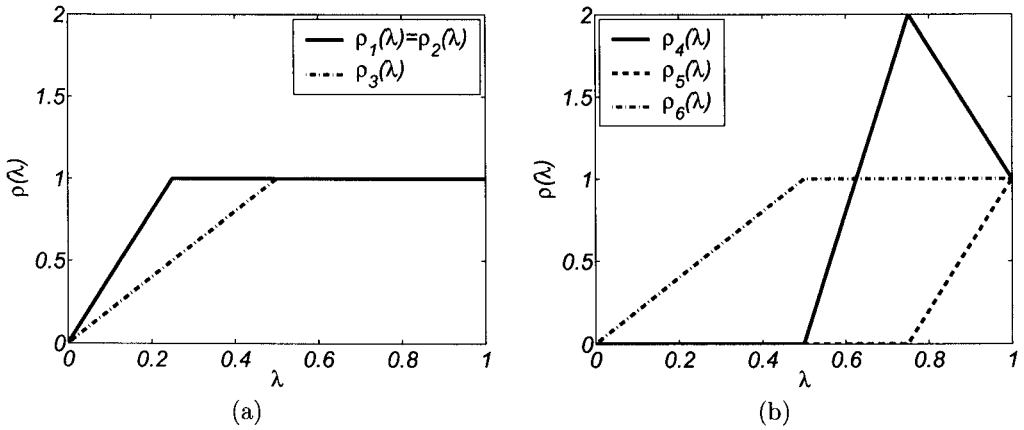


FIG. 4. Plots of the weighting functions used in $H(X, \lambda)$ for (**a**) template terms and (**b**) target terms.

overcoming the large energy barriers in the dihedral potentials for template-target pairs whose lowest energy conformations contain dihedral angles of opposite sign corresponding to the same subchain.

In the second half of the homotopy, contributions from E^1_{dih1} and E^1_{dih2} are introduced into H, but at different rates. We do this because for subchains containing at most one neutral particle, E_{dih} has two local minima with relatively high energy values, and we would like to avoid such minima if possible. Figure 5 shows the contribution of E^1_{dih1} and E^1_{dih2} to H for several values of λ. At $\lambda = 0.5$, the potential is zero. As λ is increased from 0.5 to 0.75, the contribution of the low frequency dihedral terms is increased to $2 \times E^1_{dih1}$. This helps bias towards conformations with $\phi = \pm\pi$, avoiding the local minima of E^1_{dih}. As λ increases from 0.75 to 1, the high frequency terms are gradually included, leading to the true dihedral potential for the target chain.

3.4. Perturbations

During the k^{th} iteration of HOPE, each conformation carried over from the previous iteration is perturbed to produce new conformations. In this section, we describe perturbations based on bond length, bond angle, and individual particle adjustments that have shown promise in our computations. All perturbations occur at particles whose types in the template (p^0) and the target (p^1) do not match.

3.4.1. Bond length perturbations. Starting at one end of the chain, we visit each particle, perturbing the bond length between particles X_k and X_{k+1} if $p^0_{k+1} \neq p^1_{k+1}$. The perturbed bond length between particles X_k and X_{k+1} becomes

$$\hat{r}_{k,k+1} = r_{k,k+1} + \delta r, \tag{15}$$

where δr is taken from a uniform distribution on the interval $[-a_r, a_r]$.

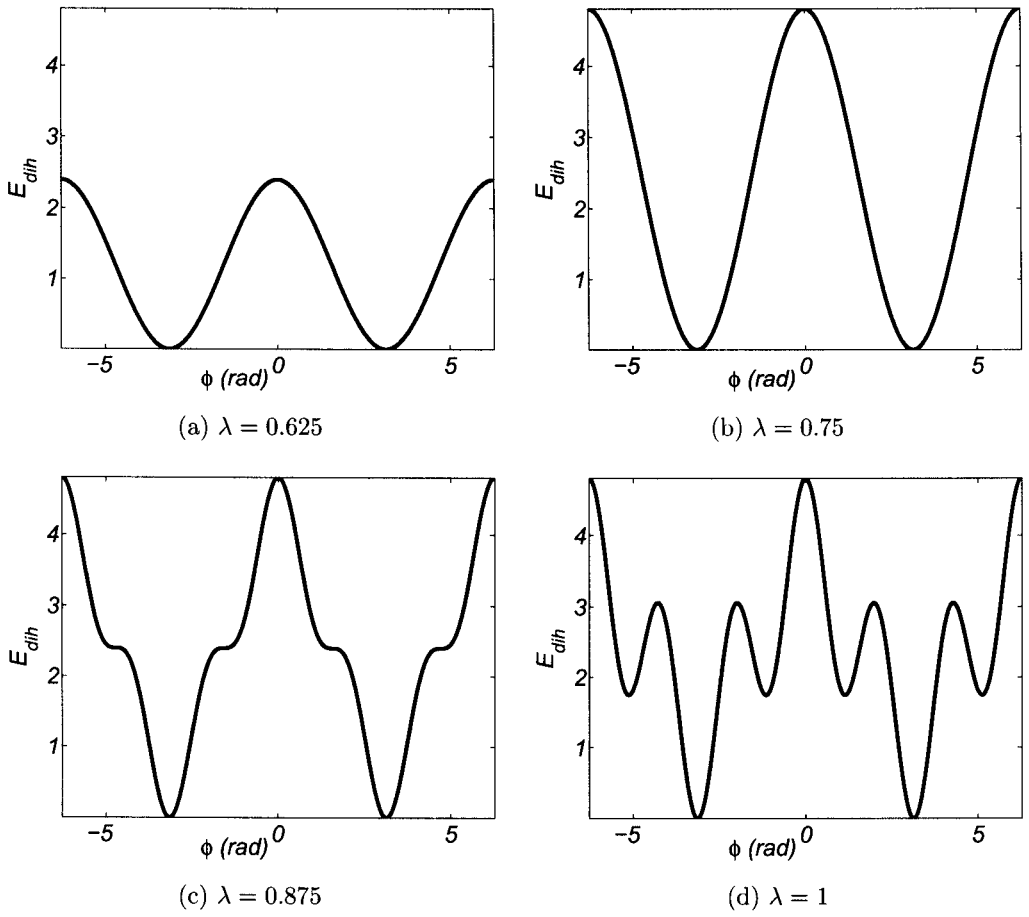
Particle X_{k+1} moves to reflect the new bond length, as shown in Fig. 6. The remainder of the chain either (a) is shifted by δr in the same direction that X_{k+1} moves or (b) remains unchanged.

3.4.2. Bond angle perturbations. Again starting at one end of the chain, we perturb each bond angle θ_k when $p^0_{k+1} \neq p^1_{k+1}$ or $p^0_{k+2} \neq p^1_{k+2}$. The perturbed angle becomes

$$\hat{\theta}_k = \theta_k + \delta\theta, \tag{16}$$

where $\delta\theta$ is taken from a uniform distribution on the interval $[-a_\theta, a_\theta]$.

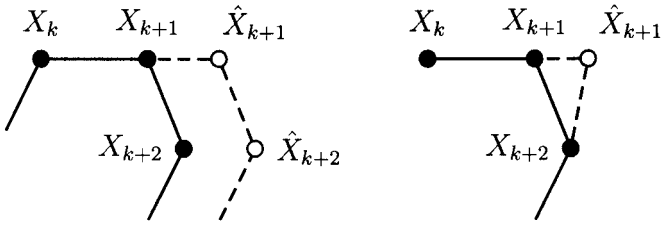
Figure 7 depicts two ways to change the bond angle to $\hat{\theta}_k$: (a) particle X_{k+2} and the remaining particles in the chain are rotated; and (b) particle X_{k+1} moves along along a ray bisecting the angle θ_k in the plane of particles X_k, X_{k+1}, and X_{k+2} and no other particles are shifted.



(a) $\lambda = 0.625$ (b) $\lambda = 0.75$

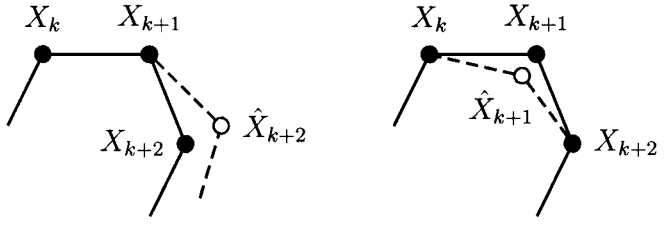
(c) $\lambda = 0.875$ (d) $\lambda = 1$

FIG. 5. Plots of the dihedral potential in $H(X, \lambda)$ for subchains with at most one neutral particle.



(a) (b)

FIG. 6. Perturbations based on bond length adjustments.



(a) (b)

FIG. 7. Perturbations based on bond angle adjustments.

TABLE 2. SEQUENCES (B, HYDROPHOBIC; L, HYDROPHILIC; AND N, NEUTRAL) AND LOWEST ENERGY VALUES OF CHAINS USED IN COMPUTATIONS

Label	*Sequence*	*Minimum Energy*
A	BBBBBBBBBNNNLBLBLBLBLB	−10.6509
B	BBBBBBBBNLNLNBLBBBLBLB	−10.9834
C	BBBBBBBBNLNNLBLBBBLBLB	−14.0423
D	BBBBBBBBBNNLNBLBLBLBLB	−11.8696
E	LBBBBBBBNBNLNBLBLBBBLB	−11.2465
F	LBBBBBBBBBNLNLNBLBBBLB	−15.7288
G	LBBBBBBBBBNLNLNBLBBBLB	−16.2159
H	LBBBBBBBBBNLNLNBLBBBLB	−16.3866
I	LBNBBBLBBBNNBBLBLBBBLB	−11.8513

3.4.3. Individual particle perturbations. Particle X_k is perturbed if $p_k^0 \neq p_k^1$ and its new position becomes

$$\hat{X}_k = X_k + \delta X, \tag{17}$$

where $\delta X \in \mathbb{R}^3$ and each element is taken from a uniform distribution on the interval $[-a_{ip}, a_{ip}]$.

4. RESULTS

We present the results of two sets of computations to show the effectiveness of HOM and HOPE in predicting the lowest energy conformations of the model proteins. These results highlight the usefulness of using the homotopy methods presented in this paper compared with some standard methods used to solve unconstrained minimization problems. Furthermore, the results show that the use of perturbations in HOPE helps increase the probability of predicting a correct target conformation over HOM.

The computations were performed using Matlab[1] under Linux on a 2.5 GHz Intel Pentium 4 processor. The potential energy function in (1) and the homotopy function in (13) were implemented in C so that the first and second derivatives could be produced using the automatic differentiation tool ADOL-C v1.8 (Griewank *et al.*, 1996).

We used the nine chains described in the experiments of Veitshans *et al.* (1997) for our test data. The sequences of the chains and the energy values of the corresponding native conformations are given in Table 2.

Each chain was used as a template (starting conformation) to predict the lowest energy conformation of the remaining eight targets, yielding a total of 72 computations. However, the lowest energy conformations match for five template-target pairs (*A–D*, *B–C*, *F–G*, *F–H*, and *G–H*). Thus, we present results for the remaining 62 computations only.

Success was measured for each predicted conformation, X^1, against the minimum energy conformation, X^*, using two metrics—the structural overlap function of Veitshans *et al.* (1997) and root mean-squared distance. The structural overlap function is the percentage of interparticle distances between nonbonded particles that differ in X^1 and X^* by more than 20% of the average bond length, $\bar{r}$. It is computed as

$$\chi(X^1) = 1 - \frac{2}{n^2 - 5n + 6} \sum_{i=1}^{n-3} \sum_{j=i+3}^{n} \Theta(0.2\bar{r} - |r_{ij}^1 - r_{ij}^*|) \tag{18}$$

where $\Theta(\cdot)$ is the Heavyside function and r^1 and r^* are distances between particles in X^1 and X^*, respectively. Note that $\chi(X^1) \in [0, 1]$, with $\chi(X^1) = 0$ meaning that X^1 is structurally equivalent to X^*.

[1]Matlab 6.5 and the Optimization Toolbox 2.2 from Mathworks, Inc.

TABLE 3. COMPARISON OF PREDICTION RESULTS USING HOM AND NEWTON-TR

Method	$\chi = 0$	*Success (%)*	$\overline{\chi}$	$\overline{\Delta}$	$\overline{t}$	$\overline{N}_f$
HOM	15	24	0.36	0.38	10	152
Newton-TR	4	6	0.45	0.55	1	20

Root mean-squared distance is computed as

$$\Delta(X^1) = \min_{S(X^1)} \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left\| X_i^1 - X_i^* \right\|^2} \tag{19}$$

where $S(X^1)$ is a rotation and translation of X^1. Thus, $\Delta(X^1)$ measures the distance between corresponding particles in the predicted and lowest energy conformations when they are optimally superimposed. For exact conformational matches, $\Delta(X^1) = 0$.

4.1. Computations using HOM

In the first set of computations, we compared HOM and a variant of Newton's method that uses a trust region to guarantee convergence (Newton-TR) (Coleman and Li, 1994, 1996). In HOM, Newton-TR was used to find local minimizers of H (Fig. 2, Step 5).

For each computation using HOM, we set the number of steps in λ to be $m = 10$, making $\Delta\lambda = 0.1$. The minimization routine was stopped when either the change in function value between iterates dropped below 10^{-6}, the maximum change in any of the variables in X between successive iterates dropped below 10^{-12}, or the number of iterates reached the maximum number of iterations allowed (1,000 for Newton-TR and 60 for each minimization performed in HOM).

Table 3 shows the results for the two methods. The second column shows the number of computations in which $\chi(X^1) = 0$, i.e., the method predicted the correct conformation. HOM has a success rate of 24% (column 3), almost four times better than Newton-TR. Columns 4 and 5 present the average structural overlap, $\overline{\chi}$, and root mean-squared distance, $\overline{\Delta}$, respectively, of the 62 computations. These results show that HOM predicts better structures than Newton-TR on average, even when an exact match of the lowest energy conformation was not predicted. The main drawback for HOM is that it is more computationally expensive than Newton-TR. The last two columns in the table show $\overline{t}$, the average clock time in seconds, and $\overline{N}_f$, the average number of function evaluations per computation. Even though HOM requires more work than Newton-TR, the tradeoff in success rate shows the benefit of using a homotopy method over a standard minimization algorithm.

4.2. Computations using HOPE

In the second set of computations, we compared HOPE and ensemble-based, basin-hopping simulated annealing (Basin-SA)—a combination of the methods of Wales and Scheraga (1999) and Salamon *et al.* (2002). The implementation of Basin-SA was developed using SA Tools v1.03 (Salamon *et al.*, 2002).

In Basin-SA, the move class, the set of possible conformations produced by perturbing a given conformation, consists solely of local minimizers of $E^1(X)$. Specifically, candidate conformations are found using Newton-TR started at a perturbed version of each conformation in the ensemble.

We chose to allow both methods to compute an equivalent number of local minimizers in the course of each computation. Specifically, we tested HOPE using $m = 10$ steps in λ and maximum ensemble sizes of $c_{max} = 2, 4, 8,$ and 16. Also, we set $\hat{c} = 1$, allowing only one perturbed version to be generated for each ensemble conformation. This yielded upper limits of $20, 40, 80,$ and 160 local minimizers to be computed for the values of c_{max}, respectively. In Basin-SA, we used ensembles of size c_{max} at each of $m = 10$ steps of the annealing schedule,[2] starting at $T = 10^5$. Therefore, the number of calls to Newton-TR in

[2]The **berkeley** schedule in SA Tools was used.

Basin-SA matched the corresponding upper limit of those allowed in HOPE. The main difference between the methods is the function being minimized—the homotopy function in HOPE and the potential energy function of the target chain in Basin-SA.

Computations were performed for both methods using the perturbations in Section 3.4, with maximum amounts of perturbation $a_r = 1$, $a_\theta = 40°$, and $a_{ip} = 1$. It is typical in simulated annealing methods that the perturbations allowed in the move class are functions of the temperature, T, defined by the annealing schedule. The following function was used to specify the maximum perturbation in the Basin-SA computations:

$$a_{max} = \alpha e^{-\frac{1}{\gamma T}}$$

where $\gamma \in (0, 1)$ and $\alpha = a_r, a_\theta$ or a_{ip}, depending on the perturbation method used. Note that this function passes through the point $(\hat{T}, \hat{a})$ when

$$\gamma = -\frac{1}{\hat{T} \ln\left(\hat{a}/\alpha\right)}.$$

The point $(\hat{T}, \hat{a}) = (1, 0.9)$ was used in Basin-SA so that the amount of maximum perturbation allowed was equivalent to that allowed in HOPE for almost all values of T except those very close to zero.

Table 4 shows averages of the results of 10 runs of each of the 62 computations for HOPE and Basin-SA using the bond length perturbations where the remainder of the chain is shifted. The results using bond angle and individual particle perturbations were comparable and thus are not presented here. The results for increasing values of c_{max} (column 2) show the trends of the success rate and computational effort of each method as more local minimizers are generated. Note that the best results for each of the methods (shown in bold) are those corresponding to the largest maximum ensemble size, $c_{max} = 16$. Clearly, HOPE outperforms Basin-SA in terms of successful predictions of the native conformations for each value of c_{max}, monotonically increasing to an average success rate of 95% at $c_{max} = 16$.

The computational effort required by the two methods is reflected in the last two columns of the table. HOPE required less time and computational effort than Basin-SA to produce better results. Although the two methods were allowed to make the same number of calls to Newton-TR to generate candidate conformations, HOPE actually made far fewer calls. The average time and number of function evaluations for Basin-SA are more than twice those for HOPE, and this effort increases at a faster rate with respect to c_{max} for Basin-SA than for HOPE, as can be seen in Fig. 8. This difference is due to the dynamic nature of the ensembles produced in HOPE; the size of the ensembles were often less than c_{max} due to pruning of duplicate conformations.

HOPE was the only method that successfully predicted the native conformations in at least one of the 10 runs performed for all of the 62 computations. Figure 9 presents the results with $c_{max} = 16$ for each of the template-target pairs using (a) HOPE and (b) Basin-SA. The size of each circle represents the percentage of runs where a target was successfully predicted starting at a given template. HOPE predicted the correct target conformations in all 10 runs for 44/62 (71%) template–target pairs. In contrast, Basin-SA

TABLE 4. COMPARISON OF PREDICTIONS RESULTS USING HOPE AND BASIN-SA (AVERAGED OVER 10 RUNS)

Method	c_{max}	$\chi = 0$	*Success (%)*	$\overline{\chi}$	$\overline{\Delta}$	$\bar{t}$	$\overline{N}_f$
HOPE	2	33.4	54	0.14	0.17	35	539
	4	43.1	70	0.08	0.11	65	992
	8	54.6	88	0.03	0.04	115	1732
	16	**59.0**	**95**	**0.01**	**0.02**	**200**	**2981**
Basin-SA	2	13.1	21	0.27	0.36	52	753
	4	20.8	34	0.19	0.26	107	1576
	8	28.5	46	0.13	0.19	229	3174
	16	**40.2**	**65**	**0.08**	**0.12**	**434**	**6358**

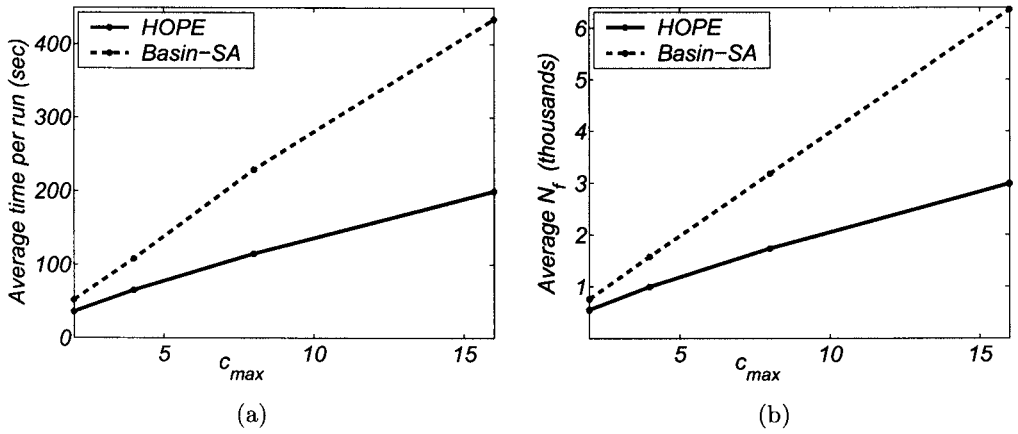


FIG. 8. Average cost per computation: (**a**) clock time in seconds and (**b**) number of function evaluations.

predicted all 10 target conformations correctly for only 14/62 (23%) pairs. More importantly, Basin-SA was not able to correctly predict the target conformations for three pairs (B–F, B–H, and F–B) in any of the 10 runs.

5. DISCUSSION

We have presented the use of HOPE in predicting the native conformation of a model protein by potential energy minimization. In the computations that use perturbations involving bond lengths, bond angles, and individual particles, HOPE outperforms a simulated annealing method in terms of successful prediction of native conformation of targets and required computation.

We plan to extend HOPE to predict the structures of target proteins found in the Protein Data Bank (PDB) (Berman *et al.*, 2000). The goal is to predict the PDB structure of a target protein given the PDB structure of a homologous template protein. A suitable potential energy function must be chosen and a homotopy function must be designed to take into account the features of that particular energy function that will facilitate HOPE the most. Because the computations presented in this paper show promise for

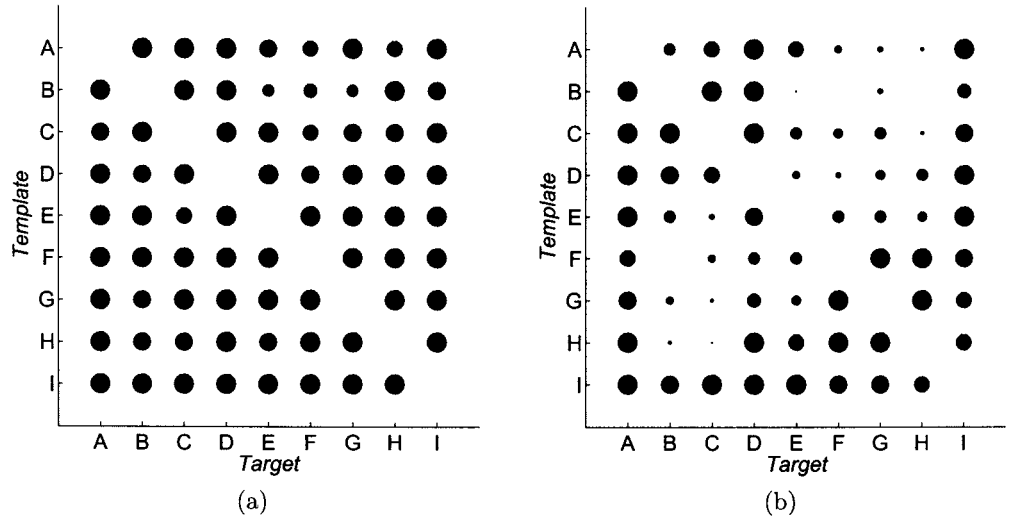


FIG. 9. Success of (**a**) HOPE and (**b**) Basin-SA with $c_{max} = 16$ for each template-target pair. The size of each circle represents the percentage of successful predictions over 10 runs.

protein-like models, we expect that HOPE can be adapted to obtain native structures for more realistic energy functions.

The most challenging aspect in using HOPE is the choice of an appropriate homotopy function. It is unclear that a general homotopy function mapping the global minimizers onto each other exists for all template-target protein pairs in the PDB. To some extent, the use of perturbations in HOPE solves this problem.

When working with proteins in the PDB, we conjecture that the individual particle perturbations may more realistically reflect the dynamics of individual atoms than the bond length and bond angle perturbations. However, results presented in this paper involving the latter perturbations were slightly better than those using the individual particle perturbations. An alternative to perturbing individual particles would be to perturb individual particles as well as those particles sharing a bond with it. The particles sharing a bond could be moved in the same direction as the perturbed particle but to a lesser extent. This could be carried out to subsequent particles to which those particles are bonded, dampening the amount of position change for particles further away in the chain. In this way, the perturbations would more closely reflect the dynamics of chemically bonded atoms. Further development and testing of this type of perturbation used in HOPE is planned.

ACKNOWLEDGMENTS

We thank Ron Unger of Bar-Ilan University in Israel for helpful discussions that took place while he was a visiting researcher at the University of Maryland Institute for Advanced Computer Science (UMIACS). D.D. was supported in part by National Library of Medicine Bioinformatics Individual Fellowship F37LM008162, D.O. was supported in part by NSF Grant CCR-0204084 and DOE Grant DEFG0204ER25655, and D.T. acknowledges partial support from NSE-CHE 05-14056.

REFERENCES

Ackermann, S., and Kliesch, W. 1998. Computation of stationary points via a homotopy method. *Theor. Chem. Acc.* 99, 255–264.

Allgower, E.L., and Georg, K. 2003. *Introduction to Numerical Continuation Methods*, SIAM, Philadelphia, PA.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The protein data bank. *Nucl. Acids Res.* 28, 235–242.

Brodmeier, T., and Pretsch, E. 1994. Application of genetic algorithms in molecular modeling. *J. Comput. Chem.* 15, 588–595.

Coleman, T.F., and Li, Y. 1994. On the convergence of reflective Newton methods for large-scale nonlinear minimization subject to bounds. *Math. Program.* 67, 189–224.

Coleman, T.F., and Li, Y. 1996. An interior, trust region approach for nonlinear minimization subject to bounds. *SIAM J. Optim.* 6, 418–445.

Coleman, T.F., and Wu, Z. 1994. Parallel continuation-based global optimization for molecular conformation and protein folding. Technical report MCS-P443-0694, Argonne National Laboratory.

Das, B., Meirovitch, H., and Navon, I. M. 2003. Performance of enriched methods for large unconstrained optimization as applied to models of proteins. *J. Comput. Chem.* 24, 1222–1231.

Dunlavy, D.M., and O'Leary, D.P. 2005. HOPE: A homotopy method for unconstrained minimization. In preparation.

Fiser, A., and Sali, A. 2003. Comparative protein structure modeling. In Chasman, D., ed., *Protein Structure: Determination, Analysis, and Applications for Drug Discovery*, 167–206. Marcel Dekker, New York, NY.

Griewank, A., Juedes, D., and Utke, J. 1996. Algorithm 755: ADOL-C: A package for the automatic differentiation of algorithms written in C/C++. *ACM Trans. Math. Software* 22, 131–167.

Hunjan, J.S., Sarkar, S., and Ramaswamy, R. 2002. Global optimization on an evolving energy landscape. *Phys. Rev. E* 66, 046704.

IUPAC-IUB, Commission on Biochemical Nomenclature (CBN). 1970. Abbreviations and symbols for the description of the conformation of polypeptide chains. *Biochemistry* 9, 3471–3479.

Kawai, H., Kikuchi, T., and Okamoto, Y. 1989. A prediction of tertiary structures of peptide by the Monte Carlo simulated annealing method. *Protein Eng.* 3, 85–94.

Le Grand, S.M., and Merz, Jr., K.M. 1993. The application of the genetic algorithm to the minimization of potential energy functions. *J. Global Optim.* 3, 49–66.

Li, Z., and Scheraga, H. A. 1987. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. USA* 84, 6611–6615.

Ngo, J.T., and Marks, J. 1992. Computational complexity of a problem in molecular structure prediction. *Protein Eng.* 5, 313–321.

Salamon, P., Sibani, P., and Frost, R. 2002. *Facts, Conjectures, and Improvements for Simulated Annealing*, SIAM, Philadelphia, PA.

Schelstraete, S., Schepens, W., and Verschelde, H. 1998. Energy minimization by smoothing techniques: A survey. In Balbuena, P.B., and Seminario, J.M., eds., *Molecular Dynamics. From Classical to Quantum Methods*, 129–185. Elsevier, New York.

Shin, J.K., and Jhon, M.S. 1991. High directional Monte Carlo procedure coupled with the temperature heating and annealing as a method to obtain the global energy minimum structure of polypeptides and proteins. *Biopolymers* 31, 177–185.

Veitshans, T., Klimov, D., and Thirumalai, D. 1997. Protein folding kinetics: Timescales, pathways, and energy landscapes in terms of sequence-dependent properties. *Fold. Des.* 2, 1–22.

Wales, D.J., and Scheraga, H.A. 1999. Global optimization of clusters, crystals, and biomolecules. *Science* 284, 1368–1372.

Watson, L.T., and Haftka, R.T. 1989. Modern homotopy methods in optimization. *Comput. Methods Appl. Mech. Eng.* 74, 289–304.

Wilson, S.R., and Cui, W. 1990. Applications of simulated annealing to peptides. *Biopolymers* 29, 225–235.

Wilson, S.R., Cui, W., Moskowitz, J.W., and Schmidt, K.E. 1988. Conformational analysis of flexible molecules: Location of the global minimum energy conformation by the simulated annealing method. *Tetrahedron Lett.* 29, 4373–4376.

Xie, D., and Schlick, T. 1999. Efficient implementation of the truncated-Newton algorithm for large-scale chemistry applications. *SIAM J. Optim.* 10, 132–154.

Address correspondence to:
Daniel M. Dunlavy
Applied Mathematics and Scientific Computation Program
1103 A.V. Williams Building
College Park, MD 20770

E-mail: ddunlavy@cs.umd.edu