# A Method for Parameter Optimization in Computational Biology

J. B. Rosen,* A. T. Phillips,† S. Y. Oh,‡ and K. A. Dill§

*Computer Science and Engineering Department, University of California at San Diego, San Diego, California 92093 USA; †Department of Computer Science, University of Wisconsin–Eau Claire, Eau Claire, Wisconsin 54702 USA; ‡Department of Mathematics, Chungnam National University, Taejon 305-764, Korea; and §Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94118 USA

ABSTRACT Models in computational biology, such as those used in binding, docking, and folding, are often empirical and have adjustable parameters. Because few of these models are yet fully predictive, the problem may be nonoptimal choices of parameters. We describe an algorithm called ENPOP (energy function parameter optimization) that improves—and sometimes optimizes—the parameters for any given model and for any given search strategy that identifies the stable state of that model. ENPOP iteratively adjusts the parameters simultaneously to move the model global minimum energy conformation for each of *m* different molecules as close as possible to the true native conformations, based on some appropriate measure of structural error. A proof of principle is given for two very different test problems. The first involves three different two-dimensional model protein molecules having 12 to 37 monomers and four parameters in common. The parameters converge to the values used to design the model native structures. The second problem involves nine bumpy landscapes, each having between 4 and 12 degrees of freedom. For the three adjustable parameters, the globally optimal values are known in advance. ENPOP converges quickly to the correct parameter set.

## INTRODUCTION

There are many models in computational chemistry, biology, and materials science, in which parameterized energy functions are used to predict three-dimensional structures of molecules. Folding, threading, docking, protein-protein recognition, and loop refinement methods are examples from computational biology. It is seldom clear whether failures of such models are attributable to the form of the model's mathematical functions, or to poor choices of the parameters used in them. Such models are usually so computationally expensive that it is impossible to be systematic about finding the "optimal" parameters, i.e., those parameters that minimize some measure of total structural error.

To find the optimal parameters, say for a folding algorithm, it would be necessary to 1) compute the folded structures for many proteins, 2) determine the errors, 3) change the parameters, and then iterate this whole process for many different sets of parameters to find the best ones. This has not been computationally feasible before. Instead, model parameters are usually chosen to varying degrees by physical estimates, guesswork, and arbitrary trial and error. Such efforts involve small searches through large parameter spaces. As with other types of search problems, parameter optimization can depend on the order in which the parameters are chosen, and it can get caught in traps from which it cannot escape.

There are methods for optimizing parameters in certain classes of problems (Esposito and Floudas, 1998; Maiorov and Crippen, 1994; Koretke et al., 1998; Hao and Scheraga, 1996). Two examples are threading (Goldstein et al., 1992; Hendlich et al., 1990; Maiorov and Crippen, 1994; Thomas and Dill, 1996; Mirny and Shakhnovich, 1996; Huber and Torda, 1998; Koretke et al., 1996) and lattice models of folding (Hao and Scheraga, 1996; Goldstein et al., 1992; Shrivastava et al., 1995; Govindarajan and Goldstein, 1995). In both cases, the ability to find optimal parameters is a direct consequence of the facts that 1) the conformational space is discrete, and 2) the global optimum is guaranteed to be among the conformations searched. Whenever global optima can be found through finite searches, there are methods that can be used to learn parameters that can distinguish correct from incorrect structures. This is arguably the principal advantage of threading versus folding algorithms: the former involve discrete searches, so parameters can be improved systematically.

As far as we know, no algorithm yet exists that can find optimal parameters for models having continuous degrees of freedom. Consider protein folding. To improve the parameters in a folding model, you need to recompute the lowest energy structure many times, once after each iteration of small changes in parameters. This means many minimizations. The main problem in optimizing parameters for continuum models is that the typical minimization methods—Monte Carlo, simulated annealing, or molecular dynamics—find only local minima and fall into different energy wells each time, so there is no unique and reproducible mapping from a given set of model parameters to a given model native structure. This lack of reproducibility is the primary reason that parameter optimizations are difficult in continuum models.

We describe here a computational method, called ENPOP (energy parameter optimization), that searches for globally optimal parameters for continuous models. It takes as input

*Received for publication 9 December 1999 and in final form 7 July 2000.*

Address reprint requests to Dr. Ken A. Dill, Department of Pharmaceutical Chemistry, University of California at San Francisco, Suite 415, Laurel Heights Campus, 3333 California St., San Francisco, CA 94118. Tel.: 415-476-9964; Fax: 415-476-1508; E-mail: dill@maxwell.ucsf.edu.



0006-3495/00/12/2818/07 $2.00

1) a given model and its associated adjustable parameters, 2) the (usually incorrect) best structure for each of $m$ different molecules that is predicted by the starting parameters, 3) the $m$ correct (true) structures that the model should produce, and 4) some measure of structural error between predicted structures and true, known, or correct best structures. The technology that enables us to circumvent the reproducibility limitation and to systematically optimize parameters is the recently developed CGU (convex global underestimator) method that finds global minima—or at least reproducible minima—of energy landscapes, at least for short enough chains (Dill et al., 1997a,b).

Our approach is general and guarantees improved, and sometimes even globally optimal, parameters. It should be applicable to a wide range of models in computational biology and chemistry. It allows any differentiable measure of "structural error" to be minimized over the continuous space of both potential function parameters and molecular conformations, while placing no restrictions on the form of the potential function other than that it be differentiable. Our method simultaneously tracks the global minimum for each of the proteins in the model as the parameters are changed, while providing a guaranteed reduction in the structural error between model-native and true-native states. For this we do not require any additional global minimization (beyond that needed to get the initial global minima), so the method is computationally efficient. Our strategy then is to learn a set of parameters for one set of proteins with known structures and apply those parameters to the prediction of other structures. Our hope is that by finding better parameters for any given model, this method will ultimately allow computational biologists to develop improved models for folding, binding, docking, etc.

## ENERGY FUNCTION IMPROVEMENT BY PARAMETER OPTIMIZATION

We consider some model for an energy landscape, $F(\Phi, \alpha)$. $F(\Phi, \alpha)$ is the conformational energy or free energy as a function of the $n$ degrees of freedom (coordinates or bond angles, etc.) that are given by the vector $\Phi \in \Re^n$ and as a function of the $k$ parameters of the model that are given by the parameter vector $\alpha \in \Re^k$. The parameters might include Lennard-Jones parameters, steric terms, hydrogen bond or hydrophobic interaction strengths, coefficients of bond angle energies, etc. There is no limitation on the functional forms of the terms. Although the method is general, we will make the discussion more concrete by focusing on protein folding. The predicted native state is given by the global minimum vector $\Phi_G$, which has a free energy $F(\Phi_G)$. If the model energy function were perfect, it would predict the true native structure; that is, we would have $\Phi_G = \Phi_N$, where $\Phi_N$ is the correct dihedral angle vector for the native state of this protein.

We will consider the prediction accuracy in terms of the error in the degrees of freedom $\|\Phi_G - \Phi_N\|^2$ for each protein. Other measures, such as RMS, could be used with minor modification. To make explicit the dependence of $F$ on the parameters, we let $\alpha$ be a vector of $k$ parameters of the energy function that are common for all proteins. In the energy functions we test here, we typically have $k \leq 15$.

Suppose we wish to improve the native structure predictions for a set of $m$ proteins, with energy functions $F^{(j)}(\Phi^{(j)}, \alpha)$ and native states $\Phi_N^{(j)}$, $j = 1, \ldots, m$. Note that each potential function $F^{(j)}$ and set of $n_j$ independent variables $\Phi^{(j)}$ will be different and will depend on the number and sequence of beads in the $j$th protein. But the vector $\alpha$ is independent of $j$ and will be the same for all proteins. For any fixed value of $\alpha$, we can use the CGU algorithm to find the corresponding global minimum $\Phi_G^{(j)}(\alpha)$ for each protein. Each such global minimum is characterized by the conditions

$$F^j(\Phi_G^{(j)}, \alpha) \leq F^{(j)}(\Phi^{(j)}, \alpha), \tag{1}$$

$$\nabla_\Phi F^{(j)}(\Phi_G^{(j)}, \alpha) = 0, \quad \text{and} \tag{2}$$

$$H^{(j)}(\Phi_G^{(j)}, \alpha) \text{ positive definite}, \tag{3}$$

where $H^{(j)}$ is the $(n_j \times n_j)$ Hessian matrix with respect to $\Phi$ of $F^{(j)}$ $(\Phi^{(j)}, \alpha)$. While unlikely, it is possible that for some initial choice $\alpha_I$, some Hessian $H^{(j)}$ may only be positive semidefinite at the corresponding global minimum. If this occurs, a different value of $\alpha_I$ should be chosen. The total conformational error is defined by

$$\rho(\alpha) \sum_{j=1}^{m} w_j \rho_j(\alpha), \tag{4}$$

where $w_j \geq 0$ are any arbitrary weighting factors that the user wants to include, and $\rho_j(\alpha)$ is the $j$th molecule conformational error, given by

$$\rho_j(\alpha) \|\Phi_G^{(j)}(\alpha) - \Phi_N^{(j)}\|^2. \tag{5}$$

The weights $w_j$ can be selected to give more or less importance to certain proteins, for example, based on known accuracies or reliabilities of their structures, but otherwise might normally be set equal to one. The parameter adjustment method proposed here can now be stated as a $k$-dimensional minimization problem, with simple bounds on the parameters

$$\min_\alpha \rho(\alpha) \quad \text{subject to } \alpha_{min} \leq \alpha \leq \alpha_{max}. \tag{6}$$

The bounds $(\alpha_{min})_j$ and $(\alpha_{max})_j$ should be chosen to appropriately restrict the range on the parameter $\alpha_j$ to values that are meaningful to the problem.

The key computational expense in this method is computing the function $\rho(\alpha)$ and its gradient $\nabla_\alpha \rho$. So we use an efficient minimization method that requires relatively few

function and gradient calls, the Sequential Quadratic Programming (SQP) method (Gill et al., 1986).

We need to track the changing value of $\Phi_G^{(j)}(\alpha)$ as $\alpha$ is adjusted. We impose the constraint that $\Phi_G^{(j)}(\alpha)$ must remain a local minimum of $F^{(j)}$ ($\Phi^{(j)}$, $\alpha$) as $\alpha$ is modified. This is done by requiring that the system of $n_j$ nonlinear equations in Eq. 2 remains satisfied as $\alpha$ changes, starting with an initially chosen parameter vector $\alpha_I$. Note that this does not guarantee that the updated value of $\Phi_G^{(j)}(\alpha)$ will also remain the global minimum of $F^{(j)}(\Phi^{(j)}, \alpha)$, although this is easily checked at the conclusion of the method by another CGU global optimization step.

$\Phi_G^{(j)}(\alpha)$ can be updated without recomputing the global minimum, by using the implicit function theorem, provided that the Hessians $H^{(j)}$ are positive definite. This condition is satisfied at $\Phi_G^{(j)}$ by Eq. 3. The implicit function theorem uses a linearized approximation to Eq. 2 at its current values of $\Phi$ and $\alpha$ and therefore gives the required small change $\Delta\alpha$ to balance any small change $\Delta\Phi$. It is valid if the neglected terms in $\|\Delta\alpha\|^2$ and $\|\Delta\Phi\|^2$ are much smaller than the first-order term in $\|\Delta\alpha\|$ and $\|\Delta\Phi\|$. For a small change $\Delta\Phi_G^{(j)}$, the corresponding change $\Delta\alpha$ required to keep Eq. 2 satisfied is given approximately by

$$H^{(j)}(\Phi_G^{(j)}, \alpha)\Delta\Phi_G^{(j)} + J_\alpha^{(j)}(\Phi_G^{(j)}, \alpha)\Delta\alpha = 0, \quad (7)$$

where $J_\alpha^{(j)} = J_\alpha^{(j)}(\Phi_G^{(j)}, \alpha)$ are the $n_j \times k$ Jacobians of $\nabla_\Phi F^{(j)}(\Phi_G^{(j)}, \alpha)$. Because the Hessians $H^{(j)} = H^{(j)}(\Phi_G^{(j)}, \alpha)$ are nonsingular, we can write this as

$$\Delta\Phi_G^{(j)} = -[H^{(j)}]^{-1} J_\alpha^{(j)}\Delta\alpha. \quad (8)$$

Because $\nabla_\alpha\rho_j(\alpha) = -[J_\alpha^{(j)}]^T[H^{(j)}]^{-1}\nabla_\Phi\rho_j(\alpha)$, it then follows that the desired gradient $\nabla_\alpha\rho(\alpha)$ of the total conformation error as a function of parameters is given by

$$\nabla_\alpha\rho(\alpha) = -2\sum_{j=1}^{m} w_j[J_\alpha^{(j)}]^T[H^{(j)}]^{-1}(\Phi_G^{(j)}(\alpha) - \Phi_N^{(j)}). \quad (9)$$

Corresponding to each $\Phi_G^{(j)}$, good approximations to the Hessians $H^{(j)}$ are obtained directly from the SQP implementation of the local minimization, a key phase of the CGU method. Also, the Jacobians $J_\alpha^{(j)}$ are relatively easy to compute because $J_\alpha^{(j)}$ has only $k$ columns.

## THE ENPOP ALGORITHM

We call our method ENPOP (energy function parameter optimization). Here is the general algorithm:

1. Guess an initial $\alpha_I$.

2. Run the CGU method successively on the functions $F^{(j)}$, $j = 1, \ldots, m$ with $\alpha = \alpha_I$ fixed. Denote the result of each CGU run by $\Phi_G^{(j)}$.

3. Using each $\Phi_G^{(j)}$ from step 2, perform the minimization over $\alpha$ described in Eq. 6. To perform this step, use $\alpha_I$ as a starting point and use Eq. 9 for the gradient of $\rho$.

To compute the change $\Delta\Phi_G^{(j)}$ corresponding to altered parameters (at each step of the local minimization) $\Delta\alpha$, we use the implicit function theorem, which states that (see Eq. 7) $\Delta\Phi_G^{(j)} = -[H^{(j)}]^{-1}J_\alpha^{(j)}\Delta\alpha$. This step will result in a new parameter set $\alpha_{new}$ with corresponding global minima $\Phi_G^{(j)}(\alpha_{new})$, $j = 1, \ldots, m$.

The ENPOP algorithm is quite general and can be applied to any set of molecular structures with a given model and potential function. All that is required is a suitable representation for the Hessian $H^{(j)}$ and Jacobian $J_\alpha^{(j)}$ specific to the model.

## GUARANTEEING REDUCTION IN STRUCTURAL ERROR

The method described above optimizes parameters while enforcing the requirement that the initial global minimum remain at least a local minimum of the model energy function. But what guarantees that the original global minimum will not shift to become just a local minimum? In this section, we describe criteria that ensure that the original minimum remains global. The purpose of such a criterion is computational efficiency: when the global minimum is no longer ensured, the CGU can be run again to check whether the current minimum is global. The goal is to run the CGU only the minimum possible number of times, to save computational cost.

When attempting to improve the parameter values for the set of potential functions $F^{(j)}(\Phi_G^{(j)}(\alpha), \alpha)$, for $j = 1, \ldots, m$, it is useful to have prior information about the amount of reduction in the error $\rho(\alpha)$ that can be expected. We now show that a lower bound on the decrease in $\rho(\alpha)$ can be given, based on information available at the initial value $\alpha = \alpha_I$. That is, we show that we can always obtain a parameter vector $\alpha_1$ such that

$$\rho(\alpha_1) \le \rho(\alpha_I) - \Delta\rho, \quad (10)$$

where $\Delta\rho$ is given by Eq. 14. Because $\Delta\rho$ is a guaranteed lower bound, the actual decrease in $\rho(\alpha)$ obtained by ENPOP may typically be much greater.

To determine $\Delta\rho$, we need to know several quantities that are available once the global minima $\Phi_G^{(j)}$ have been computed for each protein by the CGU method. For this purpose, we define the gradient in parameter space, $g = \nabla_\alpha\rho(\alpha_I)$, as given by Eq. 9, and the initial energy gaps $\Delta F^{(j)}$ for $j = 1, \ldots, m$ and $\alpha = \alpha_I$, by

$$\Delta F^{(j)} = F_{LM}^{(j)} - F_G^{(j)} > 0. \quad (11)$$

In Eq. 11, $F_G^{(j)} \equiv F^{(j)}(\Phi_G^{(j)}(\alpha_I), \alpha_I)$ represents the global minimum energy found by the CGU method, using the parameter vector $\alpha_I$, and $F_{LM}^{(j)}$ is the corresponding next lowest local minimum energy, both for the $j$th protein. If the gradient $\|g\| = 0$, then $\alpha_I$ is already a stationary point of $\rho(\alpha)$, and a new value of $\alpha_I$ must be selected. Similarly, if

$\Delta F^{(j)} = 0$, for any $j$, then any change in $\alpha$ may cause the coordinates of the global minimum for $F^{(j)}(\Phi(\alpha), \alpha)$ to move discontinuously from its current conformation $\Phi_G^{(j)}$ to the conformation corresponding to the alternate global minimum (because $F_{LM}^{(j)} = F_G^{(j)}$). In this case, the $j$th protein should be replaced (or removed) in the test set. Therefore, we can assume that Eq. 11 holds for all $j = 1, \ldots, m$, and that $\|g\| > 0$.

In addition, we will require the quantities $\Delta F_{min} \equiv \min_j \Delta F^{(j)} > 0$, the Hessian $H_\rho(\alpha)$ of $\rho(\alpha)$, and the curvature of $\rho$ in the direction of $g$ as given by

$$\lambda = g^T H_\rho g / g^T g. \tag{12}$$

The value of $\lambda$ is bounded by $\lambda_{min} \le \lambda \le \lambda_{max}$, where $\lambda_{min}$ and $\lambda_{max}$ are the minimum and maximum eigenvalues of $H_\rho(\alpha_I)$. Because $H_\rho(\alpha)$ may be indefinite at $\alpha_I$, $\lambda$ may be negative.

Finally, we need a uniform bound on the gradient of $F(\Phi(\alpha), \alpha)$ with respect to $\alpha$:

$$\|\nabla_\alpha F(\Phi(\alpha), \alpha)\| \le \beta. \tag{13}$$

In terms of these quantities, it can be shown that

$$\Delta\rho = \begin{cases} \|g\|\Delta F_{min}/2\beta & \lambda \le 0 \\ \|g\|\Delta F_{min}/4\beta & \lambda > 0 \text{ and } \lambda\Delta F_{min} \le 2\beta\|g\| \\ \|g\|^2/2\lambda & \lambda > 0 \text{ and } \lambda\Delta F_{min} > 2\beta\|g\| \end{cases}. \tag{14}$$

In the first two bounds in Eq. 14, the change in $\alpha$ is limited by the possibility that the current global minimum is replaced by one of the other local minima (it is assumed that a new global minimum will arise only from existing local minima). In the third bound, the decrease in $\rho(\alpha)$ is at least that obtained by a single steepest descent step in the direction $-g$ in $\alpha$-space.

We now test ENPOP on two very different problems. The first test involves three short protein molecules, in two-dimensional conformational space, using a simple energy function with four parameters. ENPOP reduces the total conformation error from its initial value $\rho(\alpha_I) = 1012.3$ to its minimum $\rho(\alpha_N) = 0.87$. The second test problem is one for which we know in advance the optimum parameter vector $\alpha_N$, such that $\Phi_G(\alpha_N) = \Phi_N$. We do not, of course, use this knowledge in ENPOP, but we can verify that ENPOP computes the correct parameter vector $\alpha_N$. This test problem consists of nine different bumpy energy landscapes, each with a different number of degrees of freedom (ranging from 4 to 12), but with a common set of three parameters. We find that ENPOP always finds the optimum parameter vector $\alpha_N$, starting with different initial vectors $\alpha_I$.

## TEST PROBLEM 1: 2D PROTEIN FOLDING MODEL

We first consider a problem involving three short chain molecules, in two-dimensional conformation space, using a simple energy function. The energy function consists of four terms, a bond length penalty term, a Lennard-Jones (LJ) attraction/repulsion term, a hydrophobic attraction term, and a hydrogen bond term. The energy function contains four adjustable parameters. The test molecules have from 12 to 37 beads and differ from each other in the specification of which beads are hydrophobic and which pairs are hydrogen bonded. The native state conformations are created in advance by arbitrarily choosing bond angles. The four parameters are adjusted by ENPOP so that the global minimum of the energy function, for each molecule, gives a corresponding conformation as close as possible to its known native conformation.

To simplify both the calculation and the discussion, the degrees of freedom are the $(x, y)$ coordinates of each bead, $(x_i, y_i)$, $i = 1, \ldots, n$, $n = 12, 25, 37$. For each molecule, bead 1 is fixed at the origin. We let $\mathbf{X}$ denote the vector with $2n$ elements $(x_i, y_i)$, $i = 1, \ldots, n$. The conformation of a molecule is then specified by the vector $\mathbf{X}$. Because the distance $r_{ij}$ between beads $i$ and $j$ is given by $r_{ij}^2 = (x_i - x_j)^2 + (y_i - y_j)^2$, the vector $\mathbf{X}$ completely specifies all pairwise distances between beads. We also let $\alpha \in \Re^4$ and denote the parameter vector specifying the four parameter values as $\alpha_i$, $i = 1, \ldots, 4$.

The energy function is

$$F(\mathbf{X}, \alpha) = 50 \sum_{i=1}^{n-1} [r_{i,i+1} - 1.0]^2 + \sum_{i=1}^{n-1} \sum_{j>i} L(r_{ij}, \alpha_1, \alpha_2) - \alpha_3 \sum_{i,j \in H} r_{ij}^{-6} - \alpha_4 \sum_{i,j \in HB} r_{ij}^{-6}, \tag{15}$$

where $\alpha_i > 0$ and

$$L(r, \alpha_1, \alpha_2) = \alpha_1[-2r^{-6} + 0.1\alpha_2 r^{-12}]. \tag{16}$$

$H$ denotes the set of hydrophobic beads, and $HB$ denotes the set of hydrogen bonded pairs. Note that the minimum in the LJ term occurs at $r_{min} = (0.1\alpha_2)^{1/6}$, where it has the value $(-10\alpha_1)/\alpha_2$.

For each test molecule the sets of hydrophobic and hydrogen-bonded beads are chosen to be different, while the parameter values are the same for all molecules. There is one energy function for all proteins, but folds are different because proteins have different monomer sequences. We represent the energy function for the $j$th test molecule by $F_j(\mathbf{X}_j, \alpha)$, $j = 1, 2, 3$. For each test molecule we have a known native state conformation $\mathbf{X}_{Nj}$, $j = 1, 2, 3$.

Molecule 1 consists of 12 beads, with four of them hydrophobic: $H = (1, 2, 11, 12)$; and there are three hydrogen-bonded pairs: $HB = (1, 4), (5, 8), (9, 11)$. Molecule 2 consists of 25 beads, with seven of them hydrophobic: $H = (3, 6, 7, 15, 21, 24, 25)$; and four hydrogen-bonded pairs: $HB = (3, 6), (9, 12), (15, 18), (19, 22)$. Finally, molecule 3 consists of 37 beads, with 12 of them hydrophobic: $H = (1, 5, 8, 9, 12, 17, 20, 21, 24, 29, 32, 33)$; and seven hydrogen-

bonded pairs: $HB$ = (1, 4), (3, 6), (7, 10), (17, 20), (19, 22), (23, 31), (34, 36).

We start with an arbitrary initial choice $\alpha_I$ of the parameter vector. Corresponding to $\alpha_I$ are global minimum energy conformations $\mathbf{X}_{Ij}$, obtained by minimizing $F_j(\mathbf{X}_j, \alpha_I), j = 1, 2, 3$. These initial and native conformations are shown in Fig. 1, as the first and last of the four conformations for each molecule. Let $\mathbf{X}_j(\alpha)$ denote the minimum energy conformation for any $\alpha$. We have $\mathbf{X}_j(\alpha_I) = \mathbf{X}_{Ij}$. We aim to find $\alpha_N$, so that the conformation error $\rho(\alpha)$ is minimized, where

$$\rho(\alpha) = \sum_{j=1}^{3} \|\mathbf{X}_j(\alpha) - \mathbf{X}_{Nj}\|^2. \quad (17)$$

To illustrate the process of parameter optimization, we explore trajectories in parameter space, from the initial to final parameter vectors. We follow the steepest descent path in parameter space to find the native parameter vector $\alpha_N$, such that $\rho(\alpha_N)$ is a minimum. At each descent step $\Delta\alpha$ we reduce $\rho(\alpha)$ by a small amount. The corresponding changes in the function values $F_j$ and corresponding conformations $\mathbf{X}_j(\alpha + \Delta\alpha)$ are then obtained by local minimizations of $F_j(\mathbf{X}_j, \alpha + \Delta\alpha)$, starting with $\mathbf{X}_j(\alpha)$. The SQP code NPSOL (Gill et al., 1986) is used for this local minimization, and because the conformation change due to $\Delta\alpha$ is small, this computation is fast. A typical steepest descent calculation in parameter space requires several hundred steps and takes 10–30 min on a current workstation. Note that the more efficient SQP minimization method could also have been used in parameter space, but it would have given much less information about the nature of the contours on the $\rho(\alpha)$ surface.

One steepest descent calculation is shown in Figs. 1 and 2. For this example, the conformation changes for each of the three molecules as $\alpha$ changes from $\alpha_I$ to $\alpha_N$ are shown in Fig. 1. For each of the three test molecules, the initial conformation (given by $\mathbf{X}_j(\alpha_I), j = 1, 2, 3$) is shown on the left of the figure. The conformations on the right of the figure show $\mathbf{X}_j(\alpha_N)$, $j = 1, 2, 3$. The two intermediate conformations, for each molecule, show the $\mathbf{X}_j(\alpha)$ for two intermediate values of $\alpha$ along the path in parameter space, as $\rho(\alpha)$ was minimized. The red beads are hydrophobic, and hydrogen bonding is between selected pairs of beads, as stated above. The initial values of the parameters were $\alpha_I$ = (1.0, 0.5, 0.1, 50). The parameter values that minimized $\rho(\alpha)$ were $\alpha_N$ = (1.7, 0.85, 0.73, 35). The initial value of the conformation error was $\rho(\alpha_I)$ = 1012.3, and its final value was $\rho(\alpha_N)$ = 0.87. The difference between the desired native conformations $\mathbf{X}_{Nj}$ and the final computed conformations $\mathbf{X}_j(\alpha_N)$ are so small that they cannot be distinguished in Fig. 1.

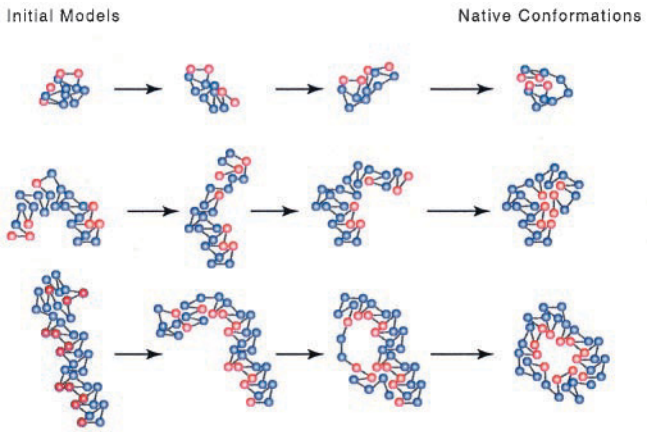


FIGURE 1 Three two-dimensional model proteins. The three structures on the right are the true natives, which are the global minimum conformations for the native parameter vector $\alpha_N$. The three structures on the left are the global minima for an incorrect initial parameter vector $\alpha_I$. ENPOP finds the correct parameters $\alpha_N$ through an iterative procedure starting with $\alpha_I$.

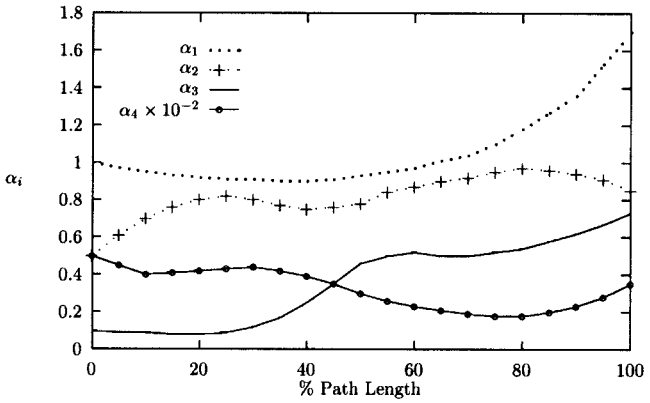


FIGURE 2 Changes in the parameter values as ENPOP steps from $\alpha_I$ to $\alpha_N$, through a steepest descent path in parameter space.

The conformations in Fig. 1 show that initially the hydrogen bond term dominates relative to the hydrophobic term, but that in the native state the hydrophobic term essentially determines the structure. This is because $\alpha_3$ (hydrophobic) increases from 0.1 to 0.73, while $\alpha_4$ (hydrogen bond) decreases from 50 to 35. The manner in which the four parameters changed along the steepest descent path is shown in Fig. 2, which shows the trajectory in parameter space in moving from $\alpha_I$ to $\alpha_N$, along the steepest descent path for $\rho(\alpha)$. Because we are following a steepest descent path in parameter space, the value of $\rho(\alpha)$ decreases monotonically as $\alpha$ is changed from $\alpha_I$ to $\alpha_N$. However, as shown in Fig. 2, the values of the four parameters, $\alpha_i$, $i = 1, 2, 3, 4$, do not change monotonically as a function of the path length in parameter space. Thus, for example, $\alpha_1$ and $\alpha_3$ are initially decreasing, even though they eventually increase along the steepest descent path. This shows that while minimizing $\rho(\alpha)$ in parameter space is a nontrivial problem, it can be successfully accomplished, as illustrated by this example.

## MODEL PROBLEM 2: BUMPY LANDSCAPES WITH KNOWN OPTIMAL PARAMETERS

For the second test problem, we sought a set of bumpy energy landscapes for which we can know the optimal parameter set in advance. The degrees of freedom are $\Phi \in \Re^n$, and the energy function is $F(\Phi, \alpha)$, which depends in a smooth way on $\Phi$ and on the parameter vector $\alpha \in \Re^k$. In this case we chose the function

$$F(\Phi, \alpha) = \Psi - \mu \cos(w\Psi), \tag{18}$$

where $\Psi = \Psi(\Phi, \alpha) = \frac{1}{2}(\Phi - A\alpha)^T D(\Phi - A\alpha)$, $\mu > 0$ and $w > 0$ are constants, $A$ is an $n \times k$ matrix of rank $k$, and $D$ is an $n \times n$ positive diagonal matrix. For any fixed value of the parameter $\bar{\alpha}$, $F(\Phi, \bar{\alpha})$ has many local minima but attains its unique global minimum at $\Phi = \Phi_G(\bar{\alpha}) = A\bar{\alpha}$, with the value

$$F(\Phi_G(\bar{\alpha}), \bar{\alpha}) = -\mu. \tag{19}$$

Thus, the solution $\Phi_G(\alpha)$ to

$$\text{global min}_{\Phi} F(\Phi, \alpha) \tag{20}$$

for any fixed $\alpha$ is known a priori. The dependence of $F(\Phi, \alpha)$ on $\Phi$ is illustrated in Fig. 3 for $k = n = 2$, $\mu = 20$, $w = 5$, and

$$A = \begin{pmatrix} 0.069 & 0.501 \\ -0.399 & 0.034 \end{pmatrix},$$
$$D = \begin{pmatrix} 15.318 & 0.0 \\ 0.0 & 10.452 \end{pmatrix}, \quad \text{and} \tag{21}$$
$$\alpha = \begin{pmatrix} -0.038 \\ 0.364 \end{pmatrix}.$$

$F(\Phi, \alpha)$ has a large number of local minima, but the global minimum is at

$$\Phi_G = A\alpha = \begin{pmatrix} 0.180 \\ 0.028 \end{pmatrix}. \tag{22}$$

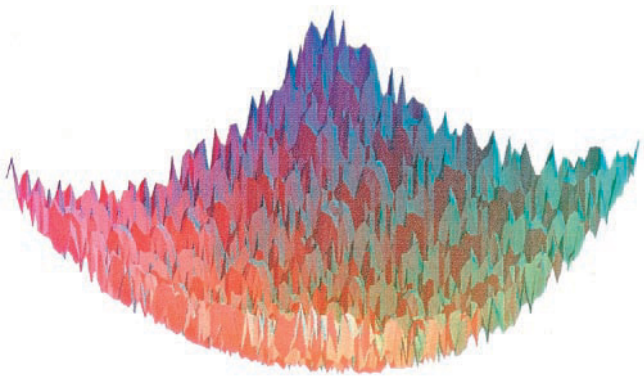

FIGURE 3 The potential function $F(\Phi, \alpha)$ for model problem 2 with $k = n = 2$.

This form of the potential function is characterized by a rugged energy landscape with numerous kinetic traps, energy barriers, and narrow pathways to the native state. Although this potential function is artificial, it shares many of the characteristics of real protein folding energy landscapes (Bryngelson and Wolynes, 1987, 1989; Chen and Dill, 1998; Dill and Chan, 1997; Leopold et al., 1992).

To extend this test problem to the more general case of $m$ different molecules with native states $\Phi_N^{(j)}, j = 1, \ldots, m$, the energy function $F^{(j)}$ for the $j$th such molecule will be given by Eq. 18 with $D = D^{(j)}$ and $A = A^{(j)}, j = 1, \ldots, m$. If the $A^{(j)}$ are different, we can usually only make $\Phi_G^{(j)} = \Phi_N^{(j)}$ for at most one landscape, where $\Phi_G^{(j)}$ is the global minimum vector $\Phi$ of Eq. 20 for the $j$th landscape. In this case, we set the weights $w_j$ given in Eq. 4 to $w_j = 1$ for $j = 1, \ldots, m$. Then the general problem of determining the vector $\alpha$ that minimizes the average angular error $\rho(\alpha)$, over all $m$ landscapes, previously defined by Eqs. 4–6, simply reduces to

$$\min_{\alpha} \rho(\alpha) = \sum_{j=1}^{m} \|\Phi_G^{(j)}(\alpha) - \Phi_N^{(j)}\|^2. \tag{23}$$

For the specific energy function in Eq. 18, we know that

$$\Phi_G^{(j)}(\alpha) = A^{(j)}\alpha. \tag{24}$$

Therefore, Eq. 23 is the simple least-squares problem for $\alpha$ given by

$$\min_{\alpha} \sum_{j=1}^{m} \|A^{(j)}\alpha - \Phi_N^{(j)}\|^2. \tag{25}$$

Its solution $\alpha^*$ is given by the solution to the system of linear equations

$$\sum_{j=1}^{m} (A^{(j)T}A^{(j)})\,\alpha = \sum_{j=1}^{m} A^{(j)T}\Phi_N^{(j)}. \tag{26}$$

Furthermore, the gradient of $F$ with respect to $\Phi$ is given by

$$\nabla_{\Phi} F = [1 + \mu w \sin(w\Psi)]D(\Phi - A\alpha). \tag{27}$$

This gradient is zero at $\Phi = A\alpha$ and at all points where $1 + \mu w \sin(w\Psi) = 0$. If the Hessian of $F$ is positive at such a point, it is a local minimum of $F$.

Because of the special form of $F(\Phi, \alpha)$, as given by Eq. 18, we know in advance the Hessian and Jacobian matrices that are needed to compute the gradient $\nabla_{\alpha}\rho(\alpha)$. As a result, we were able to use an efficient SQP local minimization directly in the parameter space to minimize $\rho(\alpha)$, as given by Eq. 23.

Test problem 2 consists of nine different landscapes, each with a different number of degrees of freedom (ranging from 4 to 12). Therefore, the vector $\Phi^{(j)}, j = 1, \ldots, 9$, for each molecule has a different dimensionality, with $\Phi^{(j)} \in \Re^{n_j}$, $n_j = j + 1$, and $j = 1, \ldots, 9$. The value of $k = 3$ is

chosen so that there are three parameters to be determined. For each molecule, an $(n_j \times 3)$ matrix $A^{(j)}$ was generated with linearly independent columns. The native state, denoted by $\Phi_N^{(j)}$, for each molecule was obtained by choosing a value $\bar{\alpha}$ and then computing $\Phi^{(j)} = A^{(j)}\bar{\alpha}$, $j = 1, \ldots, 9$. Each $\Phi^{(j)}$ was then randomly perturbed to give a native state $\Phi_N^{(j)}$. This was done so that no value of $\alpha$ exists for which $\Phi_G^{(j)}(\alpha) = \Phi_N^{(j)}$; that is, $\rho(\alpha)$ in Eq. 23 cannot be zero. From the solution to Eq. 26, we know a priori the optimal value of $\alpha^*$ and its corresponding minimum error $\rho(\alpha^*) = 2.19$. This is necessary for validating that the method can find globally optimal parameters.

A randomly chosen $\alpha_I \in \Re^3$ was used as the starting value for the parameters. This gives the initial error $\rho(\alpha_I) = 58.76$. Corresponding to this initial choice for $\alpha$, each landscape has an initial global minimum conformation $\Phi_G^{(j)}(\alpha_I)$, computed by the CGU global minimization algorithm. This value is, of course, also given directly by $A^{(j)}\alpha_I$, but we did not use this information for our tests. The ENPOP algorithm was then applied to improve $\alpha$ so as to simultaneously bring all of the global minima $\Phi_G^{(j)}(\alpha)$ as close as possible to their corresponding native conformations $\Phi_N^{(j)}$. ENPOP required five iterations to reduce $\rho$ from $\rho(\alpha_I) = 58.76$ to its minimum value $\rho(\alpha^*) = 2.19$. Each iteration gives a reduced value of $\rho$ and corresponding global minimum conformations $\Phi_G^{(j)}(\alpha)$. The corresponding values of $\rho(\alpha)$ obtained by the algorithm at each iteration were 58.76, 11.47, 9.36, 3.56, 2.78, and 2.19. The final value $\rho(\alpha^*) = 2.19$ is the known minimum possible value of $\rho(\alpha)$ for this test problem.

## CONCLUSIONS

We have described a method, called ENPOP, for optimizing the parameters in models used in computational biology, such as in folding and docking, where there is a unique structure at a global minimum of energy. ENPOP iteratively refines the parameters by enforcing the requirement that the energy minimum that represents the best predicted structures move systematically closer to the true native structures. We validate the method on two very different test problems. One is a two-dimensional short-chain protein folding problem. The other involves bumpy energy landscapes for which the optimal parameters are known in advance, to check that the method can identify the unique globally optimal parameters. While these test problems are relatively simple, they show that the ENPOP method is computationally efficient and can improve and optimize the parameters in models of the type that are commonly used in computational biology.

The authors thank the San Diego Supercomputer Center for computational resources and the National Institutes of Health (grant GM34993) and the National Science Foundation (grant DBI-9996165) for support for this research.

## REFERENCES

Bryngelson, J. D., and P. G. Wolynes. 1987. Spin-glass and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA.* 84: 7524–7528.

Bryngelson, J. D., and P. G. Wolynes. 1989. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.* 93:6902–6915.

Chan, H. S., and K. A. Dill. 1998. Protein folding kinetics from the perspectives of simple models. *Proteins Struct. Funct. Genet.* 30:2–33.

Dill, K. A., and H. S. Chan. 1997. From Levinthal to pathways to funnels. *Nature Struct. Biol.* 4:10–19.

Dill, K. A., A. T. Phillips, and J. B. Rosen. 1997a. Protein structure and energy landscape dependence on sequence using a continuous energy function. *J. Comp. Biol.* 4:227–239.

Dill, K. A., A. T. Phillips, and J. B. Rosen. 1997b. Molecular structure prediction by global optimization. *In* Developments in Global Optimization. I. M. Bomze, T. Csendes, R. Horst, and P. M. Pardalos, editors. Kluwer, Dordrecht, the Netherlands. 217–234.

Esposito, W. R., and C. A. Floudas. 1998. Global optimization in parameter estimation of nonlinear algebraic models via the error-in-variable approach. *Ind. Eng. Chem. Res.* 37:1841–1858.

Gill, P. E., W. Murray, M. A. Saunders, and M. H. Wright. 1986. User's guide for NPSOL (version 4.0): a Fortran package for nonlinear programming, Technical Report, Department of Operation Research, Stanford University.

Goldstein, R. A., Z. Luthy-Schulten, and P. G. Wolynes. 1992. Protein tertiary structure recognition using optimized Hamiltonians with local interactions. *Proc. Natl. Acad. Sci. USA.* 89:9029–9033.

Govindarajan, S., and R. A. Goldstein. 1995. Searching for folding protein structures using optimized energy functions. *Biopolymers.* 36:43–51.

Hao, M. H., and H. A. Scheraga. 1996. How optimization of potential functions affects protein folding. *Proc. Natl. Acad. Sci. USA.* 93: 4984–4989.

Hendlich, M., M. Lackner, S. Weitckus, and H. Floeckner. 1990. Identification of native protein folds amongst a large number of incorrect models—the calculation of low energy conformations from potentials of mean force. *J. Mol. Biol.* 216:167–180.

Huber, T., and A. E. Torda. 1998. Protein fold recognition without Boltzmann statistics or explicit physical basis. *Protein Sci.* 7:142–149.

Koretke, K. K., Z. Luthy-Schulten, and P. G. Wolynes. 1996. Self-consistency optimized statistical mechanical energy functions for sequence structure alignment. *Protein Sci.* 5:1043–1059.

Koretke, K. K., Z. Luthey-Schulten, and P. G. Wolynes. 1998. Self-consistently optimized energy function for protein structure prediction by molecular dynamics. *Proc. Natl. Acad. Sci. USA.* 95:2932–2937.

Leopold, P. E., M. Montal, and J. N. Onuchic. 1992. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA.* 89:8721–8725.

Maiorov, V. N., and G. M. Crippen. 1992. Contact potential that recognizes the correct folding of globular proteins. *J. Mol. Biol.* 227: 876–888.

Maiorov, V. N., and G. M. Crippen. 1994. Learning about protein folding via potential functions. *Proteins Struct. Funct. Genet.* 20:167–173.

Mirny, L. A., and E. I. Shakhnovich. 1996. How to derive a protein folding potential? A new approach to an old problem. *J. Mol. Biol.* 264: 1164–1179.

Shrivastava, I., S. Vishveshwara, M. Cieplak, A. Maritan, and J. R. Banavar. 1995. Lattice models for rapidly folding protein-like heteropolymers. *Proc. Natl. Acad. Sci. USA.* 92:9206–9209.

Thomas, P. D., and K. A. Dill. 1996. An iterative method for extracting energy-like quantities from protein structure. *Proc. Natl. Acad. Sci. USA.* 93:11628–11633.