# MOPED: Method for Optimizing Physical Energy Parameters Using Decoys

**ARTICLE** *in* JOURNAL OF COMPUTATIONAL CHEMISTRY · JANUARY 2003

Impact Factor: 3.59 · DOI: 10.1002/jcc.10124 · Source: PubMed



**4 AUTHORS**, INCLUDING:

John Damon Chodera
Memorial Sloan-Kettering Cancer Center
**64** PUBLICATIONS **3,672** CITATIONS
SEE PROFILE

Ken A Dill
Stony Brook University
**392** PUBLICATIONS **28,354** CITATIONS
SEE PROFILE

# MOPED: Method for Optimizing Physical Energy Parameters Using Decoys

CHAOK SEOK,[1] J. B. ROSEN,[2] JOHN D. CHODERA,[3] KEN A. DILL[1]
[1]*Department of Pharmaceutical Chemistry, University of California in San Francisco, San Francisco, California 94118*
[2]*Computer Science and Engineering Department, University of California in San Diego, San Diego, California 92093*
[3]*Graduate Group in Biophysics, University of California in San Francisco, San Francisco, California 94118*

*Received 23 January 2002; Accepted 26 April 2002*

**Abstract:** We present a method called MOPED for optimizing energetic and structural parameters in computational models, including all-atom energy functions, when native structures and decoys are given. The present method goes beyond previous approaches in treating energy functions that are nonlinear in the parameters and continuous in the degrees of freedom. We illustrate the method by improving solvation parameters in the energy function EEF1, which consists of the CHARMM19 polar hydrogen force field augmented by a Gaussian solvation term. Although the published parameters for EEF1 correctly discriminate the native from decoys in the decoy sets of Levitt et al., they fail on several of the more difficult decoy sets of Baker et al. MOPED successfully finds improved parameters that allow EEF1 to discriminate native from decoy structures on all protein structures that do not have metals or prosthetic groups.

 J Comput Chem 24: 89–97, 2003

**Key words:** MOPED; computational models; EEF1

## Introduction

Predicting the folded structure of a protein or the docked conformation of a ligand with a protein depends on having a suitable model of the energetics of folding or binding. An essential property of a good energy model is the ability to distinguish the native from non-native structures. Such energy models have been created in various ways. First, there are physical potentials that are derived by fitting to *ab initio* quantum mechanics and to experimental data on small molecules. Among the various classes of energy models, it is believed that the physical models have the advantage of being the most transferable from one prediction problem to the next. The disadvantages are that the physical models are complex, and they have a large number of parameters. Second, there are low-resolution models combined with simple potentials, such as united-residue models with contact potentials. Statistical potentials are derived from contact frequencies in databases of known protein structures and have been widely used because of their simplicity.

Third, another method has recently emerged for deriving low-resolution energy scoring functions, based on "decoys."[1–6] Decoys are non-native protein-like conformations that lie in low-energy regions of conformation space, often generated by threading methods or other folding algorithms. Parameters are learned by maximizing an energy gap between the native and decoy structures.[7,8] Scoring functions derived in this way can often discriminate decoys better than some of the statistical potentials.[9]

Decoy-based parametrization has been enabled by two important simplifications: the linearity of the model energy as a function of parameters and the discreteness of the conformational search. Linearity arises because contact potentials are simple products of the numbers $n_{ij}$ of contacts between monomer types $i$ and $j$, and a coefficient with units of energy $\epsilon_{ij}$; that is, the energy, $\Sigma_{ij}\epsilon_{ij}n_{ij}$, is linear in the parameters $\epsilon_{ij}$. The discreteness of the conformational search results from the modest requirement that the energy parameters merely need to be good enough to give a lower energy for the native structure than for any of the decoys in a small set of static structures. Because of these simplifications, standard linear programming methods can rapidly converge on a set of parameters that can succeed at this discrimination task.

***Correspondence to:*** K. A. Dill; e-mail: dill@zimm.uscf

Contract/grant sponsor: NSF ITR program

Contract/grant sponsor: Howard Hughes Institute Graduate Fellowship (to J.D.C.)

The main problem with decoy-based parametrization lies not in the parametrization philosophy itself but in the potentials that it is used to parametrize—contact potentials cannot perfectly discriminate decoys from native for large numbers of decoys, for any parameter set.[9–12] The thermodynamic hypothesis states that the native state of a protein is the state with the lowest free energy.[13] Therefore, a truly correct free energy function should be able to discriminate the native from a nearly infinite list of decoys. The failure of contact potentials implies that the energy functions often used are not sufficiently physical. All-atom force fields, which are usually taken to be the most physical energy functions, are intended to model nature's free energy function. In calculations, the protein conformational and vibrational entropies are often neglected. The remaining quantity is sometimes called the "effective energy";[14] here, we refer to it as the "energy." The energy of the native conformation is also expected to be lower than any nonnative conformations.[14]

Decoy-based parametrization has not, so far, been applicable to physical energy functions, such as CHARMM or AMBER, or other all-atom force fields, for two reasons. First, physical energies are not always linear in parameters: there are nonlinear torsion parameters, exponents in power-law terms, products of charges in electrostatic terms, as well as nonlinear relationships between internal and Cartesian degrees of freedom. Second, folding and binding involve continuum degrees of freedom. As a parameter optimization method iterates to generate a new set of parameters, the original native and decoy structures may no longer be local minima in the energy as a function of conformational degrees of freedom. In the optimization of contact potentials, computed scores are not very sensitive to small changes in configuration; hence, the native and decoy structures can be held fixed. With all-atom physical energy functions, perturbations in parameters result in changes in the positions of local minima, so the energy evaluated at a static conformation can vary rapidly, introducing noise. A better strategy would recognize that every change in parameters should also lead to a change in the decoys, to be at least at local minima. Discriminating the minimized native from the minimized decoys is more challenging than holding the structures fixed as parameters are varied.

Here we combine the advantages of physical models with the advantages of a decoy-based parameter optimization approach in a method we call MOPED: Method for Optimizing Physical Energy parameters using Decoys. Using a physical energy function and an initial parameter set, we modify the parameters in an attempt to satisfy the constraint that, for a training set of proteins, the energy of the minimized native conformation be lower than the minimized energies of a set of decoys (see Fig. 1). We recognize that this is a necessary (but not sufficient) requirement for creating an energy function with its global minimum at the true native structure. Parameters thus obtained are not guaranteed to be optimal, because there can be decoys not included in the training set that have lower energy than the native. Nevertheless, if the size and quality of the decoy set is sufficient, this strategy can at least lead to improved parameters.

In a similar spirit, Meirovitch et al.[15,16] optimized solvation parameters associated with solvent accessible surface areas in all-atom physical energy functions. However, their search of parameter space is not systematic, and applied to only a few parameters: three parameters in one case[15] and six in another.[16] In our approach, parameter search is directed by effectively performing a gradient descent in parameter space. Our strategy can thus handle much larger number of parameters (76 parameters are adjusted in this study), and also can be applied to larger proteins and larger training sets.

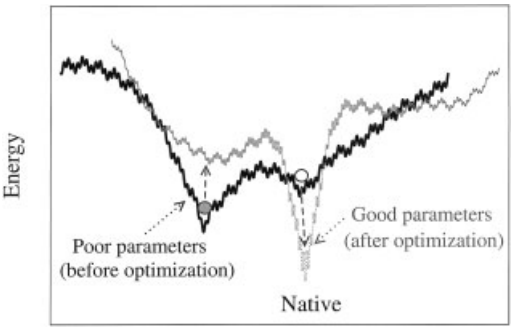


**Figure 1.** Changing parameters in a model can cause a change in the relative energies of local minima. MOPED aims to find parameters for which $E_{\text{Native}} < E_{\text{Decoy}}$.

We apply MOPED to refining the solvation parameters of the effective energy function EEF1, developed by Lazaridis and Karplus, which consists of the CHARMM19 polar hydrogen force field augmented by a pairwise solvation free energy term derived from a Gaussian solvent-exclusion model.[17,18] EEF1 is computationally efficient: the solvation contribution takes only 50% of the computer time for the vacuum CHARMM19 energy. EEF1 has previously been demonstrated to successfully discriminate native from decoy structures among the "4state_reduced" decoy sets of Levitt et al.[19] We find here, however, that EEF1 fails to discriminate some natives from decoys in the more challenging ROSETTA decoy sets produced by Baker et al.[20] We, therefore, use ROSETTA decoys as a test problem to see if our optimization strategy can give an improved parameter set for discriminating both the Levitt and Baker decoys. We find that it succeeds at this task, and that the parameters obtained from learning on only a few proteins appear to be transferable to other proteins.

## Methods

### *The MOPED Method for Parameter Optimization*

We consider an energy function of the form $E(\mathbf{R}; \alpha, \mathbf{s})$, where $\mathbf{R}$ is the protein conformation represented by the coordinates of atoms or interaction centers, $\alpha$ is the vector describing the parameters to be optimized, and $\mathbf{s}$ is the amino acid sequence. Given a functional form for the energy function, the goal is to find a parameter set $\alpha$ that satisfies the following requirement,

$$E(\mathbf{Y}; \alpha, \mathbf{s}) > E(\mathbf{X}; \alpha, \mathbf{s}), \quad (1)$$

for all non-native conformations $\mathbf{Y}$ and native conformations $\mathbf{X}$, for all the proteins in the training set.

What is an appropriate condition for choosing a parameter set $\alpha$? One strategy that has been widely used is to choose parameters that maximize an energy gap between the native and decoys. However, we prefer not to maximize an energy gap because it can lead to nonphysical, hence possibly nontransferable, parameters. Because our interest here is in trying to keep the energy function as physical as possible, we use a different strategy. We only require that the energy gap be larger than some small positive value $\Delta E_{\min}$. Ultimately, we choose the value of $\Delta E_{\min}$ to give a high transferability, i.e., so that we learn parameters from the smallest possible set of proteins to succeed in predicting structures of the largest possible set of proteins.

We define the energy minimized natives $\hat{\mathbf{X}}^{(i)}$ for proteins $i = 1, 2, 3, \ldots$ and the energy minimized decoy conformations $\hat{\mathbf{Y}}_j^{(i)}$ for decoys $j = 1, 2, 3, \ldots$ of protein $i$. These are the conformations that are at local energy minima for a given set of parameters $\alpha$, $\hat{\mathbf{X}}^{(i)}(\alpha)$ and $\hat{\mathbf{Y}}_j^{(i)}(\alpha)$. We now redefine the energy constraints as

$$E(\hat{\mathbf{Y}}_j^{(i)}(\alpha); \alpha, \mathbf{s}^{(i)}) - E(\hat{\mathbf{X}}^{(i)}(\alpha); \alpha, \mathbf{s}^{(i)}) >= \Delta E_{\min}, \qquad (2)$$

for all proteins $i$ and decoys $j$.

The energies in eq. (2) are evaluated at the corresponding local energy minimum conformation $\hat{\mathbf{X}}^{(i)}(\alpha)$ and $\hat{\mathbf{Y}}_j^{(i)}(\alpha)$, not at the fixed initial conformations $\mathbf{X}^{(i)}$ and $\mathbf{Y}_j^{(i)}$. The local energy minima depend on the choice of the initial conformations for energy minimizations. We compute the local energy minima for the current parameter set by performing energy minimizations from the fixed initial native and decoy conformations because the standard procedure for validating potentials by decoy discrimination uses this method. Unfortunately, the local minimum do not change as a continuous function of parameters because the initial conformations may belong to different local energy basins of attraction as the energy surface changes with parameters. Tracking local energy minima in successive energy minimizations as parameters are varied, instead of minimizing energy starting from the fixed conformations, can lead to fewer problems with discontinuities in energy. However, the resulting local minima can depend upon the path taken through the parameter space.

Because of the discontinuities in local minimum energy with respect to parameters, we use an iterative method, where each iteration involves changing the parameters slightly while holding the conformations fixed. At fixed conformation $\mathbf{R}$, $E(\mathbf{R}; \alpha, \mathbf{s})$ is a well-behaved function of $\alpha$, so optimization can be performed efficiently. The conformations are then minimized in energy at the resulting parameters, and this procedure is repeated.

We formulate the parameter optimization problem at fixed conformations as an optimization problem with nonlinear constraints as follows: maximize $\gamma$ with respect to the parameter set $\alpha$ and $\gamma$ itself under the constraints

$$E(\hat{\mathbf{Y}}_j^{(i)}(\alpha'); \alpha, \mathbf{s}^{(i)}) - E(\hat{\mathbf{X}}^{(i)}(\alpha'); \alpha, \mathbf{s}^{(i)}) >= \gamma, \qquad (3)$$

for all $i$ and $j$, where $\hat{\mathbf{Y}}_j^{(i)}(\alpha')$ and $\hat{\mathbf{X}}^{(i)}(\alpha')$ are local energy minima at parameter $\alpha'$, and thus independent of $\alpha$. The constraints are nonlinear because parameters are, in general, nonlinear in the energy function. The nonlinear constraints always have a feasible solution for a sufficiently negative choice of the initial value of $\gamma$. We solve this nonlinear optimization problem with the package SNOPT,[21] which uses an efficient Sequential Quadratic Programming method.

The algorithm is as follows:

1. Start with the initial parameter set $\alpha_k$ ($k = 0$).
2. Perform energy minimizations on native and decoy conformations to obtain local minima $\hat{\mathbf{R}}(\alpha_k)$.
3. Fix the conformations at $\hat{\mathbf{R}}(\alpha_k)$ and maximize $\gamma$ with respect to $\alpha$ and $\gamma$, starting from $\alpha_k$. In the optimization, the parameter range is set to $[\alpha_k - \delta\alpha, \alpha_k + \delta\alpha]$ with small $\delta\alpha$ so that fixing the conformations is not a bad approximation. Let the resulting parameter set be $\alpha_{k+1}$.
4. Increase $k$ by 1 and repeat steps 2 and 3 until eq. (2) is satisfied.

In our calculations, $\delta\alpha$ is taken to be $0.02(\alpha_{\text{upper}} - \alpha_{\text{lower}})$, where the total parameter range $[\alpha_{\text{lower}}, \alpha_{\text{upper}}]$ is set to $[-30, 5]$ kcal/mol for $\Delta G^{\text{free}}$, [2, 10] Å for $\lambda$, [0, 30] Å$^3$ for $V$, and [0.5, 3] Å for $R$. (See later for descriptions of the various solvation parameters.) $\Delta E_{\min}$ is set to 5 kcal/mol because this value appears to result in good transferability.

For each protein in the training set, only 20 (out of 1000) lowest energy decoys for the initial parameter set are included in the iterative parameter optimization procedure. At each SNOPT optimization step (step 3), only 10 out of the 20 lowest energy for the previous parameter set are included. In the constrained optimizations (both outer iteration and SNOPT optimization) only those decoys with the lowest energy determine the active constraints. Furthermore, it is only the active constraints that affect the solution (inactive constraints can be removed without changing the optimal solution). The choice of 20 and 10 is a compromise between the desire for a fast solution and the need to include all the active constraints.

### *Monte Carlo Optimization*

To assess relative performance of the MOPED, we have compared it with a Monte Carlo method for optimizing parameters. A subset of elements of the current parameter vector is perturbed, and the change is accepted or rejected with the Metropolis criterion based on the change in the minimum energy gap $\min_{i,j}[E(\hat{\mathbf{Y}}_j^{(i)}(\alpha); \alpha, \mathbf{s}^{(i)}) - E(\hat{\mathbf{X}}^{(i)}(\alpha); \alpha, \mathbf{s}^{(i)})]$. The temperature parameter, $kT$, is set to 1 kcal/mol to allow fluctuations in the energy gap on the order of 1 kcal/mol. Whether each parameter is perturbed or not is independently decided with probability $P$. Perturbations are drawn uniformly over the interval $[\alpha_{\text{current}} - \delta\alpha, \alpha_{\text{current}} + \delta\alpha]$, where $\delta\alpha = \eta(\alpha_{\text{upper}} - \alpha_{\text{lower}})$. $P = 0.1$ and $\eta = 0.05$ gave an acceptance ratio of about 50% and resulted in better energy gaps than several other choices tried.

### *EEF1 Energy Function and Parameters*

For simplicity as a test of our parameter optimization methodology, we do not vary all the parameters in the EEF1. We only vary 76 solvation parameters. The other parameters could also have been varied with only a moderate increase in computational costs.

We have implemented the EEF1 energy function as described in ref. 17, except that a switching function between 7 and 9 Å is

**Table 1.** Lazaridis–Karplus (LK) Parameters for the EEF1 Solvation Free Energy.

| Atom types | $\Delta G^{ref}$ | $\Delta G^{free}$ | $V$ | $\lambda$ | $R$ |
|---|---|---|---|---|---|
| H | 0.000 | 0.00 | 0.0 | 3.5 | 0.800 |
| HC | 0.000 | 0.00 | 0.0 | 3.5 | 0.600 |
| CR | −0.890 | −1.40 | 8.3 | 3.5 | 2.100 |
| C | 0.000 | 0.00 | 14.7 | 3.5 | 2.100 |
| CH1E | −0.187 | −0.25 | 23.7 | 3.5 | 2.365 |
| CH2E | 0.372 | 0.52 | 22.4 | 3.5 | 2.235 |
| CH3E | 1.089 | 1.50 | 30.0 | 3.5 | 2.165 |
| CR1E | 0.057 | 0.08 | 18.4 | 3.5 | 2.100 |
| N | −1.000 | −1.55 | 0.0 | 3.5 | 1.600 |
| NR | −3.820 | −4.00 | 4.4 | 3.5 | 1.600 |
| NH1 | −5.950 | −8.90 | 4.4 | 3.5 | 1.600 |
| NH2 | −5.450 | −7.80 | 11.2 | 3.5 | 1.600 |
| NH3 | −20.000 | −20.00 | 11.2 | 6.0 | 1.600 |
| NC2 | −10.000 | −10.00 | 11.2 | 6.0 | 1.600 |
| O | −5.330 | −5.85 | 10.8 | 3.5 | 1.600 |
| OC | −10.000 | −10.00 | 10.8 | 6.0 | 1.600 |
| OH1 | −5.920 | −6.70 | 10.8 | 3.5 | 1.600 |
| S | −3.240 | −4.10 | 14.7 | 3.5 | 1.890 |
| SH1E | −2.050 | −2.70 | 21.4 | 3.5 | 1.890 |

added to the solvation free energy instead of a cutoff at 9 Å, as in ref. 17, to obtain continuous gradients, and thus better behavior during energy minimization. The introduction of the switching function does not significantly alter the behavior of the energy function in discriminating the decoy sets.

The Lazaridis and Karplus (LK) solvation free energy term in EEF1 is approximated by the reference solvation free energy minus the excluded volume contribution from the neighboring atoms. In EEF1, it is assumed that the solvation free energy $\Delta G^{solv}$ is expressed as the sum of group contributions:

$$\Delta G^{solv} = \sum_i \Delta G_i^{solv}. \quad (4)$$

The solvation free energy contribution for group $i$, $\Delta G_i^{solv}$ is obtained by correcting the reference solvation free energy, $\Delta G_i^{ref}$, which is the solvation free energy for a fully solvent-exposed group, by the contribution from the volume occupied by other groups as follows:

$$\Delta G_i^{solv} = \Delta G_i^{ref} - \sum f_i(r_{ij})V_j, \quad (5)$$

where $V_j$ is the volume of group $j$ and $f_i(r_{ij})$ is the solvation free energy density of group $i$ at distance $r_{ij}$. Solvation effects other than solvent exclusion are treated by using a distance-dependent dielectric constant and neutralizing ionic side chains. The solvation free energy density $f_i(r_{ij})$ is taken to be a Gaussian function of the form

$$f_i(r_{ij}) = \frac{\Delta G_i^{free}}{2\pi\sqrt{\pi}\,\lambda_i r_{ij}^2} \exp\left[-\left(\frac{r_{ij} - R_i}{\lambda_i}\right)^2\right], \quad (6)$$

where $\Delta G_i^{free}$ is the solvation free energy of the isolated free atom group $i$, $\lambda_i$ is a correlation length, and $R_i$ is the radius of group $i$.

The parameters in this solvation model are $\Delta G_i^{ref}$, $\Delta G_i^{free}$, $V_i$, $\lambda_i$, and $R_i$ for each group $i$. LK considered 19 types of atomic groups in EEF1. They are listed in Table 1, with their parameter values. For $\Delta G_i^{ref}$, LK use the values obtained by a linear decomposition analysis on experimentally measured hydration free energies for small, extended molecules.[22] LK added a small correction of 0.2 kcal/mol to account for the long-range Lennard–Jones interactions. $\Delta G_i^{free}$ is obtained by requiring that the solvation free energy of deeply buried groups is approximately zero. $\Delta G_i^{ref}$ and $\Delta G_i^{free}$ for charged groups are arbitrarily assigned large negative values to produce a large desolvation penalty for those groups. LK take the van der Waals radius of group $i$ for $R_i$ and obtain the volume $V_i$ from this radius by subtracting average overlapping volumes due to the neighboring atoms. $\lambda_i$ is taken to be 3.5 Å, which is about the thickness of the first hydration shell, except for charged atom groups for which a larger value of 6 Å is used.

We use MOPED to improve the solvation parameters, fixing all other nonsolvation parameters, including partial charges, at the canonical values given by EEF1. The quantities $\Delta G_i^{ref}$ are also fixed because their contribution is conformation independent. We vary all other parameters $\Delta G_i^{free}$, $V_i$, $\lambda_i$, and $R_i$ for 19 atom types, yielding a total of 76 free parameters. ($V_i$ and $R_i$ are varied independently.) The improved parameter set was obtained by training on the decoy sets 1orc and 1utg (from Baker decoy sets) and 4pti (from Levitt decoy sets), and we refer to it as SD. The SD parameter values are shown in Table 2. It is shown in Figures 2–5, that with the SD parameters, the native protein energy is lower than all the corresponding decoys for every protein with available decoys.

**Table 2.** Improved Parameter Set SD for the EEF1 Solvation Free Energy.

| Atom types | $\Delta G^{free}$ | $V$ | $\lambda$ | $R$ |
|---|---|---|---|---|
| H | 0.0000 | 0.0000 | 3.5000 | 0.8000 |
| HC | 0.0000 | 0.0000 | 3.5000 | 0.6000 |
| CR | 0.0000 | 8.2925 | 3.5000 | 2.1500 |
| C | 0.0000 | 15.4038 | 3.5000 | 2.1000 |
| CH1E | −0.0886 | 25.8419 | 4.6249 | 2.2650 |
| CH2E | 4.5190 | 17.7931 | 4.8611 | 2.6350 |
| CH3E | 3.0494 | 30.0000 | 3.3144 | 2.5736 |
| CR1E | −1.4413 | 25.7571 | 3.1602 | 2.0000 |
| N | −1.4451 | 0.0000 | 2.6198 | 1.5500 |
| NR | −6.0513 | 4.4000 | 3.6008 | 1.9500 |
| NH1 | −8.3401 | 6.4843 | 2.9691 | 1.5000 |
| NH2 | −7.8305 | 11.3706 | 3.1955 | 1.5508 |
| NH3 | −16.8852 | 12.0042 | 4.1677 | 1.2500 |
| NC2 | −8.9949 | 10.4103 | 3.5885 | 1.1000 |
| O | −8.6925 | 10.2000 | 2.9156 | 1.6848 |
| OC | −10.7229 | 7.4402 | 4.4528 | 1.8002 |
| OH1 | −12.4117 | 12.4926 | 3.9414 | 2.3000 |
| S | −5.0149 | 14.7122 | 2.6198 | 1.4400 |
| SH1E | −4.4273 | 21.5410 | 2.9966 | 2.0900 |

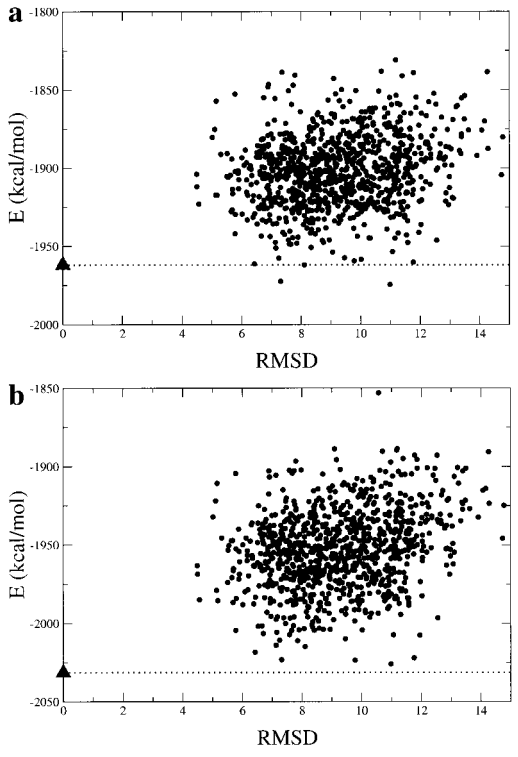


**Figure 2.** (a) Minimized energy at the LK parameters vs. $C_\alpha$ RMSD (in Å) from the native structure for the ROSETTA decoy set 1orc. Native energy is marked as a triangle at zero RMSD, and decoys are represented by dots. (b) Minimized energy at the SD parameters versus $C_\alpha$ RMSD (in Å) for the ROSETTA decoy set 1orc. 1orc is one of the training proteins on which the SD parameters are learned.

### *Decoy Sets and Energy Minimization*

ROSETTA all-atom decoy sets[20,23] were used to improve and test parameters. We only used those proteins that have X-ray structures. We did not consider proteins having NMR structures because we felt that they would require more sophistication in handling ensembles, and more computational refinement than crystal structures. We also excluded proteins with metals or prosthetic groups due to the lack of a functional form to describe their interaction. As a result, we used 21 out of 92 total ROSETTA decoy sets. Each protein set has 1000 or 2000 decoy structures. Most of the ROSETTA decoy structures have several truncated end residues, so we model those missing residues by attaching the conformations of the crystal structure and minimizing the energy.

The improved parameters that were learned by MOPED were then also tested on the Levitt decoy sets:[24,19] "4state_reduced," "fisa," "fisa_casp3," "lmds," and "lattice_sshift." Each set has between 200 and 2000 decoy structures.

In all cases, we modeled missing atoms such as polar hydrogens, and added disulfide bonds. Both the crystal structures and the decoy structures are then initially minimized in energy with the LK parameter set, and the resulting structures are taken as the initial structures for parameter optimization by MOPED. This speeds up the energy minimization in the parameter optimization because bad contacts in the models are removed in advance. Energy minimizations are performed in Cartesian coordinates with the L-BFGS-B algorithm,[25] using a maximum gradient component termination criterion of 1 kcal/mol Å.

## Results and Discussion

### *Tests of the LK Parameter Set*

Lazaridis and Karplus previously demonstrated that the EEF1 model can perfectly discriminate the "4state_reduced" Levitt decoys from their corresponding native structures.[18] We refer to this initial parametrization as the LK parameter set. We have tested the same parameter set on more extensive sets of decoys, as described earlier. All the native and decoy structures are minimized in energy and compared. The number of decoys with lower energy than the native conformation, $n_d$, and the energy gap between the native and the lowest lying decoy, $\Delta E$, are shown for various decoy sets in the second columns of Tables 3, 4, and 5. The values of the energy gaps that we found are not necessarily the same as those calculated as in ref. 18 because (1) our implementation of EEF1 employs a switching function for the Gaussian solvation term, and (2) different energy minimization termination criteria are used, as explained earlier.

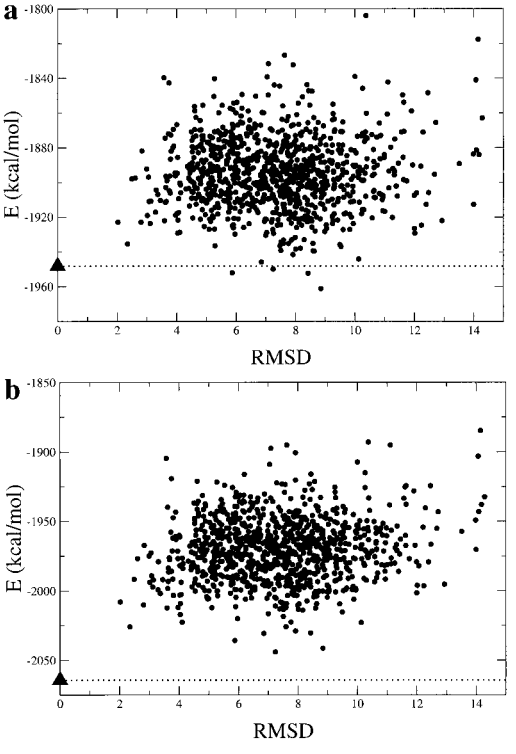


**Figure 3.** (a) Minimized energy at the LK parameters vs. $C_\alpha$ RMSD (in Å) for the ROSETTA decoy set 1r69. (b) Minimized energy at the SD parameters vs. $C_\alpha$ RMSD (in Å) for the ROSETTA decoy set 1r69. 1r69 is not included in the training set for the SD parameters.

Tables 4 and 5 confirm that the LK parameter set works perfectly ($n_d = 0$) on all of the Levitt decoys except for the proteins having bound metals or other prosthetic groups. We do not consider the latter cases to be failures of LK because the specialized intermolecular interactions are not taken into account due to the lack of a proper functional form.

Although the LK parameter set succeeds on the Levitt decoys, it fails on 4 out of 21 proteins in the ROSETTA decoy sets. In these cases, at least one decoy conformation is lower in energy than the native ($n_d > 0$ and $\Delta E < 0$) with the LK parameter set, as shown in Table 3. This suggests that the ROSETTA decoy sets are more challenging than the Levitt decoys. For example, the LK parameter set succeeds on the "4state_reduced" Levitt decoys for the protein 1r69, but it fails on the ROSETTA decoys for the same protein.

Figures 2a, 3a, and 4a show three instances where the LK parameters fail, and Figure 5a shows one instance in which the LK parameters succeed.

### *Testing the Optimized Parameters*

Our strategy here is to focus on the few failures of the LK set, learn a new parameter set that succeeds on these cases, then check that the new parameter set also remains successful on the many other molecules for which the LK set is adequate. This is equivalent to including more proteins in the training set if the resulting parameters succeed in decoy discrimination for those proteins. Parameters that satisfy eq. (2) with $\Delta E_{\min} = 5$ kcal/mol were obtained in a small number of iterations (~20). Test results of the trained parameters for different training sets are summarized in Table 6. For all cases, the proteins in the training set have energy gap $\Delta E >$ 5 kcal/mol, as it should be.

MOPED parameters that are learned from the two proteins, 1utg and 1orc, are successful on all the ROSETTA decoy sets, including the four proteins for which the LK parameter set fails. But we found that those parameters then failed on 4pti from the "4state_reduced" Levitt decoy set. Hence, we added that protein to the training set and relearned these three proteins. We call this final set the SD parameter set; these parameters are listed in Table 2. The difference between the LK and SD parameters is $\|\alpha - \alpha'\|/\sqrt{\|\alpha\|\|\alpha'\|} = 0.3$ for $\Delta G^{\text{free}}$, 0.25 for λ, 0.16 for $V$, and 0.16 for $R$. The improved parameter set now succeeds at decoy discrimination for previously problematic proteins (see Figs. 2b, 3b, and 4b) and performs somewhat better than the LK parameters on proteins omitted from the training set: on average, $n_d$ is smaller and $\Delta E$ is larger (see Tables 3, 4, 5, and 6).

Analysis of the energy components shows that the contribution of the solvation free energy to the total energy has increased in SD relative to LK approximately by 10%. Further analysis of the solvation free energy for each atom type shows that the top two atom types that contribute most to the increase in the energy gap

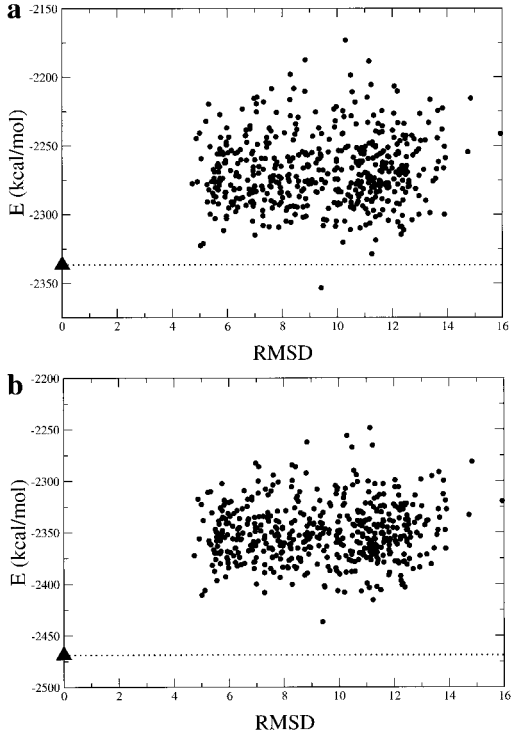


**Figure 4.** (a) Minimized energy at the LK parameters versus $C_\alpha$ RMSD (in Å) for the "fisa" Levitt decoy set 4icb. (b) Minimized energy at the SD parameters vs. $C_\alpha$ RMSD (in Å) for the "fisa" Levitt decoy set 4icb. 4icb is not included in the training set for the SD parameters.

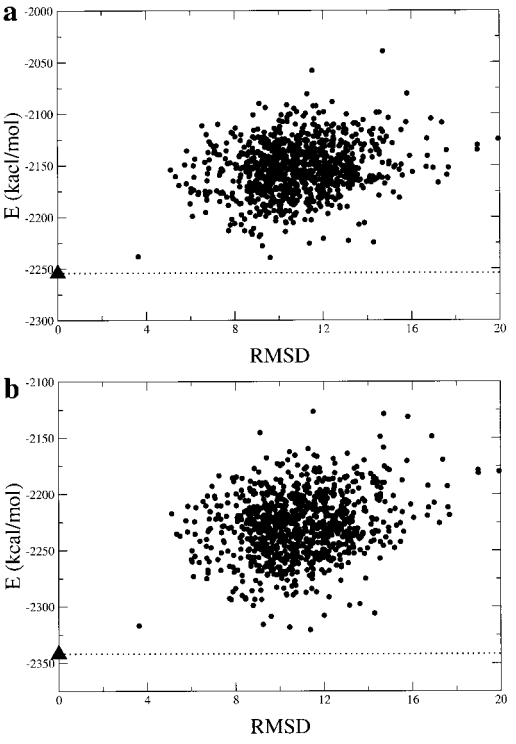


**Figure 5.** (a) Minimized energy at the LK parameters vs. $C_\alpha$ RMSD (in Å) for the ROSETTA decoy set 1ail. (b) Minimized energy at the SD parameters vs. $C_\alpha$ RMSD (in Å) for the ROSETTA decoy set 1ail. 1ail is not included in the training set for the SD parameters.

for the 10 lowest-energy decoys are {CH2E, CH3E} for 1orc, 1vls, and 1r69 and {CH2E, NH3} for 1utg. Therefore, the increase in the solvation free energy for carbon atoms contributes to correct decoy discrimination the most. This also explains the observation that trained parameters on the 1orc set are transferable to 1vls and 1r69, but not to 1utg (see Table 6).

We also performed a Monte Carlo search in parameter space, to compare to the MOPED method. Three Monte Carlo runs were performed with the training set consisting of 1utg and 1orc. Two of the three runs gave $\Delta E > 5$ kcal/mol after about 440 iterations, but the other run could not reach the desired energy gap even after 1000 iterations. When the Monte Carlo method does converge on good parameters, it takes about 17 times longer than MOPED. Hence, the MOPED method appears to be relatively efficient, and much better than Monte Carlo.

### *Simplifying the Solvation Model*

The existence of a fast parameter optimization method such as MOPED allows us to seek simpler energy functions that might have the same discrimination power. In general, if the minimum energy gap $\min_{i,j}[E(\hat{\mathbf{Y}}_j^{(i)}(\alpha);\ \alpha,\ \mathbf{s}^{(i)}) - E(\hat{\mathbf{X}}^{(i)}(\alpha);\ \alpha,\ \mathbf{s}^{(i)})]$ is a nonlinear function of parameters, there can be multiple feasible regions in the parameter space. Therefore, different feasible regions could be found by starting from different initial parameter sets. In our case, we were interested to see if we could simplify the EEF1 solvation model and still achieve the same discrimination power in the reduced parameter space. We now use a modified functional form for the Gaussian solvation free energy density eq. (6) from earlier.

$$f_i^{(0)}(r_{ij}) = \frac{\Delta G_i^{\text{free}}}{2\pi\sqrt{\pi}\,\lambda_i r_{ij}^2}\exp\left[-\left(\frac{r_{ij}}{\lambda_i}\right)^2\right]. \tag{7}$$

We were motivated to choose this form because it is much simpler than the original for integrating the solvation free energy density with respect to the coordinates. This function is equivalent to setting the parameter $R_i$ for all atom types to zero, so that it is no longer part of the parameter optimization. The initial parameter values for $\Delta G_i^{\text{free}}$ and $\lambda_i$ are obtained by matching the solvation free energy densities for $f_i(r_{ij})$ and $f_i^{(0)}(r_{ij})$ at $r_{ij} = R_i$ and $r_{ij} = r_{1/2}$, where $R_i$ is the van der Waals radius and $r_{1/2}$ is defined by $r$ at which the radial solvation free energy density $4\pi r_{ij}^2 f_i(r_{ij})$ takes half of its value at $r_{ij} = R_i$. This initial parameter set is referred to as $\text{LK}_r$, and is shown in Table 8. There are now 57 free variables, rather than 76, because the $R_i$s are fixed.

MOPED was run to obtain an improved set of parameters, starting from the $\text{LK}_r$ set. Different sets of training proteins were tried, and a parameter set that succeeds on all the cases for which $\text{LK}_r$ fails were obtained for the training set of the two proteins

**Table 3.** Test Results of LK (Lazaridis–Karplus Parameters), SD (Improved Parameters by MOPED), $\text{LK}_r$ (Initial Parameters for the Reduced Model), and $\text{SD}_r$ (Improved Parameters by MOPED for the Reduced Model) Parameter Sets on the ROSETTA Decoy Sets.

| Test proteins | Parameter set | | | |
|---|---|---|---|---|
| | LK | SD | $\text{LK}_r$ | $\text{SD}_r$ |
| 1utg | 1 (−3) | 0 (6) | 0 (3) | 0 (4) |
| 1orc | 3 (−13) | 0 (5) | 4 (−7) | 0 (6) |
| 1r69 | 4 (−13) | 0 (20) | 5 (−10) | 0 (17) |
| 1vls | 1 (−16) | 0 (15) | 1 (−11) | 0 (43) |
| 1lis | 0 (10) | 0 (19) | 1 (−14) | 0 (6) |
| 1lfb | 0 (65) | 0 (74) | 0 (2) | 0 (29) |
| 1pgx | 0 (74) | 0 (79) | 0 (0) | 0 (7) |
| 1aa2 | 0 (44) | 0 (66) | 0 (55) | 0 (85) |
| 1acf | 0 (77) | 0 (68) | 0 (64) | 0 (130) |
| 1aho | 0 (93) | 0 (99) | 0 (74) | 0 (112) |
| 1ail | 0 (15) | 0 (21) | 0 (16) | 0 (35) |
| 1csp | 0 (38) | 0 (50) | 0 (24) | 0 (43) |
| 1erv | 0 (124) | 0 (151) | 0 (193) | 0 (253) |
| 1gvp | 0 (39) | 0 (45) | 0 (56) | 0 (58) |
| 1kte | 0 (60) | 0 (93) | 0 (72) | 0 (98) |
| 1lz1 | 0 (225) | 0 (144) | 0 (150) | 0 (225) |
| 1msi | 0 (45) | 0 (65) | 0 (43) | 0 (61) |
| 1pdo | 0 (107) | 0 (143) | 0 (220) | 0 (270) |
| 1ris | 0 (48) | 0 (20) | 0 (36) | 0 (79) |
| 1tul | 0 (80) | 0 (71) | 0 (76) | 0 (96) |
| 1who | 0 (62) | 0 (79) | 0 (65) | 0 (103) |

Number of decoys with lower energy than the native conformation ($n_d$) is shown together with the energy difference between the lowest energy decoy and the native ($\Delta E$) in parentheses.

**Table 4.** Test Results of LK, SD, $\text{LK}_r$, and $\text{SD}_r$ Parameter Sets for EEF1 and Tobi-Elber (TE) Parameter Set for Their Simple Energy Function on the Levitt Decoy Sets.

| | LK | SD | $\text{LK}_r$ | $\text{SD}_r$ | TE |
|---|---|---|---|---|---|
| | | | 4state_reduced | | |
| 1r69 | 0 (26) | 0 (56) | 0 (23) | 0 (48) | 0 |
| 2cro | 0 (4) | 0 (18) | 0 (22) | 0 (22) | 0 |
| 4pti | 0 (23) | 0 (6) | 0 (19) | 0 (17) | 6 |
| | | | fisa | | |
| 2cro | 0 (19) | 0 (58) | 0 (22) | 0 (54) | 0 |
| | | | fisa_casp3 | | |
| 1jwe | 0 (36) | 0 (61) | 3 (−10) | 0 (25) | 0 |
| | | | lattice_ssfit | | |
| 1beo | 0 (186) | 0 (166) | 0 (169) | 0 (189) | — |
| 1nkl | 0 (22) | 0 (20) | 0 (30) | 0 (40) | 0 |
| 1pgb | 0 (96) | 0 (94) | 0 (104) | 0 (129) | 0 |
| | | | lmds | | |
| 1dtk | 0 (14) | 0 (6) | 0 (18) | 0 (24) | 4 |
| 1igd | 0 (11) | 0 (22) | 1 (−2) | 0 (3) | 1 |
| 2cro | 0 (31) | 0 (86) | 0 (32) | 0 (80) | 0 |
| 2ovo | 0 (19) | 0 (26) | 0 (24) | 0 (31) | 0 |
| 4pti | 0 (12) | 0 (13) | 0 (18) | 0 (31) | — |

Proteins with bound extra molecules except for water are not included, and they are shown separately in Table 5. Meaning of the numbers in each entry is the same as in Table 3.

**Table 5.** Test Results of LK, SD, $LK_r$, and $SD_r$ Parameter Sets for EEF1 and Tobi-Elber (TE) Parameter Set for Their Simple Energy Function on the Levitt Decoy Sets.

| | LK | SD | $LK_r$ | $SD_r$ | TE |
|---|---|---|---|---|---|
| | | | 4state_reduced | | |
| 1ctf | 0 (46) | 0 (50) | 0 (44) | 0 (52) | 0 |
| 1sn3 | 0 (64) | 0 (67) | 0 (45) | 0 (52) | 5 |
| 3icb | 0 (7) | 0 (21) | 0 (16) | 0 (29) | — |
| 4rxn | 37 (−48) | 27 (−52) | 35 (−50) | 83 (−63) | 15 |
| | | | fisa | | |
| 1fc2 | 94 (−28) | 31 (−18) | 224 (−31) | 164 (−30) | 15 |
| 1hdd | 0 (18) | 0 (29) | 0 (18) | 0 (31) | 0 |
| 4icb | 1 (−17) | 0 (32) | 3 (−30) | 0 (8) | — |
| | | | fisa_casp3 | | |
| 1bg8 | 2 (−7) | 0 (12) | 1 (−13) | 0 (6) | 2 |
| 1bl0 | 0 (158) | 0 (161) | 0 (164) | 0 (205) | 2 |
| | | | lattice_ssfit | | |
| 1ctf | 0 (65) | 0 (66) | 0 (66) | 0 (75) | 0 |
| 1dkt | 0 (7) | 0 (45) | 1 (−1) | 0 (18) | 1 |
| 1fca | 0 (22) | 0 (25) | 0 (15) | 0 (14) | 35 |
| 1trl | 0 (50) | 0 (45) | 0 (52) | 0 (64) | 0 |
| 4icb | 0 (18) | 0 (35) | 0 (0) | 0 (8) | — |
| | | | lmds | | |
| 1b0n | 11 (−18) | 3 (−8) | 10 (−10) | 9 (−11) | — |
| 1ctf | 0 (40) | 0 (60) | 0 (30) | 0 (51) | 0 |
| 1fc2 | 0 (2) | 0 (18) | 0 (1) | 0 (12) | 13 |
| 1shf | 0 (61) | 0 (68) | 0 (42) | 0 (65) | 0 |

Those proteins excluded from Table 4 because of bound metals or other molecules are shown here. Meaning of the numbers in each entry is the same as in Table 3.

**Table 6.** Test Results of Parameters Trained on Several Combinations of Problem Proteins in the ROSETTA Decoy Sets.

| Training proteins | Test proteins | | | |
|---|---|---|---|---|
| | 1utg | 1orc | 1r69 | 1vls |
| none (LK) | 1 (−3) | 3 (−13) | 4 (−13) | 1 (−16) |
| 1utg | *0 (8) | 9 (−17) | 1 (−4) | 1 (−3) |
| 1orc | 5 (−12) | *0 (8) | 0 (18) | 0 (22) |
| 1r69 | 1 (−0.3) | 3 (−9) | *0 (9) | 0 (9) |
| 1vls | 1 (−0.2) | 3 (−9) | 0 (9) | *0 (11) |
| 1utg/1orc | *0 (5) | *0 (6) | 0 (30) | 0 (23) |
| 1utg/1r69 | *0 (7) | 5 (−13) | *0 (7) | 0 (11) |
| 1utg/1vls | *0 (8) | 6 (−14) | 0 (4) | *0 (7) |
| 1orc/1r69 | 3 (−18) | *0 (8) | *0 (17) | 0 (23) |
| 1orc/1vls | 4 (−12) | *0 (7) | 0 (16) | *0 (22) |
| 1r69/1vls | 1 (−0.3) | 3 (−9) | *0 (9) | *0 (9) |

Those cases where the result is for a protein in the training set are marked with *. Meaning of the numbers in each entry is the same as in Table 3.

**Table 7.** Comparison of the Parameter Sets LK, SD, $LK_r$, and $SD_r$.

| | LK | SD | $LK_r$ | $SD_r$ |
|---|---|---|---|---|
| $\min(\Delta E)$ | −16 | 5 | −14 | 3 |
| No. $n_d > 0$ | 4 | 0 | 6 | 0 |
| No. $n_d > 0$* | 5 | 3 | 6 | 3 |
| No. $\Delta E <= \Delta E_{LK}$ | 34 | 7 | 15 | 7 |
| No. $\Delta E <= \Delta E^*_{LK}$ | 18 | 3 | 13 | 5 |

The second row shows the smallest energy gap among all training and test proteins except those in Table 5. Third and fourth rows show the number of decoy sets that $n_d > 0$, and the last two rows show the number of sets with $\Delta E$ not greater than with LK parameter set. Those cases marked with * are for the proteins in Table 5, and the total number of proteins considered is 18. The total number of proteins for the other rows is 34.

from the ROSETTA decoy sets, 1orc and 1lis. This improved and reduced parameter set is referred to as $SD_r$ (see Table 9). The scaled difference between $LK_r$ and $SD_r$ are 0.16 for $\Delta G^{free}$, 0.23 for $\lambda$, and 0.04 for $V$. The test results on all the test proteins are compared with results on the LK set in Tables 3, 4, and 5. The parameter sets are compared in Table 7. The initial parameter set $LK_r$ is less successful than the LK set, but the improved $SD_r$ set is as successful as SD.

Tables 4 and 5 also show a comparison with a model of Tobi and Elber (TE),[9] a discrimination function that has also been applied to the Levitt decoy sets that we test here. The TE discrimination function fails for 3 out of 11 test proteins (vs. 0 out of 13 for SD) in Table 4, and 8 out of 14 test proteins (vs. 3 out of 18 for SD) in Table 5. We do not regard the three failures of SD in Table 5 as failures because interaction energy of the proteins with the extra bound molecules are not taken into account due to the fact that EEF1 is missing appropriate parameters for those molecules.

**Table 8.** Parameter Set $LK_r$ for the EEF1 Solvation Free Energy.

| Atom types | $\Delta G^{free}$ | $V$ | $\lambda$ |
|---|---|---|---|
| H | 0.0000 | 0.0000 | 4.3562 |
| HC | 0.0000 | 0.0000 | 4.1587 |
| CR | −2.5350 | 8.3000 | 5.4687 |
| C | 0.0000 | 14.7000 | 5.4687 |
| CH1E | −0.4819 | 23.7000 | 5.6687 |
| CH2E | 0.9723 | 22.4000 | 5.5715 |
| CH3E | 2.7586 | 30.0000 | 5.5184 |
| CR1E | 0.1449 | 18.4000 | 5.4687 |
| N | −2.4803 | 0.0000 | 5.0698 |
| NR | −6.4008 | 4.4000 | 5.0698 |
| NH1 | −14.2418 | 4.4000 | 5.0698 |
| NH2 | −12.4816 | 11.2000 | 5.0698 |
| NH3 | −26.7520 | 11.2000 | 7.6852 |
| NC2 | −13.3760 | 11.2000 | 7.6852 |
| O | −9.3612 | 10.8000 | 5.0698 |
| OC | −13.3760 | 10.8000 | 7.6852 |
| OH1 | −10.7214 | 10.8000 | 5.0698 |
| S | −7.0552 | 14.7000 | 5.3048 |
| SH1E | −4.6461 | 21.4000 | 5.3048 |

Table 10 compares the statistical potentials taken from ref. 9 with the physical models tested here. Although the model of Tobi and Elber, for example, were trained on 33 million decoys from 674 proteins, nevertheless our CHARMM-based strategy appears to be more successful, and SD and $SD_r$ sets are the best parameters for the discrimination problem we posed here. It is not surprising that EEF1 with both LK and our improved parameters are more successful than statistical potentials, because they are more physical.[26,27]

## Conclusions

We have described a method, called MOPED, for improving parameters in models in computational biology that may have nonlinear dependencies on the parameters and are continuous functions of the degrees of freedom. The method relies on having good decoys. The MOPED method iterates to find parameters that can discriminate natives from decoys, with local energy minimizations at each step. As a proof of principle, we demonstrated the method by improving the solvation parameters in EEF1, an implicit solvent model combined with CHARMM19, that we have tested on protein folding decoys from the Baker and Levitt groups. Our method does not give a unique parameter set, but for the current decoy sets, all the feasible parameter sets that we have obtained are equally good at satisfying the discrimination problem we have posed. Although there is no guarantee that parameters optimized for our application will be optimal for other application, our improved parameters are not specific to folding decoys, and seem to be transferable to other problems such as loop modeling (unpublished results). Our aim here has been to demonstrate a parameter optimization method given a fixed set of decoys. An equally challenging problem that remains is how to improve the generation of decoys so that they can be used with methods like MOPED to produce better energy models.

**Table 9.** Improved Parameter Set $SD_r$ for the EEF1 Solvation Free Energy.

| Atom types | $\Delta G^{\text{free}}$ | $V$ | $\lambda$ |
|---|---|---|---|
| H | 0.0000 | 0.0000 | 4.3562 |
| HC | 0.0000 | 0.0000 | 4.1587 |
| CR | −1.6862 | 8.3000 | 6.2949 |
| C | 0.0000 | 15.3000 | 5.4687 |
| CH1E | 4.2681 | 23.1000 | 7.8283 |
| CH2E | 3.7976 | 23.0105 | 8.0981 |
| CH3E | 5.0000 | 30.0000 | 7.7260 |
| CR1E | −0.2182 | 20.8000 | 5.0287 |
| N | −3.1803 | 0.0000 | 4.2150 |
| NR | −6.4008 | 4.4356 | 7.0798 |
| NH1 | −12.7610 | 4.4000 | 3.9539 |
| NH2 | −13.1816 | 11.8000 | 5.9812 |
| NH3 | −28.3439 | 11.2000 | 6.4264 |
| NC2 | −14.0383 | 10.6000 | 5.8782 |
| O | −10.6933 | 10.8000 | 4.1869 |
| OC | −11.9760 | 10.8000 | 9.2593 |
| OH1 | −9.3214 | 10.8000 | 6.3759 |
| S | −7.0552 | 14.7000 | 4.3762 |
| SH1E | −4.6461 | 21.4000 | 5.3048 |

**Table 10.** Comparison of the Three Parameter Sets LK, SD, $LK_r$, and $SD_r$ for the EEF1 and Different Statistical Potentials Taken from ref. 9: TE (Tobi and Elber), MJ (Miyazawa and Jernigan), GKS (Godzik, Kolinski, and Skolnick), BT (Betancourt and Thirumalai), HL (Hinds and Levitt), and BJ (Bahar and Jernigan).

| | LK | SD | $LK_r$ | $SD_r$ | TE | MJ | GKS | BT | HL | BJ |
|---|---|---|---|---|---|---|---|---|---|---|
| No. Failures | 0 | 0 | 4 | 0 | 3 | 5 | 5 | 6 | 5 | 2 |
| No. Failures* | 3 | 2 | 2 | 2 | 8 | 9 | 11 | 10 | 12 | 8 |

Numbers of Levitt decoy sets for which $n_d > 0$ are shown. The total numbers of proteins considered in the second and third row are 11 and 14, respectively. The last row marked with * is for the proteins in Table 5.

## Acknowledgments

We gratefully acknowledge discussions with Prof. E. Coutsias. J.B.R. thanks Prof. P. Bourne for helpful discussions regarding the use of the SNOPT package.

This paper is dedicated to the memory of Peter Kollman, a warm and wonderful colleague, mentor, and friend.

## References

1. Maiorov, V. N.; Crippen, G. M. J Mol Biol 1992, 227, 876.
2. Koretke, K. K.; Luthey-Schulten, Z.; Wolynes, P. G. Proc Natl Acad Sci USA 1998, 95, 2932.
3. Mirny, L. A.; Shakhnovich, E. I. J Mol Biol 1996, 264, 1164.
4. Hao, M. H.; Scheraga, H. A. Curr Opin Struct Biol 1999, 9, 184.
5. Chiu, T.-L.; Goldstein, R. A. Proteins 2000, 41, 157.
6. Huber, T.; Torda, A. E. Protein Sci 1998, 7, 142.
7. Sali, A.; Shakhnovich, E.; Karplus, M. J Mol Biol 1994, 235, 1614.
8. Bryngelson, J. D.; Wolynes, P. G. J Phys Chem 1989, 93, 6902.
9. Tobi, D.; Elber, R. Proteins 2000, 41, 40.
10. Vendruscolo, M.; Domany, E. J Chem Phys 1998, 109, 11101.
11. Vendruscolo, M.; Najmanovich, R.; Domany, E. Proteins 2000, 38, 134.
12. Tobi, D.; Shafran, G.; Linial, N.; Elber, R. Proteins 2000, 40, 71.
13. Anfinsen, C. B. Science 1973, 181, 223.
14. Lazaridis, T.; Karplus, M. Curr Opin Struct Biol 2000, 10, 139.
15. Baysal, C.; Meirovitch, H. J Phys Chem B 1997, 101, 7368.
16. Das, B.; Meirovitch, H. Proteins 2001, 43, 303.
17. Lazaridis, T.; Karplus, M. Proteins 1999, 35, 133.
18. Lazaridis, T.; Karplus, M. J Mol Biol 1999, 288, 477.
19. Samudrala, R.; Levitt, M. Protein Sci 2000, 9, 1399.
20. Simons, K. T.; Bonneau, R.; Ruczinski, I. I.; Baker, D. Proteins 1999, 37, 171.
21. Gill, P. E.; Murray, W.; Saunders, M. A. User's Guide for SNOPT 5.3; UCSD, 1998.
22. Privalov, P. L.; Makhatadze, G. I. J Mol Biol 1993, 232, 660.
23. http://depts.washington.edu/bakerpg/
24. http://dd.stanford.edu/
25. Zhu, C.; Byrd, R. H.; Lu, P.; Nocedal, J. L-BFGS-B; Northwestern Univ, 1996.
26. Vorobjev, Y. N.; Almagro, J. C.; Hermans, J. Proteins 1998, 32, 399.
27. Dominy, B. N.; Brooks, C. L., III. J Comput Chem 2002, 23, 147.