JCTC Journal of Chemical Theory and Computation

Assessment of Detection and Refinement Strategies for de novo Protein Structures Using Force Field and Statistical Potentials

Michael S. Lee*,†,‡ and Mark A. Olson‡

Computational and Information Sciences Directorate, U.S. Army Research Laboratory, Aberdeen Proving Ground, Maryland 21005, and Department of Cell Biology and Biochemistry, U.S. Army Medical Research Institute of Infectious Diseases, Frederick, Maryland 21702

Received June 12, 2006

Abstract: De novo predictions of protein structures at high resolution are plagued by the problem of detecting the native conformation from false energy minima. In this work, we provide an assessment of various detection and refinement protocols on a small subset of the second-generation all-atom Rosetta decoy set (Tsai et al. *Proteins* **2003**, *53*, 76−87) using two potentials: the all-atom CHARMM PARAM22 force field combined with generalized Born/surface-area (GB-SA) implicit solvation and the DFIRE-AA statistical potential. Detection schemes included DFIRE-AA conformational scoring and energy minimization followed by scoring with both GB-SA and DFIRE-AA potentials. Refinement methods included short-time (1-ps) molecular dynamics simulations, temperature-based replica exchange molecular dynamics, and a new computational unfold/refold procedure. Refinement methods include temperature-based replica exchange molecular dynamics and a new computational unfold/refold procedure. Our results indicate that simple detection with only minimization is the best protocol for finding the most nativelike structures in the decoy set. The refinement techniques that we tested are generally unsuccessful in improving detection; however, they provide marginal improvements to some of the decoy structures. Future directions in the development of refinement techniques are discussed in the context of the limitations of the protocols evaluated in this study.

1. Introduction

Protein structure prediction is becoming an increasingly important part of the biologist's toolkit as the number of protein-encoding DNA sequences from genomic studies vastly outnumbers the available experimentally obtained protein structures. Structure prediction has been tackled by a variety of strategies depending on the similarity of a target amino acid sequence to known protein structures. Comparative modeling is used when the target sequence is very close to one or more known protein structures.[1] Fold prediction and threading are employed when the sequence can be matched through profile similarities with one or more known structures.[2] Finally, with little perceived similarity to known folds, de novo algorithms generate protein structures either by united-residue folding simulations[3] or fragment assembly.[4]

The Rosetta program from the Baker group[4] is considered one of the top methods for de novo structure predictions. Traditionally, de novo folding has been used as a last resort for protein structure prediction. The Rosetta protocol has proven to be very powerful for predicting structures where the fold and its subsequent template alignment can be guessed, but the fold prediction is less than certain.[5] Rosetta can generate structures of low to medium resolution in many cases, although detecting such structures as being near-native is frequently difficult.[6−8] Near-native, in the context of this work, refers to structural models whose root-mean-square-

* Corresponding author e-mail: michael.lee@amedd.army.mil.
† U.S. Army Research Laboratory.
‡ U.S. Army Medical Research Institute of Infectious Diseases.

10.1021/ct600195f
Published on Web 12/14/2006

deviation (rmsd) of their alpha-carbon backbone (C_α) are within 2−3 Å of the experimentally determined structure. Often for a given protein target, between 10 000 to 100 000 models must be generated for a few models to be near-native structures. Also, the more near-native structures that are generated, the greater the likelihood an atom-based scoring function will be able to detect one or more of the near-native structures as the best scoring. Two criteria must be satisfied, however, to make successful detection and refinement possible. First, the scoring function should score the native as lowest in energy compared to any misfolded structures. In addition, it is necessary that as the native structure is approached, as can be measured by various native-biased metrics such as rmsd or fraction of native contacts, the scores trend toward the native value. This requirement, which we will call a "scoring funnel" is analogous to the folding funnel, whereby real proteins move on a folding trajectory that take on the native fold in a finite time due to some leaning, however slight, toward the lowest free-energy basin.[9] One caveat in the connection between the scoring funnel and the folding funnel is that the scoring function often lacks some or all of the entropy contributions.[10]

Several refinement protocols have been considered in the literature, although the problem remains largely unsolved.[11] Presently, a grand challenge problem is to consistently refine low- to medium-resolution protein structure predictions (e.g., C_α rmsd > 4 Å) to the accuracy necessary for drug-based design (e.g., C_α rmsd < 2.5 Å.) Recent efforts have included the work of Lu and Skolnick,[12] which evaluated the effect of short simulations (∼50 ps) using force field and knowledge-based potentials. Misura and Baker[7] outlined a scheme of making random perturbations to the original Rosetta models, which works well in tandem with their homology-based enrichment procedure.[8] Fan and Mark[13] investigated the use of long-time molecular dynamic simulations (>10 ns) in improving initial models. In cases, where only small segments of a protein need to be "refined" (i.e., nonconserved regions of a homology model), configurational enumeration techniques can be quite successful.[14−16] Nevertheless, larger nonconserved regions (e.g., number of residues, N_{res} > 11) are still difficult to model because the number of plausible conformations increases exponentially with the number of residues.

A priori knowledge of which protein conformations in a large set of structures are near-native is an unsolved problem because of three reasons. First, the side-chain packing may not be correct, even if the backbone is near-native. In this case, the high-resolution scoring function will often fail. Second, the best structures may not be within the "radius of convergence" of the native basin for a given energy function.[8] Finally, the high-resolution energy function may sometimes assign a lower energy to a non-native structure compared to the native or near-native conformation.

The potential or scoring functions to discriminate and refine protein structures are currently based on three methods: force-field based[10,17,18] and knowledge-based[19] and hybrids of the two.[6,7,20] Force-field based detection functions often employ a standard parameter set such as CHARMM PARAM22[21] or AMBER[22] and an implicit solvent function such as generalized Born(GB)[23] or Poisson−Boltzmann.[24] One of the goals of this work is to compare two different but exemplary scoring functions, PARAM22/GB-SA[17,25] and the all-atom distance-scaled ideal-gas reference state (DFIRE-AA) statistical potential,[19,26] for detection and refinement. The SA denotes a simple solvent-accessible, surface area-based treatment of the hydrophobic effect. PARAM22/GB-SA exhibited one of the best detection capabilities among several force-field based functions in an assessment of CASP4 protein structures, where the specific model of GB was GBMV2.[17] GBMV2 is a molecular-volume dielectric boundary implicit solvent model which does a good job in mimicking the results of more expensive Poisson solvation calculations.[25]

DFIRE-AA, on the other hand, is very good at distinguishing the native structure from non-native conformations for a wide variety of decoy sets.[27] Statistical potential approaches have also been successfully employed in the drug-docking problem to detect native poses and estimate binding affinities.[27,28] Also studied is the ability of such functions to detect near-native structures[3,29,30] or optimal alignments of structural templates in homology models.[31,32] Statistical potentials are developed from the growing database of crystal structures in the Protein Data Bank (PDB).[33] The traditional method involves analyzing the probability distributions and subsequently the potentials of mean force along the distances between pairs of atoms.

In this work, we introduce a hybrid force field for molecular dynamics (MD) simulations that combines a continuous version of the DFIRE-AA statistical potential with the internal energies and van der Waals interactions of a united-atom force field.[12] Interestingly, MD simulations using this hybrid potential quickly condense the protein and trap it in a local minimum. To take advantage of the rapidity of condensing a protein structure, we developed a method that quickly unfolds and refolds a protein model, thereby generating hundreds of new protein models which can be scored by the DFIRE-AA or any other discriminating energy function. The hope is that some of the newly generated protein models will be lower in energy and closer to the native structure.

We first perform a standard comparison between the all-atom PARAM22/GB-SA potential[25] and the DFIRE-AA statistical potentials[34] for detection of native and near-native protein structures using several sets of Rosetta-generated protein conformations. We then perform replica exchange simulations using separately the all-atom potential and an MD-adapted form of the statistical potential. Replica exchange entails running several parallel simulation windows spanning a range of temperatures[35] whereby periodically exchanges of temperature between windows are performed based on a Metropolis Monte Carlo criterion. As a final method, we look at unfolding/refolding of model structures using the hybrid force-field/statistical potential.

2. Theory and Methods

2.1. Potentials. The DFIRE-AA statistical potential is one of several knowledge-based potentials described in the literature.[29,36] It is defined as[19]

$$u_{\mathrm{DFIRE}}(i,j,r) = -\eta k_B T \ln\left[\frac{N_{\mathrm{obs}}(i,j,r)}{\left(\frac{r}{r_{\mathrm{cut}}}\right)^{\alpha}\frac{\Delta r}{\Delta r_{\mathrm{cut}}}N_{\mathrm{obs}}(i,j,r_{\mathrm{cut}})}\right] \quad (1)$$

where i and j are non-hydrogen atom types, r is a pairwise distance, r_{cut} is the cutoff beyond which pairwise interactions are neglected, Δr is the histogram bin size, N_{obs} is a cumulative histogram of the observed occurrence of pairs as a function of the pairwise distance, α is set to 1.61 based on an empirical analysis of hard-sphere protein-like spatial distributions,[37] and k_B and T are the Boltzmann constant and absolute temperature, respectively. The parameter, η, is an arbitrary constant that can be modified either to estimate free-energy differences[27] or to tune the strength of the DFIRE energy term versus other energy terms. The histograms N_{obs} in this work were obtained from analysis of a culled set of 1836 PDB structures from the PISCES server[38] which had better than 1.8-Å resolution and were less than 30% homologous to each other. We deviated from the original DFIRE protocol by assigning $\Delta r = 0.5$ Å at all distances and having r range from 0.25 Å to 14.75 Å, such that $r_{\mathrm{cut}} =$ 15 Å.

Like many statistical potentials, the DFIRE model is not suitable by itself for exploring the energy landscape without some sort of restraints or constraints.[12] In the case of Monte Carlo exploration, one can sample different dihedral rotamers of the backbone and side chains, where each conformation is forced to obey standard bond lengths and bond angles. In our case, where we desire to run molecular dynamics, a further issue is that the DFIRE potential needs to be smoothed out. We employed cubic interpolation[39] to smooth out the potential so that the first derivatives are continuous. An example of this procedure is illustrated in Figure 1. Our complete dynamics potential, denoted here as DFIRE-MD, consists of the standard PARAM19 internal energy and van der Waals attraction/repulsion terms and the smoothed DFIRE-AA potential with η set to 0.25. As compared to DFIRE-AA, DFIRE-MD only incorporates smoothed statistical potential energies from nonbonded list pairs which include intraresidue pairs beyond 1−4 interactions. In minor contrast, typical DFIRE-AA includes all pairs of atoms up to precisely 15 Å excluding all intraresidue pairwise interactions. Electrostatics and solvation were omitted in DFIRE-MD as they were considered analogous to the contributions of DFIRE-AA. The van der Waals term was retained so that short-range steric interactions were properly modeled. Besides the obvious issue of overcounting in this energy model, it is questionable whether a statistical potential that defines a free energy should be used as a potential for molecular dynamics. Nonetheless, we are mainly concerned in this work with the exploration of a scoring-function surface, and not thermodynamics.

We also employ the all-atom PARAM22 force field[21] combined with GBMV2[25] implicit solvent model. A linear surface-area-based hydrophobic term of 30 cal/(mol·Å^2)[17] was also included using the SASA-1 approximation.[25] As indicated in the Introduction, this combination potential, which we will refer to as PARAM22/GB-SA, was one of the best performers in a previous protein structure detection study using the CASP4 predictions as decoy sets.[17] We believe that alternative implicit solvent models might lead to a modest decrease in accuracy but being considerably more computationally efficient may outweigh this.

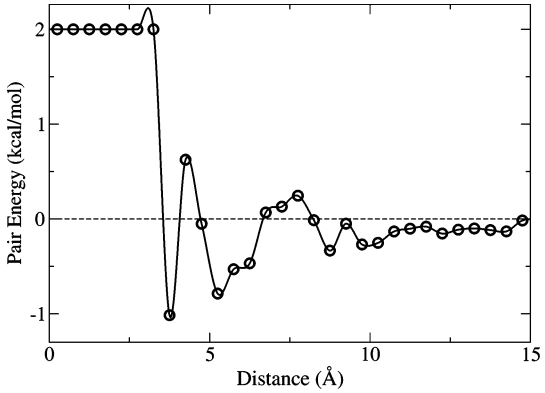

Figure 1. Regular and smoothed DFIRE-AA potential for the pairwise interaction of two alanine C_αs. The circles denote the regular DFIRE-AA potential values at the bin centers. The solid curve is the cubic-interpolated version suitable for MD simulations.

Table 1. Features of the Nine Protein Decoy Sets Used in This Work

PDB ID	N_{res}[a]	% alpha	% beta	no. of decoys in set	best rmsd rmsd	best rmsd % nc[b]	best % nc[b] rmsd	best % nc[b] % nc[b]
1ail	67	85	0	1807	2.0	55	2.0	55
1csp	64	0	53	1809	3.2	43	3.9	46
1ctf	67	52	19	1922	2.7	57	3.5	64
1pgx	57	25	46	1851	1.5	63	1.5	63
1r69	61	64	0	1733	1.4	64	1.4	69
1tif	59	17	37	1849	2.6	56	2.6	56
1utg	62	79	0	1897	3.4	36	5.4	53
1vif	48	0	50	1896	0.4	56	1.2	86
5icb	72	57	6	1870	3.0	58	3.1	59

[a] Number of residues in protein. [b] % nc − percentage of native contacts.

2.2. Protein Model Sets. The specific interest of this work is to assess detection and refinement of de novo-generated protein structure models created by the Baker lab using the Rosetta program.[4] We looked at nine proteins in this study, with the following PDB identifiers:[33] 1ail, 1csp, 1ctf, 1pgx, 1r69, 1tif, 1utg, 1vif, and 5icb (see Table 1). These proteins were chosen based on their diversity of secondary structure, availability of online Rosetta decoy sets (which we call *Rosetta2*, denoting the second generation),[6] availability of X-ray crystal native structures, and overlap with previous detection and refinement studies.[7,8,40] Each one of the decoy sets contains approximately 1800 models, which consist of ∼1000 decoys from the original Rosetta decoy set, ∼400 somewhat near to the native, and ∼400 of the lowest C_α rmsd from an exhaustive 200 000 model Rosetta run.[6] The enrichment of low rmsd structures in these sets is certainly an influence on our results and cannot be fairly compared to a prediction protocol where far less than 200 000 Rosetta models are generated. This issue is considered more in the Discussion section.

For the first statistical potential detection trial, the all-atom decoys were used as-is. For all of the other methods, the Rosetta models were converted to a PARAM19 format using the Multiscale Modeling Tools for Structural Biology (MMTSB) *convpdb.pl* command and minimized modestly to remove steric clashes (MMTSB *minCHARMM.pl* command interfaced with CHARMM[41]): 50 steps with a steepest descent algorithm followed by 100 steps with an adopted basis Newton−Raphson protocol. The energy function for minimization used a distance-based dielectric electrostatic term with a coefficient of 4.[17]

2.3. Clustering. While results may vary for other sets of Rosetta-generated models, low C_α-rmsd structures often show up in the Rosetta2 decoy sets as seen in Table 1. Also, the population of these low C_α-rmsd models may be diminishingly small.[6] In general, we observe that at the collection phase after a Rosetta run, it is imperative not to discard structures solely by score, because they could actually be the best models, i.e., nearest-native. However, some amount of filtering needs to take place before any computationally intensive refinement procedure such as replica exchange or Z-fold (both described below). In this work, we hierarchically cluster Rosetta-generated decoy structures[6] to obtain a diverse set of structures using the MMTSB command *cluster.pl* with the *-jclust* option. Our nondefault clustering parameters included a maximum of four subclusters per parent cluster (*-maxnum* option) and minimum of four elements per subcluster (*-minsize* option.) The clusters were selected from the fourth hierarchical level, such that in each decoy set, at least 16 clusters could be identified in all of the protein sets. The average DFIRE-AA scores from each cluster were ranked, and the lowest-energy conformers from each of the top 16 clusters were defined as the diversity set. Note that the PARAM22/GB-SA scores could have been used instead for ranking.

2.4. Replica Exchange. The replica exchange method (ReX)[42] is a state-of-the-art technique for sampling an energy landscape. It has been used successfully in studies of protein folding,[43] loop structure prediction,[44] and lattice-based protein structure prediction.[45] The concept behind the method is to run multiple simultaneous molecular dynamics or Monte Carlo simulations with a spectrum of biases and/or temperatures. The principle of using ReX in this study is to allow for automatic unfolding of worse scoring structures and refolding of better scoring structures. In this work, a range of temperature windows is used, and we looked at the performance of separately the PARAM22/GB-SA and DFIRE-MD potential. After a specified block simulation time, τ, windows a and b exchange temperatures with a probability, W:[46]

$$W(a \leftrightarrow b) = \begin{cases} 1 & \Delta_{ab} \le 0 \\ \exp(-\Delta_{ab}) & \Delta_{ab} > 0 \end{cases}$$

$$\Delta_{ab} = (\beta_a - \beta_b)(E_a - E_b) \qquad (2)$$

where β is $1/k_B T$ and E is the potential energy of a particular replica. We used 16 temperature windows ranging exponentially from 298 to 650 K for the DFIRE-MD simulations and 298 to 500 K for the PARAM22/GB-SA runs. The different temperature ranges selected for each potential reflected the fact that we tried to ramp up the temperature for the DFIRE-MD simulations as high as possible to counter the strong collapsing propensity of this potential, while retaining some energy overlap between windows. The initial structures placed in each window corresponded to the 16-member diversity set described above. Block simulation times, τ, were set to 0.4 ps. A total of 2500 exchange steps were carried out, for a cumulative simulation time of 1 ns. Molecular dynamics simulations were performed with the CHARMM software package,[41] and the replica exchange method was performed with the MMTSB *aarex.pl* program.[47]

Even though ReX enhances sampling, some accuracy will be lost simply by having to filter out a small number of structures to create a diversity set. Therefore, we decided to also run every minimized decoy with 298 K molecular dynamics for a small amount of simulation time, 1 ps, to compare with the ReX simulations. With such short runs, the relevant question was whether a small amount of refinement could improve detection.

2.5. Z-Fold Method. Noting the strong attractive nature of a pairwise statistical potential during a MD run, we decided to utilize this feature to refold protein structures with the aim of generating a diversity of conformations in the vicinity of a given model structure. The *Z-fold* method starts by temperature unfolding (400 K) a protein model over a short time with secondary structure restraints and only the vdW and internal energy terms turned on. This is followed by refolding with the DFIRE-MD potential retaining the secondary structure restraints. In this work, the unfolding simulations were performed for 10 ps, and refolding simulations were performed for 6 ps. For each starting model, 10 unfolding simulations with different random seeds were performed. For each unfolded structure, there were then 10 refolds performed, for a total of 100 refolded structures per starting model. The secondary structure restraints were obtained via the DSSP[48] program evaluated on the original model. Secondary structure restraints, E_{ss}, of the form

$$E_{ss} = K \times \max\left[0, \mathrm{abs}\left(\theta - \frac{\pi\theta_{\min}}{180^\circ}\right) - w\right]^2 \qquad (3)$$

were used to restrict the backbone dihedral angles of the identified secondary structure elements to plus or minus the width, w, from $\theta_{\min}$. The force constant, K, was set to 100 kcal/mol/rad,[2] w is the width of the potential, and θ corresponds to either the ϕ- or φ-dihedral angles. For the α helix restraints, the parameters were $w = 7^\circ$, $\phi_{\min} = -64^\circ$, and $\varphi_{\min} = -41^\circ$. For the beta-strand restraints, the parameters were $w = 40^\circ$, $\phi_{\min} = -120^\circ$, and $\varphi_{\min} = +120^\circ$. The 16-member diversity sets for each protein were also the starting models in this part of the study. After generation, each refolded structure was minimized and rescored using the PARAM22/GB-SA detection protocol described above.

2.6. Analysis Techniques. A popular measure of the similarity of a model structure with the native conformation is rmsd. In this work, rmsd is defined for the C_α protein backbone versus the native X-ray structure in units of Å. A common evaluation of scoring functions is the Z-score, which normalizes the score of the native, E_{native}, relative to the mean,

$\bar{E}$, and standard deviation, σ, of the scores of the decoy set:

$$Z_{\text{ener}} = \frac{E_{\text{native}} - \bar{E}}{\sigma} \quad (4)$$

The Z-score is a useful measure of the depth of the scoring funnel, whereby greater negative values indicate deeper funnels. Nonetheless, detecting a near-native structure from a set of models can only be reliably achieved when there is some propensity of the scoring function to favor structures as they become more and more nativelike. Therefore, we are concerned as well in this work with other criteria: the rmsd of the lowest scoring structure (excluding the native); the best rmsd of the top five scoring structures; the 15 × 15% enrichment score;[6] and statistical correlation between rmsd and score. The 15 × 15% enrichment score measures the number of structures which are both in the top 15% of scores and top 15% of RMSDs to native divided by the number of structures one would expect by chance to satisfy these two criteria. Summa et al.[36] show that other measures of the usefulness of a scoring potential are correlated significantly with the ones we use here.

While C_α rmsd is a popular measure of similarity of a conformation to the native structure, it is sometimes helpful to look at other similarity measures, such as fraction of native contacts. The definition for fraction of native contacts is as follows. First, for a given native structure, the native contacts are identified as all side-chain center-of-mass pairs, (i,j): $j > i + 1$, whose distances are less than 6.5 Å.[49] Then for each model conformation, the fraction of native contacts is the number of native contacts in the model divided by the total number of native contacts in the X-ray structure. In this work, to conform to the directionality of rmsd scatter plots, we take one minus the result.

3. Results

The following section considers separately detection and refinement using two distinct scoring functions: DFIRE-AA and PARAM22/GB-SA. In the detection subsection, we consider the ability of these two scoring functions to find near-native structures from large decoy sets of de novo-generated conformations. In the refinement subsection, we first ask whether short-time molecular dynamics enhances the detection capabilities of the force field-based score. Then we test the two scoring functions in a standard replica exchange protocol to see whether small subsets of the decoy sets can be induced toward the native state. Finally, noting the collapsing propensities of the DFIRE-AA as a sampling function, we evaluate the above-described unfold/refold method with the same small subsets of decoys.

3.1. Detection. Table 1 outlines some of the features of the decoy sets we chose. The best structures in each set have C_α RMSDs of ~3 Å and below. In contrast, the structures with the best percentage of native contacts have only between 50% and 65% similarity. This means that 35−50% of the native contacts are missing even in the best decoy structures. Therefore, it might be conjectured that scoring functions with atomic resolution may fail to detect the structures that are closer to native, because they are still some distance away in contact space. Finally, in only three of the nine protein sets is the best rmsd structure also the closest to the native in contact space.

Table 2. Summary of Results for Detection of Structures Using the DFIRE-AA Potential Score on the Original Decoy Structures

PDB ID	Z_{ener}	rmsd of top scoring structure	best rmsd of top 5 scoring structures	enrichment (15 × 15%)	av rmsd of top cluster	best av rmsd of top 5 clusters
1ail[a]	−2.3	8.7	4.5	0.69	9.2	7.1
1csp[a]	−3.2	4.3	4.3	2.82	6.0	6.0
1ctf	−3.5	3.3	3.3	1.85	4.8	4.8
1pgx	−4.4	5.9	2.4	2.45	5.5	4.1
1r69	−3.6	2.2	1.5	3.49	3.5	3.5
1tif	−5.1	7.8	3.9	0.96	5.1	5.0
1utg[a]	−1.3	10.7	6.7	0.54	6.2	6.2
1vif	−2.8	0.6	0.6	4.80	3.8	3.8
5icb[a]	−2.2	4.3	4.3	2.19	5.7	5.7
avg[b]	−3.2	5.3	3.5	2.20 (1.39)	5.5	5.1

[a] Native structure was not detected as the lowest in energy. [b] Standard deviation in parentheses.

Table 3. Summary of Results for Detection of Structures Using the DFIRE-AA Potential Score on the Minimized Decoy Structures[a]

PDB ID	Z_{ener}	rmsd of top scorer	best rmsd of top 5 scoring structures	enrichment (15 × 15%)	av rmsd of top cluster	best av rmsd of top 5 clusters
1ail[b]	−1.4	9.7	7.7	0.64	9.2	6.6
1csp[b]	−3.1	4.3	3.9	2.78	6.0	6.0
1ctf	−3.3	3.3	3.3	1.87	4.8	4.8
1pgx	−3.6	5.9	2.4	2.52	4.2	4.1
1r69	−3.5	2.2	1.5	3.80	3.5	3.5
1tif	−3.8	7.8	3.8	1.08	5.1	5.0
1utg[b]	−0.5	5.4	5.4	0.30	6.2	6.2
1vif	−2.2	0.6	0.6	4.71	3.8	3.8
5icb[b]	−1.8	4.4	4.2	2.04	5.7	5.6
avg[c]	−2.6	4.8	3.6	2.19 (1.45)	5.4	5.1

[a] Structures were minimized using the protocol specified in the Methods section. [b] Native structure was not detected as the lowest in energy. [c] Standard deviation in parentheses.

The PARAM22/GB-SA potential is marginally better than DFIRE-AA at detecting a low-rmsd structure using score alone, as seen in Tables 2−4. Both potentials perform roughly the same in detection if the top five scoring conformations are considered. One can also see that the average Z-score for the PARAM22/GB-SA is slightly superior to the DFIRE-AA one. Furthermore, DFIRE-AA fails to detect the native X-ray crystal structure for four proteins (even with minimization), while PARAM22/GB-SA fails for only two proteins. The 15 × 15% enrichment scores for both potentials are on average roughly the same, while the standard deviation of these scores suggest DFIRE-AA can be either better or worse than PARAM22/GB-SA for a specific protein. For example, the DFIRE-AA potential fares worse than chance (i.e., enrichment scores less than 1) for three proteins, while the PARAM22/GB-SA enrichment values are above chance in each protein case. Using a clustering scheme to choose structures or sets of structures is somewhat worse than single conformation detection for DFIRE-AA (Tables 2 and 3) and significantly worse for PARAM22/GB-SA (Table 4). In principle, clustering should

Table 4. Summary of Results for the Detection of Structures Using the PARAM22/GB-SA Potential on the Minimized Decoy Structures[a]

PDB ID	Z_{ener}	rmsd of top scoring structure	best rmsd of top 5 scoring structures	enrichment (15 × 15%)	av rmsd of top cluster	best av rmsd of top 5 clusters
1ail	−3.3	10.7	4.0	2.58	6.6	6.6
1csp	−4.2	4.5	4.3	1.82	6.0	6.0
1ctf	−4.9	3.7	3.3	2.22	4.8	4.8
1pgx	−5.5	2.4	2.4	1.32	5.5	4.5
1r69	−5.8	2.4	1.7	2.59	3.5	3.5
1tif	−5.1	4.4	4.0	1.35	6.0	5.0
1utg[b]	−2.1	4.7	4.6	1.55	6.2	6.2
1vif	−3.2	0.5	0.4	4.22	3.8	3.8
5icb[b]	−1.8	4.1	4.0	1.66	8.4	5.6
avg[c]	−4.0	4.2	3.2	2.15 (0.92)	5.6	5.1

[a] Structures were optimized using the protocol specified in the Methods section. [b] Native structure was not detected as the lowest in energy. [c] Standard deviation in parentheses.

help smooth out noise in the scoring function. In practice, lingering clashes in specific structures are more penalized in the PARAM22/GB-SA results, likely leading to worse overall average cluster energies. Furthermore, cluster populations at this stage are unlikely to be fruitful, given that they are dependent on the "thermodynamic" sampling of the lower-resolution Rosetta united-residue function.

It is interesting to compare the results of the original paper associated with the decoys we used.[6] Tsai et al. reports an average Z-score of −4.5 and enrichment value of 1.86 for all 78 proteins using their single unified α/β scoring potential. This is not a fair comparison between our results and theirs as we are using a small manually selected subset of proteins from their large collection. However, it shows that our force-field results are in line with their analyses, which had used a different atomic resolution potential.

In Figure 2, we show two examples of the detection problem using the Rosetta decoy sets: an easy case and a difficult case. In the easy situation, 1vif (Figure 2a,b), there are several very near-native structures generated. Also, as the structures approach the native, there is a downward slope in energy. Structures below 1 Å in rmsd are detectable versus the rest of the set using the PARAM22/GB-SA potential. Furthermore, the lowest energy structure for this potential is nearly the lowest rmsd.[8] In contrast, in the difficult case, the 1ail (Figure 2c,d) decoy set has few structures that are better than 4 Å and only one structure better than 2 Å. Visually, one might consider the group of structures in Figure 2c at ∼4.5 Å have on average a better score than the other group, which suggests that by clustering a ∼4 Å conformation could be selected out. In reality, though, no such lower-scoring cluster was identified (Table 4). Figure 2c also illustrates how single structure detection can fail sometimes, as it picks out the low-scoring conformation at 10.7 Å. Figure 2d shows that DFIRE-AA cannot discern the native structure as lowest in energy. In addition, there are no visible trends in this scatter plot.

Figure 3 illustrates the point that even if many 2 and 3 Å structures are in the decoy set, they may have a lot of missing native contacts. This provides some evidence of why atomic resolution scoring functions may not detect these lower rmsd structures. Figure 3b shows very little funnel-like behavior, likely due to the large gap in native contact space between the best decoys and the native structure. In Figure 3c, the

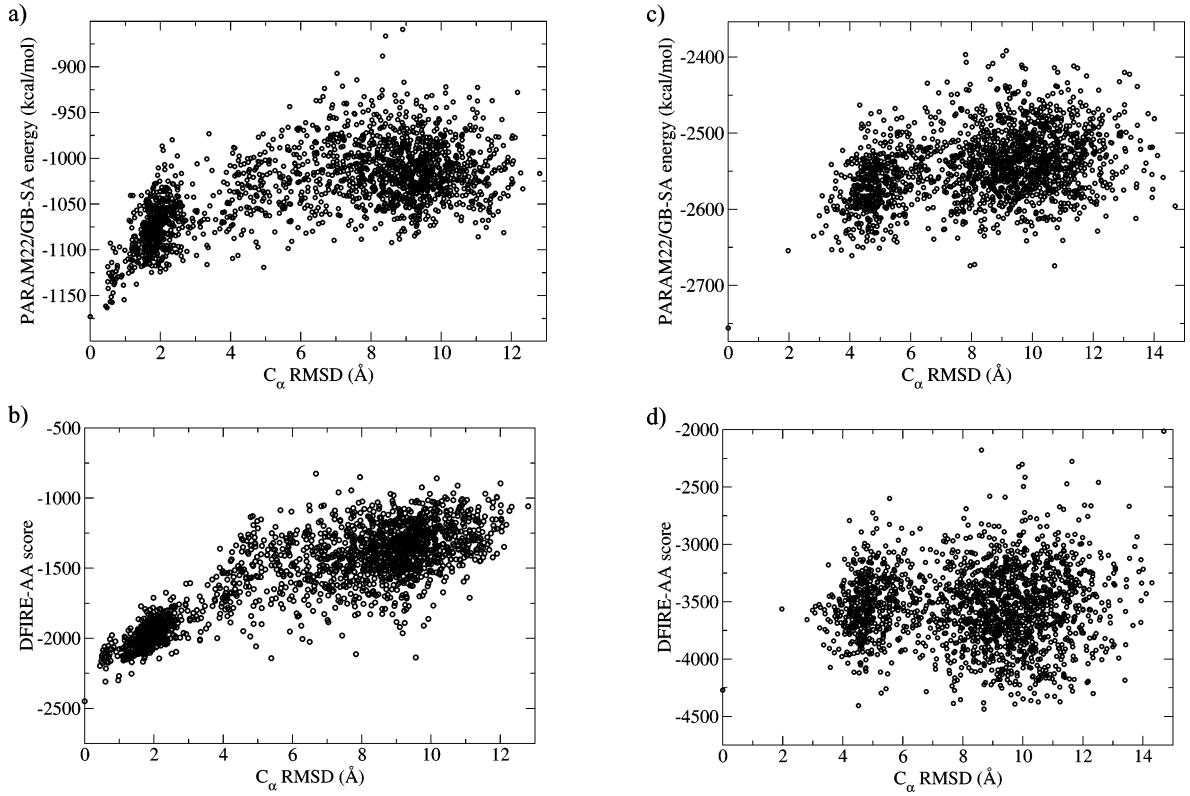

Figure 2. Scatter plots of the PARAM22/GB-SA and DFIRE-AA potentials vs C_α rmsd to native: (**a-b**) 1vif, an easy test case for detection and (**c-d**) 1ail, a difficult test case for detection.

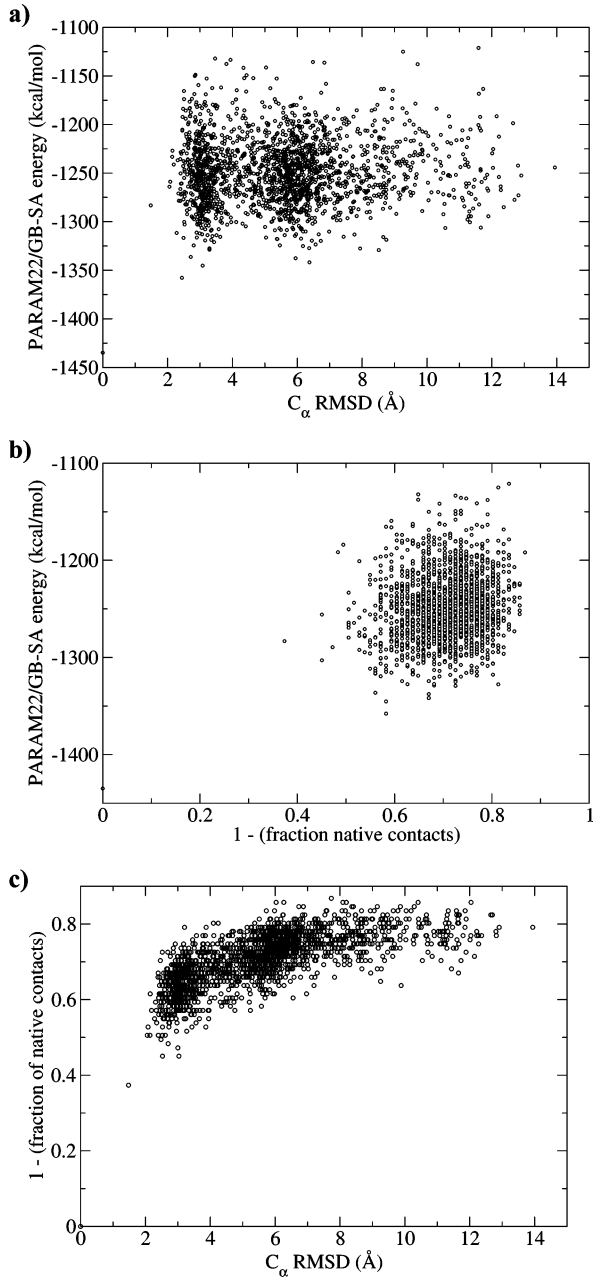

Figure 3. For the 1pgx decoy set, comparison of (**a**) PARAM22/GB-SA score with C_α rmsd, (**b**) PARAM22/GB-SA score with fraction of native contacts, and (**c**) fraction of native contacts with C_α rmsd.

structures between 2 and 4 Å begin to have some slope toward more native contacts than the continuum of structures in the set.

In Table 5, the funnel-like behavior of the two scoring functions is further quantified by looking at the correlation coefficient of the score to the rmsd of decoys which are close to the native.[7,50] In most proteins, small correlations do exist between score and rmsd. However, in some notable cases, such as 1tif and 1utg for DFIRE-AA and 1pgx for PARAM22/GB-SA, the correlations are nearly zero or negative, indicating no funnel-like behavior. Since the Rosetta-generated decoys do not completely span the conformation space of our test proteins, the correlations are probably, in general, underestimated. In fact, protein decoy sets obtained by perturbation of the native structure tend to show improved correlation between score and rmsd at various rmsd ranges (results not shown).[51]

Table 5. Correlation Coefficient of DFIRE-AA and PARAM22/GB-SA Scores vs RMSD as a Function of Different RMSD Ranges of Conformations

PDB ID	DFIRE-AA			PARAM22/GB-SA		
	<4 Å	<6 Å	all	<4 Å	<6 Å	all
1ail	−0.05	0.20	0.03	0.32	0.34	0.25
1csp	0.07	0.22	0.48	0.06	0.12	0.11
1ctf	0.06	0.22	0.43	0.17	0.30	0.23
1pgx	0.44	0.39	0.41	−0.03	−0.01	0.04
1r69	0.48	0.51	0.51	0.32	0.38	0.24
1tif	−0.09	0.08	0.39	−0.16	0.13	0.06
1utg	−0.09	−0.22	0.11	0.27	0.17	0.18
1vif	0.74	0.87	0.85	0.51	0.67	0.65
5icb	0.22	0.27	0.44	0.07	0.16	0.16
avg	0.20	0.28	0.41	0.17	0.25	0.21

Table 6. Summary of Results for Detection/Refinement of Structures Using Short-Time Molecular Dynamics (1 ps) with the PARAM22/GB-SA Potential

PDB ID	rmsd[a] of top scorer[b]	best rmsd[a] of top 5 scorers[b]	enrichment (15 × 15%)
1ail	10.8	4.0	2.53
1csp	7.7	4.5	1.87
1ctf	4.0	3.6	1.94
1pgx	11.8	2.5	2.21
1r69	1.7	1.7	3.92
1tif	4.0	3.4	1.44
1utg	11.0	4.5	1.41
1vif	0.9	0.9	4.29
5icb	4.1	4.2	1.90
avg[c]	6.2	3.3	2.39 (1.04)

[a] rmsd defined with respect to the final structures of the dynamics trajectories [b] Score defined as the average potential energy over the short-time dynamics simulation. [c] Standard deviation in parentheses.

3.2. Refinement. Short-run MD results on every decoy structure are presented in Table 6. The goal here was to obtain quick refinement of all of the decoy structures in the hopes that poor side-chain contacts in good rmsd structures might be rectified and detection would be improved. Unfortunately, single structure detection results were 2 Å worse on average than minimization alone. Side-by-side comparisons with the optimized structure results show that dynamics increased detection errors for a few of the difficult cases and 1pgx. Using the top five scoring conformations criteria, the dynamics results are on par with simple optimization. Finally, enrichment scores are overall enhanced somewhat by short-time dynamics. While these simulations lack equilibration at 298 K, which could be a source of error, there is a practical compromise with simulation runtime when thousands of structures must be simulated.[8]

Tables 7 and 8 summarize the results of ReX simulations on a diversity set of conformations ($N = 16$) for each protein. Each small set includes at least one structure of ∼3 Å rmsd quality. The sampling nature of ReX−MD simulations permits us to look at clusters and their respective populations

Table 7. Summary of Results for Detection and Refinement of Structures from 1-ns Replica Exchange Molecular Dynamics Simulations Using the DFIRE-MD Potential

PDB ID	best rmsd in diversity set	lowest rmsd (298 K)	lowest rmsd (all T)	lowest av energy cluster (rmsd)[a,b]	most populated cluster (rmsd)[a,b]
1ail	5.3	5.3	4.9	10.9	10.9
1csp	3.6	3.5	3.4	8.0	5.7
1ctf	3.6	3.4	3.4	10.4	5.2
1pgx	1.5	1.9	1.0	8.8	8.8
1r69	1.5	1.1	1.1	3.5	1.5
1tif	4.1	4.0	3.5	4.9	4.9
1utg	4.8	5.7	4.8	8.4	10.4
1vif	0.6	0.6	0.6	9.4	3.7[c]
5icb	4.3	4.1	3.3	4.8	4.4
avg	3.3	3.3	2.9	7.7	6.2

[a] Structures and energies obtained from the final step of the simulation block in the 298 K window. [b] rmsd was averaged over the structures in the specified cluster. [c] Averaged over two clusters tied for first, with RMSDs of 2.8 Å and 4.6 Å.

Table 8. Summary of Results for Detection and Refinement of Structures via 1-ns Replica Exchange Molecular Dynamics Simulations with the PARAM22/GB-SA Potential

PDB ID	best rmsd in diversity set	lowest rmsd (298 K)	lowest rmsd (all T)	lowest av energy cluster (rmsd)[a,b,c]	most populated cluster (rmsd)[a,b]
1ail	5.3	5.2	4.8	6.8	11.8
1csp	3.6	3.4	3.4	7.0	4.2
1ctf	3.6	3.1	3.1	11.3	10.9
1pgx	1.5	1.7	1.6	6.5	7.1
1r69	1.5	1.3	1.3	3.0	3.0
1tif	4.1	3.9	3.6	6.1	6.1
1utg	4.8	4.4	4.1	5.7	10.0
1vif	0.6	0.7	0.7	2.0	1.8
5icb	4.3	4.2	3.4	5.0	5.0
avg	3.3	3.1	2.9	5.9	6.7

[a] Structures and energies were obtained from the final step of each simulation block in the 298 K window. [b] rmsd was averaged over the structures in the specified cluster. [c] Clusters with less than 10 elements were filtered out.

at 298 K as analogous to free energies of these clusters. The data indicate the cluster population offered better detection than average energy for the DFIRE-MD potential but not the PARAM22/GB-SA potential. Given the limited number of proteins in this work, neither average energy nor free energy can be distinguished as better than the other. In only a few of the protein cases for both potentials did the lowest rmsd starting conformation contribute significantly to the lowest-energy cluster. This highlights the limitations of the potentials and the fact that no significant folding funnel could be discerned at such limited rmsd quality. Proteins 1ail and 1utg exemplify the latter constraint.

Interestingly, there are slight rmsd improvements, albeit undetectable via energy criteria, which take place for both potentials for most of the proteins.[7,12] At 298 K, the improvements average 0.2 Å for PARAM22/GB-SA. Over all the temperature windows, improvements average as much as 0.4 Å. In the case of DFIRE-MD, these rmsd improvements may not reflect refinement as much as compacting of the model structures. It is also noted that the best structures were not produced and preserved in the lowest temperature (298 K) window.

Some further details of a single replica exchange simulation (1pgx/PARAM22-GBSA) are presented in Figure 4. The progressions of the two lowest rmsd models, as seen in Figure 4a, are quite different. The 2.4 Å structure stabilizes and becomes lower in rmsd to about 2.0 Å, while the 1.4 Å structure gets significantly worse over time. This divergence can be explained in Figure 4b,c, where the 2.4 Å structure spent much more time in cooler temperature windows than the 1.4 Å model. Finally, as illustrated in Figure 4d, we note that in the first 600+ ps, a 7 Å model dominates the lowest temperature window. Consistent with this result, the data in Table 8 indicate that the lowest free-energy cluster had an average rmsd of ~7 Å. One might surmise that with further sampling the 2.5 Å model would dominate the 298 K window and be detected as the lowest in free energy.

Two more important points can be gleaned from Figure 4. First, the energy of a structure and not its rmsd to the native dictates how the models will percolate through the temperature windows. Hence, a poorly scoring low-rmsd structure inserted into a simulation may end up getting muddled by high temperatures. Second, the ReX simulations, as computationally intensive as they are, generally must be run for much longer simulation times than were done here (e.g., 10−100 ns) to get convergent population statistics.

Overall, the ReX results are not as remarkable as the simple detection schemes, despite the orders of magnitude more computational effort. Our PARAM22/GB-SA replica exchange on proteins of the size studied here required 2 days of computation per protein on 16 AMD Athlon 2200+ processors. In contrast, the PARAM22/GB-SA detection protocol on an entire decoy set required about 5 h on a single CPU.

Figure 5 illustrates that the DFIRE scores can be highly correlated with the compactness of the conformations. Although this trait may not be able to completely explain DFIRE-AA detection abilities, it does suggest that running on the DFIRE-MD energy surface could cause structures to become more compact. In fact, Figure 6 illustrates that DFIRE-MD tends to compress protein structures and make them more spherical in shape. This can be attributed to the fact that DFIRE-potential tends to maximize intraprotein contacts. An opposing protein contact breaker, such as a solvation term, is lacking.[36] Despite the distortions caused by DFIRE-MD, the potential is very expedient at forming contacts. In Figure 7, one can see that a partially extended conformation is quickly collapsed into a compact structure in a mere 5 ps of simulation time.

Given the quick collapsing propensities of DFIRE-MD, the Z-fold method was tested, and the results are summarized in Table 9. As one can see, the results are not much better than replica exchange on average. The lowest average energy and lowest free-energy clusters are on par with clustering results in the detection and replica exchange calculations. Most noticeably, the best rmsd structure is on average 0.3

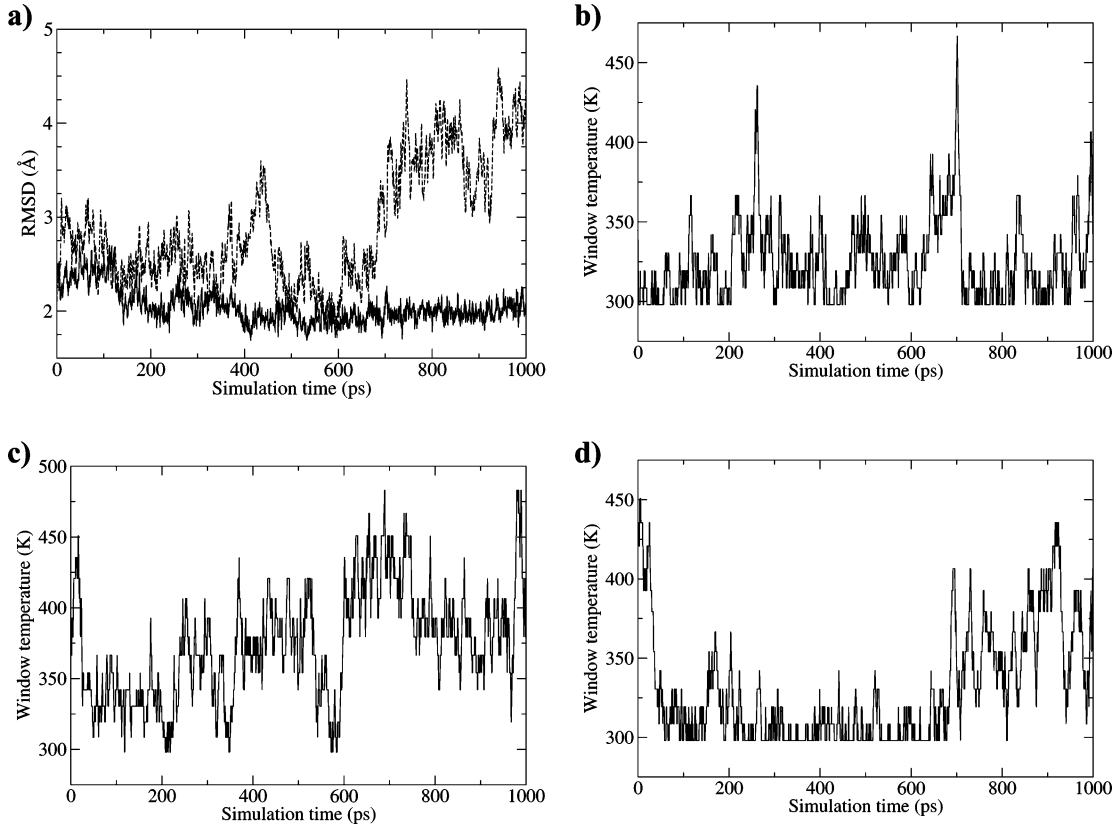

Figure 4. Replica exchange results for the 1pgx diversity set using the PARAM22/GB-SA potential. (**a**) Comparison of 2.4 Å (solid line) and 1.4 Å (dashed line) models. Temperature progressions of the (**b**) 2.4 Å, (**c**) 1.4 Å, and (**d**) 7 Å models.

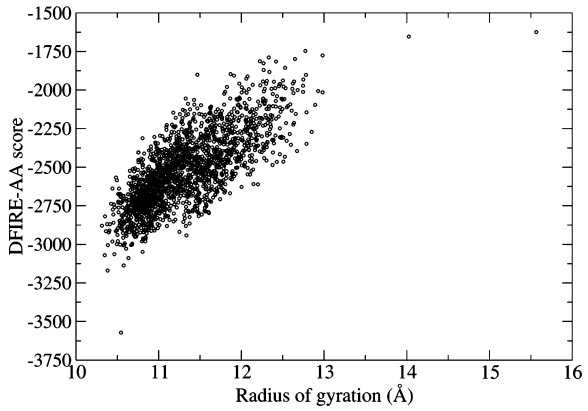

Figure 5. Comparison of the DFIRE-AA score with radius of gyration for the 1pgx decoy set.

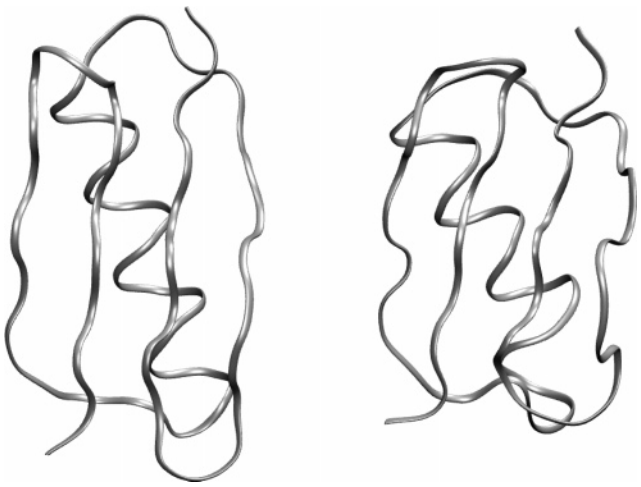

Figure 6. DFIRE-MD compresses native 1pgx protein structure over a simulation time period of 1 ns: (left) native structure and (right) after 1 ns of DFIRE-MD. Molecular graphics rendered with VMD software.[67]

Å better than the original best structure. Nevertheless, some of the rmsd improvement could be due to the compacting nature of the DFIRE-MD potential. Another issue is that neither the DFIRE-MD nor the PARAM22/GB-SA potential was able to detect the best rmsd structures. In Figure 8, it appears that improvements in rmsd were achieved for structures 2 Å and farther from the native. Sometimes rmsd improvements could be detected by DFIRE-MD as illustrated by the filled squares which lie above and below the zero line. Once again, some structure compacting may be occurring, and small rmsd improvements may not translate completely as refinements.

4. Discussion

4.1. Decoy Set Properties. The ability to detect near-native structures from a set of conformations is inevitably related to the quality of structures in the set. If enough low-rmsd structures are available, any good detection function should be able to pick up at least some of these structures as better in score than the rest. The extreme case of an easy decoy set would be one where model structures are developed from perturbations of the native. Small perturbation decoys would

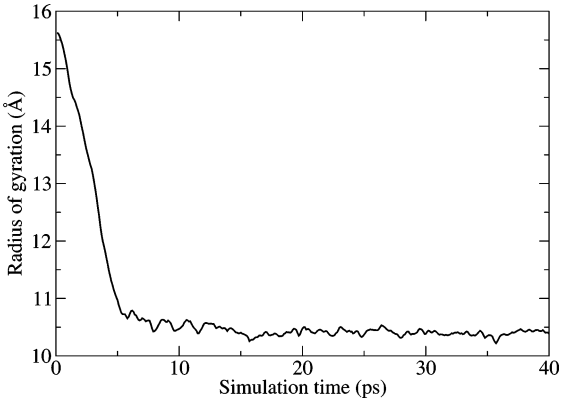

Figure 7. Radius gyration as a function of simulation time for the most extended model structure in the 1pgx decoy set using the DFIRE-MD potential at a temperature of 298 K.

be very near-native, and if the detection function labels the native as best, it would likely label near-natives better than misfolds.

Most of the low-rmsd structures in the Rosetta2 decoy sets are culled from Rosetta runs of nearly 200 000 structures per protein set. Generation of 200 000 structures for a single target is roughly an order magnitude larger than the standard automated server Robetta protocol. With today's computing, a single protein prediction would require effort on the order of CPU-weeks[52] to generate 200 000 models. Enrichment of the decoy sets with low-rmsd structures seems to increase the probability of detecting near-native structures, because the likelihood that at least one near-native structure will outscore all of the other structures increases. In addition, if clustering is performed, the near-native enrichment may provide a distinct cluster of structures from which to select. Furthermore, analyses such as the colony energy method,[14] which modifies the scores based on the presence of structural neighbors, would be biased by the enrichment protocol since as structures become closer to the native, they also become closer to each other, hence enhancing the pairwise rmsd weighting factors.

Bradley et al. shows that low-rmsd structures can often be found by using sequences homologous to the target in the Rosetta algorithm[8] where only a total of 10−20 thousand structures need to be built. Note that for the three proteins in common between their test set and ours (1tif, (1r69, and (1csp), their "Round 2" C_α rmsd results (4.1, 1.2, and 4.7 Å) are quite comparable to our simple PARAM22/GB-SA detection of their enriched decoy set (4.4, 2.4, and 4.5 Å). This favorable comparison is likely due to the fact that generating a large decoy set of 200 000 models increases the probability that there will be enough lower rmsd structures from which to detect. Moreover, it is unclear from the work of Bradley et al., whether their improvements were gained by increasing diversity or by using a computationally intensive refinement procedure (100−150 CPU-days per protein).

Despite the presence of near-native structures in every set in this work based on rmsd, other indicators such as fraction of native contacts suggest that the so-called near-natives are not near enough. With only an average of 60% native contacts for the best structures, it makes sense that the atomic resolution scoring functions may have some difficulty in detection. There are two reasons why the fraction of native contacts may be lacking. First, the side-chain prediction algorithm used for these decoy sets may not be optimal. Tests of rebuilding side chains with the SCAP method[53] led to better overall DFIRE-AA scores (results not shown). The other issue is that the fraction of native contacts may have, in analogy to a scoring function, a narrow funnel versus backbone rmsd. Most of the native contacts will collapse into place only when the protein is very close to the native in backbone rmsd space (see, for example, Figure 3c).

4.2. Scoring/Energy Functions. In this work, we looked at two diverse scoring/energy functions: one force field-based and the other statistically based. Force field-based functions are considered to be accurate but have many drawbacks. First, the standard van der Waals repulsion term is very sensitive to the positions of neighboring atoms such that structural minimization is required. Tsai et al. suggest the use of finite core repulsion terms to alleviate this issue.[6] A compromise, however, must be made to ensure that the core is repulsive enough to filter out incorrectly packed structures. Another problem with force field-based functions is that the folding funnel is trying to mimic the physical energy landscape of real proteins. As such, real proteins may have a subtle free energy gradient toward the native that requires long folding times (e.g., milliseconds to several seconds). Compared to the standard simulation times possibly using current computer resources, typically in the single-digit nanosecond range, there is a gap of several orders of magnitude.[13] A final problem with force field-based potentials is that they may be too inaccurate. Consequently, after exceptional computational effort of using them, simulations may still lead to unphysical structures.

One of the main problems with a pairwise-only statistical potential such as DFIRE-AA is the lack of a microenvironment or solvation term.[36] Many scoring functions already employ such additional terms.[6,20,36] This is needed because pairwise contacts are not statistically independent in known protein structures.[36] We believe such additions might increase the number of near-natives detected for some protein sets. The DFIRE-AA potential, like other statistical potentials, gleans information from native PDB structures. Consequently, unfolded state information is noticeably absent. This presumably leads to the large energy gradient in DFIRE-AA seen in the protein collapsing simulations (Figures 6 and 7). Atomic force fields, on the other hand, contain a relatively balanced description of unfolded and folded states. Thus, the energetic differences and subsequent propensities to drive folding are much more subtle and should be on the order of 5−15 kcal/mol, at least in terms of free energy.[54] Skolnick et al.[20] suggest parametrizing an energy function based on decoys/misfolds and near-natives. In this way, there is an enforced funnel or directionality between the two extremes which can be tuned to obtain a desired folding gradient.

The general issue regarding scoring functions is to what extent can they be optimized to achieve a significant folding funnel? Furthermore, the two key aspects of the funnel are its depth and width. Maximization of the Z-score by Tsai et

Table 9. Summary of Results for Detection and Refinement of Structures Using the *Z*-Fold Method

PDB ID	best rmsd diversity set	best rmsd	low av energy cluster	low free energy cluster	rmsd of lowest energy	best rmsd of top 5	top rescored[a]	best rmsd of top 5 rescored[a]
1ail	5.3	5.0	11.9	8.6	7.6	7.6	7.7	7.5
1csp	3.6	3.3	4.2	4.2	3.4	3.4	4.5	4.1
1ctf	3.6	2.9	8.3	4.4	3.8	3.5	3.7	3.0
1pgx	1.5	1.2	2.1	5.7	10.5	2.0	2.2	1.2
1r69	1.5	1.0	1.5	5.7	1.1	1.1	1.1	1.1
1tif	4.1	3.9	5.2	6.9	5.1	5.0	5.2	5.2
1utg	4.8	4.2	10.9	5.1	10.9	5.0	10.5	9.3
1vif	0.6	1.6	2.5	5.1	3.1	1.9	2.9	1.8
5icb	4.3	3.8	2.9	3.9	9.0	4.3	8.4	2.8
avg	3.3	3.0	5.5	5.5	6.1	3.8	5.1	4.0

[a] Rescoring potential is PARAM22/GB-SA after standard structure optimization (see text).

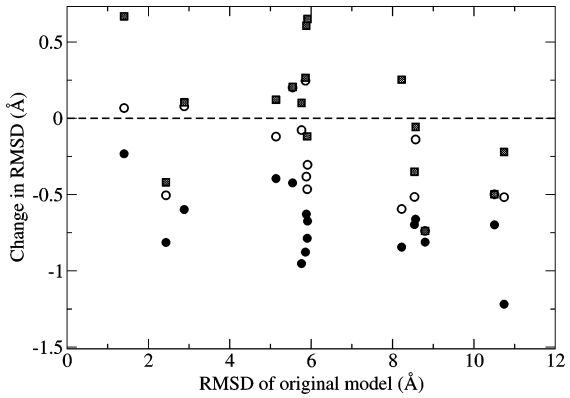

Figure 8. Refinement capability of Z-fold method as a function of rmsd of the original model for the 1pgx diversity set. Closed circles represent the lowest rmsd structures in the set, open circles denote the lowest rmsd structures out of the top five scoring conformations, and shaded squares represent the top scoring conformation.

al.[6] is an example of maximizing the depth of the scoring funnel such that the native is significantly lower in energy than any decoys. On the other hand, increasing the width of the funnel is also very important, since detection algorithms will work only if one or more model structures are within the funnel. It appears from our Z-score results, that the DFIRE-AA potential has a modestly smaller funnel depth than PARAM22/GB-SA. In addition, the overall increased enrichment scores suggest DFIRE-AA has a slightly larger funnel width. The problem with DFIRE-AA is that in some of the test sets, enrichment scores were 1 or less, suggesting that the funnel was nonexistent in the vicinity of the 15% lowest rmsd structures. The compromise to creating a wide and deep folding funnel is that, in general, the native structures of most, if not all, aqueous proteins will need to lie at or near the scoring function minimum.

4.3. Conformational Sampling. Many researchers have found that all-atom MD simulations are unable to explore diverse conformations at room-temperature despite simulation times on the order of several nanoseconds.[40] Fan and Mark[13] have suggested that even longer MD simulations on the order of hundreds of nanoseconds or even microseconds may be a viable technique for refinement. We agree that sufficiently long simulations probably would succeed some of the time. Invariably, though, simulation times of microseconds or longer are still outside most researchers' current computing capabilities.[55] Another difficulty is that the model structure to be refined may be, for all practical purposes, permanently trapped in a misfolded conformation. Misfolded proteins in vivo often require either intervention from chaperones or disposal by the cell machinery.[56,57]

In this work, we examined the ReX method which has been used successfully by Zhang et al. to sample conformational space with a sophisticated united-residue force field.[3] In fact, Misura et al.[7] commented that the addition of temperature might enhance sampling. Regrettably, the combination of an all-atom force field and ReX may not be useful without restraints, because high-temperature unfolding leads to destruction of the informational content of the original model. Furthermore, low-temperature refolding of a partially denatured structure can take an inordinately long simulation time when force-field potentials are used. In contrast, ReX simulations can be successful in loop modeling,[44] because the number of degrees of freedom are small enough to be sampled well within a feasible simulation time. In addition, the restraints of the two loop stems limit the extent of possible unfolded conformations.

Perhaps other sampling schemes such as Monte Carlo might fare better. Misura et al. performed multiple zero temperature Monte Carlo runs on small sets of decoys. They employed backbone and side-chain rotamer move sets which were able to find lower rmsd structures than the original models. One drawback was their inability to sometimes detect the lowest rmsd structures via an energy function alone. Furthermore, there was a compromise between the size of the move sets and the ability to sample rare side-chain conformations that might be crucial to achieve correct packing.[7]

The Z-fold method which entails a slight unfolding and refolding of a model conformation is a compelling alternative. It stands in contrast to simply simulating the rearrangement of a protein that is trapped in a misfolded compact state. Also, the Z-fold approach benefits from a statistical potential with a fast refolding process because the energy gradient from the partially unfolded state to a compact state is large. Nonetheless, there are several problems with using a statistical potential. First, compacting will occur at local levels causing distortion in secondary structures. This can be ameliorated somewhat through the use of secondary

structure restraints. In addition, the folds produced will be limited by the accuracy of the statistical potential. The most visible effect of this occurrence is that proteins will tend to form compact spherical structures as the competition with solvent interactions is neglected (Figure 6). Furthermore, lacking hydrogen atoms, detailed steric volume exclusions and explicit hydrogen bonding are neglected. Perhaps, a careful reweighing of energy terms and the introduction of solvation-like terms may offer the best of both worlds—a relatively fast compacting potential with diminished unphysical artifacts.

4.4. Future Directions. The fact that decoy sets with more near-native rmsd structures fared better in the detection results suggests that one should use lower-resolution models to their fullest extent before constructing and scoring all-atom models. Furthermore, all-atom model potentials are replete with local minima which hindered our dynamics-based optimization approaches. A good example of pushing the limits of united residue models is the work of Zhang et al.[45] which describes a new generation of lattice-based united residue models which can refine homology models to some degree. Furthermore, Misura et al. have shown that searches within the united-residue-based Rosetta protocol are capable of building homology models better than those created by simply constructing from a template.[58]

Given that the rmsd/score correlation values in Table 5 were suboptimal in critical rmsd ranges for most proteins, another improvement we suggest is optimizing scoring functions such as DFIRE-AA and PARAM22/GB-SA for the protein structure detection and refinement problem. For instance, the scoring funnel can be both deepened[6] and optimized to expedite folding.[59] In addition, the energy function can be smoothed[60] or transformed to enhance sampling.[61,62] Finally, hybrid strategies for conformational sampling that combine both knowledge-based and physical-based energy functions may prove to be particularly effective in refinement.[12,26]

Finally, all-atom molecular dynamics and standard replica-exchange protocols may not be the optimal methods for refinement as seen in our results. Large scale conformational changes induced by molecular dynamics are likely to be slow compared to large-scale moves possible in a Monte Carlo approach. Alternatively, enumerative sampling methods have been shown useful in small search problems such as modeling of loop regions.[15] Perhaps, local enumerative optimization could be performed on structural regions deemed to be unfavorable in energy. Regarding replica exchange, recent work of Zuckerman, et al. suggests limitations in this approach for the sole purposes of canonical sampling at 298 K.[63] Alternative sampling approaches should be considered such as genetic algorithms[64] and resolution exchange.[63]

5. Conclusion

Statistical potentials are a fast alternative to force-field-based potentials. Unfortunately, without reference to unfolded and misfolded states in their parametrization they may not be well suited to temperature-based sampling schemes.[3] Ironically, this feature makes them useful in a framework where fast refolding of structures is desired. The Z-fold method, which can produce random rearrangements of model conformations, benefits greatly from the fast refolding capabilities of the DFIRE-AA potential. Undesirably, the DFIRE-AA potential, in particular, lacks certain multibody solvent effects which will tend to cause a protein to minimize its surface area and "sphericalize" regardless of the protein's actual fold type.

The force field potential we used here includes a state-of-the-art implicit solvent model.[65] As a tool for detecting near-native structures, we believe this potential is on par with other force field potentials currently available.[66] However, there are many deficiencies in the physics of most implicit solvent force fields that still need to be addressed (e.g., charge polarization, treatment of structural waters, etc.). Deficiencies aside, the noisy nature of the energy landscape will require creative new methods in exploring conformations adjacent to the models generated by a lower-resolution potential. Temperature-based sampling schemes, such as replica exchange using different temperature windows, may not be helpful for the refinement problem on the atomic scale without additional enhancements.

Acknowledgment. We thank Drs. M. Feig and L. Carlacci for helpful discussions. M.S.L. acknowledges financial support from the Department of Defense High Performance Computing Modernization Program Office (HPCMO), the Biotechnology High Performance Computing Software Applications Institute (HSAI), and the U.S. Army Medical Research and Material Command (Project No. RIID 02-4-1R-069). We thank the Army Research Laboratory Major Shared Resource Center for computer time. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the U.S. Army.

References

(1) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucleic Acids Res.* **1997**, *25*, 3389.

(2) Miller, R. T.; Jones, D. T.; Thornton, J. M. *FASEB J.* **1996**, *10*, 171.

(3) Zhang, Y.; Kolinski, A.; Skolnick, J. *Biophys. J.* **2003**, *85*, 1145.

(4) Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D. *J. Mol. Biol.* **1997**, *268*, 209.

(5) Vincent, J. J.; Tai, C. H.; Sathyanarayana, B. K.; Lee, B. *Proteins* **2005**.

(6) Tsai, J.; Bonneau, R.; Morozov, A. V.; Kuhlman, B.; Rohl, C. A.; Baker, D. *Proteins* **2003**, *53*, 76.

(7) Misura, K. M.; Baker, D. *Proteins* **2005**, *59*, 15.

(8) Bradley, P.; Misura, K. M.; Baker, D. *Science* **2005**, *309*, 1868.

(9) Dill, K. A.; Phillips, A. T.; Rosen, J. B. *J. Comput. Biol.* **1997**, *4*, 227.

(10) Vorobjev, Y. N.; Almagro, J. C.; Hermans, J. *Proteins* **1998**, *32*, 399.

(11) Kryshtafovych, A.; Venclovas, C.; Fidelis, K.; Moult, J. *Proteins* **2005**, *61* Suppl 7, 225.

(12) Lu, H.; Skolnick, J. *Biopolymers* **2003**, *70*, 575.

(13) Fan, H.; Mark, A. E. *Protein Sci.* **2004**, *13*, 211.

(14) Xiang, Z.; Soto, C. S.; Honig, B. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7432.

(15) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J.; Honig, B.; Shaw, D. E.; Friesner, R. A. *Proteins* **2004**, *55*, 351.

(16) Rohl, C. A.; Strauss, C. E.; Chivian, D.; Baker, D. *Proteins* **2004**, *55*, 656.

(17) Feig, M.; Brooks, C. L., III *Proteins* **2002**, *49*, 232.

(18) Hsieh, M. J.; Luo, R. *Proteins* **2004**, *56*, 475.

(19) Zhou, H.; Zhou, Y. *Protein Sci.* **2002**, *11*, 2714.

(20) Skolnick, J.; Zhang, Y.; Arakaki, A. K.; Kolinski, A.; Boniecki, M.; Szilagyi, A.; Kihara, D. *Proteins* **2003**, *53* Suppl 6, 469.

(21) Mackerell, A. D., Jr.; Bashford, D.; Bellott, D. M.; Dunbrack Jr., R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, I., W.E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

(22) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999.

(23) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.

(24) Gilson, M. K.; Honig, B. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 7.

(25) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., III *J Comput. Chem.* **2003**, *24*, 1348.

(26) Zhu, J.; Xie, L.; Honig, B. *Proteins* **2006**, *65*, 463.

(27) Zhang, C.; Liu, S.; Zhu, Q.; Zhou, Y. *J. Med. Chem.* **2005**, *48*, 2325.

(28) Muegge, I.; Martin, Y. C. *J. Med. Chem.* **1999**, *42*, 791.

(29) Wang, K.; Fain, B.; Levitt, M.; Samudrala, R. *BMC Struct. Biol.* **2004**, *4*, 8.

(30) Dominy, B. N.; Brooks, C. L. *J. Comput. Chem.* **2002**, *23*, 147.

(31) Sippl, M. J.; Weitckus, S. *Proteins* **1992**, *13*, 258.

(32) Melo, F.; Feytmans, E. *J. Mol. Biol.* **1997**, *267*, 207.

(33) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J. D.; Zardecki, C. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58*, 899.

(34) Zhang, C.; Liu, S.; Zhou, Y. *Protein Sci.* **2004**, *13*, 391.

(35) Felts, A. K.; Harano, Y.; Gallicchio, E.; Levy, R. M. *Proteins* **2004**, *56*, 310.

(36) Summa, C. M.; Levitt, M.; Degrado, W. F. *J. Mol. Biol.* **2005**, *352*, 986.

(37) Zhang, C.; Liu, S.; Zhou, H.; Zhou, Y. *Protein Sci.* **2004**, *13*, 400.

(38) Wang, G.; Dunbrack, R. L., Jr. *Nucleic Acids Res.* **2005**, *33*, W94.

(39) Skeel, R. D.; Tezcan, I.; Hardy, D. J. *J. Comput. Chem.* **2002**, *23*, 673.

(40) Lee, M. R.; Tsai, J.; Baker, D.; Kollman, P. A. *J. Mol. Biol.* **2001**, *313*, 417.

(41) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminatham, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(42) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96.

(43) Zhou, R.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777.

(44) Olson, M. A.; Feig, M.; Brooks, C. L., III *J. Comput. Chem.*, submitted for publication.

(45) Zhang, Y.; Arakaki, A. K.; Skolnick, J. *Proteins* 2005.

(46) Ishikawa, Y.; Sugita, Y.; Nishikawa, T.; Okamoto, Y. *Chem. Phys. Lett.* **2001**, *333*, 199.

(47) Feig, M.; Karanicolas, J.; Brooks, C. L., III *J. Mol. Graph. Model* **2004**, *22*, 377.

(48) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.

(49) Sheinerman, F. B.; Brooks, C. L., III *J. Mol. Biol.* **1998**, *278*, 439.

(50) Stumpff-Kane, A. W.; Feig, M. *Proteins* **2006**, *63*, 155.

(51) Samudrala, R.; Levitt, M. *Protein Sci.* **2000**, *9*, 1399.

(52) Chivian, D.; Kim, D. E.; Malmstrom, L.; Bradley, P.; Robertson, T.; Murphy, P.; Strauss, C. E.; Bonneau, R.; Rohl, C. A.; Baker, D. *Proteins* **2003**, *53* Suppl 6, 524.

(53) Xiang, Z.; Honig, B. *J. Mol. Biol.* **2001**, *311*, 421.

(54) Bursulaya, B.; Brooks, C. L., III *J. Am. Chem. Soc.* **1999**, *121*, 9947.

(55) Pande, V. S.; Baker, I.; Chapman, J.; Elmer, S. P.; Khaliq, S.; Larson, S. M.; Rhee, Y. M.; Shirts, M. R.; Snow, C. D.; Sorin, E. J.; Zagrovic, B. *Biopolymers* **2003**, *68*, 91.

(56) Dalton, W. S. *Semin. Oncol.* **2004**, *31*, 3.

(57) Dobson, C. M. *Semin. Cell Dev. Biol.* **2004**, *15*, 3.

(58) Misura, K. M.; Chivian, D.; Rohl, C. A.; Kim, D. E.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 5361.

(59) Xia, Y.; Levitt, M. *Proteins* **2004**, *55*, 107.

(60) Tappura, K.; Lahtela-Kakkonen, M.; Teleman, O. *J. Comput. Chem.* **2000**, *21*, 388.

(61) Berne, B. J.; Straub, J. E. *Curr. Opin. Struct. Biol.* **1997**, *7*, 181.

(62) Fujitsuka, Y.; Takada, S.; Luthey-Schulten, Z. A.; Wolynes, P. G. *Proteins* **2004**, *54*, 88.

(63) Zuckerman, D. M.; Lyman, E. *J. Chem. Theor. Comput.* **2006**, *2*, 1200.

(64) Yang, Y.; Liu, H. *J. Comput. Chem.* **2006**, *27*, 1593.

(65) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., III *J. Comput. Chem.* **2004**, *25*, 265.

(66) Zhu, J.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2005**, *109*, 3008.

(67) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33.

CT600195F