

Can a Physics-Based, All-Atom Potential Find a Protein's Native Structure Among Misfolded Structures? I. Large Scale AMBER Benchmarking

LILIANA WROBLEWSKA, JEFFREY SKOLNICK

Center for the Study of Systems Biology, School of Biology, Georgia Institute of Technology,
Atlanta, Georgia 30318

Received 1 November 2006; Revised 23 January 2007; Accepted 8 February 2007

DOI 10.1002/jcc.20720

Published online in Wiley InterScience (www.interscience.wiley.com).

Abstract: Recent work has shown that physics-based, all-atom energy functions (*AMBER*, *CHARMM*, *OPLS-AA*) and local minimization, when used in scoring, are able to discriminate among native and decoy structures. Yet, there have been only few instances reported of the successful use of physics based potentials in the actual refinement of protein models from a starting conformation to one that ends in structures, which are closer to the native state. An energy function that has a global minimum energy in the protein's native state and a good correlation between energy and native-likeness should be able to drive model structures closer to their native structure during a conformational search. Here, the possible reasons for the discrepancy between the scoring and refinement results for the case of *AMBER* potential are examined. When the conformational search via molecular dynamics is driven by the *AMBER* potential for a large set of 150 nonhomologous proteins and their associated decoys, often the native minimum does not appear to be the lowest free energy state. Ways of correcting the potential function in order to make it more suitable for protein model refinement are proposed.

© 2007 Wiley Periodicals, Inc. J Comput Chem 00: 000-000, 2007

Key words: protein structure prediction; model refinement; all-atom potential; continuum solvent; AMBER force field

Introduction

With the improvement of protein structure prediction methods, the protein model refinement problem is becoming increasingly important. State of the art structure prediction procedures, including *TASSER*,^{1,2} *ROSETTA*,³ *PCONS*,⁴ *3D-SHOTGUN*,⁵ or *CABS*⁶ are able to assemble approximately correct structures for a significant fraction of protein sequences when a weakly homologous structure is available in the Protein Databank, PDB.⁷ In a benchmark test for proteins covering the PDB below 35% sequence identity, *TASSER* was able to predict models with a root mean square deviation from native, RMSD, <6.5 Å for ~70% of single domain proteins <200 residues in length, and ~60% proteins of <300 residues.^{2,8} Yet while these results are encouraging, the models are generally not close enough to native for use in biochemical function prediction or ligand screening as part of the drug discovery process.⁹ This fact highlights the importance of developing approaches that can refine low-resolution structures to higher resolution. A natural choice for a refinement protocol would involve a detailed atomic model and the use of all-atom physics based potentials. There has been some work in the direction of both structure ranking and refinement using all-atom potentials over the last decade. The *AMBER* potential^{10,11} assisted by different solvation

schemes was tested by Lee,¹² Hsieh,¹³ and Lee¹⁴ for their scoring ability to rank a set of structures. This work showed that *AMBER* could recognize the native structure among a variety of decoys with a good accuracy. Lazaridis et al.¹⁵ and Dominy et al.¹⁶ tested the *CHARMM*¹⁷ potential in a similar way using both decoys generated in folding experiments by other force fields, and decoys of the native fold with sequences borrowed from different proteins. In such tests, *CHARMM* also successfully scored the native structure as having the lowest energy. Similarly, the *OPLS* force field¹⁸ was shown to have the ability to find the native structure among a set of misfolded structures.¹⁹ Additionally, there are examples of successful native ranking with knowledge based and simplified all-atom potentials.²⁰⁻²²

In contrast to the promising results of structure ranking of conformations generated by alternative protocols, the case of structure refinement has seen much less success. The few reported examples include the work of Vieth,²³ Samudrala,²⁴ Simmerling,²⁵ Lee,²⁶ and Bradley²² with the best improvement being about 2 Å.²³ The common explanation for the discrepancy between the scoring and

Correspondence to: J. Skolnick; e-mail: skolnick@gatech.edu

Contract/grant sponsor: NIH; contract/grant numbers: RR-12255

refinement results is that the conformational search using an all-atom force field is computationally very demanding; thus, the requisite CPU times to achieve such an improvement are excessive. Of course, underlying this statement is the belief that extant atomic potentials are adequate and the problem is merely one of the conformational search.

For any given potential to be suitable for structure prediction or refinement, there are two conditions that need to be fulfilled: (1) the potential must score the native structure as the lowest in energy and (2) there must be a correlation of the potential energy with native-likeness (e.g. RMSD) to drive the conformational search in the direction of the native structure. By exploring the energy surfaces of 150 single domain proteins using an all-atom (*AMBER*) potential, we try to answer whether the *AMBER* potential fulfills the above conditions, and therefore whether it can be used for structure refinement. In contrast to previous studies, here we address the issue of how the conformational search, driven by the *AMBER* potential, affects the scoring results. The search is applied to both native and a set of decoy structures that span the range of significant to random relationships to the native structure. Also, we present results for the largest testing set of proteins used so far. To account for the solvation component of the free energy, we use *AMBER* with the generalized Born (GB) implicit solvation model²⁷ and also include a surface area dependent term (SA).

Using a representative set of 150 proteins and their associated protein-like decoys, we monitor the ranking of the native structure and investigate the relationship between native similarity and the energy of *AMBER/GBSA* potential as a function of search time. We conducted the tests in three different Relaxation Regimes: I. at time zero, with only minimization of all the native and decoy structures, II. after local relaxation: a 200 ps molecular dynamics (MD) search was conducted, followed by minimization, and III. after a relatively extensive search: 2 ns of MD, followed by minimization. The objective is to see how the extent of the conformational search affects the scoring results, and what is the shape of configurational free energy space for all the proteins in different relaxation windows. Finally, our goal is to answer the key question: is the search problem the main reason for the slow progress in the all-atom protein structure refinement field?

Methods

Protein Sets

In this study, we employ a previously prepared²⁸ comprehensive benchmark protein set, which includes 1489 test proteins and covers the PDB library⁷ with lengths from 41 to 200 residues at 35% sequence identity (PDB200). Both the native structure and a collection of protein-like decoys from *TASSER* are available for each protein in the set. We then randomly select a subset of 150 proteins from the PDB200 set according to following criteria: (1) *The structures do not contain large ligands, prosthetic groups, and binding partners necessary for maintaining the fold.* Such a selection is justified by the fact that we cannot include crystallization partners in the calculation; we found that most of the structures co-crystallized with large partners were not stable when subject to MD based relaxation. (2) *The structures were obtained by X-ray crystallography.* We also decided not to include most NMR struc-

tures due to their conformational ambiguity. NMR structures are usually deposited in the PDB library as a collection of models that satisfy spatial restraints from experiment. Typically, the models are composed of a structurally conserved core region and variable regions (loops and chain ends). The variable regions may cause structural differences as large as 5 Å in the $C\alpha$ RMSD from native, and the collection of models covers a large spectrum of *AMBER* energies. Since our goal is to compare the native energy with the energies of decoy structures, it is crucial to have one, well-defined native conformation. Therefore, for further calculations, we use only a few NMR structures, for which a structurally close (RMSD <2 Å) X-ray mutant or homologue structure is available in the PDB.

Some of the proteins from our testing set were crystallized as part of larger molecular assemblies but were verified by comparing different PDB entries that the same protein or a close homologue had a very similar structure despite different crystallization conditions. Our assumption was that the partners are not essential for maintaining the fold in such cases.

The set of 150 proteins will be denoted as the “150-set.” Since Regime III is computationally very demanding, we were not able to explore it for the whole 150-set. Thus, we selected 50 smaller proteins from the 150-set, termed the “50-set” in what follows. The PDB ID list for both sets is available at: <http://cssb.biology.gatech.edu/skolnick/files/all-atom/>

Decoys

The decoys used in this work come from *TASSER*-based protein structure prediction.² These decoys have protein-like topologies and interactions, yet they vary in their similarity to the native structure. *TASSER* uses a coarse-grained protein model of two interaction centers per amino acid (the $C\alpha$ and side chain center of mass, CM). All atom structures of 14,000 decoys per protein were constructed using *PULCHRA*.²⁹ Then, the decoy set for a given protein was divided into 50 intervals of descending native similarity measured by the TM-score,²⁸ and up to 20 models per each interval were chosen for further calculations (giving up to 20×50 decoys per protein). All decoy structures were minimized, and the lowest energy decoy from each TM-score interval was selected for further calculations. Separately, for a few proteins, all 14,000 decoys were minimized with *AMBER* and used for comparison. The results from both protocols are very similar, and therefore, the use of the less time consuming protocol is justified. A side chain reconstruction procedure was also applied to the native structure: the native structure was first reduced to a $C\alpha + CM$ (side chain center of mass) representation and then reconstructed with *PULCHRA*.²⁹ The native-reconstructed structure was also included in further calculations for comparison.

Structure Similarity Metrics

We use two different metrics to measure structure similarity: root-mean-square deviation between two structures, RMSD, and the template modeling score, TM-score.²⁸ While there are a variety of other structure comparison metrics that could also be used, the RMSD and TM-score metrics are chosen, as they capture most of the structural similarity features we want to monitor. The RMSD

is commonly used and well recognized and it appropriately describes the region of close structural similarity. But in the region of lower structural similarity, the information given by RMSD is very limited. For instance, a single hinge motion between two parts of a molecule (e.g. two domains) can lead to very high RMSD values, despite the structural similarity being otherwise very high. The RMSD is also protein size dependent. The TM-score, on the other hand, has no protein size dependence and finds the superposition of two structures that balances the coverage of the region of the protein with highest structural similarity and the alignment accuracy.²⁸ It weights close matches higher than distant matches. A hypothetical match of two structures that have 80% of their structures identical, but have a significantly different conformation of a terminal tail is an example. The RMSD between such structures can be very high, while the TM-score will denote a significant structural match. The TM-score ranges from (0,1], with 1 denoting identical structures. A TM-score higher than 0.4 indicates a meaningful structural similarity and lower than 0.17 means a random match.

Free Energy Function

All calculations were performed using the *AMBER* force field, ff99¹⁰ including GB/SA implicit solvation. In this approach, the solute (protein) is represented by an all-atom detailed model, while the solvent is treated as a mean electrical field approximated through GB theory.²⁷ Nonpolar solvation interactions are modeled by a term proportional to the solvent accessible surface area (SA).³⁰ The *AMBER/GBSA* free energy is then approximated as a sum of two terms: the internal energy of the protein (the molecular mechanics energy, E_{MM}) and the solvation free energy (ΔG_{solv}), that is further decomposed into polar (ΔG_{GB}), and nonpolar (ΔG_{SA}) contributions. The internal configurational entropy of the protein is neglected based on earlier predictions, that the internal entropy of a protein is similar in native, misfolded and denatured states.³¹ E_{MM} is the sum of an internal strain energy (vibration of covalent bonds and rotation of valence bond angles and torsional angles), a Van der Waals energy modeled by a Lennard–Jones potential and a protein electrostatic energy approximated as Coulomb interactions of atomic point charges.

Scoring Protocols

Three protocols were used to establish the scoring abilities of *AMBER/GBSA* potential and to monitor the dependence of the results as a function of conformational search time. The first protocol included a short, simple minimization of the native structure (termed “native-I”) and all the decoys. In the course of minimization, the structures were first relaxed, with their $C\alpha$ positions frozen for 50 steps using a distance dependent dielectric constant, to remove bad side-chain contacts that often appear in the decoy structures. Then, GB/SA solvation was turned on, and the minimization was carried out for 5000 more steps. The TM-score span of decoys for each protein was divided into 50 intervals, and the lowest energy decoys from each TM-score interval were used for scoring analysis and further calculation. This protocol includes only local relaxation of protein structures and we refer to it further in the text, as Relaxation Regime I. Then, a more crucial test was

applied: the 50 lowest energy decoys of different native structure similarity, and the native structure were subject to a 100 ps equilibration and a 100 ps MD production run with *AMBER/GBSA*. Twenty snapshots from each MD run were minimized (5000 steps) and taken for subsequent scoring analysis; we term this Relaxation Regime II. The lowest energy snapshot from native trajectory is further referred to as “native-II.” The third protocol, as it is the most computationally demanding, was performed only on a part (50) of the 150-protein set. The 50 proteins were chosen mostly randomly, with some preference to include small proteins, and proteins with a representative (average) scoring result in Relaxation Regime II. For this subset, the MD simulation was extended to 2 ns. Again, 20 snapshots from the last part of each MD run were chosen for minimization and energy versus TM-score scoring. This protocol involves thorough relaxation of native (“native-III”) and decoy structures and is denoted as Relaxation Regime III.

For all the native and decoy structures chosen for scoring, the energy gap between native structure and the lowest energy decoy is calculated to check if the force field is able to pick the native structure from a set of decoys based on energy. To test the ability of the force field to refine protein models, the correlation of the energy versus TM-score was also monitored. A funnel like energy landscape with a good energy–native likeness correlation would promote structural changes towards native during the conformational search. On the other hand, a flat and rugged energy landscape would trap decoy structures in local energy minima and prevent structural changes in the direction of the native state.

Results

Relaxation Regime I: Scoring After Minimization

Native–Decoy Energy Gap (Native Scoring)

In the least demanding test, the energies of minimized decoy structures are compared with the energy of the native-I structure (the minimized experimental structure). We find that the energy of the native-I structure is lower than the energies of all the decoys in 100% of the cases for the considered protein set. The average native-I–decoy energy gap, that is the difference between native-I energy and lowest decoy energy, is -406 kcal/mole. A representative plot of energy versus TM-score for all the structures of a given protein sequence is shown in Figure 1A for the protein lag6_. Of the X-ray proteins that we tested, the only cases in which the scoring test fails (when the native-I structure is not the lowest energy minimum) are proteins that were co-crystallized with large partners like DNA, prosthetic groups, protein ligands. These partners were not included in the energy evaluation; this is a likely reason for the failure to identify the native conformation as the lowest energy structure. Therefore, we exclude such proteins from our testing set of 150 proteins.

Interestingly, most NMR native structures were not ranked first in the set of decoys. Out of a collection of NMR models available for a given sequence in PDB, we always evaluated only the first model and it was not always the lowest energy structure of all deposited NMR models. On this basis, we decided to exclude most NMR structures from our testing 150-set. The few NMR structures

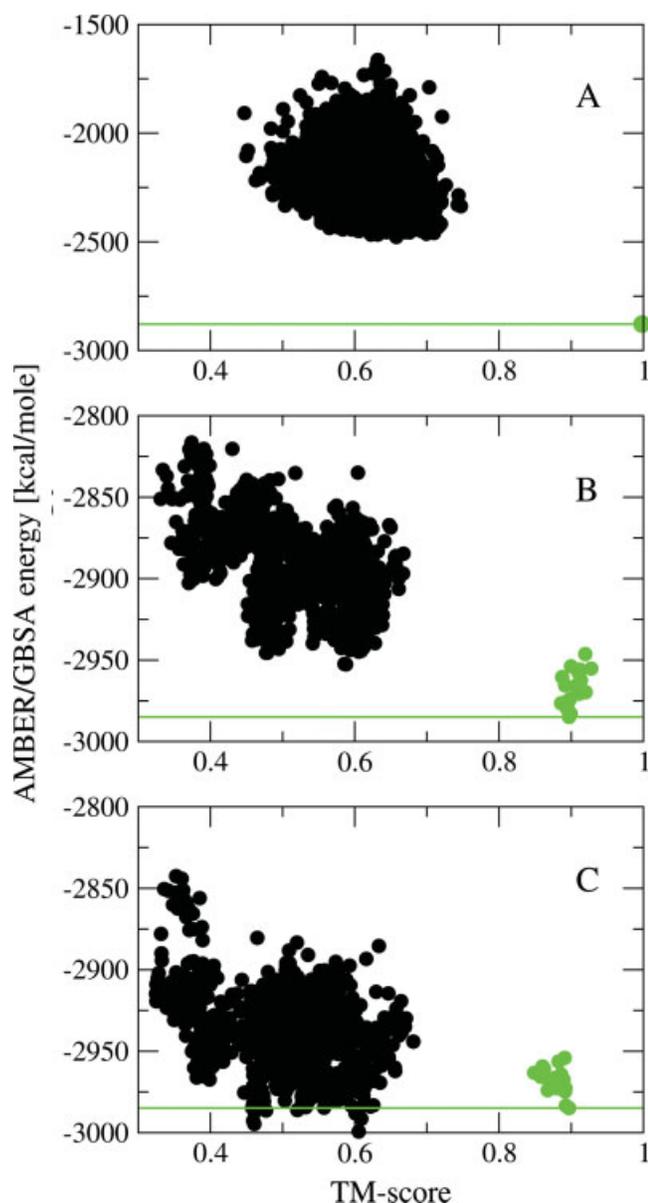


Figure 1. Representative plot of *AMBER/GBSA* energy as a function of TM-score in the three different Relaxation Regimes for protein lag6_. Green line denotes “native” energy level, black dots—decoy structures, green dots—“native structures” (minimized experimental structure or 20 minimized snapshots from MD simulation of the experimental structure). (A) Regime I; the native structure, native-I, is the minimized experimental structure, (B) Regime II; the native structure, native-II, is the lowest energy snapshot from 200 ps MD, starting from the experimental structure and (C) Regime III; the native structure, native-III, is the lowest energy snapshot from 2 ns MD, starting from the experimental structure. Note the energy scale change between Regime I, and II and III. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

that had a structurally close homologue, determined by X-ray crystallography in the PDB, passed this test (native was ranked no. 1) and therefore we use them in further calculations.

The native-I–decoy energy gap was also decomposed into the contributions from different kinds of interactions. Each individual component of the *AMBER/GBSA* native-I energy is on average lower than for the lowest energy decoy structure, with the two exceptions being the GB energy term and electrostatic interactions of bonded atoms (1–4 interactions). These two components are on average lower for the lowest energy decoy than for the native-I structure. When electrostatic interactions are considered together (electrostatics + GB + electrostatics of bonded atoms), they also favor the native-I structure.

The decoy structures that come from *TASSER* have only the coordinates of the $C\alpha$ atoms and side-chain centers of mass (CAS model) and the missing atoms were added before the *AMBER* energy was calculated, using *PULCHRA*.²⁹ We wanted to check how much the all-atom building procedure increases the energy of the native structure. We therefore converted all native structures to the CAS representation, and then applied the same reconstruction procedure as was used for decoys. Structures generated this way have on average higher energies than the original native-I structures, with an energy gap of 104.8 kcal/mole. The energy gap is significant, but much smaller than the native-I–decoy gap. These rebuilt-native structures are not considered in further calculations.

Correlation Coefficient of Energy and Native-Likeness

In the second part of the analysis, the correlation coefficients (CC) between the TM-score (RMSD) and all energy components were monitored. The purpose was to check if *AMBER/GBSA* energy components promote native like structures among decoys. Note, we exclude the native-I structures themselves from this calculation. There is practically no correlation of energy with either TM-score or RMSD observed for decoy structures. The average CC equals 0.4. The highest correlation is observed for Van der Waals energy term (0.5). For about 1/3 of the proteins, the observed CC is higher than 0.6, and in this group, the only energy term that exhibits a significant correlation with TM-score (RMSD) is again the Van der Waals term. All scoring results, also including those in Relaxation Regimes II and III, are presented in Table 1.

Decoy Scoring Behavior

When the native-I structure is compared to the decoys, the native-I always has the best energy and best TM-score. Next, we ask whether the best decoy structure can be chosen by best energy. Only in four cases does the lowest energy decoy have the best TM-score. All decoy-scoring results are presented in Table 2.

Relaxation Regime II: Scoring After 200 ps of MD

Reference Structure

In the second test, we relax all the structures with a total of 200 ps of MD with *AMBER/GBSA*, and then we repeat a similar analysis as in Relaxation Regime I. For all proteins, after 200 ps of MD there are decoys that are lower in energy than the minimized experimental structure, native-I. Also the minimized snapshots from the native MD trajectories have lower energies than the corresponding native-I structures. Clearly, the minimized experimental structure can no longer be used as a reference. We then choose 20 snapshots from each trajectory (native and decoy) in the same time

Table 1. Summary of the Results from Relaxation Regimes I–III.^a

	Regime I		Regime II		Regime III
	150 set	50 set	150 set	50 set	50 set
% of proteins with native energy ranked no. 1	100	100	70	66	20
Native–decoy energy gap	−406.05	−335.98	−25.87	−19.63	14.44
Average native energy	−4553.86	−3863.69	−4768.47	−4060.43	−4092.73
Average lowest decoy energy	−4147.81	−3527.71	−4742.60	−4040.80	−4107.17
Energy–TM-score CC	0.43	0.39	0.18	0.13	0.08

^aAll energies are given in kcal/mole and the correlation coefficients (CC) are given for decoy structures only.

frame and compare decoys to the lowest energy snapshot from the native trajectory (native-II).

Native–Decoy Energy Gap (Native Scoring)

After a short relaxation with MD, all the structures are lower in energy. However, for the decoys, the energy decrease is much more pronounced than for the native-II structures. Now, the lowest energy structure comes from the native trajectory in 70% of the cases. Thus, on average, the native-II structures are still lower in energy than the decoys, but the native-II–decoy energy gap is much smaller than in Relaxation Region I and is now −25.9 kcal/mole. A representative plot of energy versus TM-score is shown in Figure 1B for 1ag6_. In the course of MD, the native trajectory deviates from the experimental structure, and in the case of some proteins, the native-II structures are of similar quality in terms of their RMSD from the experimental structure as the best decoys. That is, the structures that started from the experimental native structure begin to drift away. We then additionally apply a cutoff for “nativeness” of a 2.5 Å RMSD from the experimental structure. All proteins with the native-II structure above the RMSD cutoff and lowest energy decoys below the cutoff are discarded. This way, we ensure the reference structure is always within 2.5 Å from the experimental structure and the decoy structure does not belong to our arbitrarily chosen native cluster. Such a filtration process leaves 118 proteins for further analysis. In this set, the lowest energy structure comes from the native trajectory in 75% of the cases, and the average native-II–decoy energy gap is −26.28 kcal/mole. Whether the filter for “nativeness” is applied or not, the two

analyses give similar results, which is also indicative of their robustness. Also decomposition of the energy gap into different energy terms is consistent in the two analyses, and the energy terms that consistently favor native-II structures are the bond, angle and Van der Waals energy terms. On the other hand, the dihedral angle energy term consistently favors decoy structures. All the other components display almost no preference towards native-II or decoy structures (they favor native-II in nearly half of the cases), but on average, the electrostatics and SA terms favor decoys and the GB term favors native-II structures.

The energy gap correlates best with the energy gaps calculated for the bond, angle and Van der Waals components. It also correlates with the number of atoms in the protein (the larger the protein, the more negative the energy gap). There is no correlation between the total energy gap and the quality (RMSD) of the lowest energy decoy. The average RMSD of the decoys that are lower in energy than the corresponding native-II structures is 6.9 Å and 1/3 of the decoys have a RMSD higher than 10 Å. For the cases of proteins when the native-II has the lowest energy, the average RMSD of the lowest energy decoy is 8.2 Å.

CC of Energy with Native-Likeness

There is no correlation observed between the RMSD of the decoys and their energy. The average CC equals 0.18, and only for two proteins is the correlation higher than 0.6. There is also very little correlation of energy with RMSD for native snapshots. The average CC is even smaller (CC = 0.03) but there are 17/150 proteins for which the correla-

Table 2. Summary of the Decoy Scoring Results for Relaxation Regimes I, II, and III.

	Regime I		Regime II		Regime III
	150 set	50 set	150 set	50 set	50 set
Average TM-score of the best decoy	0.62	0.59	0.59	0.56	0.54
No. of proteins with the best decoy ranked no. 1	4	1	0	0	0
Average TM-score of the lowest energy decoy	0.56	0.52	0.48	0.43	0.39

tion is higher than 0.6. The results are very similar, when the TM-score as a measure of native similarity is used instead of the RMSD.

Decoy Scoring

The best decoy structures after relaxation with MD have on average lower native similarity than the best starting decoys. Also the best TM-score decoy is never the lowest energy decoy.

Relaxation Regime III: Scoring After 2 ns of MD

Reference Structure

In the course of the MD simulation, lower energy states are found in both the native trajectories and in the decoy trajectories. The average improvement in energy for the minimum energy native snapshot over the native-II structure (minimum energy native snapshot from the Relaxation Regime II) is ~ 30 kcal/mole. We again use the lowest energy snapshot from last 100 ps of the native 2 ns MD simulation as the reference point, and we refer to it as native-III. Only in the case of one protein, 1ag6_, was the Relaxation Regime II snapshot lower in energy by 1.5 kcal/mole, and therefore it is used as the reference native energy.

Native–Decoy Energy Gap (Native Scoring)

In the most demanding of our three tests, the average native-III–decoy energy gap is no longer favorable for the native-III structures, and on average, it is +14 kcal/mole. For most proteins (80%), one of the higher RMSD (nonnative) decoys is the lowest energy structure. A representative result is shown in Figure 1C for 1ag6_. We also checked the results using a 2.5 Å RMSD cutoff for “nativeness,” as in Relaxation Regime II, but this does not change the results significantly. When the energy gap is decomposed into contributions from the different energy components, we again observe that the terms that most favor the decoys are the dihedral angle energy term and the electrostatics of bonded atoms (1–4 interactions). Also, the SA energy term often favors the decoys over native-III, but its contribution is very small. Bond, angle and Van der Waals interactions of bonded atoms 1–4 consistently favor native-III structures. Also, when the energy gap is calculated separately for proteins where the lowest energy decoy structure has a higher energy than the native-III structure and those whose decoys have energy lower than native-III, the largest difference between the two sets is in the Van der Waals interactions. This contribution is on average much more favorable towards native in those cases when the native-III structure is the lowest energy structure. The electrostatics and GB solvation terms favor native-like structures in approximately half of the cases, and they always contradict each other. On average, electrostatics favors decoy structures and GB favors native-III structures, but both contributions are large and they have large error bars (the addition of another protein to the testing set can completely change the average result).

CC of Energy and Native-Likeness

The energy does not exhibit any correlation with either the RMSD or TM-score; indeed for all of the proteins, the CC is lower than 0.5, and on average it is 0.1. A weak correlation (0.5–0.6) is observed only for bond, angle and Van der Waals energy terms in

the case of a few proteins. For other components, the CC is always lower than 0.5. A histogram of correlation coefficients between AMBER/GBSA energy and RMSD is presented in Figure 2.

Decoy Scoring

After 2 ns of MD, the best TM-score decoy is never the lowest energy decoy. Also, the best energy decoy has an average TM-score of 0.39, indicative of rather weak structural similarity to the native structure. Clearly, the longer the search, the lower the energy minima found for decoys that are distant from the native structure.

Drift of Decoys

We also monitor the RMSD and TM-score change of the decoys relative to the native structure in the course of the MD simulation. Using RMSD as the structural similarity measure, this analysis shows its drawbacks. While there are many structures that improved relative to the native structure with the largest improvement of ~ 10 Å towards native, most changes are actually meaningless and account only for a change in the compactness of the decoy. The RMSD improvement correlated very well with the quality of the initial decoy (the worse the decoy was initially, the larger was the change towards native). The greatest improvement was from a RMSD of 22–12 Å from native. However, the decoy structures bear no similarity to the native structure either at the beginning or the end of the MD simulation. Thus, we use the TM-score that can distinguish between meaningful and random structural changes. An improvement was found for $\sim 15\%$ of the decoys. The most significant improvement was of 0.1168 TM-score units. The accompanying RMSD improvement was only 0.2 Å.

Optimization of the AMBER Potential

In the scoring analysis, we observed that some energy components consistently favor decoy or native structures. We next checked if, by changing the relative contributions of the individual energy

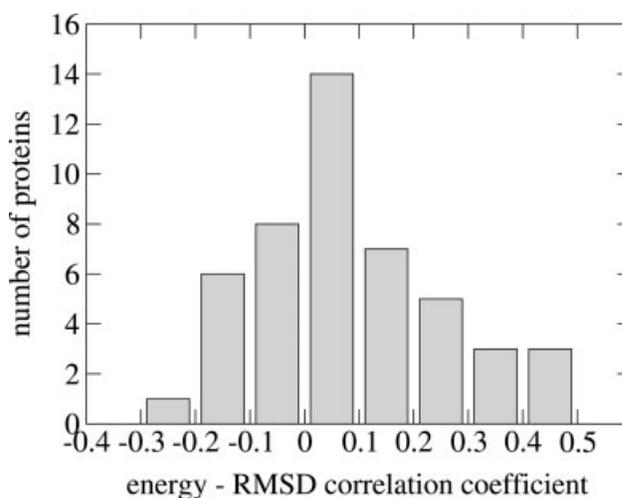


Figure 2. Histogram representation of the CCs between AMBER/GBSA energy and RMSD from the experimental structure for Relaxation Regime III.

components, we could improve the energy–native-likeness correlation. We use the CERN MINUIT program³² for optimization using the optimization function as described previously.¹ In short, the function changes the weights of the energy components to maximize the CC between energy and TM-score, and maximizes the native-decoy structure energy gap. The optimization attempt was unsuccessful, and the obtained improvement of correlation was essentially meaningless: the original average CC of 0.1 increased to 0.3 on optimization. This shows that the lack of ability of the force field to recognize native-like structures arises from the parameterization of the potential, rather than from a wrong balance of energy components.

Discussion

The ability of the *AMBER/GBSA* potential to recognize the native structure among decoys in the first scoring test, when all structures are only minimized (Relaxation Regime I), appears to be an artifact of the decoy preparation procedure. Short relaxation of all the structures with 200 ps MD (Relaxation Regime II) also does not reveal the true shape of the potential. In fact, this is the way that the scoring of all-atom potentials were previously done.^{14,16,19} Experimental structures are compact and have physical, well-minimized distances and angles. The decoy structures are not only misfolded, but they often contain unrealistic conformations of side chains and have much worse packing than experimental models. When decoys and experimental structures are only minimized prior to energy comparison, the challenge for a scoring function is mainly to recognize the most compact and best-packed structure, rather than a true native fold. This is also why the Van der Waals energy is the only contribution that correlates with native similarity in our test for Relaxation Regime I. But when all the structures are well relaxed with the scoring potential, prior to energy comparison, the differences in compactness and packing disappear, and it becomes a real challenge to select the native-like structure from a collection of alternatives. In such a test, the *AMBER/GBSA* energy fails, and the results reveal that the potential has a quite flat and rugged landscape, with many comparable minima away from the native structure. The results (average native–decoy energy gap) do not depend on the quality of the decoys used for the analysis; in the set of proteins with a decoy structure lower in energy than the native-III structure, the lowest energy decoy was $\sim 3\text{--}16$ Å away from the native structure. The results also do not depend on the secondary structural class of protein used. The average native–decoy energy gap depends however on the length and temperature of the MD simulation (results for the temperature dependence are not shown). The more thorough the search, the lower the energy minima of both nonnative and native decoys that are found. Indeed, the energy landscape seems rather flat, since no correlation is found between energy and native-likeness, even in the near native region. Previous work has also indicated some possible improvements to the accuracy of the *AMBER/GBSA* potential. These factors include optimization of the dihedral angle parameters and partial charges,^{33–39} correcting the GB solvent model,^{40–42} and the development of improved functional forms to model nonpolar solute–solvent interactions.⁴³

An attempt to optimize the weights of the *AMBER/GBSA* potential shows that a simple balancing of the different energy

terms does not significantly improve the scoring abilities of the potential. To change the potential landscape into a funnel-like shape with the minimum corresponding to the native structure, changes in the force field parameters are needed. Decomposition of native–decoy energy gaps and energy–native-likeness CCs into contributions coming from different interactions may help guide the optimization procedure. For example, the dihedral angle component consistently favors decoy structures over native. Also the electrostatic energy, that is always a large contribution, does not help to distinguish between native and decoy structures at all. An extension of such an analysis is a comparison of *AMBER* (ff99) results with the results from different force fields for the same set of proteins. Currently, we are testing *TASSER* and *DFIRE*²¹ knowledge-based potentials and the ff03 version of *AMBER* potential.³⁴ The preliminary results show that *AMBER*, ff03 performs better in ranking native structures as the lowest in energy, which is in line with our results, since one of the major changes between ff99 and ff03 was reparameterization of the backbone dihedral angle energy component. In future work, we will couple an efficient all-atom search procedure that is currently developed in our laboratory with the force field parameter optimization procedure to maximize the native–decoy energy gap and the CC between energy and native similarity.

Acknowledgments

Liliana Wroblewska thanks P. Rotkiewicz and A. Jagielska for helpful discussions and auxiliary analysis software. Part of the calculations was performed at the Pittsburgh Supercomputer Center.

References

1. Zhang, Y.; Kolinski, A.; Skolnick, J. *Biophys J* 2003, 85, 1145.
2. Zhang, Y.; Skolnick, J. *Proc Natl Acad Sci USA* 2004, 101, 7594.
3. Simons, K. T.; Bonneau, R.; Ruczinski, I.; Baker, D. *Proteins* 1999, 37 (Suppl. 3), 171.
4. Lundstrom, J.; Rychlewski, L.; Bujnicki, J.; Elofsson, A. *Protein Sci* 2001, 10, 2354.
5. Fisher, D. *Proteins: Struct Funct Genet* 2003, 51, 434.
6. Kolinski, A.; Bujnicki, J. *Proteins* 2005, 61 (Suppl. 7), 84.
7. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res* 2000, 28, 235.
8. Zhang, Y.; Skolnick, J. *Biophys J* 2004, 87, 2647.
9. Arakaki, A. K.; Zhang, Y.; Skolnick, J. *Bioinformatics* 2004, 20, 1087.
10. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J Am Chem Soc* 1995, 117, 5179.
11. Case, D. A.; Cheatham, I.; T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A., Jr.; Simmerling, C.; Wang, B.; Woods, R. J. *Comput Chem* 2005, 26, 1668.
12. Lee, M. R.; Tsai, J.; Baker, D.; Kollman, P. A. *J Mol Biol* 2001, 313, 417.
13. Hsieh, M.-J.; Luo, R. *Proteins: Struct Funct Bioinf* 2004, 56, 475.
14. Lee, M. C.; Duan, Y. *Proteins: Struct Funct Bioinf* 2004, 55, 620.
15. Lazaridis, T.; Karplus, M. *J Mol Biol* 1998, 288, 477.
16. Dominy, B. N.; Brooks, C. L., III. *J Comput Chem* 2002, 23, 147.
17. Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J Comput Chem* 1983, 4, 187.

18. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J Am Chem Soc* 1996, 118, 11225.
19. Felts, A. K.; Gallicchio, E.; Wallqvist, A.; Levy, R. M. *Proteins: Struct Funct Genet* 2002, 48, 404.
20. Summa, C. M.; Levitt, M.; DeGrado, W. F. *J Mol Biol* 2005, 352, 986.
21. Zhou, H.; Zhou, Y. *Protein Sci* 2002, 11, 2714.
22. Bradley, P.; Misura, K. M. S.; Baker, D. *Science* 2005, 309, 1868.
23. Vieth, M.; Kolinski, A.; Brooks, C. L., III; Skolnick, J. *J Mol Biol* 1994, 237, 361.
24. Samudrala, R.; Xia, Y.; Huang, E.; Levitt, M. *Proteins: Struct Funct Genet* 1999, 37 (Suppl. 3), 194.
25. Simmerling, C.; Lee, M. R.; Oritz, A. R.; Kolinski, A.; Skolnick, J.; Kollman, P. A. *J Am Chem Soc* 2000, 122, 8392.
26. Lee, M. R.; Baker, D.; Kollman, P. A. *J Am Chem Soc* 2001, 123, 1040.
27. Tsui, V.; Case, D. A. *Biopolymers* 2001, 56, 275.
28. Zhang, Y.; Skolnick, J. *Proteins: Struct Funct Bioinf* 2004, 57, 702.
29. Rotkiewicz, P.; Skolnick, J. Manuscript in preparation.
30. Sitkoff, D.; Sharp, K. A.; Honig, B. *J Phys Chem* 1994, 98, 1978.
31. Vorobjev, Y. N.; Almagro, J. C.; Hermans, J. *Proteins: Struct Funct Bioinf* 1998, 32, 399.
32. James, F. CERN Program Library Long Writeup D506; CERN Geneva: Switzerland, 1998.
33. Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J Am Chem Soc* 2002, 124, 11258.
34. Duan, Y.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A. *J Comput Chem* 2003, 24, 1999.
35. Sakae, Y.; Okamoto, Y. *J Theor Comput Chem* 2004, 3, 339.
36. Sakae, Y.; Okamoto, Y. *J Theor Comput Chem* 2004, 3, 359.
37. Mackerell, A. D.; Feig, M.; Brooks, C. L., III. *J Comput Chem* 2004, 25, 1400.
38. Lwin, T. Z.; Luo, R. *Protein Sci* 2006, 15, 2642.
39. Jagielska, A.; Skolnick, J. *J Comput Chem* 10.1002/jcc.20616.
40. Zhou, R. *Proteins: Struct Funct Gen* 2003, 53, 148.
41. Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* 2004, 55, 383.
42. Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. *J Chem Theory Comput* 2007, 3, 156.
43. Wagoner, J. A.; Baker, N. A. *Proc Natl Acad Sci USA* 2006, 103, 8331.