

Development of a Physics-Based Force Field for the Scoring and Refinement of Protein Models

Liliana Wroblewska, Anna Jagielska, and Jeffrey Skolnick

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

ABSTRACT The minimal requirements of a physics-based potential that can refine protein structures are the existence of a correlation between the energy with native similarity and the scoring of the native structure as the lowest in energy. To develop such a force field, the relative weights of the Amber ff03 all-atom potential supplemented by an explicit hydrogen-bond potential were adjusted by global optimization of energetic and structural criteria for a large set of protein decoys generated for a set of 58 nonhomologous proteins. The average correlation coefficient of the energy with TM-score significantly improved from 0.25 for the original ff03 potential to 0.65 for the optimized force field. The fraction of proteins for which the native structure had lowest energy increased from 0.22 to 0.90. Moreover, use of an explicit hydrogen-bond potential improves scoring performance of the force field. Promising preliminary results were obtained in applying the optimized potentials to refine protein decoys using only an energy criterion to choose the best decoy among sampled structures. For a set of seven proteins, 63% of the decoys improve, 18% get worse, and 19% are not changed.

INTRODUCTION

Two of the major unsolved problems in protein structure prediction involve the scoring of decoy structures, such that the most native-like structure is selected on the basis of its energy, and the refinement of low-resolution protein models to higher accuracy atomic models (1). In practice, the correct scoring of decoys is a less complex task than their systematic refinement (2). Indeed, for a significant fraction of tested proteins, many potentials correctly identify the native structure as having the lowest energy among decoys (3–10). However, only very rarely is there a correlation between energy and native similarity (11). For such typical predictions, this correlation is necessary for choosing the decoy closest to the native structure on the basis of its energy and most likely represents the physically realistic situation.

The refinement of low-resolution predicted models with a backbone root mean-square deviation (RMSD) from the native structure of ~ 6 Å to high-resolution all-atom structures whose RMSD is < 2 Å has proven to be an extremely difficult task. The solution to this problem has become more essential with the improvement of protein structure prediction methods. State of the art structure prediction procedures, including *TASSER* (11,12), *ROSETTA* (13), *PCONS* (14), *3D-SHOTGUN* (15), and *CABS* (16), generate approximately correct structures for a significant fraction of protein sequences for which a weakly homologous structure is available in the Protein Data Bank (PDB) (17). For example, in a benchmark test for proteins covering the PDB below 35% sequence identity, *TASSER* was able to predict models with an RMSD < 6.5 Å for $\sim 70\%$ of single domain proteins < 200 residues in length,

and $\sim 60\%$ of proteins of < 300 residues. However, for many important applications, such as detailed studies of interactions, molecular mechanisms, ligand screening, and drug design, more accurate structures at atomic detail are required.

For structure refinement to be routine, the correlation of energy with native-likeness has to be satisfied not only for the ranking of decoys generated in an extrinsic procedure by a different energy function but also for the collection of structures generated when the energy function drives the search. As previously demonstrated (2), the apparent correlation of energy versus native-likeness observed for one potential when the decoys are generated with another potential is often an artifact of decoy preparation. Native structures are compact with well-minimized distances and angles. Decoy structures not only are misfolded but often contain unrealistic side-chain conformations with much worse packing than experimental structures. When decoys and native structures are only minimized before energy comparison, the main challenge for a scoring function is to recognize the most compact and best-packed structure, rather than the native fold.

When all the structures are well relaxed with the scoring potential, before energy comparison, the differences in compactness and packing disappear, and it becomes a significant challenge to select the native-like structure from the sea of alternatives (2). For a set of 150 proteins, we have shown using the Amber ff99 potential (18,19) and decoys obtained with the *TASSER* force field (11,12) that a weak correlation (~ 0.4 on average) of the energy with TM-score (a measure of structural similarity that ranges from 0 to 1.0 for identical structures with a value of 0.3 for the best structural alignment of a pair of randomly related structures) is observed only for the initial set of decoys (2). Using the initial set of decoys as starting structures, after running a molecular dynamics (MD) search with the ff99 potential, this correlation decreases

Submitted September 20, 2007, and accepted for publication December 7, 2007.

Address reprint requests to Jeffrey Skolnick, Tel.: 404-407-8975; Fax: 404-385-7478; E-mail: skolnick@gatech.edu.

Editor: Ron Elber.

© 2008 by the Biophysical Society
0006-3495/08/04/3227/14 \$2.00

doi: 10.1529/biophysj.107.121947

during the course of the simulation and is lost completely after a longer search, revealing the inherent flatness of the sampled potential. Similarly, the ability of the ff99 potential to rank the native structure as the lowest energy among initial decoys for 100% of tested proteins drops to 20% after a longer conformational search.

Among the reasons that the native structure does not correspond to the global minimum of energy for many force fields and that the correlation between the energy and native similarity is low, is that during the creation of the force field, not enough information about the global shape of the energy landscape is taken into account. Such energy global landscape sculpting was employed by Zhang et al. (11) for a large set of decoys and proteins to optimize the weights of the *TASSER* force field, which employs a reduced protein model. For both sets of nonhomologous training and testing proteins, the average correlation coefficient (CC) of energy and RMSD was 0.69. A similar idea was also employed by Liwo et al. (20,21) on a much smaller set of proteins to optimize the parameters of the coarse-grained *UNRES* potential for ab initio protein structure prediction. These ideas were also employed to derive an all-atom force field (*ECEPP-5*) for the prediction of the crystal structures of organic molecules (22–24).

In this work, we explored the ability of global parameter optimization to sculpt a funnel-like landscape for the all-atom physics-based Amber ff03 (25) potential. For 58 nonhomologous proteins, we used a large number of decoys generated with the ff03 force field and optimized the relative weights of the energy components. We obtained a significant improvement in the correlation of the energy with native-likeness of the decoys and the ranking of the native structure as the lowest energy as compared to the original ff03 potential (25). Next, we showed that by adding an explicit backbone hydrogen-bond (HB) potential to the ff03 force field followed by global optimization of the combined potential, we achieved a further significant improvement in the funnel-like character of the energy landscape. We also investigated the relative contributions to the ff03 force field (supplemented by the HB potential) by turning off the electrostatic (ELE) and generalized Born (GB) solvation (26) energy components. The optimized reduced force field still scored the native structures better than the original ff03 potential and retained the improved correlation of the energy with native-likeness. Finally, we present preliminary results for protein decoy refinement using the optimized force fields and the minimum energy criterion to choose the best decoy among sampled structures. We observed refinement for 63% of decoy structures (18% of the decoys got worse, and 19% of the decoys did not change).

METHODS

Benchmarking of the ff03 force field

In this study, we used the same benchmarking protocol as previously described in detail for the evaluation of the Amber ff99 force field (2). In short,

we employed a previously prepared (12) comprehensive benchmark protein set, PDB200, which includes 1489 test proteins and covers the PDB library (17) with lengths from 41 to 200 residues at 35% sequence identity. From the set, we randomly selected 58 proteins that satisfy the following criteria: 1), the structures do not contain large ligands, prosthetic groups, or large macromolecular binding partners necessary for maintaining the fold; and 2), the structures were obtained by x-ray crystallography. For these 58 proteins (listed in Supplementary Material, Table S1), we took into consideration both the native structure and decoy structures of varying native similarity. The initial set of decoy structures were generated by *TASSER* (11,12). We chose 50 decoys per protein that span a maximal range of native similarity, and we constructed an all-atom representation of each decoy using *PULCHRA* (27).

For the native structure (i.e., the crystallographic structure) and all-atom decoys, we examined the performance of the ff03 potential (25) that includes GB and surface area (SA) dependent solvation terms (26,28) in three relaxation regimes: I) after minimization with Amber, II) after 200 ps of MD simulation, and III) after 2 ns of MD. Because of the very large CPU time cost of the above benchmarking procedure, we limited our protein set to 58 proteins. Although this protein set does not include all the proteins out of the PDB200 set that satisfy our criteria, it comprises a good representation of protein lengths and structural class. To our knowledge, it is the largest protein set ever benchmarked this extensively using an all-atom force field.

Decoys for force field optimization

To further improve the coverage of conformational space by decoys, we picked 50 low-energy decoys from MD trajectories and used them as starting structures in a thorough conformational search using the *atomic-TASSER* (*A-TASSER*) program. Finally, ~30,000 decoys per protein were collected, minimized in ff03/GB/SA potential, and used in force field optimization. We call this decoy set “Set58”. In preparing Set58, we required on average a low correlation of the decoys’ TM-score (29) to the native state with their radius of gyration to avoid the situation where the correlation is associated only with bad packing (“swollen” decoys). The average CC (CC_{ave}) of the TM-score to the native structure with the radius of gyration for Set58 was 0.40 (thus most decoys are well packed and compact but not necessarily native structures). There are six proteins (1ame_, 1em9A, 1a0b_, 1a7xA, 1bm8_, and 1a19A) in Set58 for which the correlation of the TM-score with radius of gyration was high (Table S1, column ff03). These are included to increase the diversity of the decoy set and to have some representation of less packed, “swollen” decoys in the set of structures used for parameter optimization.

Conformational search method

To search the conformational space of proteins and generate more decoy structures, we used our newly developed *A-TASSER* program. *A-TASSER* represents the protein at atomic detail and employs the replica exchange Monte Carlo (REMC) (30,31) search method with a parallel hyperbolic sampling (PHS) acceptance criterion (32) to reduce higher energy barriers. *A-TASSER* employs three types of moves that change only the torsional angles of the molecule: local “fixed end” moves (33), end moves, and the side-chain moves (Supplementary Material, Fig. S1). The rotation angle was randomly chosen within a given amplitude range. We used (–30, 30) and (0, 360) degree rotation amplitude ranges for the end moves and the side-chain torsional moves, respectively. The “fixed end” move rotates a fragment composed of a few (2–12) residues around the axis connecting the C_{α} atoms of the residues at the fragment ends, whereas the rest of the protein remains unchanged.

For each local move, the rotation amplitude is adjusted so that the backbone valence angles of the end residues of the fragment do not change beyond the statistical fluctuation range, which is ~5° (33). The amplitude of this motion typically does not exceed 30°. The end moves rotate the free ends of the molecules and involve one to five residues. The side-chain moves rotate the side-chain atoms by perturbing one or two randomly chosen torsional angles. The move types and the torsional angles to be perturbed were also randomly chosen at each step. The bond lengths and valence angles do

not change during the search (except for the backbone valence angles of the end residues of the fragment undergoing the “fixed end” move that are allowed to change within the statistical fluctuations seen in native proteins).

Force field optimization method

For each tested potential (described in the section “Types of the optimized force fields”), the energy components, E_i were multiplied by individual weights, w_i (Eq. 1), and the weights were optimized to minimize the target function, F (Eqs. 2–6).

$$E_{\text{TOT}} = \sum_i w_i E_i, \quad (1)$$

$$F = G_1 \times G_2 \times G_3, \quad (2)$$

$$G_1 = 1 - \exp(A_1 \times (CC - 1)), \quad (3)$$

$$G_2 = 1 - \frac{1}{1 + A_2 \times \chi^2}, \quad (4)$$

$$G_3 = \frac{1}{1 + \exp(A_3 \times Z\text{-score})}, \quad (5)$$

$$Z\text{-score} = \frac{\langle E_{\text{Dec}} \rangle - \langle E_{\text{Nat}} \rangle}{(\langle E_{\text{Dec}}^2 \rangle - \langle E_{\text{Dec}} \rangle^2)^{1/2}}. \quad (6)$$

During minimization of the function F , the component G_1 (Eq. 3) tends to maximize the linear CC of the total energy, ETOT, with the TM-score (29). We maximized the CC only for decoys with a TM-score to the native state in the range 1–0.4 (we remind the reader that structures with a TM-score closer to 1 are closer to the native state). Structures with a TM-score below 0.4 are usually far from the native state, and there is no reason to expect a correlation of energy with native-likeness in this regime. The energies of these structures are only expected to be higher than the energies of the structures that are closer to the native state. However, during optimization such a requirement was not explicitly enforced.

The component G_2 minimizes the deviation of the dependence of the energy versus TM-score from linearity through minimization of chi-square value χ^2 . G_3 maximizes the energy gap between the ensemble of native-like structures, $\langle E_{\text{Nat}} \rangle$ (those whose TM-score to the native structure is >0.9) and nonnative structures, $\langle E_{\text{Dec}} \rangle$, as a function of the Z-score (Eq. 6). The CC, χ^2 , and the Z-score in the function F are averaged over all proteins in the training set (described in the next section), and they depend on the weights w_i . The constants A_1 , A_2 , and A_3 were set to 2, 0.01, and 0.5, respectively, and they were chosen so that G_1 , G_2 , and G_3 all change over the same range, from 0 to 1 (or close to 1 in the case of G_1) and have a large gradient for the important ranges of the CC, Z-score, and χ^2 .

The behavior of G_1 , G_2 , and G_3 is illustrated in Fig. S2. F possesses multiple minima in parameter (w_i) space. Therefore, we used a global minimization method (34) to find the global minimum of F with respect to the weights w_i . The method is independent of the starting values of the weights and finds multiple sets of weights that minimize function F . We used the range from 0 to 5 for starting values of each weight. From this range, multiple starting weight values were sampled using the global minimization procedure (34). The final weights were unrestricted, and they could adopt values from outside the starting range. For each of 30 training subsets (described in the next section, “Training and testing protein sets”), we ran 10 independent optimization runs and collected the five lowest minima from all runs per subset. This way, we obtained 150 (30×5) sets of weights for every optimized potential. All 150 sets of weights were tested on the appropriate testing protein set.

Training and testing protein sets

From Set58, two different sets of 15 proteins were chosen randomly as training sets (Train1, Train2; Table S1). The remaining 43 proteins with respect to each of two training sets constitute the testing set (Test1, Test2). To

increase the diversity of the training set, we generated 15 subsets for Train1 and Train2 by the leave-one-out method. Thus, there were 30 (15×2) training subsets that were independently used for force field optimization.

Testing of scoring performance of the optimized force fields

The energies of decoy structures for the testing and training protein sets were calculated using the optimized force fields. To assess the scoring ability of the optimized force fields, we used the following measures: 1), the correlation coefficient of ETOT with TM-score, CC; 2), the average Z-score ($Z\text{-score}_{\text{ave}}$; Eq. 6); 3), the fraction of proteins with a CC >0.60, CC_{fr} (we considered the CC ≥ 0.60 to be a significant correlation); 4), the fraction of proteins for which the lowest energy decoy has a TM-score to the native structure >0.90, TM_{fr} ; and 5), the fraction of proteins for which the lowest energy decoy has an RMSD over C_α atoms from the native structure <2.0 Å, RMSD_{fr} .

Types of optimized force fields

Three types of the potential energy functions were used to optimize the weights: 1), the full version of the ff03 Amber potential, supplemented by GB/SA solvation, Eq. 7,

$$E_{\text{FF03}} = w_{\text{DIH}} E_{\text{DIH}} + w_{\text{VDW}} E_{\text{VDW}} + w_{\text{VDW1-4}} E_{\text{VDW1-4}} + w_{\text{ELE}} E_{\text{ELE}} + w_{\text{ELE1-4}} E_{\text{ELE1-4}} + w_{\text{GB}} E_{\text{GB}} + w_{\text{SA}} E_{\text{SA}}; \quad (7)$$

2), the ff03 potential with an explicit HB potential added, Eq. 8,

$$E_{\text{FF03/HB}} = w_{\text{DIH}} E_{\text{DIH}} + w_{\text{VDW}} E_{\text{VDW}} + w_{\text{VDW1-4}} E_{\text{VDW1-4}} + w_{\text{ELE}} E_{\text{ELE}} + w_{\text{ELE1-4}} E_{\text{ELE1-4}} + w_{\text{GB}} E_{\text{GB}} + w_{\text{SA}} E_{\text{SA}} + w_{\text{HB}} E_{\text{HB}}; \quad (8)$$

and 3), the ff03 potential with HB but without electrostatic interactions and GB (by setting the weights for those energy components to zero), Eq. 9,

$$E_{\text{FF03/HB/R}} = w_{\text{DIH}} E_{\text{DIH}} + w_{\text{VDW}} E_{\text{VDW}} + w_{\text{VDW1-4}} E_{\text{VDW1-4}} + w_{\text{SA}} E_{\text{SA}} + w_{\text{HB}} E_{\text{HB}}. \quad (9)$$

Because the sampling method keeps the bonds and valence angles unchanged, we also set the weights in front of the bond (BOND) and angle energy (ANG) components to zero.

In Eqs. 7–9, the following abbreviations are used: DIH, dihedral; VDW, van der Waals; VDW1–4, van der Waals for atom pairs separated by less than four bonds; ELE, electrostatic; ELE1–4, electrostatic for atoms separated by less than four bonds; GB, generalized Born (electrostatic component of solvation, we used the GB parameter set from Onufriev et al. (35)); and SA, the surface area-dependent term (hydrophobic component of solvation).

The hydrogen-bond potential

We tested two different approaches for the calculation of the HB energy: 1) a knowledge-based TASSER-like (11,36) HB potential, and 2) the DSSP potential (37). Although the performance in terms of native scoring and energy/native-likeness correlation of the two potentials is very similar, the DSSP energy is less computationally expensive. Therefore, the HB potential that we employed in this study follows the DSSP approach. The HB energy of the system C–O ··· H–N is calculated according to Eq. 10:

$$E_{\text{HB}} = q_1 q_2 (r(\text{ON})^{-1} + r(\text{CH})^{-1} - r(\text{OH})^{-1} - r(\text{CN})^{-1}) \times 332, \quad (10)$$

where $q_1 = 0.42e$ and $q_2 = 0.20e$, where e is the magnitude of the charge of an electron, $(q_1, -q_1)$ are point charges on C and O atoms, respectively, $(q_2, -q_2)$ are point charges on H and N atoms, respectively, $r(AB)$ is the distance between atoms A and B in Å, and E_{HB} is the energy in kcal/mol. An HB occurs when two cutoff criteria are satisfied: 1) the N-O distance is ≤ 5.2 Å, and 2) the calculated energy is < -0.5 kcal/mol. Therefore, E_{HB} is a step function; it is calculated according to Eq. 10 when the cutoff criteria are satisfied, and it is zero otherwise. Only energies for backbone HBs were calculated.

RESULTS AND DISCUSSION

Comparison of scoring performance of the ff03 and ff99 potentials

Similar to our previous study (2), we performed tests of the ff03 force field in three relaxation regimes: 1), after minimization with Amber ff03/GBSA; 2), after 200 ps of MD (followed by minimization of MD snapshot structures); and 3), after 2 ns of MD (followed by minimization of the snapshots). As in the case of the ff99 force field, we found that the initial structures, the native and decoys, are in very shallow energy minima. During the conformational search with MD, much deeper minima were found nearby, and the true shape of the potential was revealed only after a long relaxation time. The most important conclusion from this initial analysis is that the ff03 force field performs better than the ff99 potential in terms of scoring the native structure as the lowest in energy and correlation between energy and native similarity. The CC for the ff99 force field was only 0.1, whereas for ff03, it is 0.25. For the ff99 potential, native-like structures are the lowest energy among the decoys for only 20% of tested proteins. In the case of ff03, this is true for 48% of proteins, when a similar criterion for native-likeness is used (RMSD of 2 Å or less from the experimental structure). Such results are encouraging for the purpose of force field optimization, and we decided to use the ff03 potential as our base energy function in all further calculations. The optimization of the ff03 force field is required because during the MD simulations using this potential, 84% of the decoys drifted farther away from the native structure, and only 16% of the decoys improved their TM-score to the native state.

Correlation of energy with native-likeness in the original ff03 force field

For Set58, we calculated the correlation coefficient of ETOT and each energy component of the original ff03 force field (Eq. 7) with the TM-score to the native structure. The results are shown in Table 1 (the HB energy is not present in the original ff03 force field). The correlation coefficient of ETOT with TM-score is low: 0.25. Among all the energy components, the bond (BOND) and van der Waals (VDW) energies have a weak correlation with TM-score, with a CC above 0.4, whereas the remaining energy components have no correlation with native-likeness. Therefore, during optimization, one would expect the weights of these two components to dominate. Since our conformational search method fixes the bond

TABLE 1 The average correlation coefficients CC_{ave} and their standard deviations (SD) of the individual components of the original Amber ff03 potential with TM-score (rows ETOT-SA) and the CC_{ave} of the DSSP HB with TM-score for representative protein set (Set58)

Energy component	CC_{ave} (SD)
ETOT*	0.25 (0.25)
BOND [†]	0.41 (0.23)
ANG [‡]	0.26 (0.33)
DIH [§]	-0.22 (0.29)
VDW [¶]	0.52 (0.25)
VDW1-4	-0.25 (0.23)
ELE**	0.06 (0.30)
ELE1-4 ^{††}	0.05 (0.15)
GB ^{‡‡}	-0.09 (0.30)
SA ^{§§}	0.36 (0.26)
HB ^{¶¶}	0.58 (0.18)

*ETOT, total potential energy (Amber, ff03+GBSA).

[†]BOND, bond energy.

[‡]ANG, angle energy.

[§]DIH, dihedral angle energy.

[¶]VDW, van der Waals energy.

^{||}VDW1-4, short distance van der Waals energy (for atom pairs separated by less than four bonds).

**ELE, electrostatic energy.

^{††}ELE1-4, short distance electrostatic energy (for atom pairs separated by less than four bonds).

^{‡‡}GB, generalized Born solvation energy.

^{§§}SA, surface area dependent solvation energy.

^{¶¶}HB, DSSP hydrogen bond energy (not present in the original ff03 force field).

lengths and valence angles, the bond and angle energies are set to zero during optimization. It is very interesting to notice that the electrostatic interactions (ELE, ELE1-4) and GB solvation energy are completely uncorrelated with native-likeness (their CCs with TM-score are close to 0). These interactions appear to be nonspecific in recognizing similarity to the native structure. Therefore, one could expect relatively small values of the weights at those energy components during force field optimization.

There is no reason for the ELE to change monotonically with native similarity, and the native state does not have to have lower ELE than the decoys; it will be strongly protein dependent. For our decoy Set58, on average we do not observe any correlation of the ELE with native-likeness at any range of TM-score to the native state (the CC_{ave} in all ranges of TM-score are close to zero). There are only two examples of proteins with significant correlation ($CC > 0.6$) or anti-correlation ($CC < -0.6$) of the ELE with TM-score. In force fields, the “frozen” point charge approximation and the absence of polarization additionally introduce abnormally large fluctuations of the ELE, even for small changes of local geometry. In nature, the changes of electron density are smoother because large unfavorable electrostatic interactions in some conformations are quenched by the polarization of electron density as well as screening by counterions.

The GB solvation energy also has an electrostatic character and suffers from the same large nonphysical fluctuations as

the ELE, possibly caused by the point charge approximation. The solvation energy is usually favorable for extended structures, and for some proteins it may be weakly anticorrelated with native similarity, as the structures become more compact and less solvated. For Set58, the average correlation of GB energy with TM-score is close to zero at each range of TM-score to the native state, and it is negative and insignificant for most proteins. Only seven proteins have some noticeable correlation of GB energy with TM-score, among which five show a weak anticorrelation ($CC < -0.4$).

The dihedral energy (DIH) and short-distance VDW interactions on average appear to be weakly anticorrelated with native-likeness; however, their CC values are practically negligible. The dihedral energy landscape is almost flat for a wide range of the native-likeness. However, for the near-native region ($RMSD < 2 \text{ \AA}$), there is a noticeable average anticorrelation of the DIH with TM-score. There is no physical reason for the DIH to anticorrelate with native similarity. The DIH component, due to its anticorrelation with native-likeness, seems to be a reasonable candidate for improvement to increase the correlation of the ff03 total energy with native similarity.

Optimized ff03 force field

We applied the optimization procedure, described in the section “Force field optimization method” to optimize the weights of the energy components of the ff03 force field, E_{FF03} (Eq. 8). We used the training protein decoy sets described in the section “Testing and training protein sets.” The weights of the bond and angle energy components were set to 0, and the remaining weights were optimized without restraints. The results for the best set of weights (Wgt-0) are presented in Table 2. The optimized force field (column ff03 optimized Wgt-0) has a much higher CC_{ave} between the en-

ergy and TM-score compared to the original potential (column ff03). On average, over the entire Set58 (column Set58), the CC increased from 0.25 to 0.62 for the original ff03 and optimized ff03 force fields, respectively. The values of the CCs of the energy with TM-score for each protein for the original and optimized ff03 force fields are given in Table S1.

Besides the CC value, we also analyzed the values of the $Z\text{-score}_{ave}$ (Eq. 6), the CC_{fr} , TM_{fr} , and the $RMSD_{fr}$ described earlier in the section “Testing of scoring performance of the optimized force fields.” The more positive the Z-score, the better the energy separation between the native and nonnative decoys clusters. The force field optimization improved the $Z\text{-score}_{ave}$ from 0.16 to 2.30 for the original and optimized ff03 force fields, respectively. The fraction of proteins with a significant correlation coefficient, CC_{fr} , also greatly increased: from 0.12 to 0.48 for the original and optimized ff03 force fields, respectively. This means that for $\sim 48\%$ of the proteins, selecting the lowest energy decoys guarantees that the decoys are closest to the native structure. TM_{fr} and $RMSD_{fr}$ describe the ability of a force field to pick the native-like structure among decoys by an energy criterion ($TM\text{-score} > 0.90$) and to indicate by energy the near-native cluster ($RMSD < 2.0 \text{ \AA}$). The TM-score, unlike RMSD, is chain-length independent, so the two measures cannot be directly compared; but for our set of proteins and decoys, a TM-score of 0.9 roughly corresponds to an average RMSD of 1.4 \AA . The TM_{fr} value increased after optimization of the force field from 0.22 to 0.86, and the $RMSD_{fr}$ increased from 0.48 to 0.89. It is important to notice that the potential optimized on the training protein set is well transferable to the testing set.

Additional illustration of the performance of the optimized ff03 potential as compared with the original one is given in Fig. 1, A–E. Fig. 1 A presents the values of the CC of the energy versus TM-score for each protein after optimization of

TABLE 2 Comparison of scoring performance of the unoptimized and optimized force fields

Scoring performance measures	ff03*		ff03 optimized [‡]			ff03/HB optimized [§]		
	ff03/HB [†]	ff03/HB [†]	Wgt-0		Wgt-1	Wgt-1		
	Set58 [¶]	Set58 [¶]	Train	Test ^{**}	Set58 [¶]	Train	Test ^{**}	Set58 [¶]
$CC_{ave}^{\dagger\dagger}$	0.25	0.31	0.63	0.61	0.62	0.67	0.64	0.65
$Z\text{-score}_{ave}^{\dagger\dagger\dagger}$	0.16	0.23	2.65	2.18	2.30	2.59	2.19	2.29
$CC_{fr}^{\S\S}$	0.12	0.14	0.47	0.49	0.48	0.60	0.65	0.64
$TM_{fr}^{\P\P}$	0.22	0.26	0.93	0.84	0.86	1.00	0.86	0.90
$RMSD_{fr}^{\R\R\R}$	0.48	0.55	0.93	0.88	0.89	1.00	0.88	0.91

*Original unoptimized ff03 potential.

[†]Unoptimized ff03/HB potential (ff03 supplemented by hydrogen bond potential).

[‡]Optimized ff03 potential, weight set Wgt-0.

[§]Optimized ff03/HB potential (ff03 with added hydrogen bond potential, weight set Wgt-1).

[¶]Set58, the entire set of 58 proteins.

^{||}Train, training protein set.

^{**}Test, testing protein set.

^{††} CC_{ave} , average correlation coefficient of the energy with TM-score.

^{†††} $Z\text{-score}_{ave}$, average Z-score between native cluster and the remaining decoys (native cluster is defined by $TM\text{-score} \geq 0.9$).

^{§§} CC_{fr} , fraction of proteins with correlation coefficient of energy with TM score > 0.6 .

^{¶¶} TM_{fr} , fraction of proteins for which the lowest energy structure had the TM-score to the native state > 0.90 .

^{|||} $RMSD_{fr}$, fraction of proteins for which the lowest energy structure had the RMSD to the native state $< 2 \text{ \AA}$.

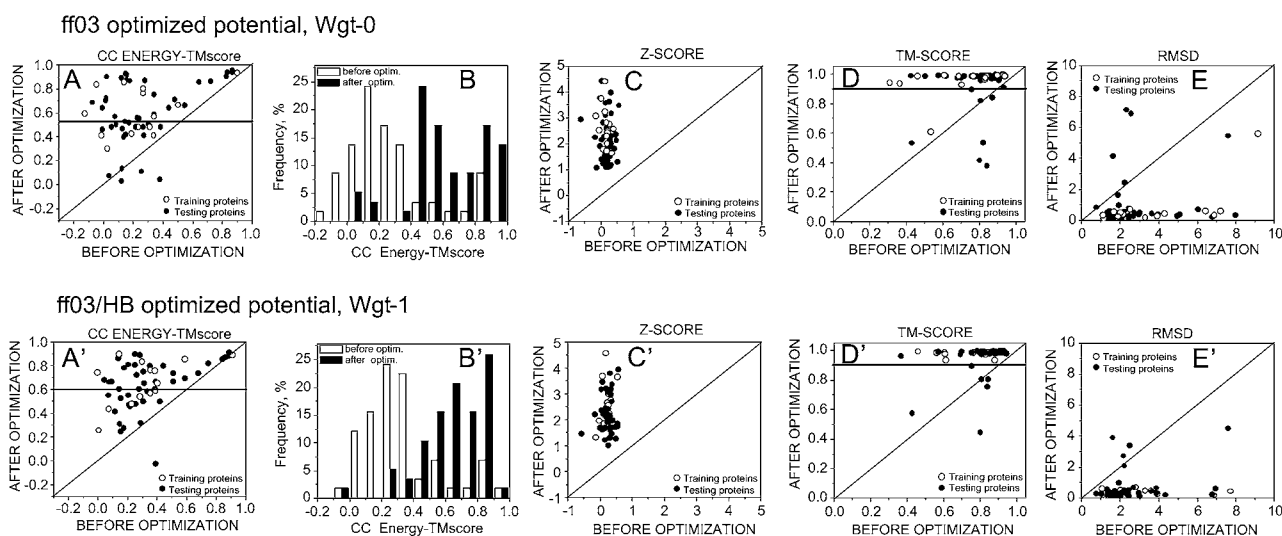


FIGURE 1 Comparison of the performance of the optimized ff03 (weight set Wgt-0) and ff03/HB (ff03 with added HB potential, weight set Wgt-1) force fields for Set58. (A–E) The results for the optimized ff03 potential. (A'–E') the results for the optimized ff03/HB potential. (A, A') CCs of the energy with TM-score to the native structure after optimization with respect to the values before optimization. (B, B') Distribution of CCs of the energy with TM-score to the native structure before (*open bars*) and after (*black bars*) optimization of the force fields. (C, C') Z-score after optimization with respect to the values before optimization. (D, D') TM-score to the native state of the lowest energy decoy after optimization with respect to the values before optimization. (E, E') C α atom RMSD to the native state of the lowest energy decoy after optimization with respect to the values before optimization. (*Open circles*) Results for the training protein set. (*Black circles*) Results for the testing protein set.

the force field with respect to the values before optimization. The CC values improved for most of the proteins (*points* above the diagonal) for both the training (*open circles*) and testing (*black circles*) sets. The improvement of the CC is also shown in Fig. 1 B for different intervals of the CC values, where the bars represent the percentage of the proteins with the CC in a given interval. The black bars represent the distribution after optimization of the force field, and the open bars represent the distribution before the optimization. There is a visible shift of the distribution toward the significant range of the CC values. Fig. 1, C–E, show the values of the Z-score, TM-score of the lowest energy structure, and the RMSD of the lowest energy structure, respectively, for each protein after optimization of the force field with respect to the values before optimization. The Z-score values increased for all proteins (Fig. 1 C). The TM-score and the RMSD to the native structure of the lowest energy decoy improved for the majority of the proteins (Fig. 1 D, *points* above the diagonal for TM-score; Fig. 1 E, *points* below the diagonal for RMSD).

Influence of explicit hydrogen-bond potential on the correlation of the energy with native-likeness and the scoring of the native structure

When the explicit HB potential (Eq. 10) that implicitly contains the angular dependence of the HB energy is added to the original ff03 force field (with weight 1), the performance of the force field improves. In Table 2, columns ff03 and ff03/HB compare the values of the CC_{ave} , the Z-score, CC_{fr} , TM_{fr} , and $RMSD_{fr}$ for the original ff03 and for the ff03 with the HB

potential included (unoptimized). All the control values improve after adding the HB potential. However, the CC_{ave} of the total energy with TM-score increases from 0.25 to only 0.31, whereas the CC_{ave} of the HB energy alone with TM-score (Table 1, HB) is much larger, 0.58. Therefore, optimization of ff03/HB should allow further improvement in the accuracy of the force field.

It is important to notice that the DSSP formulation of HB potential is very similar to the electrostatic potential of interaction between C=O and N-H groups in the original ff03 force field, with some difference in the point charges on the C, O, N, and H atoms, and the use of cutoffs in the DSSP potential. The addition of the DSSP-like energy (on top of the already existing HB description in the ff03 force field) evidently helps in native scoring and improves the ETOT correlation with native similarity. This may reflect either the ability of the DSSP potential to somehow better score the correctly oriented hydrogen bonds (due to its cutoffs) or the problems in the balance between the HB and other energy components (the HB energy in the original ff03 force field may be overwhelmed by other dominating energy terms, e.g., the remaining electrostatics).

Optimized ff03/HB force field

Optimization greatly improves the accuracy of the combined ff03/HB force field. The values of the CCs of the energy with TM-score for each protein for the unoptimized and optimized ff03/HB force fields are given in Table S1. The optimized ff03/HB force field (called Wgt-1) also outperforms the op-

timized ff03 potential (Table 2). The CC_{ave} for the optimized ff03/HB Wgt-1 potential is higher than that for the ff03 optimized potential (0.65 compared to 0.62; Table 2, column Set58). The fraction of proteins with a significant CC increased from 0.48 to 0.64, and the recognition of the native-like structure (TM_{fr}) and native cluster ($RMSD_{fr}$) is also better: 0.90 compared to 0.86 and 0.91 compared to 0.89,

respectively. Fig. 1, $A'-E'$, show a graphic representation of the performance of the optimized ff03/HB Wgt-1 force field. Fig. 1 A' presents the values of CC of the energy versus TM-score for each protein after optimization of the force field with respect to the values before optimization.

The CC improved for almost all the proteins, and the improvement is on average larger than that for the optimized ff03

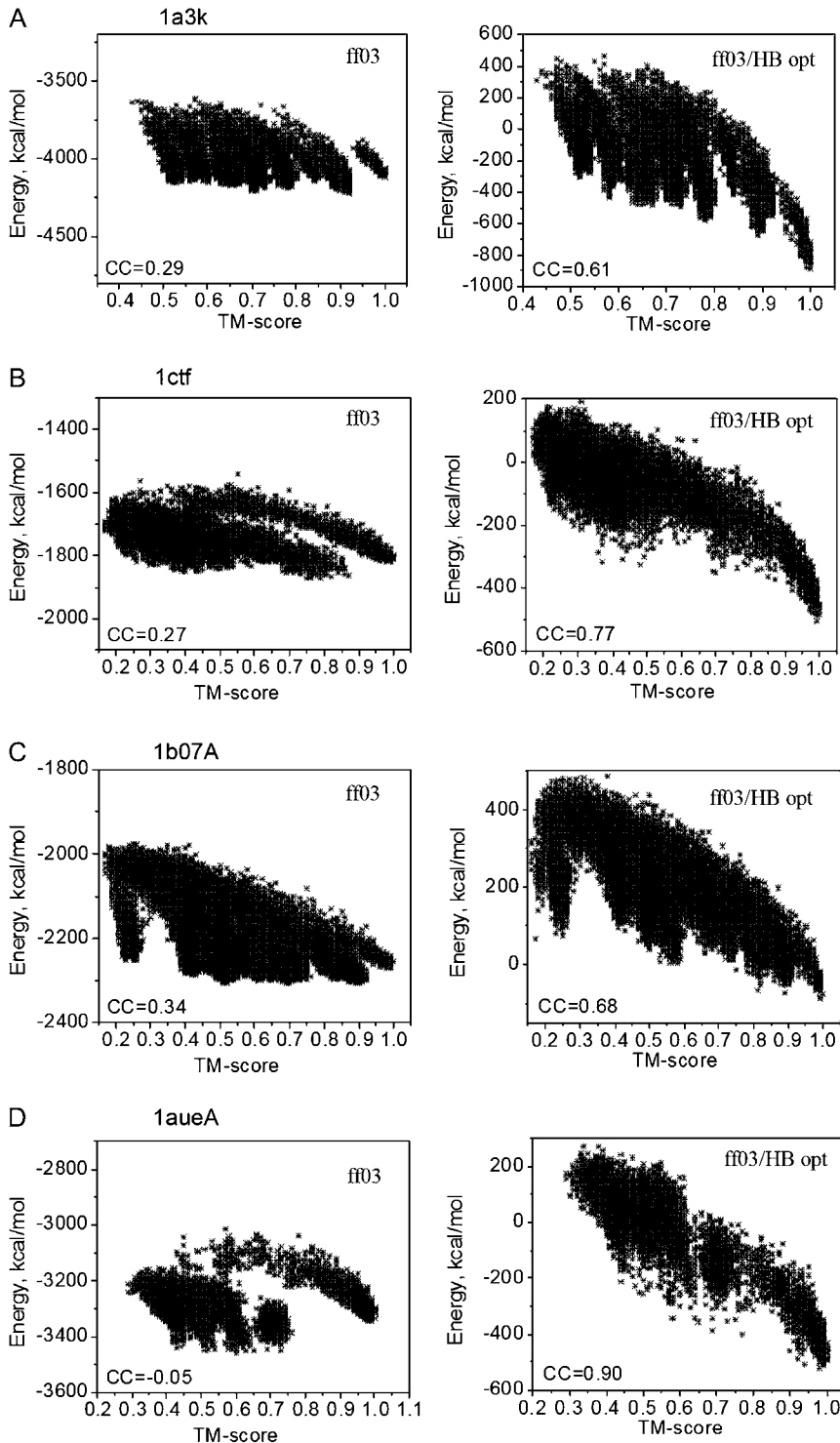


FIGURE 2 Scatter plots of the energy versus TM-score for decoy structures for the original unoptimized ff03 force field (ff03, weight set Wgt-0) and optimized ff03/HB potential (ff03/HB opt, weight set Wgt-1). CC-energy/TM-score correlation coefficient.

force field. Also, the distribution of the CC moved toward larger values, significantly more than for the optimized ff03 force field (compare Fig. 1 *B'* with Fig. 1 *B*). The Z-score improved for all the proteins (Fig. 1 *C'*), and the TM-score (Fig. 1 *D'*) and RMSD (Fig. 1 *E'*) to the native state of the lowest energy structure improved for the great majority of the proteins. For additional illustration, in Fig. 2, *A–D*, we show examples of the plots of energy versus TM-score for the original ff03 (unoptimized) potential and the optimized ff03/HB Wgt-1 potential. Fig. 2, *A–C*, illustrates the average improvement of the CC, and Fig. 2 *D* shows an example of a very large improvement of the CC.

These results show the importance of an accurate HB scheme for improving the correlation of the energy with native-likeness of protein decoys. The increase of the relative weight of hydrogen bond potential in the force field optimization process may indicate that this potential in the original ff03 force field may be dominated by other energy components, and it is relatively too small (i.e., other energy components may be too large). Hydrogen bonding was previously shown to be a necessary requirement for the generation of protein-like structures (38). The HB potential that contains an angular dependence of the HB energy is sensitive to small changes of the angular orientation of the atoms that form a hydrogen bond. This is reflected in the continuous increase of the energy of the structures as their hydrogen bonding deviates from the perfect pattern and the good correlation of HB energy with native-likeness, even in the region close to the native structure. Many well-packed, but misfolded structures with distorted hydrogen bonding become higher in energy. Such a potential can help to recognize misfolded structures among well-packed decoys that are sometimes difficult to distinguish by van der Waals energy alone.

As in the case of the optimized pure ff03 potential, the optimized ff03/HB shows good transferability between the training (Train) and testing (Test) protein sets (Table 2, ff03/HB optimized Wgt-1).

Weights for optimized ff03/HB force field

Among many sets of weights obtained during the optimization procedure that minimize the target function F in Eq. 2, the best performance in decoy scoring showed the sets with some of the weights negative for both the ff03 (Table 3, Wgt-0) and ff03/HB (Table 3, Wgt-1) force fields. The performance of these weight sets was discussed above. In the best weight set for the ff03/HB potential (Table 3, Wgt-1), the VDW, short-distance VDW, and HB energies have positive and relatively large weights. The remaining weights, of the DIH, electrostatic (ELE, ELE1-4), GB solvation, and SA energy terms, are negative. The occurrence of the negative weights for these terms indicates that they are not individually useful in generating a funnel-like shape of the potential. By assigning negative, nonphysical weights, the optimization procedure creates a linear combination of the energy terms that has larger

TABLE 3 Relative weights of energy components for the optimized force fields

Energy Component	ff03 optimized*		ff03/HB optimized [†]		ff03/HB reduced optimized [‡]
	Wgt-0 [§]	Wgt-1 [§]	Wgt-2 [¶]	Wgt-3	Wgt-R [¶]
BOND**	0	0	0	0	0
ANG ^{††}	0	0	0	0	0
DIH ^{‡‡}	-1.25	-1.17	-0.32	0.28	-0.42
VDW ^{§§}	1.00	1.00	1.00	1.00	1.00
VDW1-4 ^{¶¶}	1.04	0.88	0.56	0.56	4.33
ELE	-0.27	-0.40	-0.25	0.03	0
ELE1-4 ^{***}	-0.16	-0.23	-0.22	0.17	0
GB ^{†††}	-0.22	-0.23	-0.14	0.18	0
SA ⁺⁺⁺	-0.51	-2.07	0.14	3.39	0.51
HB ^{§§§}	0	6.25	1.32	2.56	4.26

All weights were scaled so that the weight for VDW energy = 1 for easier comparison.

*Optimized original ff03 force field.

[†]Optimized ff03/HB force field (ff03 with added hydrogen bond potential).

[‡]Optimized reduced ff03/HB force field (ff03 with added hydrogen bond potential, and with electrostatic (ELE and ELE1-4) and generalized Born solvation (GB) energy components turned off).

[§]Wgt-0 and Wgt-1, the best weight set for ff03 and ff03/HB potentials respectively (no restriction on the sign of the weights).

[¶]Wgt-2 and Wgt-R, the weight sets with allowed negative weights for dihedral (DIH), and for [¶]Wgt-2 also electrostatic (ELE, ELE1-4), and generalized Born solvation (GB) energies.

^{||}Wgt-3, the weight set with all the weights positive.

**BOND, bond energy.

^{††}ANG, angle energy.

^{‡‡}DIH, dihedral angle energy.

^{§§}VDW, van der Waals energy.

^{¶¶}VDW1-4, short distance van der Waals energy (for atom pairs separated by less than four bonds).

^{|||}ELE, electrostatic energy.

^{***}ELE1-4, short distance electrostatic energy (for atom pairs separated by less than four bonds).

^{†††}GB, generalized Born solvation energy.

⁺⁺⁺SA, surface area dependent solvation energy.

^{§§§}HB, hydrogen bond energy.

correlation with native-likeness than do the individual components.

Although there is no reason for any energy component alone to have a correlation with native-likeness and although this correlation is expected for the total energy, analysis of such individual correlations can help us to interpret the meaning of the weights in the optimized potential. In the case of the DIH, the negative weight most likely reflects its initial anticorrelation with TM-score (Table 1, DIH) (discussed earlier in the section “Correlation of energy with native-likeness in the original ff03 force field”). The weak anticorrelation of the DIH reflects distortion of the backbone torsional angles in the ff03 force field from their experimental values, especially for near-native conformations.

The electrostatic (ELE, ELE1-4) and GB energies are completely uncorrelated with TM-score, and their weights are relatively small. Therefore, the negative sign of the weights for these energy components does not introduce an unphysical

dependence of these terms on native-likeness—they remain uncorrelated and small. As expected, the weights of the energy terms that had initial low CC (ELE, ELE1-4, GB) are relatively smaller than the weights of the terms showing larger initial correlation of energy with TM-score (VDW, SA, HB, VDW1-4, and DIH).

The negative weight for the SA energy (Tables 1 and 3) is partly an artifact of the optimization procedure and reflects the fact that the average correlation of the SA energy with TM-score is weak for our decoy set ($CC = 0.36$). The SA energy landscape is flat for a wide range of native similarity up to an RMSD from native >8 Å, reflecting the low dependence of our decoy set on the radius of gyration and the compactness of the decoys. Only in the near-native region does the SA have a noticeable correlation with TM-score. The values of the SA energy are smaller than other energy components (roughly two orders of magnitude smaller than the ELE), and assigning it a negative weight probably helps to balance some deficiencies of the correlation of the other energy terms. For a physical potential, we require a positive weight for the SA energy term, since it represents the hydrophobic energy, and it should energetically favor the transition from the unfolded conformation to a more globular one, not the opposite. Including more unfolded decoys in the force field optimization process should help to obtain a positive weight for the SA-dependent energy term.

Although the linear combinations of the components with some negative weights may produce a potential that correctly scores compact decoy structures, such a potential may not be useful for applications associated with the generation of the new structures (e.g., the refinement of the protein decoys). The most important future goal is to use the optimized force field for the refinement of protein models. For this purpose, we need a potential with the smallest number of negative weights but which still performs very well in decoy scoring with a good energy/native-likeness correlation. Restricting more weights to positive values decreases the performance of the potential; therefore we chose the weight set that is a compromise between the number of negative weights and the performance, set ff03/HB Wgt-2, and we compared its performance with the best unrestricted ff03/HB Wgt-1 set (Table 4). Set Wgt-2 is the best performing weight set under the requirement that the weights of the VDW, short-distance VDW (VDW1-4), SA, and HB energies are positive (Table 3, Wgt-2). The negative weights for the ELE, ELE1-4, and GB solvation have no physical meaning, because these energy components are uncorrelated with native-likeness and their relative weights are small.

We also allowed small negative values of the weights for the dihedral angle energy, because this energy is weakly anticorrelated with native-likeness and can be partially compensated for by long- and short-distance VDW interactions. As shown in Table 4, the performance of the Wgt-2 set is slightly worse than that of the Wgt-1 set; however, it is still significantly better than that of the unoptimized force field

(Table 2, ff03/HB). The CC_{ave} between the energy and TM-score is 0.61, the CC_{fr} is 0.54, and the ability to indicate the native-like structure (TM_{fr}) and native cluster ($RMSD_{fr}$) remains very high: above 0.90. These results mean that using the ff03/HB Wgt-2 potential should allow the refinement of decoy structures for $\sim 54\%$ of proteins.

We also analyzed the performance of the force field with only positive weights. The best potential with all the weights positive, ff03/HB Wgt-3, performs slightly worse than do the Wgt-1 and Wgt-2 potentials (Table 4, column Wgt-3). The CC_{ave} between the energy and TM-score is 0.57, CC_{fr} is 0.47, and the ability to indicate the native-like structure (TM_{fr}) and native cluster ($RMSD_{fr}$) is still good: above 0.70. These results mean that using the ff03/HB Wgt-3 potential should allow the refinement of decoy structures for $\sim 47\%$ of proteins. The weights for this potential are listed in Table 3, Wgt-3.

Comparison of the performance of the Wgt-1, Wgt-2, and Wgt-3 potentials is shown in Fig. S3. The CC is still improved for the great majority of proteins, and the CC distribution is shifted toward the significant values for both Wgt-2 (Fig. S3, *A'* and *B'*) and Wgt-3 (Fig. S3, *A''* and *B''*), compared with the unoptimized ff03/HB potential. The Z-score improved and is positive for all the proteins for both sets (Fig. S3, *C'* and *C''*). The scoring of the native structure and of the native cluster for the Wgt-2 is as good as for the Wgt-1 (Fig. S3, *D'* and *E'*) and becomes a bit worse for Wgt-3 (Fig. S3, *D''* and *E''*).

Reduced optimized ff03/HB force field

As discussed in previous sections, the electrostatic and GB energy terms have a very low correlation with the TM-score and do not show specificity in recognizing the native structure. The magnitude of the electrostatic and GB solvation energies is larger than that of the other energy components (roughly by an order of magnitude) and introduces a noisy uncorrelated background. During optimization, the weights of these terms tend to decrease, resulting in the decrease of the background noise and increase of the relative contribution to the total potential of the remaining energy terms. For the purpose of protein structure refinement, which is our ultimate future goal, it may be reasonable to turn off the electrostatics and GB solvation energy. These components do not help drive the structure toward the native state, and they are the most time consuming to calculate.

Following these arguments, we optimized the ff03/HB force fields with the weights set to zero for the ELE, short-distance electrostatic (ELE1-4), and GB components of energy. As previously, the weights for BOND and ANG were also set to zero. The optimization procedure was the same as described in the section “Force field optimization method”. We chose the best performing weight set, Wgt-R (Table 3) with the requirement for positive weights for the VDW, VDW1-4, SA, and HB energy terms, allowing the DIH to have small negative weight. In Table 5, we compare the

TABLE 4 Comparison of scoring performance of the ff03/HB-optimized force fields with different weight sets

Scoring performance measures	Wgt-1*			Wgt-2 [†]			Wgt-3 [‡]		
	Train [§]	Test [¶]	Set58	Train [§]	Test [¶]	Set58	Train [§]	Test [¶]	Set58
CC ^{ave} **	0.67	0.64	0.65	0.65	0.59	0.61	0.62	0.55	0.57
Z-score ^{ave} ††	2.59	2.19	2.29	2.49	1.86	2.02	2.12	1.49	1.65
CC ^{fr} ‡‡	0.60	0.65	0.64	0.73	0.47	0.54	0.60	0.42	0.47
TM ^{fr} §§	1.00	0.86	0.90	1.00	0.86	0.90	0.73	0.72	0.72
RMSD ^{fr} ¶¶	1.00	0.88	0.91	1.00	0.88	0.91	0.80	0.79	0.79

*Wgt-1, the best weight set (no restriction on the sign of the weights).

[†]Wgt-2, the weight set with allowed negative weights for dihedral (DIH), electrostatic (ELE, ELE1-4), and generalized Born solvation (GB) energies.

[‡]Wgt-3, the weight set with all the weights positive.

[§]Train, training protein set.

[¶]Test, testing protein set.

^{||}Set58, the entire set of 58 proteins.

**CC^{ave}, average correlation coefficient of the energy with TM-score.

††Z-score^{ave}, average Z-score between native cluster and the remaining decoys.

‡‡CC^{fr}, fraction of proteins with correlation coefficient of energy with TM score >0.6.

§§TM^{fr}, fraction of proteins for which the lowest energy structure had the TM-score to the native state >0.90.

¶¶RMSD^{fr}, fraction of proteins for which the lowest energy structure had the RMSD to the native state <2 Å.

performance of the reduced ff03/HB (Wgt-R) potential with the full ff03/HB (Wgt-2) force field, optimized under similar restrictions of positive weights for VDW, VDW1-4, SA, and HB energy terms. For the Wgt-R, there is a slight decrease of performance compared to Wgt-2, visible in the change of CC^{ave} from 0.61 to 0.58, the decrease of the CC^{fr} from 0.54 for optimized ff03/HB to 0.43, and slightly worse recognition of the native-like structure (TM^{fr}) and native cluster (RMSD^{fr}). Comparison of the performance of the Wgt-2 and Wgt-R is shown in Fig. S4. Restricting weights to only positive values does not change the results significantly (results not shown).

The reduced force field should be able to refine structures for 43% of the proteins and to find the native structure by

energy criterion for over 70% of the proteins, and it is less computationally demanding than the full potential. With electrostatics and GB solvation energies turned off, the dominating weights are those for the short-distance van der Waals (VDW1-4) and HB energies (Table 3, Wgt-R).

Refinement of protein decoy structures using optimized force fields

One important application of the optimized force field is the scoring of protein decoy structures and the selection by energy of the structure closest to the native. We demonstrated in this work that the optimized physics-based force field shows

TABLE 5 Comparison of scoring performance of the ff03/HB-optimized and ff03/HB-reduced optimized force fields

Scoring performance measures	ff03/HB optimized* Wgt-2			ff03/HB reduced [†] optimized Wgt-R		
	Train [‡]	Test [§]	Set58 [¶]	Train [‡]	Test [§]	Set58 [¶]
CC ^{ave}	0.65	0.59	0.61	0.58	0.58	0.58
Z-score ^{ave} **	2.49	1.86	2.02	1.69	1.51	1.56
CC ^{fr} ‡‡	0.73	0.47	0.54	0.53	0.40	0.43
TM ^{fr} §§	1.00	0.86	0.90	0.67	0.72	0.71
RMSD ^{fr} ¶¶	1.00	0.88	0.91	0.73	0.88	0.84

Both potentials were optimized under similar conditions, allowing a negative weight at the DIH. In Wgt-2, the electrostatic (ELE and ELE1-4) and GB energies also had negative weights, and in Wgt-R the corresponding weights are set to zero.

*Optimized ff03/HB potential, Wgt-2.

[†]Optimized ff03/HB reduced potential, Wgt-R (with electrostatic (ELE and ELE1-4) and generalized Born solvation (GB) energy components turned off).

[‡]Train, training protein set.

[§]Test, testing protein set.

[¶]Set58, entire set of 58 proteins.

^{||}CC^{ave}, average correlation coefficient of the energy with TM-score.

**Z-score^{ave}, average Z-score between native cluster and the remaining decoys.

‡‡CC^{fr}, fraction of proteins with correlation coefficient of energy with TM score >0.6.

§§TM^{fr}, fraction of proteins for which the lowest energy structure had the TM-score to the native state >0.90.

¶¶RMSD^{fr}, fraction of proteins for which the lowest energy structure had the RMSD to the native state <2 Å.

reasonable progress toward addressing these goals. Another goal, more complicated to achieve, is to refine protein decoy structures using the optimized force field, i.e., to bring them closer to the native state. We conducted preliminary tests of the optimized force fields in protein decoy refinement, using the REMC procedure described in the section “Conformational search method”. For seven proteins (in the range of 55–77 residues that are β and α/β proteins, chosen randomly from the test set) (Table 6), we picked 100 decoy structures per protein from our optimization decoy set; so they span a wide range of TM-score (0.2–1) to the native structure. For each decoy, we performed 300 swaps between replicas and 200 steps of PHS between each swap. Then, the lowest energy structure from each decoy trajectory was compared with the starting decoy structure.

The results of the refinement for all seven proteins, obtained with the reduced ff03/HB force field, Wgt-R, are presented in Fig. 3, *A* (TM-score) and *B* (RMSD). The circles represent the TM-score (RMSD) to the native structure of the lowest energy decoy from each refinement trajectory with respect to the TM-score (RMSD) of the starting decoy. We define refinement as an improvement of the TM-score (RMSD) to the native state. The structure refines for the majority of the decoys (63% improve, 19% do not change, and 18% get worse, when the TM-score is used, and 67% improve, 2% do not change, and 31% get worse, when the RMSD is used). For 16% of the structures, the improvement is >0.05 in TM-score units (for 13% of structures the improvement is >0.5 Å RMSD). The largest refinement, measured as a TM-score increase, was for

TABLE 6 The TM-score and RMSD to the native structure of the lowest energy decoy after refinement of the native and protein decoy structures

PDB ID	Number of amino acids	Secondary structure	TM-score*	RMSD [Å] [†]
1fccC	56	α/β	0.95	0.61
1cskA	58	β	0.76	2.80
1c9oA	66	β	0.83	1.72
1ctf	68	α/β	0.98	0.44
1c6vX	55	β	0.83	2.06
1bxyA	60	α/β	0.96	0.54
1clyB	77	α/β	0.81	2.03

The ff03/HB-reduced optimized potential with Wgt-R weight set was used for refinement of the native and decoy structures. *The TM-score to the native structure over C α atoms of the lowest energy decoy obtained during refinement of the set of decoy structures; [†]the RMSD to the native structure over C α atoms of the lowest energy decoy obtained during refinement of the set of decoy structures (in angstroms).

1c6vX, a β protein, which improved from a TM-score of 0.44–0.61 and from a RMSD of 4.32–2.77 Å to the native structure (Fig. 4).

The energy of the decoys after refinement shows a good correlation with native-likeness for most proteins; an example of the energy versus TM-score (RMSD) plot is given in Fig. 3 *C* (*D*), for 1c6vX. We also ran the conformational search starting from the native structure of each protein. After the refinement, the lowest energy structure from the combined 100 trajectories of decoys and the trajectory of the native structure is within an ~ 2 Å RMSD to the native structure

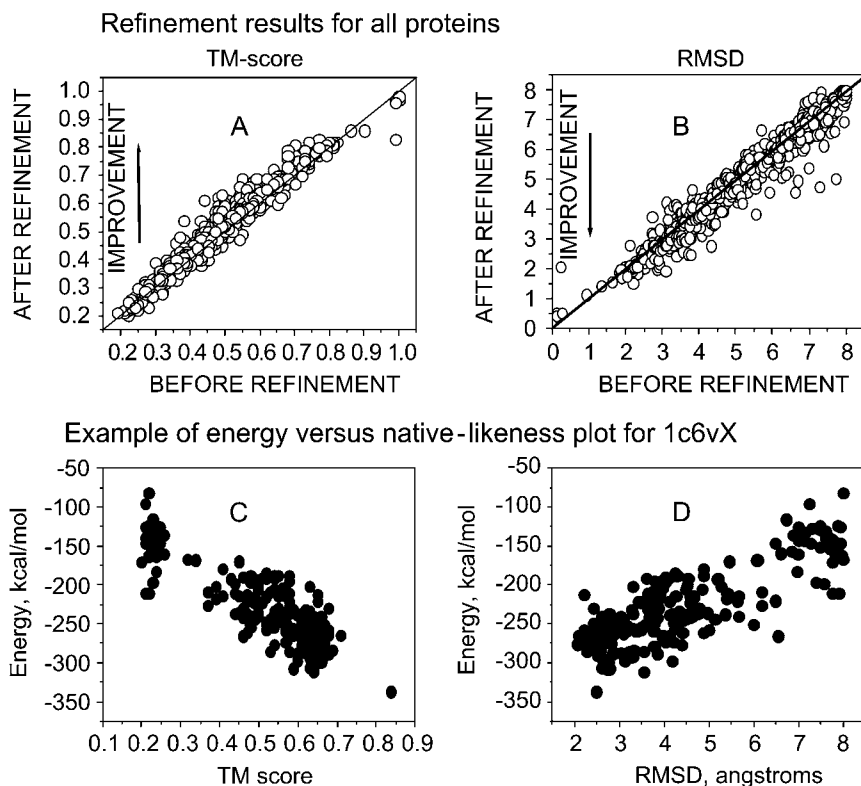


FIGURE 3 Results of the refinement of decoy structures (100 decoys per protein, seven proteins, weight set Wgt-R). (*A*) The TM-score to the native structure of the lowest energy decoy after refinement with respect to the TM-score of the initial structure for each refinement trajectory. (*B*) the C α atom RMSD to the native structure of the lowest energy decoy after refinement with respect to the RMSD of the initial structure for each refinement trajectory. (*C*) Example (for 1c6vX) of the energy versus TM-score cloud obtained during refinement. (*D*) Example (for 1c6vX) of the energy versus RMSD cloud obtained during refinement.

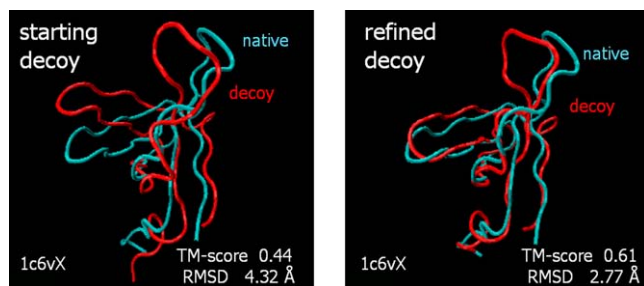


FIGURE 4 Example of starting and refined decoy (*red*) for 1c6vX, superimposed with the native structure (*blue*).

(TM-score >0.80) for most proteins (Table 6). Only for 1cskA does the lowest energy structure have an RMSD of 2.8 Å from native (TM-score = 0.76). We consider these refinement results very promising, and we are currently testing our optimized force fields and refinement method on a large set of proteins, A. Jagielska, L. Wroblewska, and J. Skolnick, unpublished.

CONCLUSIONS

In this work, we explored the applicability of the global optimization method based on a large set of protein decoy structures for many proteins to generate a funnel shape of the energy to the native structure for an Amber ff03 based, all-atom potential. Such potentials should enable the refinement of decoy structures toward the native state. We demonstrated that by including global energetic and structural data for a large set of protein decoy structures and by optimizing the relative weights of energy components of physics-based all-atom potential, it is possible to significantly improve the correlation of the energy with native-likeness and scoring of the native structure as the lowest in energy. Using such an approach to optimize the ff03/HB force field (the original Amber ff03 force field with an added explicit HB potential), we improved the CC_{ave} of the energy with TM-score from 0.25 (for the original ff03 potential) to 0.65, and the scoring of the native structure as the lowest in energy from 22% (for the original ff03 potential) to 90% of proteins, for a representative Set58. Reaching an average correlation of 0.69 of energy with TM-score for the *TASSER* coarse-grained potential, developed earlier in our laboratory, allowed systematic refinement of the reduced protein models (11). This gives us a reason to expect that our optimized atomic potentials having a similar average energy-TM score correlation will show systematic refinement ability.

We also showed that the *DSSP* (37) hydrogen bond potential that implicitly contains the angular dependence of the HB energy can significantly improve the correlation of the energy with native-likeness and the recognition of the native structure as the global energy minimum.

For a large protein decoy sample, we observed that the electrostatic and GB solvation energy components are un-

correlated with native similarity and do not show any specificity in recognizing the native state. The behavior of the electrostatic energy with native-likeness is protein dependent, and there is no reason for the electrostatics to change monotonically with native similarity. In force fields, the “frozen” point charge approximation and absence of polarization additionally introduces unnaturally large fluctuations of the ELE, even for small changes of local geometry. The GB solvation energy also has an electrostatic character and suffers from the same large nonphysical fluctuations as the electrostatics, caused by the point charge approximation. The solvation energy is usually favorable for extended structures and for some proteins is weakly anticorrelated with native similarity, as structure becomes more compact and less solvated. The electrostatic and GB solvation energy comprise a noisy uncorrelated background to the other energy components. As a result of optimization, the weights of these energy components decrease, suggesting the limited role of electrostatic energy and the electrostatic component of solvation in directing the already approximately assembled structure toward the native state. In contrast, a stronger initial correlation of energy with native-likeness is observed for the van der Waals and the hydrogen bond energy. The weights of these energy components become relatively larger after force field optimization.

The dihedral energy appears to be weakly anticorrelated with native-likeness, which results in a negative, but small weight of this energy term in some of our optimized potentials. There is no physical reason for the dihedral energy to anticorrelate with native similarity. The source of the observed anticorrelation can be either some inaccuracy of the dihedral parameters or some imbalance of the relative magnitude of the energy components in the ff03 force field. This observation suggests that the DIH term may require further reoptimization. We also observed earlier (38) the tendency of the ff03 force field to distort the dihedral angles from their gas phase equilibrium values for short helices and strands of polypeptides (from the quantum mechanical calculations). However these results may be partly justified as the dihedral parameters of the ff03 force field were developed in the condensed phase, not in the gas phase.

Since the ELE, ELE1-4, and GB solvation energy components acquire small weights during optimization, we explored the use of a reduced potential with the electrostatic and GB solvation terms turned off. The scoring performance of the optimized reduced ff03/HB force field (Wgt-R) is worse than the performance of the optimized full ff03/HB (Wgt-1) by 5% for the CC_{ave} of energy with TM-score, 20% for the fraction of proteins with a significant correlation coefficient, CC_{fr} , and 21% for the proteins for which the native-like structure has the lowest energy, TM_{fr} . Therefore, the loss of performance of the optimized reduced potential compared to the fully optimized ff03/HB force field is not very large, and for 43% of proteins the correlation coefficient of energy with TM-score is larger than 0.60, allowing correct decoy scoring.

The reduced optimized potential is significantly better than the fully unoptimized ff03 and ff03/HB force fields.

The ultimate goal of global optimization of the force fields is not only the correct scoring of protein decoys but also the refinement of structures closer to the native state. In the initial test of the optimized force field in protein structure refinement (seven proteins, 100 decoys per protein, ff03/HB Wgt-R reduced potential), we obtained an improvement in the structure for 63% of cases, with 16% showing improvements >0.05 in TM-score (13% showing improvements in RMSD >0.5 Å) to the native structure. The largest observed refinement, measured as an increase of the TM-score, was from 0.44 to 0.61 TM-score to the native structure (the RMSD decreased from 4.32 to 2.77 Å). These results are promising and highlight the need for testing the optimized force fields in refinement tests on a much larger set of proteins.

SUPPLEMENTARY MATERIAL

To view all of the supplemental files associated with this article, visit www.biophysj.org.

Calculations were conducted partly using the resources of the National Science Foundation Teragrid Project and the Terascale Computing System at the Pittsburgh Supercomputer Center. The authors thank P. Rotkiewicz for the PULCHRA program, useful suggestions about the hydrogen-bond potential, and help in figure preparation and S. B. Pandit for helpful discussions about force field optimization methods.

This research was supported in part by National Institutes of Health grant RR-12255.

REFERENCES

- Chen, J., and C. L. Brooks III. 2007. Can molecular dynamics simulations provide high-resolution refinement of protein structure? *Proteins*. 67:922–930.
- Wroblewska, L., and J. Skolnick. 2007. Can a physics-based, all-atom potential find a protein's native structure among misfolded structures? I. Large scale AMBER benchmarking. *J. Comput. Chem.* 28:2059–2066.
- Lee, M. C., and Y. Duan. 2004. Distinguishing protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized Born solvent model. *Proteins*. 55:620–634.
- Hsieh, M.-J., and R. Luo. 2004. Physical scoring function based on AMBER force field and Poisson-Boltzmann implicit solvent for protein structure prediction. *Proteins*. 56:475–486.
- Lazaridis, T., and M. Karplus. 1998. Discrimination of the native from misfolded protein models with an energy function including implicit solvation. *J. Mol. Biol.* 288:477–487.
- Dominy, B. N., and C. L. Brooks III. 2002. Identifying native-like protein structures using physics-based potentials. *J. Comput. Chem.* 23:147–160.
- Felts, A. K., E. Gallicchio, A. Wallqvist, and R. M. Levy. 2002. Distinguishing native conformations of proteins from decoys with an effective free energy estimator based on the OPLS all-atom force field and the surface generalized Born solvent model. *Proteins*. 48:404–422.
- Summa, C. M., M. Levitt, and W. F. DeGrado. 2005. An atomic environment potential for use in protein structure prediction. *J. Mol. Biol.* 352:986–1001.
- Zhou, H., and Y. Zhou. 2002. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci.* 11:2714–2726.
- Bradley, P., K. M. S. Misura, and D. Baker. 2005. Toward high-resolution de novo structure prediction for small proteins. *Science*. 309:1868–1871.
- Zhang, Y., A. Kolinski, and J. Skolnick. 2003. TOUCHSTONE II: a new approach to ab initio protein structure prediction. *Biophys. J.* 85:1145–1164.
- Zhang, Y., and J. Skolnick. 2004. Automated structure prediction of weakly homologous proteins on a genomic scale. *Proc. Natl. Acad. Sci. USA*. 101:7594–7599.
- Simons, K. T., R. Bonneau, I. Ruczinski, and D. Baker. 1999. Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins*. 37(Suppl. 3):171–176.
- Lundstrom, J., L. Rychlewski, J. Bujnicki, and A. Elofsson. 2001. Pcons: a neural-network-based consensus predictor that improves fold recognition. *Protein Sci.* 10:2354–2362.
- Fisher, D. 2003. 3D-SHOTGUN: a novel, cooperative, fold-recognition meta-predictor. *Proteins*. 51:434–441.
- Kolinski, A., and J. Bujnicki. 2005. Generalized protein structure prediction based on combination of fold-recognition with de novo folding and evaluation of models. *Proteins*. 61:84–90.
- Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.
- Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A 2nd generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.
- Case, D. A., T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr, A. Onufriev, C. Simmerling, B. Wang, and R. Woods. 2005. The Amber biomolecular simulation programs. *J. Comput. Chem.* 26:1668–1688.
- Scheraga, H. A., A. Liwo, S. Oldziej, C. Czaplewski, J. Pillardy, D. R. Ripoll, J. A. Vila, R. Kazmierkiewicz, J. A. Saunders, Y. A. Arnautova, A. Jagielska, M. Chinchio, and M. Nancias. 2004. The protein folding problem: global optimization of force fields. *Front. Biosci.* 9(Suppl. S.):3296–3323.
- Oldziej, S., J. Lagiewka, A. Liwo, C. Czaplewski, M. Chinchio, M. Nancias, and H. A. Scheraga. 2004. Optimization of the UNRES force field by hierarchical design of the potential-energy landscape. 3. Use of many proteins in optimization. *J. Phys. Chem. B*. 108:16950–16959.
- Arnaudova, Y. A., A. Jagielska, J. Pillardy, and H. A. Scheraga. 2003. Derivation of a new force field for crystal-structure prediction using global optimization: nonbonded potential parameters for hydrocarbons and alcohols. *J. Phys. Chem. B*. 107:7143–7154.
- Jagielska, A., Y. A. Arnautova, and H. A. Scheraga. 2004. Derivation of a new force field for crystal-structure prediction using global optimization: nonbonded potential parameters for amines, imidazoles, amides, and carboxylic acids. *J. Phys. Chem. B*. 108:12181–12196.
- Arnaudova, Y. A., A. Jagielska, and H. A. Scheraga. 2006. A new force field (ECEPP-05) for peptides, proteins, and organic molecules. *J. Phys. Chem. B*. 110:5025–5044.
- Duan, Y., S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. A. Kollman. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24:1999–2012.
- Tsui, V., and D. A. Case. 2001. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers*. 56:275–291.
- Rotkiewicz, P., and J. Skolnick. 2008. Fast procedure for reconstruction of full-atom protein models from reduced representations. *J. Comput. Chem.* In press. doi: 10.1002/jcc.20906.

28. Sitkoff, D., K. A. Sharp, and B. Honig. 1994. Accurate calculation of hydration free energies using macroscopic solvation models. *J. Phys. Chem.* 98:1978–1988.
29. Zhang, Y., and J. Skolnick. 2004. Scoring function for automated assessment of protein structure template quality. *Proteins*. 57:702–710.
30. Hansmann, U. H. E. 1997. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* 281:140–150.
31. Swendsen, R. H., and J. S. Wang. 1986. Replica Monte Carlo simulation of spin glasses. *Phys. Rev. Lett.* 57:2607–2609.
32. Zhang, Y., D. Kihara, and J. Skolnick. 2002. Local energy landscape flattening: parallel hyperbolic Monte Carlo sampling of protein folding. *Proteins*. 48:192–201.
33. Betancourt, M. R. 2005. Efficient Monte Carlo trial moves for polypeptide simulations. *J. Chem. Phys.* 123:174905.
34. Csendes, T. 1988. Nonlinear parameter estimation by global optimization—efficiency and reliability. *Acta Cybernetica*. 8:361–370.
35. Onufriev, A., D. Bashford, and D. A. Case. 2004. Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins*. 55:383–394.
36. Yang, J. S., W. W. Chen, J. Skolnick, and E. I. Shakhnovich. 2007. All-atom ab initio folding of a diverse set of proteins. *Structure*. 15:53–63.
37. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 22:2577–2637.
38. Zhang, Y., I. A. Hubner, A. K. Arakaki, E. I. Shakhnovich, and J. Skolnick. 2006. On the origin and highly likely completeness of single-domain protein structures. *Proc. Natl. Acad. Sci. USA*. 103:2605–2610.