JCTC Journal of Chemical Theory and Computation

Improved Energy Selection of Nativelike Protein Loops from Loop Decoys

Matthew S. Lin[†] and Teresa Head-Gordon*[,†,‡]

UCSF/UCB Joint Graduate Group in Bioengineering, Berkeley, California 94720, and Department of Bioengineering, University of California, Berkeley, California 94720

Received October 29, 2007

Abstract: We demonstrate the performance of a new implicit solvent model on native protein loop prediction from a large set of loop decoys of 4- to 12-residue in length. The physics-based energy function combines a hydrophobic potential of mean force (HPMF) description with a Generalized Born model for polarization of protein charge by the high dielectric solvent, which we combine with AMBER force field for the protein chain. The novelty of our energy function is the stabilizing effect of hydrophobic exposure to aqueous solvent that defines the HPMF hydration physics, which in principle should be an important stabilizing factor for loop conformations of a protein that typically are more solvent exposed. While our results for short loop decoy sets are comparably good to existing energy functions, we find demonstrable superiority for loop lengths of 8-residue and greater, and the quality of our predictions is largely insensitive to the length of the target loop on a filtered set of decoys. Given that the current weakness in loop modeling is the ability to select the most nativelike loop conformers from loop ensembles, this energy function provides a means for greater prediction accuracy in structure prediction of homologous and distantly related proteins, thereby aiding large-scale genomics efforts in comparative modeling.

1. Introduction

The loop regions of a protein are known to be important for its structure as they determine sequence reversals that allows the chain to collapse and fold[1] as well as for functions such as guiding or gating ligand binding, aiding protein complexation, and for enzymatic activity.[2] The ability to predict the nativeness of loops is thus an important goal, especially since functional differences between homologous proteins differ mostly in their loop conformations.[3,4] However, prediction of native loop structure is a far more difficult problem than prediction of native β-strands and α-helices, since the structures are more highly diverse, and the sequence correlations are even weaker, compared to these other standard secondary structure categories.[5,6]

As is true in prediction of overall protein structure,[7] the techniques that have been used in the prediction of loop structure can be divided into knowledge-based[8−12] and physics-based[13−17] approaches. Knowledge-based approaches rely typically on protein structural databases such as the PDB to derive sample loop conformations directly and typically use empirical criteria for native discrimination against misfolds. A library of known loop fragment structures with the same length as the target loop is fit between the backbone atoms of the residues that precede and succeed the target loop sequence (stem residues), which are then screened by a scoring function to generate plausible candidates, which are sometimes further refined with energy minimization. By contrast, physics-based approaches perform loop generation and selection based on first principles concepts. Ab initio chain growth techniques such as inverse kinematics have been widely implemented[18−21] and combined with more exhaustive conformational sampling to generate loop conformations that interpolate between the two stem residues. Typically a physically motivated energy function that aims

* Corresponding author e-mail: tlhead-gordon@lbl.gov. Corresponding author address: Department of Bioengineering, Stanley Hall 274, University of California, Berkeley, Berkeley, CA 94720.

† UCSF/UCB Joint Graduate Group in Bioengineering.

‡ University of California.

10.1021/ct700292u CCC: $40.75
Published on Web 02/21/2008

to discriminate native loops from misfolded decoys is used to rank the degree of nativeness of a given loop configuration.

Strict categorization of any loop optimization method into these two general approaches is somewhat arbitrary.[22] Until recently, the primary difference between the knowledge-based and physics-based methods is the generation of long loop structures (>8 residues) since they appeared less frequently in the PDB than shorter ones.[6] However, high-throughput structural genomics efforts are diminishing this difference, and nativelike loops, within 2.0 Å for long loops, can be found with high probability in structural database.[11,23] Regardless of whether loops are generated by database searches or ab initio techniques, current evidence suggests that sufficient sampling of loop conformations, even long loops of 8 or more residues, does not appear to be a limiting factor.[11,16,23,24] Instead, the energy or statistical functions used for the discrimination of native and nativelike loop structures against misfolded decoys mostly restrict the prediction accuracy.[11,16,25−27] In fact, a very recent study[26] found that four popularly used commercial software packages rarely selected the most nativelike loop conformer of their generated ensemble, revealing that selection of nativelike loop conformers is the *primary* weakness of the current state-of-the-art approaches for loop modeling. In summary, better energy functions are key to both categories of loop structure prediction in order to push homology modeling efforts toward greater prediction accuracy.

In this work, we demonstrate the performance on native protein loop prediction of our recently developed energy function[28] and compare the prediction accuracy with other physics-based and knowledge-based scoring functions. We then further discuss the limitations of using the raw experimental native loop conformation as the gold standard in the testing procedures and place our results in the context of current state of the art in native loop selection available in several commercial packages. We hope that our energy function provides a better means for loop selection and optimization that should aid experimental structural refinement and large-scale genomics efforts in comparative modeling.

2. Materials and Methods

2.1. Energy Function. We have recently developed a new implicit solvent model[28] to describe a given protein and its aqueous solvent free energy surface. The energy function combines the AMBER ff99 protein force field developed by Wang and co-workers[29] (V_{Protein}), the Generalized Born (GB) description of the electrostatic component of solvent free energy (V_{GB}) developed by Onufriev and co-workers,[30] and the newly developed implicit solvent model, hydrophobic potential of mean force (HPMF), to describe the hydrophobic solute−solute interaction induced by water[31,32] (V_{HPMF})

$$V = V_{\text{Protein}} + V_{\text{GB}} + V_{\text{HPMF}} \quad (1)$$

$$V_{\text{HPMF}} = \sum_{i \in SA_i > A_c}^{N_c} \tanh(SA_i) \sum_{j \in SA_j > A_c}^{N_c} \tanh(SA_j) \times \sum_{k=1}^{3} h_k \exp(-[(r_{ij} - c_k) \times w_k]^2) \quad (1a)$$

$$\tanh(SA) = \frac{1}{2}(\tanh[\text{SLOPE} \times (SA - \text{OFFSET})] + 1) \quad (1b)$$

SLOPE and OFFSET are constants set to 1000.0 and 6.0, respectively. We refer the reader to our previous work[28] for the functional form and the parameter details of the model. The novelty of this energy function is the stabilizing effect of hydrophobic exposure to aqueous solvent that defines the HPMF hydration physics and its apparent improvement over solvent accessible surface area models that penalize hydrophobic exposure. When tested on an extensive number of protein decoy sets, which allows us to compare our performance to other scoring functions for native fold, we find that our energy function outperforms other similarly tested energy and statistical functions with both substantial improvements in native ranking and Z-score drops of 0.5−2.5 units.[28]

In this publication, the entire energy function (AMBERff99+GB+HPMF) will be simply abbreviated as HPMF unless indicated otherwise.

2.2. Decoy Set Selection. Here, we demonstrate the new energy function's performance on native protein loop prediction. We have restricted our comparison to the loop decoys generated by Jacobson et al.,[16] which we and others[12] have found to be more difficult decoys than the RAPPER[14,15] decoy sets. We then compare our results to the physics-based energy function, OPLS/SGB-NP,[16,33,34] and the statistical scoring function, DFIRE,[12] which have been tested on the same loop decoy sets.

Table 1 lists the characteristics of the loop decoys for 4-, 6-, 8-, 10-, 11-, and 12-residues generated by Jacobson et al.[16] As is true in protein structure, some of the loop decoy sets have complications due to exceptional features of the native protein structure that make the testing ambiguous. To combat this problem, Jacobsen et al. provide a filtered list of decoys that (1) eliminates proteins that were crystallized at high or low pH, (2) removes proteins in which target loops have explicit interactions with heteroatoms such as metals or ions, and (3) omits proteins with low-resolution crystal structures in the target loop region which have large measured B-factors.[16] We consider both the filtered and unfiltered decoy sets in this work. Additional problems we encountered were proteins missing from the download source,[35] proteins with segments of missing structure, and nonstandard amino acid in the sequence (see the Supporting Information). Another individual case that we faced was 1DAD_1, which is part of the 11-residue filtered set. It is a dimer protein, and the target loop is at the interface between the two monomers. Since there is only one monomer used in the decoys, and it is ambiguous whether the destabilization caused by the absence of the other monomer would affect the prediction on the target loop, we decided to remove it from the filtered set but keep it in the unfiltered group.

We also consider a set of protein loop decoys recently generated by Rossi et al.[26] In that study, they evaluated the performance of the four commonly used commercial public packages, Prime, Modeller, Sybyl, and ICM, on protein loop modeling; the former two use ab initio loop generation methods coupled with an energy function, and the later ones

Table 1. Jacobson Decoy Set[16] [a]

loop length decoy set	number of protein in unfiltered set	number of protein in filtered set	average number of decoys per protein	decoy RMSD range
4	37	32	300	0.22−1.35
6	141	94	400	0.24−2.98
8	93	60	600	0.32−4.64
10	49	27	900	0.32−5.80
11	31	14	1200	0.27−6.93
12	17	9	1100	0.36−8.96

[a] The RMSD range is the average over the most nativelike decoy as the lower bound and over the least nativelike decoy as the upper bound for each length set.

use knowledge-based techniques to generate loop fragments along with a statistical scoring function. For each tested protein loop, the four modeling packages were instructed to output their top six predictions and to separate the influence of the sampling methods from the scoring functions; two performance measurements were calculated. Top-rank-RMSD is the RMSD of the decoy assigned the lowest energy by the packages' scoring functions, and best-RMSD is the RMSD of the most nativelike decoy among the six predicted models. The latter variable was used to evaluate how well each sampling technique could search around the native basin, and the former one was used to measure if each scoring function could select the most nativelike models. We collected these decoys structures generated by the commercial packages from Rossi and co-workers[36] to test whether our energy function could better detect the most nativelike decoy as the one with the lowest energy compared to the four commercial loop packages.

In the end, we considered over 350 different protein loops with anywhere between 300 and 1200 decoys each; therefore, we believe that we have made a fair comparison to the previous works.[12,16,26,37,38]

2.3. Energy Minimization Procedure. Each loop decoy is locally minimized using eq 1 with the BFGS (Broyden-Fletcher-Goldfarb-Shanno)[39] limited memory quasi-Newton method.[40] During energy minimization, only the atoms at the target loop are allowed to fluctuate, and the rest of the structure is held fixed. After convergence, the backbone heavy atoms (N, Cα, C, O) of the target loop are used to calculate the RMSD between the experimental native and each energy-minimized decoy structure, while two structures are aligned along the nontarget-loop region. We also consider the backbone RMSD values calculated as done in PLOP/Prime based on only three of the backbone heavy atoms (N, Cα, C).[41] All of the RMSD calculations were conducted using MMTSB Tool Set.[42]

2.4. Evaluation Procedure. For the decoys generated by Jacobson et al., we followed the same criteria for success used previously,[12,16] which is, for each protein loop, to take the decoy with the lowest energy ranking and determine its RMSD with respect to the native structure. These RMSD values are then averaged over all proteins in the given loop size set. Note that in some cases the lowest ranked structure is sometimes the native loop in our study, although we follow the procedure of previous studies of not including it in the averages.

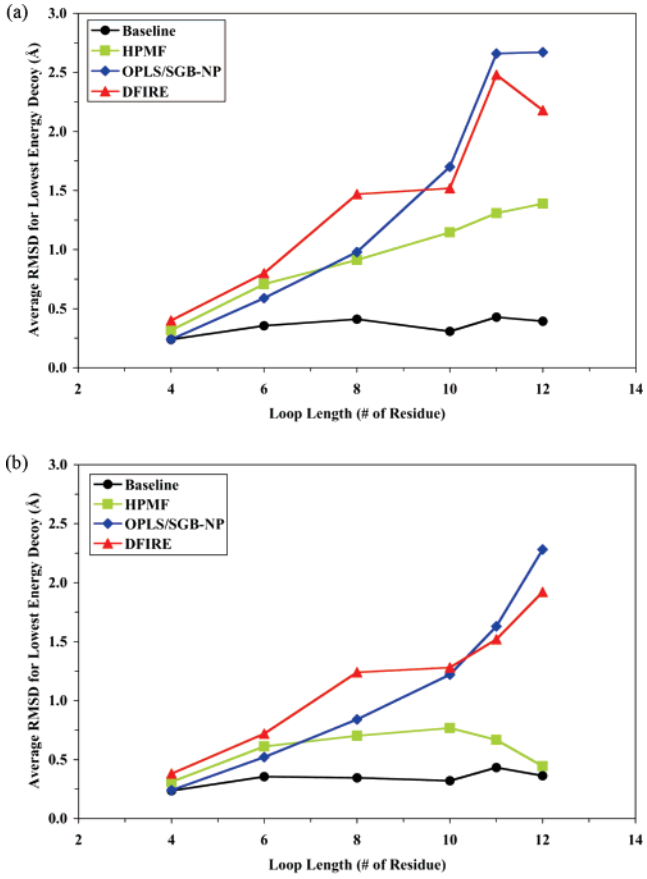

Figure 1. Performance of DFIRE[12] (triangle), OPLS-SGB/NP[16] (diamond), and HPMF[28] (square) energy functions on native loop prediction compared to a baseline of experimental uncertainty (circle): (a) unfiltered decoys and (b) filtered decoys. For comparative purposes we follow the same process for assessing the prediction results as the previous work.[12,16]

For the test on the structures generated by Rossi et al., we adopted the same two variables used in their study to examine our energy function's ability for selecting the most nativelike models. However, we took a slightly different approach when we studied the structures generated by Sybyl and ICM. The structures generated by these two methods sometimes have steric clashes, and those structures were removed if the steric collision could not be resolved using energy relaxation with constrains. Among the remaining structures, we calculated the new average top-rank-RMSD and the new average best-RMSD and used them for the evaluation.

3. Results

3.1. Jacobson Decoy Results. Figure 1 presents our results compared to the OPLS/SGB-NP[16] and DFIRE[12] energy functions on the decoy sets for 4- through 12-residue loops. Figure 1a shows the unfiltered decoy results in which it is evident that the different energy functions perform comparably at short loop lengths, with the OPLS/SGB-NP showing the best performance, 24% and 16% lower than HPMF at 4-residue and 6-residue sets, respectively. For 8-residue and longer loops, the HPMF energy function clearly improves the ability to discriminate against non-nativelike decoys. All

Table 2. Comparison of DFIRE,[12] OPLS-SGB/NP,[16] and HPMF[28] in Regards to Backbone RMSD of the Lowest Energy Decoy against the Experimental Native Structure Averaged over All Proteins in the Loop Decoy Sets of Different Length[a]

	unfiltered decoys (average lowest RMSD in Å)			filtered decoys (average lowest RMSD in Å)			
						HPMF	
loop length decoy set	DFIRE	OPLS SGB-NP	HPMF	DFIRE	OPLS SGB-NP	N,Ca,C,O	N,Ca,C
4	0.40	0.24	0.32 (0.13)	0.38	0.24	0.31 (0.12)	0.25
6	0.80	0.59	0.71 (0.58)	0.72	0.52	0.61 (0.46)	0.52
8	1.47	0.98	0.91 (0.81)	1.24	0.84	0.70 (0.57)	0.62
10	1.52	1.70	1.15 (1.36)	1.28	1.22	0.77 (0.71)	0.68
11	2.48	2.66	1.24 (1.09)	1.52	1.63	0.67 (0.51)	0.59
12	2.18	2.67	1.39 (1.31)	1.92	2.28	0.39 (0.20)	0.36

[a] The results in parentheses show the RMSD over backbone heavy atoms for decoys compared to an energy relaxed native loop structure.

Table 3. Demonstration of the Significant Contribution of HPMF on the Overall Performance of the Energy Function on 11- and 12-Residue Filtered Set

	filtered decoys (average lowest RMSD in Å)	
loop length decoy set	AMBER99+GB	AMBER99+GB+HPMF
11	0.93	0.67
12	0.40	0.39

energy functions show the same behavior of increasing RMSD error with increasing loop size, but for the largest 12-residue loop set the performance of our energy function provides improvements of about 35−50% over the earlier efforts, dropping from an average RMSD deviation from native prediction of ∼2.20−2.65 Å to ∼1.40 Å. The results are even more dramatic on the filtered set (Figure 1b). Even though the DFIRE and OPLS/SGB-NP energy functions drop about 10−40% in RMSD in the filtered set relative to their performance on the unfiltered set, their trends still show a larger average RMSD as the loop size increases. Our energy function improves from 10 to 70% on the filtered set, and furthermore it shows no strong dependence on the quality of results with the length of target loops. We believe that HPMF outperforms other energy functions at longer loops because of greater solvent exposure. The exact values are tabulated in Table 2. In the same table we also report our results when compared to the minimized native structure and when the backbone RMSD is calculated over only three of the backbone heavy atoms (N, Cα, C) as done in Prime/PLOP.[41]

To demonstrate the significant contribution from HPMF, Table 3 shows the comparison between the results on the 11- and 12-residue filtered loop set of two energy functions, AMBER99+GB and AMBER+GB+HPMF. For the 11-residue loop set, the average lowest RMSD is 0.93 Å and 0.68 Å for AMBER+GB and AMBER+GB+HPMF, respectively. The performance of the energy function including HPMF is 28% better than the other. However, there is no notable difference between the accuracy of the two energy functions for the 12-residue loop set. The subtle difference raises the question as to the sample size of the 12-residue loop set. More 12-residue loop proteins and longer loop sets are needed to further demonstrate the important contribution of HPMF in the overall performance of the energy function.

3.2. Loop Modeling Decoy Results. We tested our energy function on the long loops, 10-, 11-, and 12-residue, reported in Rossi et al.'s work[26] to demonstrate that, for modeling long protein loop, our energy function can consistently select a more nativelike model than the other energy and scoring functions in the popular modeling packages they tested. Similar to the complications that we encountered in the Jacobson decoy set, we needed to remove 1LUC from the 12-residue set and 1CVL and 2ENG from the 11-residue set due to missing fragment structures. We also removed 1FUS from the 11-residue set due to a nonstandard amino acid.

Figure 2 reports the results of our energy function for selection of nativelike loops compared to the predictions of Prime, Modeller, Sybyl and ICM. Because Sybyl and ICM do not filter out structures with steric conflicts, we eliminated those decoys in which energy relaxation (with harmonic constraints on all atom positions) did not resolve the steric collisions to define a sensible energy for comparison. The results show that our energy function can select more nativelike decoys as the lowest energy structures for each length set, such that the average RMSD is 10−40% lower than the predictions made by the standard loop modeling packages (Figure 2a). In addition, the prediction accuracy, which is measured when the most nativelike decoy is ranked as the lowest energy model, is significantly improved by our energy function (Figure 2b).

3.3. Challenges to X-ray Structures. Having established the improved performance of our energy function relative to previous reported physics and statistical potentials, we probe the relevance of these results. X-ray crystallography refinement combines structure factor data with a basic physiochemical model for chain connectivity and excluded volume to derive atom placements in a reported protein three-dimensional structure. The question is how good is this one native structure for assessing loop structure prediction accuracy? This implicit modeling aspect of the experimental data can be tested by optimizing the loop of the native structure, which we do under the usual assumption of a rigid context of the remaining protein. Table 2 shows the results when minimized loops are compared to the minimized native

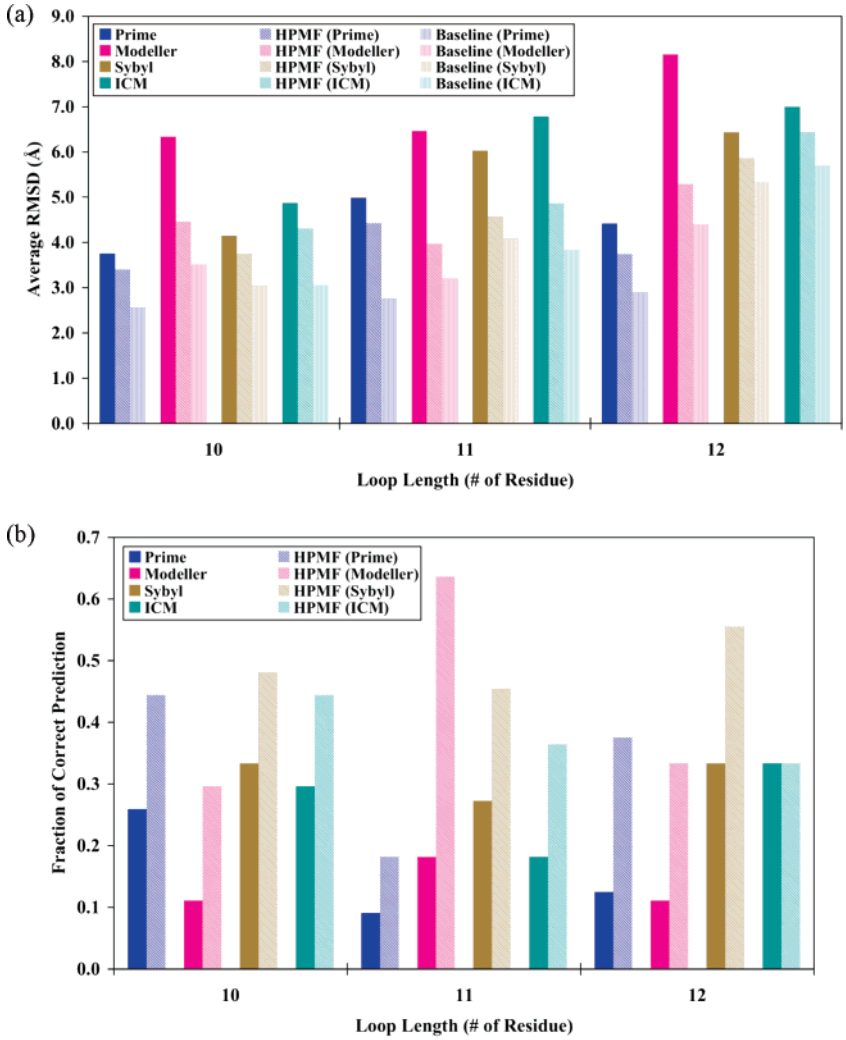

Figure 2. Results of HPMF on the decoy structures (10-, 11-, and 12-residue set) generated by Rossi et al.[26] compared with the modeling packages. The colors blue, pink, brown, and cyan are the decoy models generated by Prime, Modeller, Sybyl, and ICM, respectively. The patterns in the plots, filled, striped, and spotted, represent the average results of each commercial package, HPMF, and the baseline, respectively. (a) It shows the average top-rank RMSD of each energy function. (b) It shows the fraction of correct prediction, which is the percentage of loops that the most nativelike model is ranked the lowest energy.

loop, holding the rest of the protein fixed. Regardless of loop size, and regardless of whether unfiltered or filtered, the average RMSD change between the unoptimized and the optimized native loop structure is 0.4−0.5 Å. Correspondingly, the decoys show an average RMSD change between unoptimized and optimized loop structures of ∼0.1 Å. Thus we regard the native state as being uncertain to with ∼0.3−0.4 Å due to model effects in deriving the X-ray crystallography structure. We emphasize this point by defining a baseline function as the difference between the RMSD of the unoptimized and the optimized native structure and the RMSD of the unoptimized and the optimized lowest energy decoy, for each decoy set size, as we show in Figure 1. In fact for 4- and 12-residue filtered decoys, the prediction is as good as the baseline function. A similar strategy was developed by Rossi and co-workers[26] in which they determined a permissible range of RMSD values for defining the native loop.

4. Conclusion

In this work we have demonstrated the performance of a new protein and implicit solvent model on native protein loop discrimination from non-native decoys. The novelty of our energy function is the stabilizing effect of hydrophobic exposure to aqueous solvent that defines the HPMF hydration physics, which in principle should be an important stabilizing factor for loop conformations of a protein that typically are more solvent exposed. While our results for short loop decoy sets are comparably good to existing energy functions, we find substantial improvements for loop lengths of 8-residues and greater and find that the quality of our predictions are largely insensitive to the length of the target loop on the filtered decoy sets.

Recent work[37] reported a modification of the OPLS/SGB-NP energy by adding an ad hoc correction to the nonpolar solvation term to model the favorable free energy gained by placing hydrophobic side chains of the target loop in the hydrophobic space created when the target loop is removed,

a term that has been successfully used in ligand-protein binding. This is conceptually similar to the more physically grounded hydrophobic interaction model developed in our recent work,[28] and Zhu and co-workers showed demonstrable improvement in loop prediction on a filtered collection of 11- to 13-residue loops over the standard OPLS/SGB-NP energy function.[37] However aspects of the energy model could stand improvement. While they report an average RMSD of 1.00 Å and 1.15 Å for *filtered* 11- and 12-residue loop decoys, our average RMSD on the *unfiltered* decoy set is comparable to these results and clearly superior on similarly filtered decoy sets, reducing the error in native loop selection by over a factor of 2 (Table 2).

More recently, Zhu et al.[38] have improved their energy model by varying the protein internal dielectric constant in the electrostatics energy calculation based on the environment of interacting atoms and by further optimizing the hydrophobic term that they introduced in their previous work.[37] In their latest work,[38] they have achieved the prediction accuracy of 0.39 Å, 0.68 Å, 0.80 Å, and 1.00 Å RMSD for 6-, 8-, 10-, and 13-residue loops, respectively. However the performance criteria was changed from previous studies since they calculate the RMSD of the predictions with respect to the energy-minimized native structures. This concurs with the baseline function used here or the range of permissible native RMSDs used in Rossi et al., that physics-based predictions are better compared to energy relaxed experimental X-ray crystallographic structures.

In the study by Nayeem and co-workers,[43] the authors compared seven homology model building software packages that encompass both knowledge- and physics-based approaches for sequence alignment through to structural loop generation models. One of the tested commercially available software packages, Prime, which uses the ab initio loop generation and physically based energy function[16] was found to be one of the best performers in the study. While Prime showed comparable performance to other software packages for proteins that exhibit sequence identity with known proteins of greater than 50%, Prime showed demonstrable superiority over other approaches as sequence identity decreased.[43] However, a more recent study[26] found that all commercial software packages, including Prime, returned a best energy ranked loop that was rarely the most nativelike of their generated ensemble, revealing a weakness in the energy functions used in selecting nativelike loop conformers. Given the improvement of our energy function over all versions of the OPLS/SGB-NP results,[16,37,38] this result emphasizes that existing commercially available software packages would benefit from improvements in scoring of more nativelike loops by use of the energy function we have developed[28] and which we have tested further for loop prediction in this work.

While our results are "better" than other reported results, we raise questions about the testing procedures that assume the experimental native loop structure is the relevant gold standard for prediction. This is consistent with the recent recognition of the importance of structural ensembles for reporting X-ray crystallographic structures to describe the functional native state, as opposed to the dogma of one native structure.[44] Better measures of structural ensembles consistent with the experimentally derived structure factors could more meaningfully discriminate for or against prediction success, especially for the unfiltered decoys where experimental procedure is clearly a source of uncertainty. A tractable energy function in turn might usefully aid in the experimental refinement procedure for deriving atomic protein models from structure factor data. In a recent study,[45] loop prediction methods such as Prime have been demonstrated to improve the protein structure refinement protocol in NMR experiments. The results of the study evidently indicate that our energy function coupled with a sampling technique not only can be used in theoretical research but also can facilitate experimental studies in solving protein structures.

One of the goals of the structural genomics initiative is to rapidly obtain native structural models from X-ray crystallography and NMR spectroscopy for representative members of protein families.[46] Once the protein structure of a family member is solved, in principle it can serve as a structural template to develop comparative structural models of other family members.[46,47] However, comparative modeling structures that have a *functional* value typically require working with a structural template that exhibits greater than 40−60% sequence identity.[48,49] Below that, function begins to diverge due to structural differences on the protein surface that typically correspond to loop regions.[6,13] We hope that this energy function provides a better means for loop optimization important in recognizing the structural and functional differences between homologous and distantly related proteins to aid comparative modeling efforts.

Acknowledgment. T.H.G. gratefully acknowledges financial support under the NSF-Cyber program. M.S.L. is supported under the DOE/MICS program. We would also like to thank Dr. Karen A. Rossi for providing us the structures generated in her study and Dr. Jonathan E. Kohn's help on creating the graphic for the issue cover. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DEAC02-05CH11231.

Supporting Information Available: List of protein loops removed from this work due to missing from the download source, having missing fragments, or having nonstandard amino acids. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Viguera, A. R.; Serrano, L. *Nature Struct. Biol.* **1997**, *4*, 939−946.

(2) Fetrow, J. S. *FASEB J.* **1995**, *9*, 708−717.

(3) Liu, J.; Tan, H.; Rost, B. *J. Mol. Biol.* **2002**, *322*, 53−64.

(4) Blouin, C.; Butt, D.; Roger, A. J. *Protein Sci.* **2004**, *13*, 608−616.

(5) Efimov, A. V. *Curr. Opin. Struct. Biol.* **1993**, *3*, 379−384.

(6) Fiser, A.; Do, R. K.; Sali, A. *Protein Sci.* **2000**, *9*, 1753−1773.

(7) Tramontano, A. *FEBS J.* **2007**, *274*, 1651−1654.

(8) Burke, D. F.; Deane, C. M. *Protein Eng.* **2001**, *14*, 473−478.

(9) Espadaler, J.; Fernandez-Fuentes, N.; Hermoso, A.; Querol, E.; Aviles, F. X.; Sternberg, M. J.; Oliva, B. *Nucleic Acids Res.* **2004**, *32*, D185−188.

(10) Espadaler, J.; Querol, E.; Aviles, F. X.; Oliva, B. *Bioinformatics (Oxford, England*) **2006**, *22*, 2237−2243.

(11) Fernandez-Fuentes, N.; Oliva, B.; Fiser, A. *Nucleic Acids Res.* **2006**, *34*, 2085−2097.

(12) Zhang, C.; Liu, S.; Zhou, Y. *Protein Sci.* **2004**, *13*, 391−399.

(13) Fiser, A.; Feig, M.; Brooks, C. L., III; Sali, A. *Acc. Chem. Res.* **2002**, *35*, 413−421.

(14) DePristo, M. A.; de Bakker, P. I.; Lovell, S. C.; Blundell, T. L. *Proteins* **2003**, *51*, 41−55.

(15) de Bakker, P. I.; DePristo, M. A.; Burke, D. F.; Blundell, T. L. *Proteins* **2003**, *51*, 21−40.

(16) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J.; Honig, B.; Shaw, D. E.; Friesner, R. A. *Proteins* **2004**, *55*, 351−367.

(17) Rohl, C. A.; Strauss, C. E.; Chivian, D.; Baker, D. *Proteins* **2004**, *55*, 656−677.

(18) Canutescu, A. A.; Dunbrack, R. L., Jr. *Protein Sci.* **2003**, *12*, 963−972.

(19) Coutsias, E. A.; Seok, C.; Wester, M. J.; Dill, K. A. *Int. J. Quantum Chem.* **2005**, *106*, 176−189.

(20) Shehu, A.; Clementi, C.; Kavraki, L. E. *Proteins* **2006**, *65*, 164−179.

(21) van den Bedem, H.; Lotan, I.; Deacon, A. M.; Latome, J.-C. *Algorithmic Found. Rob.* **2005**, 345−360.

(22) Deane, C. M.; Blundell, T. L. *Protein Sci.* **2001**, *10*, 599−612.

(23) Du, P.; Andrec, M.; Levy, R. M. *Protein Eng.* **2003**, *16*, 407−414.

(24) Monnigmann, M.; Floudas, C. A. *Proteins* **2005**, *61*, 748−762.

(25) Pellequer, J. L.; Chen, S. W. *Biophys. J.* **1997**, *73*, 2359−2375.

(26) Rossi, K. A.; Weigelt, C. A.; Nayeem, A.; Krystek, S. R., Jr. *Protein Sci.* **2007**.

(27) Smith, K. C.; Honig, B. *Proteins* **1994**, *18*, 119−132.

(28) Lin, M. S.; Fawzi, N. L.; Head-Gordon, T. *Structure* **2007**, *15*, 727−740.

(29) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.

(30) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383−394.

(31) Crivelli, S.; Eskow, E.; Bader, B.; Lamberti, V.; Byrd, R.; Schnabel, R.; Head-Gordon, T. *Biophys. J.* **2002**, *82*, 36−49.

(32) Head-Gordon, T.; Brown, S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 160−167.

(33) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. *J. Comput. Chem.* **2002**, *23*, 517−529.

(34) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.

(35) Jacobson, M. P. http://francisco.compbio.ucsf.edu/∼jacobson/decoy.htm (accessed December 3, 2006).

(36) Rossi, K. A. Private communication.

(37) Zhu, K.; Pincus, D. L.; Zhao, S.; Friesner, R. A. *Proteins* **2006**, *65*, 438−452.

(38) Zhu, K.; Shirts, M. R.; Friesner, R. A. *J. Chem. Theory Comput.* **2007**, *3*, 2108−2119.

(39) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed.; Cambridge University Press: 1992.

(40) Liu, D. C.; Nocedal, J. *Math. Programm.* **1989**, *45*, 503−528.

(41) Jacobson, M. P. http://francisco.compbio.ucsf.edu/∼jacobson/plop_manual/plop_manual.htm (accessed July 28, 2007).

(42) Feig, M.; Karanicolas, J.; Brooks, C. L., III *J. Mol. Graphics Modell.* **2004**, *22*, 377−395.

(43) Nayeem, A.; Sitkoff, D.; Krystek, S., Jr. *Protein Sci.* **2006**, *15*, 808−824.

(44) Furnham, N.; Blundell, T. L.; DePristo, M. A.; Terwilliger, T. C. *Nature Struct. Mol. Biol.* **2006**, *13*, 184−185.

(45) Rapp, C. S.; Strauss, T.; Nederveen, A.; Fuentes, G. *Proteins* **2007**, *69*, 69−74.

(46) Chandonia, J. M.; Brenner, S. E. *Science* **2006**, *311*, 347−351.

(47) Head-Gordon, T.; Wooley, J. C. *IBM Syst. J.* **2001**, *40*, 265−296.

(48) Tian, W.; Skolnick, J. *J. Mol. Biol.* **2003**, *333*, 863−882.

(49) Todd, A. E.; Orengo, C. A.; Thornton, J. M. *J. Mol. Biol.* **2001**, *307*, 1113−1143.

CT700292U