Title
Reliable protein structure refinement using a physical energy function

Journal
Journal of Computational Chemistry, 32(4)

ISSN
0192-8651

Authors
Lin, Matthew S
Head-Gordon, Teresa

Publication Date
2011-03-01

DOI
10.1002/jcc.21664

Peer reviewed

Reliable Protein Structure Refinement using a Physical Energy Function

Matthew S. Lin[1], Teresa Head-Gordon[1,2*]

[1]UCSF/UCB Joint Graduate Group in Bioengineering, Berkeley, California 94720, USA

[2]Department of Bioengineering, University of California, Berkley, California 94720, USA

* To whom correspondence should be addressed. E-mail: TLHead-Gordon@lbl.gov

Keywords: energy function, aqueous solvation, hydrophobicity, optimization, protein structure refinement

SUMMARY

Proteins carry out their biological function after folding into their tertiary structures, and the functional properties are greatly dictated by the structures. Hence, discovery of proteins' native structures impacts many diverse biological fields such as drug design and enzymatic studies. Experimental techniques such as x-ray crystallography and nuclear magnetic resonance spectroscopy (NMR) are able to solve proteins' structures at high atomic resolution. However, these methods are very time consuming and labor intensive, involve various difficulties. The current computational methods can consistently predict a protein's tertiary structure at medium resolution to high resolution if similar proteins with known structures have been revealed previously. However consistent predictions at atomic high resolution still remains as one of the greatest challenges. In this work, we demonstrate that our lately designed physical energy function is able to reliably refine medium-resolution protein models closer to the native structures. In addition to the improvement to computational methods, the energy function can also be potentially used during the refinement phase of x-ray crystallography and NMR experimental processes to further increase the accuracy of atomic positions of proteins.

ABSTRACT

In the past decade, significant progress has been made in protein structure prediction. However, refining models to a level of resolution that is comparable to experimental results and can be used in studies like enzymatic activity still remains a major challenge. We have previously demonstrated that our modular protein-solvent energy function, uniquely involving a potential of mean force description for hydrophobic solvation, works well in protein globular structure prediction and loop modeling. In this work, we couple protein-solvent energy function with our global optimization method Stochastic Perturbation with Soft Constraints and use them to refine a collection of template models from submitted predictions to recent CASP blind prediction contests. A prediction protocol based on a selection of structures with the lowest energy is able to successfully refine all of the test proteins, and more importantly, our energy function does not show degradation in prediction when sampling is exhausted.

INTRODUCTION

Knowledge of protein structures at atomic resolution is vital for biological studies on enzymatic activity and ligand binding. While nuclear magnetic resonance spectroscopy (NMR) and X-ray crystallography are commonly used to solve protein structures, they are typically labor-intensive, costly and can be limited by protein size or the ability to grow a well-ordered crystal. The advancement of high-throughput genomic sequencing projects is giving rise to an exponentially widening gap between the sequencing data and the number of solved protein structures. There is a pressing need for computational methods that can model protein structures with an atomic accuracy comparable to the experimental results.

The recent 5th, 6th, 7th and 8th Critical Assessment of Techniques for Protein Structure Prediction (CASP) [1-3] have shown significant progress in both the categories of template-based prediction, which uses sequence homology or structural homology based modeling of existing structures in the Protein Databank (PDB), and free modeling (or *de novo* modeling), which is primarily used when the structure to be predicted has no known experimental analogy. Several state-of-the art template-based modeling techniques can reliably generate models with quality as good as the experimental results if the templates used have sequence identity of 50% or higher to the target protein [4-9]. But as sequence or structural similarity decreases, prediction model errors are known to increase, generating predicted structures that can range between a root mean square deviation (RMSD) of 3.0Å and upwards to 10.0Å.

De novo modeling techniques compared to template-based methods require extensively more computational power, and prediction models generally have lower resolution. A major milestone is the work of Bradley and co-workers [10], in which they were able to predict *de novo* 5 of 16 small single-domain proteins (<85 residues) to a resolution <1.5 Å using an all-atom physical energy force field and a Monte Carlo global search method optimizing backbone torsion angles and side-chain packing. However, a majority of template-based or *de novo* prediction models are typically low-resolution, limiting their applicability to important biological studies like enzymatic activity and drug design [11].

The area of structure refinement seeks to overcome the extremely difficult challenge of refining medium to poor resolution structures which are initial guesses that have merit but still lack sufficient resolution detail for refined biological questions. Over the past decade, much progress has been made in the structural refinement field [10,12-21], with success depending on the resolution of the starting structure, the quality of the protein and solvent energy function, and the global optimization approach. Levitt et. al [22,23] studied a collection of both physics-based potential energy functions and statistically derived knowledge-based scoring functions, in which they found that when the test proteins are simulated in vacuo (no solvent representation), the knowledge-based functions gave superior refined protein models than physically motivated energy functions. However, when the physics-based functions were coupled with an implicit solvent model, they clearly outperformed any of the knowledge-based functions tested, signifying the importance of including solvent in protein structure refinement. Recently Skolnick et. al developed a protein refinement algorithm that included an optimized energy function composed of weighted components of the AMBERff99 protein force field and an implicit solvent model that was sampled with a replica exchange Monte Carlo global search method, in which they successfully refined 70% of their tested proteins [20,21].

In previous work, we have developed a physics based energy function that includes a novel hydrophobic potential of mean force term, and demonstrated its success on protein globular structure prediction [24], in which we were able to discriminate between each protein's native fold from a large set of misfolded decoy structures. We have also demonstrated the energy function's application in loop prediction, and the results show that it can reliably model the nativeness of the target loops, regardless of the length of the loops [25].

In this work, we present our energy function's application in the protein structural refinement regime. Because we primarily focus on the energy function, we do limited global optimization around a given start state, leaving it to future work to couple it to more extensive global sampling. To follow the same regulation as the Critical Assessment of Structure Prediction evaluations, we select the best five refined models, which have the lowest energy at the end of the refinement algorithm for the performance evaluation. The results show average improvements of ~0.5Å RMSD, with our best result bringing 1WHZ 1.32Å closer to the native

state. More importantly, our energy function does not show degradation in prediction when sampling is exhausted.

MATERIALS AND METHODS

Energy Function. The energy function we use in structural refinement is very similar to our previous studies, but is now based on the following components:

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

where $V_{Protein}$ is updated from our previous studies by replacing AMBERff99 [26] with the AMBERff99SB protein force field developed by Simmerling and co-workers [27]. The next two terms of Eq. (1) are taken from our previous model, where V_{GB} is the Generalized Born (GB) description of the electrostatic component of solvent free energy [28], and V_{HPMF} is the hydrophobic potential of mean force (HPMF) to describe the hydrophobic solute-solute interaction induced by water [29,30]. We refer the reader to our previous work [24,25] for the functional form and the parameter details in regards these terms. Finally we have added an energy term developed by Kabsch and Sander [31] to provide a means for improving the geometry of protein backbone hydrogen bond formation.

$$V_{HB} = 332 \times q_1 q_2 \left(\frac{1}{R_{ON}} + \frac{1}{R_{CH}} + \frac{1}{R_{OH}} + \frac{1}{R_{CN}} \right) \quad (2a)$$

$$V_{HB} = V_{HB} \times \frac{1}{2}[tanh(100 \times [-R_{ON} + 5.2]) + 1] \quad (2b)$$

$$V_{HB} = V_{HB} \times \frac{1}{2}[tanh(100 \times [-V_{HB} - 0.5]) + 1] \quad (2b)$$

In Eq. (2a), q_1 and q_2 are 0.42e and 0.20e, respectively, e being the unit electron charge and R_{AB} is the distance between atom A and B in angstrom. A hydrogen bond interaction will be considered only if two criteria are fulfilled that the distance between O and N atoms is less than or equal to 5.2Å, and the energy is less than -0.5 kcal/mol. Due to these two conditions, the model becomes a step function; therefore, we implement the hyperbolic tangent function in Eq. (2b) and (2c) to convert the model to a continuous function. This V_{HB} model has been previously proven to form intact secondary structures during dynamics simulations and increase the discrimination of a protein's native from misfolds [21].

With the two modifications implemented, we validate the modified energy function as described in Eq. (1) with the two decoy sets, 4-state-reduced [32] and LMDS [33] from the collection that we used in the previous work [24]. The results of native ranking and native z-score produced by both the old function and the modified function are shown in Table I. Compared to the previous version of the energy function, the new model shows a improvement in both of these performance metrics. It is able to correctly rank the native in 15 out of the 16 tested cases, which includes 3ICB from 4-state-reduced set that the previous energy model could only rank the native as the second lowest-energy structure among the decoys. 1B0N-B (LMDS set) is the only case in which the new energy function is not able to correctly rank. It is a troublesome case that may be due to its being a designed, non-naturally occurring protein, and its small size may need other subunits to stabilize it. In terms of native z-score, the new energy function systematically shows superiority in the ability to discriminate each protein's native from the decoys, except for 2CRO, 2OVO and 4PTI in LMDS set. On average, the new function improves the z-score by 0.151 for the 4-state-reduced set, 0.055 for the LMDS set, and 0.097 overall. These results emphasize the transferability of the components of our energy model, since we did not re-optimize any components of the energy function to improve the protein aspect of the force field. However, the added hydrogen-bonding function is not the focus here. It deals with a problem that happens in repeated optimization without solvent that secondary structure degrades, and this term was chosen intentionally to keep secondary structure reasonably refined without it having a strong impact on energy ranking, which is already optimized using the GB and HPMF term.

Selection and Preparation of Proteins. For evaluating the energy function's application in refinement, we select proteins from the targets in the recent CASP competitions, and 1CTF in the 4-state-reduced set for benchmark purposes. Since energy minimization is a computationally demanding calculation, we select proteins with a manageable size below 100 amino acids. We also restrict the proteins to be refined to possess experimental native structures solved by X-ray crystallography and with no missing backbone heavy atoms. For each of the chosen proteins, we select four initial structures among the participants' predicted models that fall between 3.0Å and 5.7Å in RMSD with respect to the experimental native state, and the models must contain at least the backbone heavy atoms. Since our energy model requires hydrogen atom positions to be specified, we use the CHARMM modeling package [34] to build the positions of the hydrogen

atoms. The information concerning the selected proteins and the RMSD of the start models are listed in Table II.

Energy Minimization Procedure. Each structure is locally minimized using Eq. (1) and its Cartesian derivatives using the BFGS (Broyden-Fletcher-Goldfarb-Shanno) [35] limited memory quasi-Newton method [36]. Due to the computational cost and the number of structures subject to energy minimization, we minimize the structures for 1000 steps instead of reaching full convergence. All of the RMSD calculations are conducted using the MMTSB Tool Set [37], and they represent the values of Cα RMSD unless indicated otherwise.

Structural Refinement Algorithm. Protein structural refinement can be divided into two regimes, local and global. Local refinement focuses on topologically correct predictions that are deficient in localized regions of a protein structure such as loops and side chains, while global refinement considers the wholesale reorganization of secondary structure and/or their repacking in tertiary form. Most of the proteins we consider belong to the local class, but still are refinement targets, and thus we generated a fairly conservative global optimization strategy relative to *de novo* prediction. We introduce random sampling of all backbone dihedral angles by some degree of perturbation, which we optimize on seven proteins in the 4-state-reduce set [32] that we used in our previous work [24]. Figure 1 shows that 1° perturbation generates very little variation from the start configuration, while dihedral angle variations >5° perturbation generates a majority of structures that move further away from the native state relative to the starting structure. The optimal appear to be ~3° perturbation which gives diverse structures, half of which are closer to the native than the start model.

The overall structural refinement algorithm is illustrated in Figure 2. For each start model, we generated 10,000 structures per round for each of the test proteins to do local minimization. The five structures with the lowest energy are chosen as the initial seeds for the next search round, where each of them is used to generate another set of 10,000 structures. We repeat this procedure for a total of 5 rounds, and the final RMSD with respect to the native state of the five lowest-energy structures collected at the end of the fifth round of the global optimization are compared to the RMSD of the initial structure to evaluate the algorithm's performance.

RESULTS

The best RMSD and average RMSD over the 5 lowest-energy structures at the end of each round for each start model are listed in Table III. During the early stage of the refinement process, the energy function could not consistently pick out five structures that have lower RMSD than the start models. But as the refinement progresses to the later rounds, the number of structures that have lower RMSD increases, and in the final round, all of the five structures for each start model, which the energy function assigned the lowest energy, show improvement in RMSD, i.e. move closer to the native state. The start model that has the least improvement is the start state #2 of 3DED, which moves 0.045 Å closer to the native comparing to the starting model. However we also consider this a good result in the sense that the energy function discriminates against thousands of competing misfolds, essentially indicating that the sampling is exhausted in the localized region around the start state.

On the other hand, the largest improvement happens on 1WHZ (start state #3 with a start RMSD of 5.32 Å), and after going through five rounds of refinement, the average RMSD of the five lowest-energy structures is a little less than 4.00 Å, which shows ~1.32 Å difference or 25% of improvement. Figure 3 shows the superposition between 1WHZ native (green), the start model (cyan) and the lowest-energy structure at the end of the refinement process (magenta). There is significant improvement on every element of secondary structure, in which the refinement process improves both the hydrogen bonding between the second and third strand in the β-sheet, while the three α-helices in the refined model align better with the ones in the native. In fact, the alignment between the first α-helix in the refined model and the native is almost perfect. We also made significant progress on the refinement of 1CTF, 3D7I, 2IWP, and 1WHZ, in which all start models with RMSD of 3.3-3.7 Å for these proteins had final RMSD values less than 3.0 Å, which shows a promising achievement in refining templates of protein structures toward the high-resolution scale.

We also calculate the RMSD of the 10,000 structures in each global round for each start model before and after the energy minimization, and we then average the change in RMSD of each structure over the 10,000 structures. The resulting average ΔRMSD values are listed in Table IV. During the energy minimization in each global round for each start model, on average, the energy function is able to consistently bring the structures closer to the corresponding native state, with the range of improvement of ~0.10-0.30 Å. Among each of the 10,000-structure set, the energy function can minimally guide ~50% of them in the direction to the native, and maximally, it can direct 99% of them closer to the native. The measurements indicate that during minimization the energy function is capable of moving the majority of the structures closer to its corresponding native.

Finally, in Table V we include other metrics to measure prediction quality of our model, including the global distance test score (GDT-TS) used in the community-wide CASP and CAFASP experiments[38], a template modeling score (TM-score)[39], and side chain RMSD. GDT-TS counts the number of Cα pairs that have a distance of 1.0Å, 2.0Å, 4.0Å, and 8.0Å to the reference structure after superposition, while TM-score counts all residue pairs using the Levitt–Gerstein weight and is more sensitive to the global topology than local variations. It is evident that ~65-70% of the time that the energy model discriminates toward better GDT-TS and TM-scores. Since our model always improves backbone RMSD, the side chain RMSD suggest that the search strategy could be greatly improved by optimizing side chain rotomers after backbone randomization.

DISCUSSION

In our previous work[24,25], we developed a solvation energy model (HPMF) that captures the solvent-induced hydrophobic interactions, and we coupled it with the AMBER ff99 protein force field and the GB model to use in protein globular structure prediction and local structure refinement or loop modeling. Here, we update the AMBER protein force field from ff99 [26] to ff99sb [27] and add an energy term [31] that mediates the backbone hydrogen bond formations. Based on the test on 4-state-reduced and LMDS decoy sets, the new energy function allows for better native ranking than the previous version, and the native z-score is systematically improved. This result also demonstrates the transferability of the solvation components of our energy model in

that it can be coupled with another empirical protein force field and still remain effective. In this work we have demonstrated the solvation model's application in global structure refinement, validated with the targets selected from the recent CASPs. The function is able to consistently refine all of the start models closer to the native state with average improvement from 0.04 Å to 1.32 Å, and with some results suggesting that refining protein structures to the high-resolution regime is possible with better search strategies than explored here., especially optimization of side chain rotomers. Compared to structure prediction, structure refinement requires even more stringent sampling[40]. We have not attempted to optimize this component of the refinement strategy, instead focusing on the development of the energy model.

In summary, with the current computational methods such as template-based and *de novo* techniques, models with low to medium resolution can consistently be predicted; however, getting models to high-resolution still remains one of the great challenges in the field. We hope that the energy function that we present here can provide means for refining medium resolution models to high resolution.

ACKNOWLEDGEMENTS

We gratefully acknowledge financial support under the DOE-ASCR program. We would also like to thank Brian Carnes for his work on building the POISE molecular modeling package. 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.

REFERENCES

1. Moult J, Fidelis K, Kryshtafovych A, Rost B, Hubbard T, Tramontano A. Critical assessment of methods of protein structure prediction - Round VII. Proteins-Structure Function and Bioinformatics 2007;69:3-9.
2. Moult J, Fidelis K, Rost B, Hubbard T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP) - Round 6. Proteins-Structure Function and Bioinformatics 2005;61:3-7.
3. Moult J, Fidelis K, Zemla A, Hubbard T. Critical assessment of methods of protein structure prediction (CASP)-round V. Proteins-Structure Function and Genetics 2003;53(6):334-339.
4. Zhang Y, Kolinski A, Skolnick J. TOUCHSTONE II: A new approach to ab initio protein structure prediction. Biophysical Journal 2003;85(2):1145-1164.
5. Zhang Y, Skolnick J. Automated structure prediction of weakly homologous proteins on a genomic scale'. Proceedings of the National Academy of Sciences of the United States of America 2004;101(20):7594-7599.
6. Simons KT, Bonneau R, Ruczinski I, Baker D. Ab initio protein structure prediction of CASP III targets using ROSETTA. Proteins-Structure Function and Genetics 1999:171-176.
7. Lundstrom J, Rychlewski L, Bujnicki J, Elofsson A. Pcons: A neural-network-based consensus predictor that improves fold recognition. Protein Science 2001;10(11):2354-2362.
8. Fischer D. 3D-SHOTGUN: A novel, cooperative, fold-recognition meta-predictor. Proteins-Structure Function and Genetics 2003;51(3):434-441.
9. Kolinski A, Bujnicki JM. Generalized protein structure prediction based on combination of fold-recognition with de novo folding and evaluation of models. Proteins-Structure Function and Bioinformatics 2005;61:84-90.
10. Bradley P, Misura KMS, Baker D. Toward high-resolution de novo structure prediction for small proteins. Science 2005;309(5742):1868-1871.
11. Baker D, Sali A. Protein structure prediction and structural genomics. Science 2001;294(5540):93-96.
12. Lee MR, Tsai J, Baker D, Kollman PA. Molecular dynamics in the endgame of protein structure prediction. Journal of Molecular Biology 2001;313(2):417-430.
13. Lee MR, Baker D, Kollman PA. 2.1 and 1.8 angstrom average C-alpha RMSD structure predictions on two small proteins, HP-36 and S15. Journal of the American Chemical Society 2001;123(6):1040-1046.
14. Simmerling C, Strockbine B, Roitberg AE. All-atom structure prediction and folding simulations of a stable protein. Journal of the American Chemical Society 2002;124(38):11258-11259.
15. Fan H, Mark AE. Refinement of homology-based protein structures by molecular dynamics simulation techniques. Protein Science 2004;13(1):211-220.
16. Zhang Y, Skolnick J. The protein structure prediction problem could be solved using the current PDB library. Proceedings of the National Academy of Sciences of the United States of America 2005;102(4):1029-1034.
17. Misura KMS, Chivian D, Rohl CA, Kim DE, Baker D. Physically realistic homology models built with ROSETTA can be more accurate than their templates. Proceedings of

the National Academy of Sciences of the United States of America 2006;103(14):5361-5366.
18. Chen JH, Brooks CL. Can molecular dynamics simulations provide high-resolution refinement of protein structure? Proteins-Structure Function and Bioinformatics 2007;67(4):922-930.
19. Misura KMS, Baker D. Progress and challenges in high-resolution refinement of protein structure models. Proteins-Structure Function and Bioinformatics 2005;59(1):15-29.
20. Jagielska A, Wroblewska L, Skolnick J. Protein model refinement using an optimized physics-based all-atom force field. Proceedings of the National Academy of Sciences of the United States of America 2008;105(24):8268-8273.
21. Wroblewska L, Jagielska A, Skolnick J. Development of a physics-based force field for the scoring and refinement of protein models. Biophysical Journal 2008;94(8):3227-3240.
22. Chopra G, Summa CM, Levitt M. Solvent dramatically affects protein structure refinement. Proceedings of the National Academy of Sciences of the United States of America 2008;105(51):20239-20244.
23. Summa CM, Levitt M. Near-native structure refinement using in vacuo energy minimization. Proceedings of the National Academy of Sciences of the United States of America 2007;104(9):3177-3182.
24. Lin MS, Fawzi NL, Head-Gordon T. Hydrophobic potential of mean force as a solvation function for protein structure prediction. Structure 2007;15(6):727-740.
25. Lin MS, Head-Gordon T. Improved energy selection of nativelike protein loops from loop decoys. Journal of Chemical Theory and Computation 2008;4(3):515-521.
26. Wang JM, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? Journal of Computational Chemistry 2000;21(12):1049-1074.
27. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins-Structure Function and Bioinformatics 2006;65(3):712-725.
28. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins-Structure Function and Bioinformatics 2004;55(2):383-394.
29. Crivelli S, Eskow E, Bader B, Lamberti V, Byrd R, Schnabel R, Head-Gordon T. A physical approach to protein structure prediction. Biophysical Journal 2002;82(1):36-49.
30. Head-Gordon T, Brown S. Minimalist models for protein folding and design. Curr Opin Struct Biol 2003;13(2):160-167.
31. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22(12):2577-2637.
32. Park B, Levitt M. Energy functions that discriminate X-ray and near-native folds from well-constructed decoys. Journal of Molecular Biology 1996;258(2):367-392.
33. Keasar C, Levitt M. A novel approach to decoy set generation: Designing a physical energy function having local minima with native structure characteristics. Journal of Molecular Biology 2003;329(1):159-174.
34. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. Journal of Computational Chemistry 1983;4(2):187-217.

35. Press WH, Flannery BP, Teukolsky SA, Vetterling WT. Numerical Recipes in C: The Art of Scientific Computing: Cambridge University Press; 1992.
36. Liu DC, Nocedal J. On the Limited Memory Bfgs Method for Large-Scale Optimization. Mathematical Programming 1989;45(3):503-528.
37. Feig M, Karanicolas J, Brooks CL, 3rd. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. J Mol Graph Model 2004;22(5):377-395.
38. Read RJ, Chavali G. Assessment of CASP7 predictions in the high accuracy template-based modeling category. Proteins 2007;69(S8):27-37.
39. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. Proteins 2004;57(4):702-710.
40. Das R, Qian B, Raman S, Vernon R, Thompson J, Bradley P, SKhare S, Tyka MD, Bhat D, Dhivian D, Kim DE, Sheffler WH, Malmstrom L, Wollacott AM, Wang C, Andre I, Baker D. Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@home. Proteins-Structure Function and Bioinformatics 2007;69(S8):118-128.

FIGURE CAPTIONS

Figure 1. Degree of perturbation. One-degree of perturbation does not generate structures with a great diversity. Five and higher degree of perturbation generate too many structures that are less native-like than the initial model, moving away from the native state. Three-degree of perturbation gives diverse structures, and about half of them are closer to the native than the start model.

Figure 2. Refinement protocol flowchart. For each initial structure, we select five structures using the energy function at the end of the first round. The five models are then used as the initial seeds in the following round, and the procedure repeats for three times. We then collect the five lowest-energy structures at the end of the fifth global round for the performance evaluation.

Figure 3. Results of 1WHZ start state #3. The structure in green is the native, and the cyan is the start model with the initial RMSD of 5.32 Å, and the magenta is the lowest-energy structure with the RMSD of 3.98 Å at the end of the refinement process. (a) It gives the overview of the superposition between the three structures. (b) It focuses on the first α-helix. The native and the refined model are aligned almost perfectly. (c) It shows the 3-strand β-sheet. The left and middle strands of the final structure show a good alignment with the native, and the two strands form better hydrogen bonds than the start model. The right strand moves closer to the native, but it still requires further refinement. (d) It shows the second (middle) and the third (left) α-helix in the structure. After the refinement process, the alignment between the refined model's second α-helix and native's is better than the ones between the start model and the native. However, one

hydrogen bond is broken at the beginning of the α-helix in the refined model. The axial direction of the third α-helix in the native and the final model is aligned, but more refinement is still needed.