Received: 1 May 2019 | Revised: 11 July 2019 | Accepted: 12 July 2019

DOI: 10.1002/prot.25784

RESEARCH ARTICLE

PROTEINS WILEY

High-accuracy refinement using Rosetta in CASP13

Hahnbeom Park[1] | Gyu Rie Lee[1] | David E. Kim[1,2] | Ivan Anishchenko[1] | Qian Cong[1] | David Baker[1,2]

[1]Department of Biochemistry and Institute for Protein Design, University of Washington, Seattle, Washington

[2]Howard Hughes Medical Institute, University of Washington, Seattle, Washington

Correspondence
David Baker, Howard Hughes Medical Institute, University of Washington, Seattle, WA.
Email: dabaker@uw.edu

Funding information
National Institutes of Health, Grant/Award Number: R01GM092802

Abstract

Because proteins generally fold to their lowest free energy states, energy-guided refinement in principle should be able to systematically improve the quality of protein structure models generated using homologous structure or co-evolution derived information. However, because of the high dimensionality of the search space, there are far more ways to degrade the quality of a near native model than to improve it, and hence, refinement methods are very sensitive to energy function errors. In the 13th Critial Assessment of techniques for protein Structure Prediction (CASP13), we sought to carry out a thorough search for low energy states in the neighborhood of a starting model using restraints to avoid straying too far. The approach was reasonably successful in improving both regions largely incorrect in the starting models as well as core regions that started out closer to the correct structure. Models with GDT-HA over 70 were obtained for five targets and for one of those, an accuracy of 0.5 Å backbone root-mean-square deviation (RMSD) was achieved. An important current challenge is to improve performance in refining oligomers and larger proteins, for which the search problem remains extremely difficult.

KEYWORDS
energy function, homology modeling, protein conformational search, rotein structure prediction

1 | INTRODUCTION

The CASP refinement category challenges predictors to improve on the best models submitted for a subset of targets. In principle, with an accurate energy function and sufficient conformational search, any protein structure prediction problem including refinement should be directly solvable by searching for the lowest energy conformation. We recently showed that with progress in energy function accuracy, refinement by unrestrained large-scale search for the lowest energy nearby structure could improve low-resolution homology models.[1] However, when applied to starting models closer to the actual structure, the approach more often reduced rather than increased the accuracy of the starting models.[2,3] We reasoned that improved refinement of closer to native starting models might be achievable if the structural operators used in the search—replacing backbone torsion angles and perturbing the rigid-body orientations of secondary structures—were restrained (regularized) based on the starting structures.

Several ways of utilizing input structure information have been proposed for restraining conformational search.[4-9] Seok and collaborators used Bayesian inference models with ambiguous restraints[4,5] to restrain coordinates; for high-accuracy modeling problems such Cartesian restraints provide the most control over structure.[5,7] In this approach, once any subset of restraints derived from the input model are satisfied the rest of the restraints do not affect energy evaluation, which is advantageous for refinement scenarios where the challenge is to reconstruct potentially inaccurate local regions while preserving the core of the structure. In this CASP, we incorporated such

ambiguous restraints into our energy-guided refinement approach for the two-thirds of the starting models with GDT-HA > 50, and succeeded in a number of these cases in improving accuracy (for the remaining one-third of targets with starting GDT-HA < 50, we carried out unrestrained energy-guided search as in CASP12).

2 | METHODS

2.1 | Overview of the approach

A schematic overview of our CASP13 refinement protocol is shown in Figure 1A. Overall, the framework is similar to that in our previous studies[1,2]; here we provide a brief description. Starting from an input structure, regions with local errors were predicted through running short MD simulations[10] in Rosetta and were reconstructed by fragment assembly. An initial pool of 50 selected low energy structures were then subjected to iterative refinement; at each iteration new models were built from the current pool through recombining secondary structure chunks and replacing torsional angles with values from a generic fragment library.[11] Next generation structures were selected from the entire pool (including both parents and newly generated structures) based on Rosetta all-atom energy[12] and distance from other members of the new pool. Once the iterations were completed, the lowest energy structure sampled was identified, and conformations close to this structure were structurally averaged to output a single refined model, which was subjected to restrained MD[7] using the AMBER suite[13] to improve modeling of explicit water molecule dependent structure features, followed by structural averaging and geometry optimization.[14,15]

In our earlier work, the above framework was applied to a subset of CASP challenges with starting model GDT-HA less than 50. For this CASP we developed two variations of the overall approach for high- and low-resolution refinement challenges. The differences between the two strategies are in the restraints and the scheduling of the iterations. The low-resolution strategy consists of 50 generations with very weak distance restraints only applied to residue pairs whose distance did not vary within the first generation parents,[1] and no explicit restraints derived from the input model. The high-resolution strategy utilizes ambiguous coordinate restraint sets derived from the input models. It consists of 10 regular iterations without restraints to explore the surrounding space, followed by 10 iterations tethered through ambiguous coordinate restraints to the input structure to cause regions of the structure that have not found lower energy states in the initial unrestrained search to move back toward the starting structure (those that have fallen into lower energy states can effectively resist the applied restraints).

The details of the restraints are explained in the following *Restraints* section. In the high-resolution strategy, the weight on the restraint set relative to the Rosetta energy gradually increased from the 10th to the 15th iteration and remained the same for the final five iterations. The rationale for this restraint scheduling is as follows. During the first 10 iterations of unrestrained simulation, regions that were

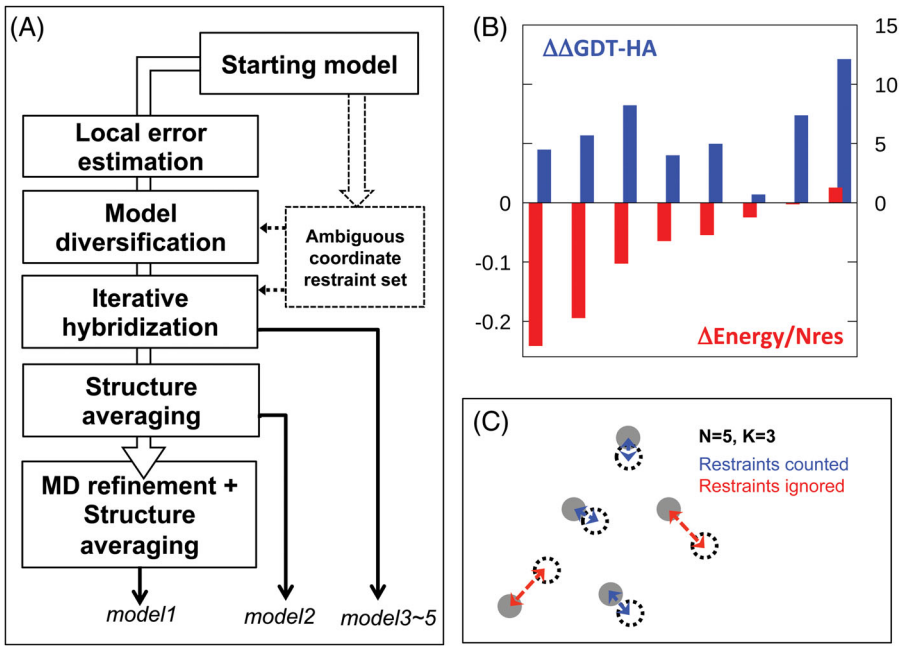

FIGURE 1 CASP13 refinement protocol. (A) Schematic overview of the protocol; see main text for brief description.[1] (B) Regularized conformation search results in better quality structures with lower energies. Eight targets from previous CASP rounds were refined with and without restraints. The difference in the improvements over the starting models between regularized and unregularized runs is in blue bars (axis values on right); higher values indicate improved performance with regularized search. The difference in energy of the lowest energy conformation sampled, normalized by the number of residues, is in red bars (axis values on left). Lower values indicate regularized search found a lower energy conformation. (C) Graphical description of ambiguous coordinate restraint set applied in this study on a simple model system with N (number of residues) = 5 and K = 3. During restraint evaluation, the violation of each restraint is measured, and the penalties associated with the K smallest violations (blue) are summed while the remainders (red) are ignored

correctly modeled in the starting structure may be already in an energy minimum and hence, stay relatively converged. As the restraints were ramped up, these regions become even more converged, and hence, the energy differences between conformations become almost entirely due to the more variable parts of the structure, enabling search to focus on minimizing the energy of those regions.

We found on a benchmark set that search with ambiguous restraints sampled lower energy models than the original unrestrained search—this is notable given that restraining search in general should hamper optimization—and also resulted in more accurate models for most of the targets (Figure 1B). Regularization of conformational search thus can help the refinement procedure find more efficient paths towards the native structure with equivalent computational cost.

In CASP13, we utilized *a priori* knowledge of starting model quality, provided by the organizers, to decide whether to use the low- or high-resolution strategies for each target. Iterative refinement with regularization (high-resolution strategy) was applied to the 20 medium-to-high accuracy starting models having GDT-HA[16] above or equal to 50. The low-resolution strategy was applied to the remaining 11 targets having starting model GDT-HA less than 50. The final MD-refined model was submitted as model1, the pre-MD model—which is an averaged structure after iterative Rosetta refinement—as model 2, and the rest of the models from the last refinement iteration pool. Models generated by the automated procedure shown in Figure 1A were submitted as a group BAKER-AUTOREFINE. These models were visually investigated, further refined if necessary, and submitted as models for a separate human group BAKER, following the above scheme for designating model1 to model5. When the best template for a target was a homo-oligomer, the whole refinement procedure was carried out with symmetry operations derived from this template. In the following sections, more details of the restraints, symmetric refinement, and human interventions are described.

2.2 | Derivation of restraints from starting models

Restrained protein structure prediction can be viewed as a search for the structure with the highest probability given the available data *D* (p[x|D] where x is the structure).[4,5] With Bayes' theorem, the problem can be formulated as:

$$\begin{aligned}\log\ p(x|D) &= \log p(x) + \log p(D|x) - \log p(D) \\ &= \left(-\Sigma E_{Rosetta} + w\,\Sigma_i^{K} E_{restraints,i}\,(x - x_0)\right) - C\end{aligned} \quad (1)$$

where the *prior term* p(x) is the Boltzmann weight of the structure (in our case with the Rosetta energy function $p(x) \sim \exp(-\Sigma E_{Rosetta}(x))$, the *likelihood term* p(D|x) is a function of the consistency of the input data (in this case, the input coordinates x_0) with the current structure x ($p(D|x) = \Sigma_i^{K} \exp(-w\Sigma_i E_{restraint,i}(x-x_0))$,[5] and the third term p(D) is irrelevant to search as it is independent of x.

In general refinement problems, not all the input information D is correct, and hence, it is desirable to aim to satisfy some but not all of it. The fraction satisfied is controlled by the parameter K in Equation 1 that ranges between 0 and the total number of restraints N. In this work, we applied coordinate restraints to Cα atoms; hence N is the number of residues and K the optimal number of residues to restrain (an example depicted in Figure 1C). At energy evaluation, all the restraints are evaluated and the K least violated are included in the sum; the remaining N-K residues are unrestrained.

Two critical parameters to be determined are f = K/N (the fraction of residues restrained), and w, the weight on the restraint set with respect to Rosetta energy function in the final five iterations (the weights on iterations 10-15 were interpolated from 0 to w). As described in the methods, we parameterized f as a function of input model accuracy (as provided by organizers) and molecular shape (the rationale for the second factor is that elongated models can vary more than globular models with same strength and number of restraints). With the formula we tested in this CASP (details in Supporting Information) the fraction of residues (f) restrained ranged from 0.24 to 0.81; we did not extensively explorer the procedure for chosing f and it can **likey** be further optimized in the future.

A limitation of our approach as described above is the dependence on input model GDT-HA (as in CASP refinement challenges); in the general (non-CASP) case this value has to be estimated as it is not known. Model accuracy estimation methods are improving, and can provide estimates of starting model accuracy within 5~10% of the actual values (as shown in CASP13 model accuracy estimation assessment) for both GDT and the contact accuracy metric (lDDT). In the general (non-CASP) case, such methods can be used to decide to what extent starting model restraints should be during refinement; ultimately such accuracy predictors can be incorporated into our protocol.

2.3 | Symmetric refinement

The Rosetta symmetry modeling machinery used for homo-oligomer comparative modeling[17,18] was incorporated into the refinement pipeline in this CASP. The machinery allows all the modeling operations in Rosetta refinement to be carried out on the asymmetric unit with propagation of fragment insertion, secondary structure chunk swapping, side-chain modeling, and minimization to symmetry mates. MD simulations with AMBER were run with full symmetric units instead due to explicit waters. These restrained MD simulations generally do not alter structures much from the input Rosetta-refined models, therefore any single chain of the averaged model across the MD trajectory (chain A by default) was submitted if a single chain was requested for submission. Otherwise, a re-symmetrized model of that single chain structure was submitted. Targets were chosen for symmetric refinement if the best template had heavy atom oligomeric contacts; three targets were selected for automated predictions (R0977-D4, R0979, and R0981-D4), and two more for human-guided predictions (R0981-D5 and R0989-D1).

2.4 | Human intervention

Human interventions included (a) detecting additional regions to reconstruct not identified by the automatic procedure, (b) deciding when to apply the low- and high-resolution refinement protocols, (c) applying symmetric refinement for cases missed in automatic refinement, and (d) use of co-evolution information if available for

large number of sequences (Nf^{19} > 64). Only the last of these interventions gave clear net differences from our automated submissions.

3 | RESULTS

3.1 | High-accuracy refinement through regularized conformational search

Our CASP refinement results are summarized in Figure 2 for the human submission group. Consistent and significant improvements were found over starting models for both human and automated groups; except for a few cases, the performance of the two was quite similar. The average changes in qualities of model1 over the starting models were + 3.3 and + 2.9 in GDT-HA, and + 2.6 and + 1.6 in SphereGrinder,[20] for human and automated predictions, respectively. Improvements were larger for medium-to-high accuracy starting models. Amongst human predictions, refinement of five targets (Figure 2A, arrows) resulted in high model quality as assessed by both coordinate (GDT-HA > 70) and contact (SphereGrinder [SG] > 90) measures. Over 16 targets refined with the high-resolution strategy, the average ΔGDT-HA was +4.0 (SD 6.4). This is a substantial improvement over our CASP12 results[3] in which average improvement was +1.2 (SD 3.9) on 19 targets in a similar starting model quality range.

Our "model 1" submissions were consistently the best or close to the best in GDT-HA metric among all our submitted models (Figure 2C). Structure averaging of low energy conformations in the Rosetta refinement calculations generally produced better models than the other selected single models, and the subsequent MD refinement consistently increased GDT-HA when the input Rosetta-refined model had reasonable model accuracy. This latter trend is weaker when improvements are assessed using a superposition-free metric (SG, Figure 2D); indeed most of the improvements in SG were achieved in the Rosetta-refinement stage.

Our prediction for target R0974s1 is among the most accurate in all CASP refinement challenges (Figure 3A). Model 1 from our human prediction has a GDT-HA of 93 and Cα-RMSD of 0.48 Å. Post analysis on this target indicated a complementary role of Rosetta and explicit water MD as observed previously[2,21,22]: while Rosetta successfully recovered a critical error at the C-terminus and side-chain orientations around it, MD successfully brought the positions of backbones and side-chains even closer to the crystal structure. Our MD refinement alone on the input model did not result in this level of accuracy (GDT-HA of 83 and RMSD 1.6 Å). However, when more aggressive sampling with MD was tried by the Feig group involving multiple rounds of Markov state simulations with restraints,[23] high-accuracy refined models were achieved comparable to ours. Based on our results and those of the Feig group on R0974s1, even for relatively easy refinement problems, it is critical to allow significant energy barrier hopping during search.

Three additional high-accuracy predictions are shown in Figure 3B-D for which substantial improvements were achieved over the starting models. Similar to R0974s1, all of these models improved in local regions (highlighted as arrows) initially deviating from the native structure while the detailed packing of secondary structures was also improved.

As noted earlier, the only factor that consistently improved our human predictions over our automated predictions was to add co-evolution information when it was reliable. Over the five targets for which co-evolution information were incorporated, the improvement over the starting models was greater for the human than the automated

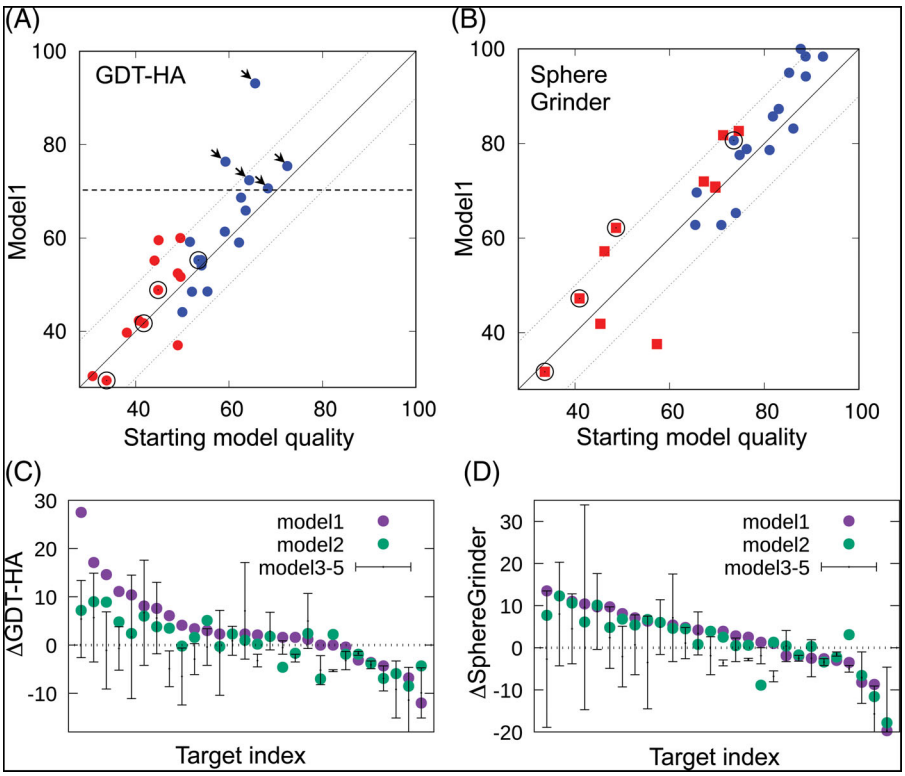

FIGURE 2 Overall results in CASP13 refinement category by group BAKER. (A, B) Scatter plots comparing the model qualities of our model1 submitted predictions (y-axis) and the starting models (x-axis). Targets refined with the high-resolution strategy (ie, regularized search) and low-resolution strategies are colored in blue and red, respectively. Dots with surrounding circles correspond to the targets refined with symmetry. (A) GDT-HA and (B) SphereGrinder. (C, D) Quality of all five models submitted. Targets are sorted based on their model 1 improvements. Model 1 and 2 are shown in purple and green dots, respectively, while the rest of five models are in error bars. (C) GDT-HA and (D) SphereGrinder

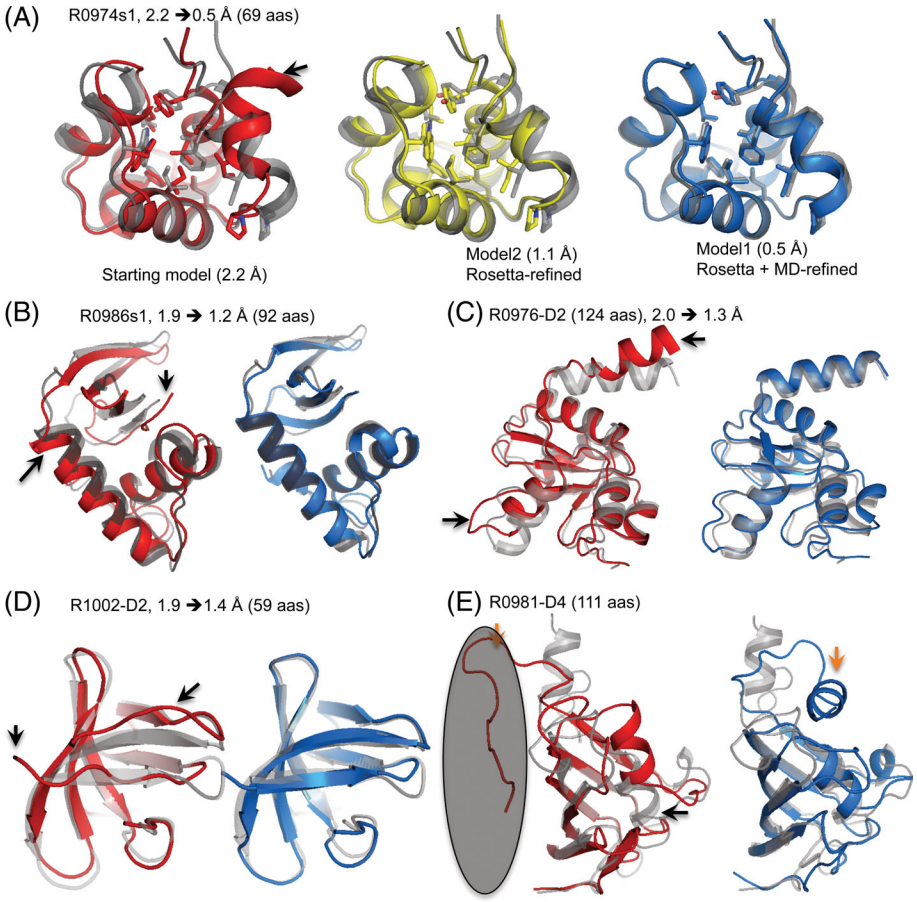

FIGURE 3 Examples of successful model refinement. In all panels, native structures, starting models, and final refined models are shown in gray, red, and blue, respectively. Regions initially wrong in the starting models are highlighted in black arrows. In each panel, the change in backbone RMSD produced by refinement is shown next to target index. (A) R0974s1, refined model is very close to the crystal structure. Wrong conformation at the C-terminus of starting model was fixed at Rosetta-refinement stage (panel at the middle, model colored in yellow), and further improved through MD simulations. Model quality improves from starting to Rosetta-stage to final MD-stage by 66 to 73 to 93 in GDT-HA, and 93 to 100 to 100 in SphereGrinder. (B-D) Other targets refined to high accuracy. Refined models for all these cases achieved GDT-HA > 70 and RMSD <1.5 Å. Floppy residues from 1 to 4 are trimmed from the native structure for the evaluation of R0976-D2. (E) R0981-D4, an example of refinement with symmetry. The protein forms a homotrimer in the template and crystal structure, the oligomeric interface of which is shown as a gray circle. The major improvement comes from fixing helical placement as highlighted in black arrow. Structural errors at the interface, highlighted by orange arrows, were not improved by refinement

submissions, average ΔGDT-HA of 5.2 and 1.1, respectively (the improvements are not distinguishable over the remaining 24 targets). The co-evolution restraints do not contribute by helping build missing contacts: all the starting models satisfy all the restraints from co-evolution. Instead, the co-evolution information further restrains the search space to a region that includes the correct structure, especially for those targets for which the more aggressive unconstrained search strategy was used (R0949 and R0997). Without restraints from co-evolution, model quality decreased using the automated protocol (ΔGDT-HA by −6 for R0949 and ΔSG by −20 for R0997).

3.2 | What did not work

It has been consistently observed both in CASP and in our previous studies[1,2,21] that refinement of proteins larger than 120 residues is very difficult due to the rapidly increasing size of conformational space. This is likely why we did not observe significant improvements with our low-resolution protocol in this CASP: there were no refinement targets shorter than 110 residues in this category (when starting model's GDT-HA <50).

We incorporated symmetric modeling into our refinement protocol in this CASP but this did not have a big effect. R0981-D4, one of the targets refined with symmetry, was considerably improved, but in regions distant from the oligomeric interface (Figure 3E); it is not clear whether refinement in oligomeric context added any value over running it on a monomeric context.

4 | DISCUSSION

4.1 | How accurate are the predictions compared to predictions in TS category?

In this CASP, great progress was achieved in regular tertiary structure prediction driven by deep learning applied to co-evolution data. The best predictors in refinement category did not utilize deep learning but were

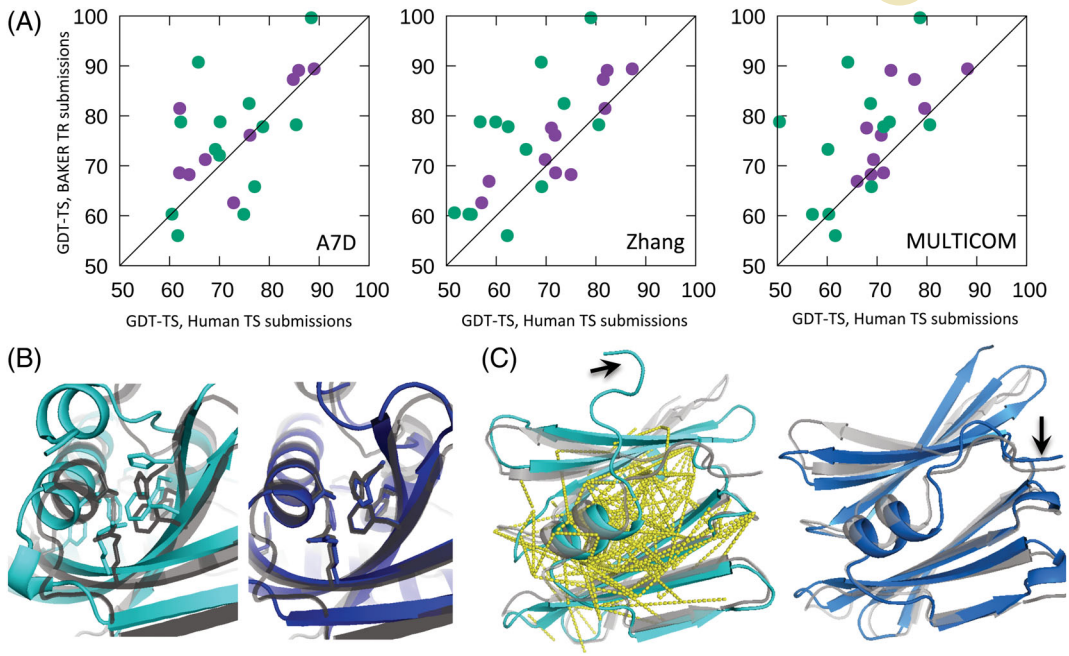

FIGURE 4 Comparison to TS human category predictions. (A) Scatter plots comparing GDT-TS values of models submitted by each of three best human predictors (x-axis, group name shown inside each panel) to those of our submissions (y-axis), on 29 targets that were in both the human tertiary prediction and refinement categories. Single- and multi-domain targets are colored in purple and green, respectively. (B) Comparison of side-chain accuracy on models for T0968s1 between the best human prediction (A7D, left panel) and our refined model (right panel). (C) Comparison of local backbone accuracy of the models for T0968s2, highlighted by black arrow, between the best human prediction (A7D, left panel) and our best refined model (right panel). Four hundred residues pairs with highest co-evolution strength estimated by GREMLIN[24] are shown in yellow dashes; note that the inaccurately modeled region is not covered by any of these

able to achieve considerable increases in accuracy. It is instructive to compare predictions from the two different categories to understand the value added from refinement efforts (it should be noted that this is not a completely fair comparison because refinement predictions always started with the knowledge of the best server model). Models from the refinement category were consistently better than the best human regular tertiary structure (TS) predictions on the 29 targets that the organizers included in both the refinement and TS categories (Figure 4A). Refinement improved both side-chain packing of the core (to make it specific for the amino acid sequence of the protein being modeled) (Figure 4B) and the backbone in regions with considerable sequence variation (Figure 4C); sequence co-evolution learned at the entire family level is not expected to capture either of these well. Modeling these aspects of protein structure will likely be an important role for refinement methods as deep learning coupled with co-evolution based structure prediction continues to advance.

We expect deep-learning techniques to contribute to protein structure refinement in the near future. Likely areas include both search and scoring, for example, more precisely identifying regions which refinement should focus on, suggesting moves to make in these regions, and ultimately optimization of protocol components or entire protocols using reinforcement learning.

ACKNOWLEDGMENTS

We thank the CASP13 organizers and the structural biologists who generously provided targets. Computing resources for this work are from Hyak supercomputer system at the University of Washington. The work was supported by the NIH.

ORCID

Hahnbeom Park https://orcid.org/0000-0002-7129-1912

REFERENCES

1. Park H, Ovchinnikov S, Kim DE, DiMaio F, Baker D. Protein homology model refinement by large-scale energy optimization. *Proc Natl Acad Sci U S A*. 2018;115:3054-3059.
2. Ovchinnikov S, Park H, Kim DE, DiMaio F, Baker D. Protein structure prediction using Rosetta in CASP12. *Proteins*. 2018;86(Suppl 1): 113-121.
3. Hovan L, Oleinikovas V, Yalinca H, Kryshtafovych A, Saladino G, Gervasio FL. Assessment of the model refinement category in CASP12. *Proteins*. 2018;86(Suppl 1):152-167.
4. MacCallum JL, Perez A, Dill KA. Determining protein structures by combining semireliable data with atomistic physical models by Bayesian inference. *Proc Natl Acad Sci U S A*. 2015;112:6985-6990.
5. Lee GR, Heo L, Seok C. Simultaneous refinement of inaccurate local regions and overall structure in the CASP12 protein model refinement experiment. *Proteins*. 2018;86(Suppl 1):168-176.
6. Wildberg A, Della Corte D, Schröder GF. Coupling an ensemble of homologues improves refinement of protein homology models. *J Chem Theory Comput*. 2015;11:5578-5582.
7. Mirjalili V, Feig M. Protein structure refinement through structure selection and averaging from molecular dynamics ensembles. *J Chem Theory Comput*. 2013;9:1294-1303.

8. Schröder GF, Levitt M, Brunger AT. Super-resolution biomolecular crystallography with low-resolution data. *Nature*. 2010;464:1218-1222.
9. Bhattacharya D. refineD: improved protein structure refinement using machine learning based restrained relaxation [internet]. *Bioinformatics*. 2019. https://doi.org/10.1093/bioinformatics/btz101.
10. Park H, Seok C. Refinement of unreliable local regions in template-based protein models. *Proteins*. 2012;80:1974-1986.
11. Gront D, Kulp DW, Vernon RM, Strauss CEM, Baker D. Generalized fragment picking in Rosetta: design, protocols and applications. *PLoS ONE*. 2011;6:e23294.
12. Park H, Bradley P, Greisen P Jr, et al. Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. *J Chem Theory Comput*. 2016;12:6201-6212.
13. Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip Rev Comput Mol Sci*. 2013;3:198-210. https://doi.org/10.1002/wcms.1121.
14. Tyka MD, Keedy DA, André I, et al. Alternate states of proteins revealed by detailed energy landscape mapping. *J Mol Biol*. 2011;405:607-618.
15. Conway P, Tyka MD, DiMaio F, Konerding DE, Baker D. Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Sci*. 2014;23:47-55.
16. Read RJ, Chavali G. Assessment of CASP7 predictions in the high accuracy template-based modeling category. *Proteins*. 2007;69(Suppl 8):27-37.
17. DiMaio F, Leaver-Fay A, Bradley P, Baker D, André I. Modeling symmetric macromolecular structures in Rosetta3. *PLoS ONE*. 2011;6:e20450.
18. Park H, Kim DE, Ovchinnikov S, Baker D, DiMaio F. Automatic structure prediction of oligomeric assemblies using Robetta in CASP12. *Proteins*. 2018;86(Suppl 1):283-291.
19. Ovchinnikov S, Park H, Varghese N, et al. Protein structure determination using metagenome sequence data. *Science*. 2017;355:294-298.
20. Antczak PLM, Ratajczak T, Blazewicz J, Lukasiak P, Blazewicz J. SphereGrinder - reference structure-based tool for quality assessment of protein structural models. 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2015. pp. 665–668.
21. Park H, DiMaio F, Baker D. CASP11 refinement experiments with ROSETTA. *Proteins*. 2016;84(Suppl 1):314-322.
22. Park H, DiMaio F, Baker D. The origin of consistent protein structure refinement from structural averaging. *Structure*. 2015;23:1123-1128.
23. Heo L, Arbour CF, Feig M. Driven to near-experimental accuracy by refinement via molecular dynamics simulations. *Proteins*. 2019;87(12):1263-1275. https://doi.org/10.1002/prot.25759.
24. Kamisetty H, Ovchinnikov S, Baker D. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. *Proc Natl Acad Sci U S A*. 2013;110:15674-15679.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Park H, Lee GR, Kim DE, Anishchenko I, Cong Q, Baker D. High-accuracy refinement using Rosetta in CASP13. *Proteins*. 2019;87:1276–1282. https://doi.org/10.1002/prot.25784