

Iterative Molecular Dynamics—Rosetta Protein Structure Refinement Protocol to Improve Model Quality

Steffen Lindert,^{†,‡,*} Jens Meiler,[§] and J. Andrew McCammon^{†,‡,||,⊥}

[†]Department of Pharmacology, University of California San Diego, La Jolla, California 92093, United States

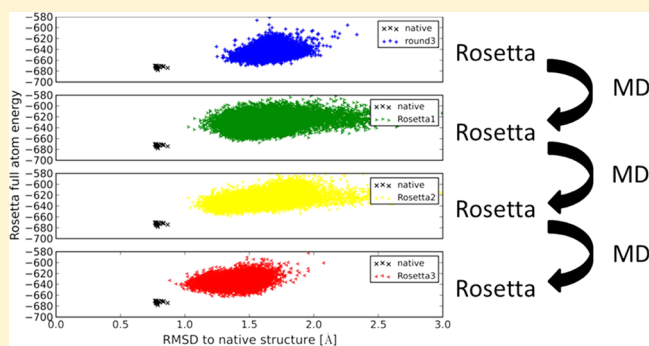
[‡]Center for Theoretical Biological Physics, La Jolla, California 92093, United States

[§]Department of Chemistry and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37212, United States

^{||}Howard Hughes Medical Institute, University of California San Diego, La Jolla, California 92093, United States

[⊥]Department of Chemistry and Biochemistry, NSF Center for Theoretical Biological Physics, National Biomedical Computation Resource, University of California San Diego, La Jolla, California 92093, United States

ABSTRACT: Rosetta is one of the prime tools for high resolution protein structure refinement. While its scoring function can distinguish native-like from non-native-like conformations in many cases, the method is limited by conformational sampling for larger proteins, that is, leaving a local energy minimum in which the search algorithm may get stuck. Here, we test the hypothesis that iteration of Rosetta with an orthogonal sampling and scoring strategy might facilitate exploration of conformational space. Specifically, we run short molecular dynamics (MD) simulations on models created by *de novo* folding of large proteins into cryoEM density maps to enable sampling of conformational space not directly accessible to Rosetta and thus provide an escape route from the conformational traps. We present a combined MD–Rosetta protein structure refinement protocol that can overcome some of these sampling limitations. Two of four benchmark proteins showed incremental improvement through all three rounds of the iterative refinement protocol. Molecular dynamics is most efficient in applying subtle but important rearrangements within secondary structure elements and is thus highly complementary to the Rosetta refinement, which focuses on side chains and loop regions.



INTRODUCTION

The last 20 years have seen an unprecedented increase in computational power available to biomedical research. One area that these vast resources have been focused on is the computational prediction and refinement of protein tertiary structures. To completely understand a protein's function and mode of action, as well as to design small molecule binders to it, knowing the protein's structure is of huge value. Also within the past few decades, experimental methods such as X-ray crystallography, nuclear magnetic resonance (NMR), or electron microscopy (EM) have helped to elucidate a large number of protein structures.¹ Despite the breathtaking achievements of experimental protein structure elucidation, there remain a considerable number of proteins for which no structure is known and also no template for comparative modeling is available. *De novo* protein structure prediction aims to elucidate the structure of such proteins. Predicting the structure of a protein from only its primary amino acid sequence is a formidable challenge, but lately, progress toward this goal has been undeniable: Protein energy landscape theory has been applied to obtain optimal energy functions for protein structure prediction.² Massively distributed computing has been used to simulate protein folding on a scale not accessible before.³ Molecular dynamics

simulations were able to probe folding pathways of very small proteins within simulation times upward of 100 μ s.⁴ The fragment replacement-based method Rosetta has become one of the prime tools to predict and refine the structures of proteins.⁵ Particular success has been demonstrated whenever the folding is aided by sparse experimental restraints. Restraints from NMR,⁶ electron paramagnetic resonance (EPR) spectroscopy⁷ as well as EM⁸ have been shown to considerably improve model quality or extend the scope of the method to much larger systems. The most recent installments of CASP (Critical Assessment of protein Structure Prediction), a biyearly community-wide blind protein structure prediction experiment, have seen Rosetta rank among the best performing programs for *de novo* protein structure prediction (free modeling category).⁹ However, despite all recent improvements, only recently algorithms became available that can sample the topology space of proteins with more than 150 residues¹⁰ without a template structure or experimental restraints.¹¹

In addition to *de novo* protein folding, Rosetta has also been shown to be able to refine small proteins to near atomic

Received: March 29, 2013

Published: July 3, 2013

resolution in favorable cases.¹² Such refinement requires the starting structures to have low root-mean-square distances (RMSDs) to the native conformation already. The Rosetta scoring function is generally able to distinguish native-like from non-native-like conformations. The biggest obstacle to successful refinement is the vast conformational space that has to be sampled to find the near-native energy minimum. If the starting structure differs substantially from the native structure, it is possible that the fragment replacement search algorithm within Rosetta will not be able to sample conformations closer to the native structure. Not only the amplitude of the difference as determined by RMSD is important but also the type of conformation change required to bridge this distance or if such a change can be achieved with fragment replacement and side chain repacking. We observed this scenario in one of our recent works, where we used Rosetta to refine models built into medium resolution cryoEM density maps.¹³ Despite reaching low RMSD values for many of the benchmark proteins, improvement stalled after two rounds of iterative loop rebuilding and refinement in Rosetta. The search algorithm became stuck in “conformational traps” from which no escape was possible using the sampling provided by Rosetta. Here, we investigate the possibility of using molecular dynamics (MD) to overcome some of these conformational traps and sample conformations from which Rosetta can more easily refine the structure. A protocol that goes through multiple iterative rounds of MD and Rosetta refinement is presented.

MATERIAL AND METHODS

System Preparation. Four different proteins were chosen for which low-RMSD models had been built *de novo* with EM-Fold and Rosetta:¹³ 1X91, 2A6B, 1ICX, 1OZ9, two of which are α -helical proteins while the other two are α - β -proteins. The proteins have between 150 and 234 residues. The best scoring model after the third round of Rosetta refinement in ref 13 for each protein was taken as input model for the iterative MD–Rosetta protocol. These systems were then prepared for molecular dynamics simulations. Tleap¹⁴ was used to neutralize the systems by adding Na⁺ or Cl⁻ counterions (1 Na⁺, 15 Na⁺, 7 Na⁺, and 4 Cl⁻ for the 1X91, 2A6B, 1ICX, and 1OZ9 systems respectively) and solvating using a TIP3P water box. The fully solvated systems contained between 25 000 and 32 000 atoms. Minimization using SANDER¹⁴ was carried out in two stages: 1000 steps of minimization of solvent and ions with the protein restrained using a force constant of 500 kcal/mol/Å², followed by a 2500 step minimization of the entire system. A short initial 20 ps MD simulation with weak restraints (10 kcal/mol/Å²) on the protein residues was used to heat the system to a temperature of 300 K.

Molecular Dynamics Simulations. All MD simulations were performed under the NPT ensemble at 300 K using AMBER¹⁴ and the ff99SBildn force field.¹⁵ Periodic boundary conditions were used, along with a nonbonded interaction cutoff of 10 Å. Bonds involving hydrogen atoms were constrained using the SHAKE algorithm,¹⁶ allowing for a time step of 2 fs. Each MD simulation during the iterative MD–Rosetta protocol was run for 2 ns.

Rosetta All Atom Refinement in Density Map. The models identified from the MD simulations were subjected to loop rebuilding and refinement within Rosetta^{5,8a} guided by the cryoEM density map. The Rosetta refinement protocol is identical to the protocol described in ref 13. In summary, regions of the models that agree least with the density map of the protein

are identified (loops_from_density.linuxgccrelease) and rebuilt (loopmodel.linuxgccrelease). Each round performs a full atom relaxation of the entire structure.

Iterative MD–Rosetta Protocol. Three rounds of iterative MD and Rosetta were run for all four benchmark proteins. The starting model for the protocol is the best scoring model after the third round of Rosetta refinement from ref 13. The protocol starts with an MD run, followed by a Rosetta run: MD1–Rosetta1–MD2–Rosetta2–MD3–Rosetta3. The MD runs are short (2 ns). After each MD run two models are picked to transition into the following Rosetta round. For this work we picked the two best models observed during the MD simulation: (a) the model with the overall lowest RMSD with respect to the native protein structure and (b) the model with the lowest RMSD over secondary structure elements (SSEs). For both these models the regions that agree least with the density map are identified and rebuilt using Rosetta, followed by an all atom refinement of the models. The Rosetta models are then sorted by score. The best scoring model is picked as input into the next MD round. Before each MD round counterions and TIP3P waters are added to the model and a short minimization and equilibration is run. All RMSDs reported are over the backbone atoms N, C α , C, and O. All RMSD calculations were done using the BCL::Quality application.^{11b}

RESULTS AND DISCUSSION

Two of the Benchmark Proteins Show Improvement through All Three Iterative MD–Rosetta Rounds. The ability of an iterative MD–Rosetta refinement protocol to minimize the models’ deviation from the native protein structure was tested on four benchmark proteins. The best scoring Rosetta

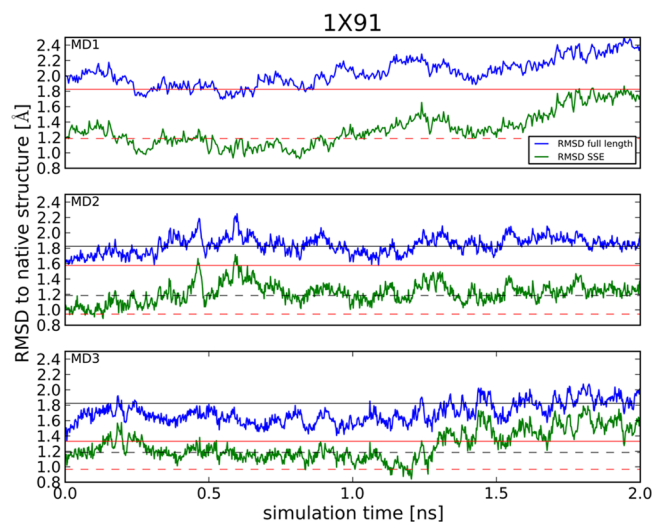


Figure 1. Model quality evolution of 1X91 during the three rounds of MD. The RMSD of the MD structure with respect to the native model is shown for all protein residues (blue) and for residues in secondary structure elements (green). RMSDs of reference models are displayed by vertical lines: the full length RMSD of the starting model (black line), the RMSD over SSEs of the starting model (dashed black line), the full length RMSD of the best scoring model from the previous Rosetta round (red line) and the RMSD over SSEs of the best scoring model from the previous Rosetta round (dashed red line). For the first round of MD, the red and black lines coincide. A successful protocol would be characterized by the blue line breaking through the red line (corresponding to MD sampling lower RMSD models than the best scoring model seen in the last Rosetta round) and the green line breaking through the dashed red line.

Table 1. Quality of the Generated Models for All Four Proteins Throughout Three Rounds of the Iterative Protocol^a

protein	start ^b	MD1 ^c	Rosetta1 ^d	MD2 ^e	Rosetta2 ^f	MD3 ^g	Rosetta3 ^h	best ⁱ
1X91	1.82(1.19)	1.76(0.93)	1.58(0.94)	1.58(0.89)	1.33(0.97)	1.34(0.84)	1.33(1.00)	0.88(0.60)
2A6B	3.17(2.56)	3.12(2.46)	3.08(2.38)	2.92(2.31)	2.86(2.22)	2.86(2.13)	2.87(2.12)	2.56(1.90)
1ICX	2.65(2.14)	2.55(1.93)	2.35(1.92)	2.35(1.67)	2.80(2.67)			2.35(1.67)
1OZ9	2.63(2.23)	2.51(2.18)	4.54(3.38)					2.51(2.18)

^aRMSDs of the models built with respect to native structure over all residues and over all residues in secondary structure elements (in parentheses). All RMSDs shown are in Ångstrom. ^bRMSDs of the starting models. ^cRMSDs of the best models seen in 2 ns of the first round of MD. ^dRMSDs of the top scoring model after the first round of Rosetta refinement. ^eRMSDs of the best models seen in 2 ns of the second round of MD. ^fRMSDs of the top scoring model after the second round of Rosetta refinement. ^gRMSDs of the best models seen in 2 ns of the third round of MD. ^hRMSDs of the top scoring model after the third round of Rosetta refinement. ⁱRMSDs of the best models ever built during the iterative MD–Rosetta refinement protocol.

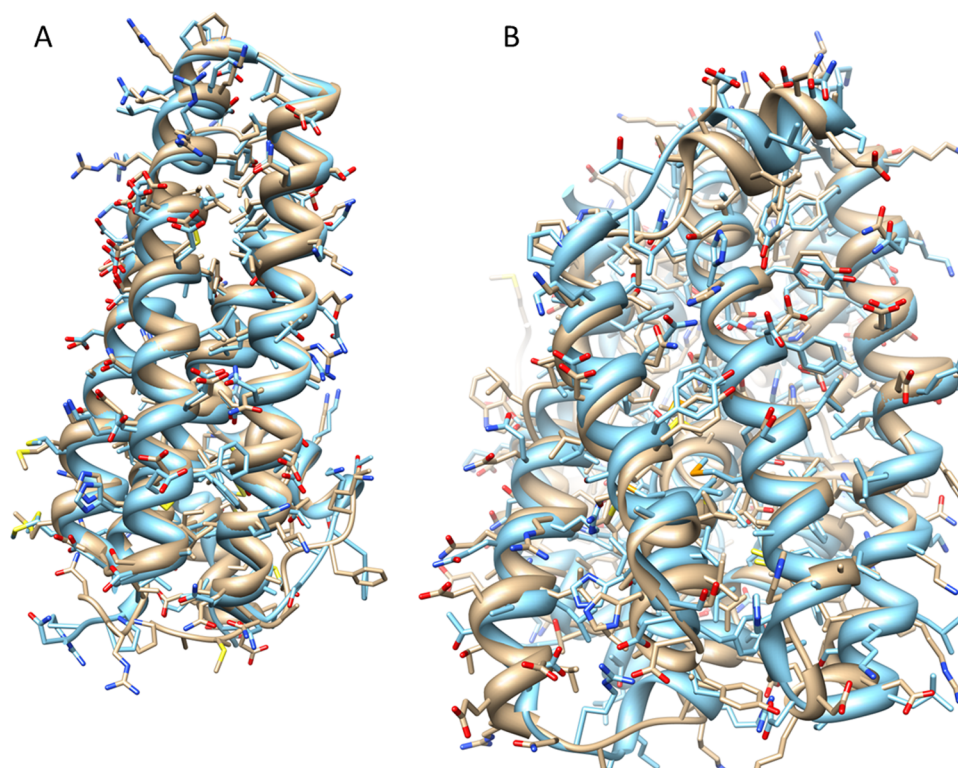


Figure 2. Lowest RMSD models after three rounds of iterative MD/Rosetta refinement for (A) 1X91 and (B) 2A6B. The native structure is shown in turquoise, while the model is shown in gold. The overall structure within secondary structure elements has been recovered in the models. Most side chain conformations within the interface of secondary structure elements have been built correctly.

structure from a previous Rosetta-only density-map-guided refinement¹³ was chosen as input model for all four proteins. Three rounds of iterative molecular dynamics followed by Rosetta refinement were performed. After each MD simulation, the two lowest RMSD models (with respect to RMSD over the entire sequence and RMSD over residues in secondary structure elements) were picked as input into a Rosetta loop rebuilding and refinement run guided by the cryoEM density map. Information from the density maps was not used during the MD simulations. Finally, the best scoring model after Rosetta refinement was used as starting model in the subsequent MD simulation. As an example, Figure 1 shows the evolution of the model quality of 1X91 during the all three rounds of MD. Frequently, the MD starting structures have slightly higher RMSDs than those of the previous-round Rosetta model. For example, the first round MD starting structure has an RMSD of 2.00 Å, compared to the RMSD of the previous round Rosetta model of 1.82 Å. These discrepancies are due to minimization/equilibration before the start of the MD, which consistently

increases the RMSD with respect to the native structure. In all three rounds of MD, models were built that surpass the previous-round best-scoring Rosetta model with respect to RMSD over secondary structure elements (the green line breaking through the dashed red line). Only during the first round of MD, models are built that also have lower full length RMSDs than the previous-round best-scoring Rosetta model. RMSDs as low as 1.34 Å over the entire protein and 0.84 Å over secondary structure elements are sampled during the third round of MD simulations.

Table 1 summarizes the quality of the generated models for all four proteins throughout all three rounds of the iterative protocol. For two of the proteins (1X91, 2A6B), there was improvement to the very end. For example, the RMSD for 2A6B improved from 3.17 Å to 2.87 Å over all protein residues and from 2.56 Å to 2.12 Å measured over residues in secondary structure elements. The improvement for 1X91 is even more considerable. Over the three rounds the RMSD over all residues improved from 1.82 Å to 1.33 Å and from 1.19 Å to

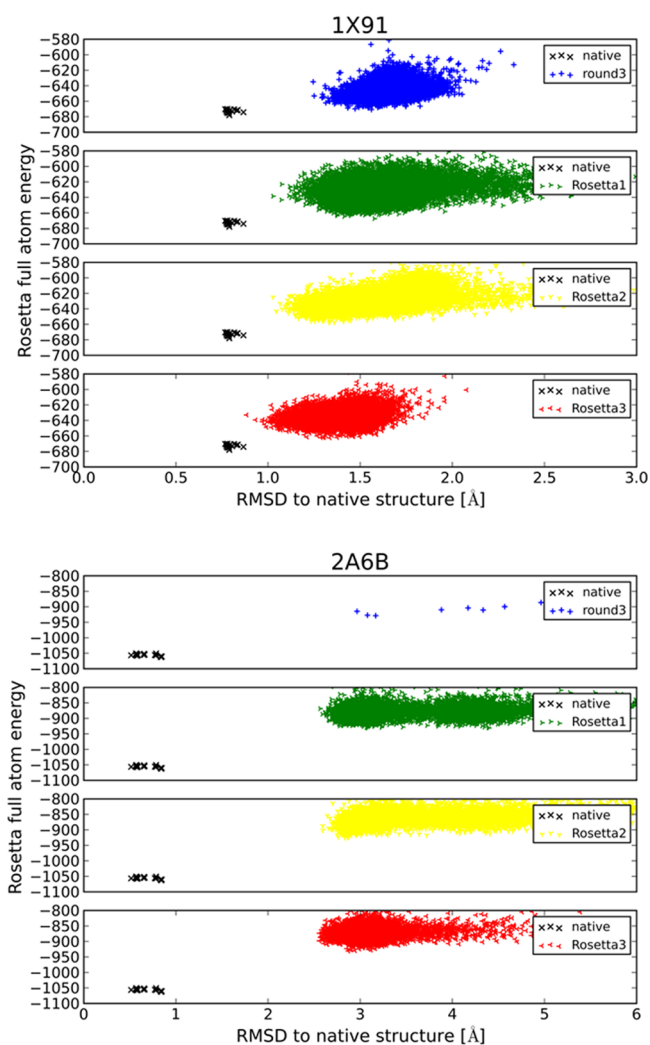


Figure 3. RMSD vs score plots for 1X91 and 2A6B. The first panel (blue, labeled round 3) shows the results of the last round of Rosetta refinement in ref 13. The other three panels show the results for the first (green), second (green), and third (red) Rosetta round of the iterative MD–Rosetta. The native structure, relaxed in the Rosetta force field, is shown in all panels (black).

1.00 Å measured over residues in secondary structure elements. The potential for model quality improvement is even more apparent when looking at the best model in the Rosetta runs of the third round of the protocol. The best 2A6B model built by Rosetta in the third round has an RMSD of 2.56 Å (with 1.90 Å RMSD over SSEs), while the best 1X91 model has an RMSD of 0.88 Å (with 0.60 Å RMSD over SSEs). While these models scored worse than the models reported, this still demonstrates the combined power of the MD–Rosetta protocol to build models of excellent quality. Figure 2 shows the best 1X91 and 2A6B models overlaid with their native structure. The prediction within the protein core is virtually perfect. For the other two proteins improvement stopped either in the first or second round. Interestingly, those proteins represented the α - β -proteins in the benchmark, while the two successful proteins were α -helical proteins. However, even for these two α - β -proteins some initial model improvement was observed. These results demonstrate that an iterative MD–Rosetta refinement protocol can improve model quality in some cases, with considerable improvement seen in selected cases.

Figure 3 shows the RMSD vs score plots of the two most successful proteins, 1X91 and 2A6B. Plots for the last round of Rosetta refinement from ref 13, alongside the three Rosetta rounds of the iterative MD–Rosetta protocol, are shown. The native structure is shown for reference in all plots. The native structure has nonzero RMSD values since it was relaxed in the Rosetta force field before being scored. Both for 1X91 and 2A6B the general distribution of RMSDs shifts to lower values with each round of the iterative MD–Rosetta protocol, suggesting incremental model quality improvement. Not surprisingly this effect is most pronounced with 1X91 due to the overall better model quality.

Molecular Dynamics is Most Efficient in Improving RMSDs over Residues in Secondary Structure Elements.

Molecular dynamics and Rosetta prove to be most efficient in improving the model quality in different parts of the proteins. From the data presented in Figure 1 and Table 1, it can be seen that the Rosetta stages of the protocol are most efficient in improving the RMSD over residues in loop regions. This is not unexpected since the Rosetta protocol includes targeted rebuilding of the loop regions guided by the cryoEM density map. However, encouragingly, MD was best at improving the RMSDs over residues in secondary structure elements. This means that MD has its strength in an area where improvement with Rosetta is more challenging. MD can thus contribute constructively to overcome some of the sampling limitations of Rosetta.

CONCLUSIONS

Here, we presented the results of a novel iterative MD–Rosetta protocol to computationally refine protein structures guided by medium resolution cryoEM density maps. It was shown that a combination of MD and Rosetta can indeed help to overcome some of the “conformational traps” in which Rosetta refinement gets trapped frequently. Molecular dynamics seems particularly helpful for improvement of model quality within secondary structure elements, thus complementing Rosetta whose strength is rebuilding and refining loop regions. The benchmark proteins all had between 150 and 234 residues. Observing models with sub-Angstrom RMSDs in a computational structure refinement protocol for proteins of that size is novel and reason for optimism. While these results are promising and demonstrate that there is the possibility of a combined MD–Rosetta protocol being more powerful than Rosetta alone, much potential for improvement remains. While full success for two out of four test cases is encouraging, future work will focus on ways of improving the success rate of the protocol. Furthermore, in the current implementation, the MD simulations were not guided by the cryoEM density map. Restraining the molecular dynamics runs with the density map may improve model quality during the MD sections of the protocol. Finally, the biggest shortcoming of the current implementation of the protocol is that, while the Rosetta models are picked based solely on score, the models from the MD rounds were picked based on low RMSD to the native structure. Clearly, outside of a benchmark scenario, this is not possible and thus currently serves more as a proof of principle. So, at this point, these calculations rely on the assumption that the best native conformation in an ensemble can be selected. Follow-up studies will focus on identifying these low RMSD models by some other metric, such as, for example, Rosetta score. Addressing these issues is the focus of ongoing research and will be discussed in subsequent publications.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 858-534-2913. Fax: 858-534-4974. E-mail: slindert@ucsd.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health, the National Science Foundation, the Howard Hughes Medical Institute, the National Biomedical Computation Resource, and the NSF Supercomputer Centers. Computational resources were supported, in part, by the National Science Foundation Grant No. PHY-0822283, the Center for Theoretical Biological Physics. Work in the Meiler laboratory is supported through the National Institutes of Health (R01 GM080403, R01 MH090192, R01 GM099842) and National Science Foundation (Career 0742762). S.L. was supported by the American Heart Association and the Center for Theoretical Biological Physics.

■ REFERENCES

- (1) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J. D.; Zardecki, C. The Protein Data Bank. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58*, 899–907.
- (2) Koretke, K. K.; Luthey-Schulten, Z.; Wolynes, P. G. Self-consistently optimized energy functions for protein structure prediction by molecular dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2932–7.
- (3) (a) Adam, L. B.; Ensign, D. L.; Jayachandran, G. K.; Khaliq, S.; Pande, V. S. Folding@home: Lessons from eight years of volunteer distributed computing. *IEEE International Symposium on Parallel and Distributed Processing*; IEEE: New York, 2009; pp 1–8; (b) Shirts, M.; Pande, V. S. COMPUTING: Screen savers of the world unite! *Science* **2000**, *290*, 1903–4. (c) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **2002**, *420*, 102–6.
- (4) (a) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334*, 517–20. (b) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. Protein folding kinetics and thermodynamics from atomistic simulation. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 17845–50.
- (5) Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **1997**, *268*, 209–25.
- (6) (a) Bowers, P. M.; Strauss, C. E.; Baker, D. De novo protein structure determination using sparse NMR data. *J. Biomol. NMR* **2000**, *18*, 311–8. (b) Meiler, J.; Baker, D. Rapid protein fold determination using unassigned NMR data. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15404–9. (c) Meiler, J.; Baker, D. The fumarate sensor DcuS: Progress in rapid protein fold elucidation by combining protein structure prediction methods with NMR spectroscopy. *J. Magn. Reson.* **2005**, *173*, 310–6. (d) Rohl, C. A.; Baker, D. De novo determination of protein backbone structure from residual dipolar couplings using Rosetta. *J. Am. Chem. Soc.* **2002**, *124*, 2723–9.
- (7) (a) Alexander, N.; Bortolus, M.; Al-Mestarihi, A.; McHaourab, H.; Meiler, J. De novo high-resolution protein structure determination from sparse spin-labeling EPR data. *Structure* **2008**, *16*, 181–95. (b) Hanson, S. M.; Dawson, E. S.; Francis, D. J.; Van Eps, N.; Klug, C. S.; Hubbell, W. L.; Meiler, J.; Gurevich, V. V. A model for the solution structure of the rod arrestin tetramer. *Structure* **2008**, *16*, 924–34. (c) Hirst, S. J.; Alexander, N.; McHaourab, H. S.; Meiler, J.

RosettaEPR: An integrated tool for protein structure determination from sparse EPR data. *J. Struct. Biol.* **2011**, *173*, 506–514.

(8) (a) DiMaio, F.; Tyka, M. D.; Baker, M. L.; Chiu, W.; Baker, D. Refinement of protein structures into low-resolution density maps using Rosetta. *J. Mol. Biol.* **2009**, *392*, 181–90. (b) Lindert, S.; Hofmann, T.; Wotzel, N.; Karakas, M.; Stewart, P. L.; Meiler, J. Ab initio protein modeling into CryoEM density maps using EM-Fold. *Biopolymers* **2012**, *97*, 669–77. (c) Lindert, S.; Staritzbichler, R.; Wotzel, N.; Karakas, M.; Stewart, P. L.; Meiler, J. EM-fold: De novo folding of α -helical proteins guided by intermediate-resolution electron microscopy density maps. *Structure* **2009**, *17*, 990–1003. (d) Lindert, S.; Stewart, P. L.; Meiler, J. Hybrid approaches: Applying computational methods in cryo-electron microscopy. *Curr. Opin. Struct. Biol.* **2009**, *19*, 218–25.

(9) Kinch, L.; Yong Shi, S.; Cong, Q.; Cheng, H.; Liao, Y.; Grishin, N. V. CASP9 assessment of free modeling target predictions. *Proteins* **2011**, *79*, 59–73.

(10) Bonneau, R.; Strauss, C. E.; Rohl, C. A.; Chivian, D.; Bradley, P.; Malmstrom, L.; Robertson, T.; Baker, D. De novo prediction of three-dimensional structures for major protein families. *J. Mol. Biol.* **2002**, *322*, 65–78.

(11) (a) Karakas, M.; Woetzel, N.; Staritzbichler, R.; Alexander, N.; Weiner, B. E.; Meiler, J. BCL::Fold—De novo prediction of complex and large protein topologies by assembly of secondary structure elements. *PLoS One* **2012**, *7*, e49240. (b) Woetzel, N.; Karakas, M.; Staritzbichler, R.; Muller, R.; Weiner, B. E.; Meiler, J. BCL::Score—Knowledge based energy potentials for ranking protein models represented by idealized secondary structure elements. *PLoS One* **2012**, *7*, e49242.

(12) Bradley, P.; Misura, K. M.; Baker, D. Toward high-resolution de novo structure prediction for small proteins. *Science* **2005**, *309*, 1868–71.

(13) Lindert, S.; Alexander, N.; Wotzel, N.; Karakas, M.; Stewart, P. L.; Meiler, J. EM-Fold: De novo atomic-detail protein structure determination from medium-resolution density maps. *Structure* **2012**, *20*, 464–78.

(14) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The AMBER biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–88.

(15) (a) 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* **2006**, *65*, 712–25. (b) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the AMBER ff99SB protein force field. *Proteins* **2010**, *78*, 1950–8.

(16) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.