Published in final edited form as:

Proteins. 2014 February ; 82(0 2): 196–207. doi:10.1002/prot.24336.

Physics Based Protein Structure Refinement through Multiple Molecular Dynamics Trajectories and Structure Averaging

Vahid Mirjalili^{1,2}, Keenan Noyes³, and Michael Feig^{2,3,*}

¹Department of Mechanical Engineering Michigan State University East Lansing, MI 48824; USA

²Department of Biochemistry and Molecular Biology Michigan State University East Lansing, MI 48824; USA

³Department of Chemistry Michigan State University East Lansing, MI 48824; USA

Abstract

We used molecular dynamics (MD) simulations for structure refinement of CASP10 targets. Refinement was achieved by selecting structures from the MD-based ensembles followed by structural averaging. The overall performance of this method in CASP10 is described and specific aspects are analyzed in detail to provide insight into key components. In particular, the use of different restraint types, sampling from multiple short simulations vs. a single long simulation, the success of a quality assessment criterion, the application of scoring vs. averaging, and the impact of a final refinement step are discussed in detail.

Keywords

CASP; structure prediction; scoring; protein; quality assessment

INTRODUCTION

Two decades of CASP (Critical Assessment of Techniques for Protein Structure Prediction) have documented significant progress with predicting the structure of proteins from their amino acid sequences.^{1–6} This can be attributed to the development of new techniques but an increasing number of structures in the Protein Data Bank (PDB)⁷ are at least an equally important factor.^{8–11} The most reliable method for protein structure prediction is template based modeling.^{6,9,12,13} The resulting models are often overall correct, but deviate from experimental structures in detail with typical root mean square deviations (RMSD) of 2–6 Å due to intrinsic errors when constructing models based on template structures.^{14,15} Therefore, recent attention has shifted towards the refinement of template-based models to improve their accuracy and generate models that are suitable for biological and pharmaceutical studies.¹⁶

A variety of methods for the refinement of template-based models have been proposed, with the majority involving some combination of sampling and scoring with an emphasis on

*Corresponding author: Michael Feig, 603 Wilson Rd, 218 BCH, East Lansing, MI 48824, 517-432-7439, feig@msu.edu.

physics-based methods, such as molecular dynamics^{17–20}. At the same time, knowledge based methods have also been proposed^{21–24}. The challenges with typical structure refinement protocols are two fold: 1) Sampling has to progress at least in part towards the native structure; and 2) improved structures generated by the sampling method have to be reliably selected. In terms of sampling, different strategies have been explored. The application of restraints on some regions of the protein judged to be of higher quality than other regions often leads to improved sampling of refined structures.¹⁷ Other strategies have involved enhanced sampling methods such as replica exchange MD simulation²⁵ and self-guided Langevin dynamics²⁶ as well as implicit and explicit solvent simulations.^{19,27} A key issue is the quality of the force field which ultimately determines whether refined structures are likely to be generated. In the past, force fields have been optimized specifically for refinement^{28,29}, but improvements in general biomolecular force fields^{30,31} are expected to also impact the ability to carry out successful structure refinement.

While sampling methods are often able to generate refined models, these are typically not found at the end of a given sampling run but instead at intermediate time points. The challenge is then to find those refined structures from the ensemble of structures generated at the sampling stage. The force fields used for sampling, while physically accurate, are often too noisy to reliably identify single structures or small subsets of structures that are most native-like. Instead, a number of statistical potential functions have been used for scoring decoy structures, such as DFIRE,^{32,33} GOAP,³⁴ DOPE,³⁵ and OPUS-PSP³⁵. All of these scoring functions have shown promise in selecting native-like structures from an ensemble, but struggle with consistently selecting refined structures.^{25,36,37}

Despite considerable efforts, effective structure refinement protocols have remained elusive. During the last round of CASP, CASP9, there were only a few groups that were able to outperform a naive prediction of simply resubmitting the initial model given by the organizers to be refined¹⁵. Furthermore, refinement progress was very modest and predictions from the most successful groups lacked consistency as some targets were refined significantly, while others were made worse. Further efforts since CASP9 include very long MD simulation by the D. E. Shaw group¹⁷. In that work, it was clearly shown that without restraints the initial models are likely to drift away from the native structure making refinement largely impossible. When restraints were applied, the sampling of refined structures became possible but the reliable selection of refined structures remained a significant obstacle. Overall, structures selected based on cluster size and/or energetic criteria were improved on average 1% in terms of GDT-TS. A similar level of performance was reported by Zhang et al.,¹⁸ in which they combined knowledge-based information with physics-based MD simulations and applied a fragment-guided method with distance restraints used on global and local structural templates from the PDB. Gront et al.³⁸ recently provided a comprehensive review of refinement methods ranging from physics based to knowledge based methods and concluded that refinement is more challenging when starting structures are already within 2–3 Å from the native structure. In that paper it was also noted that knowledge-based methods may have an advantage because they are parameterized based on experimental structures which are the target of refinement protocols vs. physics-based methods that aim at capturing the protein dynamics at the global minima of the energy landscape.

The distinction between (simulation-generated) protein dynamics and experimentally-obtained structures may become increasingly important as refinement methods aim to reproduce experimental structures at high accuracy. One particular issue is that experimental structures reflect ensemble- and time-averaged conformations rather than instantaneous snapshots. Following this idea, we have recently devised a structure refinement protocol that obtains refined structures from ensemble averages over selected subsets instead of single snapshots³⁶. When this protocol was applied to ensembles from extensive MD-based sampling with the recently updated CHARMM36 force field in combination with explicit water, significant and consistent refinement became possible when tested on CASP8 and CASP9 targets. Here, we describe the blind application of such a refinement protocol during CASP10.

In the following we will first describe the methodology before presenting and discussing results obtained during CASP10 and from subsequent post-analysis.

METHODS

The initial models from CASP10 were preprocessed by adding missing hydrogens using the HBUILD module in CHARMM.³⁹ Protonation states of His residues (if present), were determined by visual inspection. The pKa values of other titratable residues (Glu, Asp, Lys, Arg) were determined using the PROPKA web server^{40,41} followed by visual inspection. All proteins were subsequently solvated in a cubic box of water with at least 9 Å cutoff to the edge of the box. The systems were neutralized by adding Na⁺ or Cl⁻ to balance the net charge of the systems.

The solvated systems were then subjected to molecular dynamics (MD) simulations with periodic boundary conditions. The non-bonded interactions were cut off using the switching method between 8.5 to 10 Å, along with particle-mesh Ewald (PME) summation using a grid spacing of 1 Å for long range electrostatic interactions. The simulations were performed under NPT condition using Langevin dynamics at a temperature of 298 K with a Langevin piston to maintain constant pressure at 1 bar. A time step of 2 fs was used with the SHAKE algorithm to fix bonds involving hydrogen atom. The CHARMM36 force field³¹ was used to model the proteins in conjunction with the TIP3 water model⁴².

All of the simulations used some form of restraints. Two types of restraints were used for almost all of the targets; type 1 consisted of weak restraints (with a force constant of 0.05 kcal/mol/Å²) applied to all C_α atoms; type 2 involved strong restraints (with a force-constant of 1 kcal/mol/Å²) applied to C_α atoms of only the regions that were assumed to be reliable in the starting model. For targets, where CASP organizers indicated which regions to refine, we followed their suggestions. In other cases, we assumed that secondary structure elements are likely to be more reliable and applied restraints to those while leaving loops flexible. Table I shows the regions which were selected for the strong restraints. In some cases, a combination of weak and strong restraints was used by applying strong restraints on selected residues but weak restraints on the rest. Due to the presence of zinc ions in TR754, the first set (see below) was modeled with weak restraints on all C_αs except for the region around the zinc fingers.

The heating and equilibration protocol involved 10 stages: First, simulations were carried out at 50 K using C_{α} restraints according to Table I with a force constant of 2 kcal/mol/Å² and a force constant of 0.5 kcal/mol/Å² for all other C_{α} atoms. The temperatures and force constants were subsequently increased/decreased in 10 ps steps to (100 K, 2/0.5 kcal/mol/Å²), (200 K, 2/0.5 kcal/mol/Å²) (200 K, 1.5/0.2 kcal/mol/Å²), (200 K, 1/0.1 kcal/mol/Å²), (200 K, 1/0.05 kcal/mol/Å²), (250 K, 1/0.05 kcal/mol/Å²), (298 K, 1/0.05 kcal/mol/Å²), (298 K, 1/0.01 kcal/mol/Å²), and (298 K, 1/0 kcal/mol/Å²). The structure at the end of the final stage was used as the starting point for all of the production runs.

Production simulations consisted of two sets, the first set with 20 replicate MD simulations, the second set with 10 replicate simulations. Each simulation was 20 ns long and started from the same starting structure. Sets 1 and 2 (resulting in model 1 and model 2 submissions, respectively) were distinguished by using different restraint types. For most targets, we applied restraint type 2 (strong, selective restraints) if information about which residues should be refined was given by the CASP organizers or if it seemed apparent which regions are likely in need of refinement based on visual inspection. Otherwise we applied restraint type 1 (weak restraints) to set 1. In set 2 we then used the respective other type of restraint (see details in Table I). The choice of which restraint type to apply for set 1 essentially reflected our subjective assessment of what type of restraint was likely to be most successful for refining a given target. We ran multiple short simulations instead of a single long simulation to maximize sampling given limited availability of computer resources.⁴³ The cost of a single 20 ns simulation was on the order of 2,500 core hours on a recent multi-core Intel Xeon CPU. The total cost for one target was about 75,000 core hours (12 days on 256 cores). During post-analysis we also carried out single long simulations (200 ns) using the restraint types listed as set 1 (Table I).

Ensembles of structures were generated from the simulations, containing 500 snapshots for each of the MD trajectories. Structures in each replica ensemble were analyzed in terms of the RMSD from the initial model (iRMSD) and their DFIRE scores.

Following our previously established protocol³⁶ (see Fig. 1), we began by using the correlation coefficient between iRMSD vs. DFIRE as a quality assessment score. Replicas with correlation coefficients greater than 0.4 were discarded from subsequent analyses. From the remaining replicas a subset of structures with combined minimal iRMSD and DFIRE scores³⁶ were then selected. Briefly, the selection criteria is based on normalized iRMSD and DFIRE scores to be within an angle $\theta/2$ around the identity line and outside a circle of radius ρ from the center of the distribution, corresponding to the lower left corner of the scatter plot of iRMSD vs. DFIRE (see Fig. 2).³⁶ The criterion used in CASP experiment was, however, slightly different than what was used for testing the protocol on CASP8 and CASP9 targets because of additional optimization.³⁶ Here, we used $\rho=1$, and $\theta=100^{\circ}$. An average structure was then calculated from the selected subset of structures followed by a structure interpolation. This was accomplished by taking the point on the vector between the corresponding C_{α} atoms in the average and the initial model, with its distance to the initial model to be 0.55 of the vector length. The coordinates of all other atoms were copied from the initial model.

The resulting structure was then solvated again, neutralized by adding appropriate charges, and subjected to 5,000 energy minimization steps followed by 40 ps of MD simulation at 100 K with restraints on all C α s and a force constant of 100 kcal/mol/Å². The purpose of the final MD simulations was to relax structural artifacts due to the averaging procedure and generate structures that are of high stereochemical quality.

The application of the above protocol to simulation sets 1 and 2 resulted in models 1 and 2 submitted to CASP. Models 3-5 were selected from the trajectory snapshots with low DFIRE score but outside the region of the scatter plot used for averaging with the idea that some of these structures may be refined more extensively compared to models 1 and 2.

All of the molecular dynamics simulations were carried out with the NAMD molecular dynamics package in conjunction with the MMTSB tool set⁴⁴ which was also used for analysis along with custom scripts. The protein structures were visualized via the PyMol molecular visualization software.^{45,46}

RESULTS AND DISCUSSION

The MD-based refinement protocol described in the methods section was applied to 27 CASP10 refinement targets. The protocol was not applied to one target, TR722, which was modeled unsuccessfully using an entirely different procedure.

Overall CASP10 Performance

Five models were submitted for each of the targets. The first and second models resulted from ensemble averaging. The other models were selected based on favorable DFIRE scores (see methods section). Table II shows the changes upon refinement, RMSD and GDT-HA, with respect to the initial models provided by CASP for the first submitted model and the best of all five models, respectively. The average change in RMSD for the first models is -0.06 Å, and the average change in GDT-HA is 2.6. More importantly, 20 out of 27 targets improved in RMSD, and 25 targets improved in terms of GDT-HA. This performance is similar to what we found previously when testing the protocol on CASP8 and CASP9 targets.³⁶ When selecting the best out of five structures, the average improvement in terms of RMSD is -0.19 Å, and in terms of GDT-HA is 3.8. Looking at all five models, 24 targets are improved with respect to RMSD, and all targets except for TR754 are improved in GDT-HA. The overall best refinement case has an RMSD value that is improved by almost 1 Å (TR720) and GDT-HA improvements by nearly 10 units (TR723 and TR738). These results suggest that with this refinement protocol it is possible to consistently generate significantly refined structures from the initial template-based models. The only target where the predicted structure was significantly worse than the starting structure was TR754, where the presence of zinc ions presumably complicated the scoring with DFIRE.

Furthermore, Table II lists the best observed structures in terms of RMSD and GDT-HA throughout all 30×20 ns trajectories. Because the best cases were not necessarily picked out for submission, this information provides a theoretical limit of how much refinement could have been achieved with a perfect scoring function. Significant refinement of 1.85 Å in TR720 is observed, as well as several cases with improvements in GDT-HA higher than

10%. On average, improvement in RMSD is 0.55 and 7.0 for GDT-HA. On the other hand looking at the best of five models, we see that the RMSD and GDT-HA are improved by 34% and 69% of the maximum possible improvements, respectively. To our knowledge, no single-structure selection protocol can achieve such a result. Interestingly, there are a few cases where the refined structures are actually better (in terms of GDT-HA) than the best single structure from the trajectories (TR689, TR699, TR708, TR712). This indicates that the averaging procedure used here leads to additional refinement over just selecting the best structure from a given ensemble.

Figure 3 shows four of the best modeled targets, TR662, TR674, TR723, and TR738. While most of the secondary structure elements are fixed, some of the loop regions were refined to conformations intermediate between the initial model and the experimental reference. This suggests that refinement is proceeding towards the right direction but it is clear that further progress is needed to fully reach experimental accuracy.

Model Selection based on Lowest DFIRE and Highest iRMSD

For models 3-5, we selected structures with the lowest DFIRE score and higher iRMSD values. The rationale was that ideally the lowest DFIRE scores would identify the most native structures while higher iRMSD values would allow for more significant refinement but also risks larger deviations away from the native. This is in contrast to the more conservative criterion used for the subset ensemble selection based on small iRMSD values. Table II shows the best structures among the submitted models 3-5. It can be seen that while there are indeed some cases with significantly refined structures (TR720, TR723) and overall average improvement in both RMSD and GDT-HA, there are also many cases where no refinement was achieved. Although the best of the three models 3-5 were analyzed here, the results remain inferior to the single model 1 obtained from ensemble averaging.

Quality Assessment using Correlation between iRMSD and DFIRE

One aspect of our refinement protocol is to estimate whether a given set of samples likely includes significantly refined structures. As discussed in more detail in our previous paper,³⁶ we identified the correlation coefficient between DFIRE and iRMSD as a suitable metric. Correlation coefficients above 0.4 appeared to be correlated with poor refinement performance;³⁶ we applied this criterion here to discard trajectories where this condition was satisfied from further analysis. To further assess the validity of this assumption, we compare in Table III the fractions of improved structures in terms of RMSD and GDT-HA for replicas where the correlation is less than 0.4, with those where the correlation is greater than or equal 0.4. While the results vary greatly for individual targets, there is on average a modest enrichment in terms of both RMSD and GDT-HA, both by about 6%, when discarding samples from replicas where the correlation coefficient is above 0.4. This suggests that the quality assessment procedure used here adds value and it could be used in the future to guide the generation of additional trajectories for cases where refinement appears to be difficult as suggested by many replicas with correlation coefficients above the 0.4 threshold.

Restraint Choice

We used different restraints out of the following three choices: 1) weak restraints on all C_{α} s; 2) strong restraints on selected C_{α} s; 3) a combination of strong restraints on selected regions and weak restraints on the rest. The first choice is most appropriate for cases where there is specific information about which regions require refinement. In the case of CASP, this information was provided for some targets. However, for other targets – and more general applications of structure refinement methods – such information may not be available. Therefore, we evaluated how the choice of restraints affected the results. In order to compare results in a consistent fashion, we used only the first 10 replicas of each set. Some targets use strong, partial restraints for the first set with 20 replicas while for other targets weak, complete restraints were used for the first set (see Table I). Therefore, the total number of replicas that were used for each restraint type does not match among different targets. Furthermore, not all targets were run with strong, partial and weak, complete restraints. Those targets were excluded from the comparison (see Table IV).

Table IV shows the best structures in terms of RMSD (Å) and GDT-HA from all the 10 replicas for a given restraint type. Average values were calculated only for the 22 targets, which have both strong and weak restraint types. The analysis suggests that in terms of best structures that were generated, strong, partial restraints may be roughly equivalent to using weak, complete restraints. Interestingly, the RMSD seems to be improved more with strong, partial restraints (type 2) while GDT-HA scores appear to be improved more with weak, complete restraints (type 1). It is instructive to further separate the analysis into targets where the CASP organizers suggested regions to be refined vs. targets where no such information was given. We find that the degree of refinement was actually greater for the targets where no information was given, indicating that the additional information given during CASP10 was not essential for successful refinement. However, we also note that in the cases where information was available about which regions to refine, the application of partial restraints clearly outperformed weak restraints on all residues. On the other hand, targets where no information was given resulted in significantly better GDT-HA scores with weak, overall restraints than with partial restraints based on secondary structures. This suggests that an optimal strategy may be to use partial restraints if information is available which regions require refinement while applying weak restraints for all residues otherwise. While Table IV focuses on the best structures that are generated, Table V shows the result of refinement when the entire protocol is applied. The overall trends match those of Table IV.

Simulation Time: Single MD vs. Multiple Short MDs

Finally, we compared the sampling efficiency of two sets of simulations in order to assess the benefits of using multiple short simulations vs. a single long MD simulation. During CASP10 we ran multiple short simulations because of time and resource constraints. A single long simulation was run after completion of CASP for over 200 ns for each target continued from the first replica in set 1 using the same restraints as for the short simulations. We then compared the output of the 10×20 ns simulations with the results from the single 200 ns simulations. Note that the restraints can be either of strong, partial type, or weak, complete type. Figure 4 shows the cumulative minimum RMSD and cumulative maximum GDT-HA averaged over all 27 targets for the single 200 ns simulations and 10×20 ns

simulations. The results of the 10 replicas in multiple short simulations are combined at each time slot, so at each time t , the cumulative minimum RMSD and maximum GDT-HA values are calculated from the $t/10$ portion of all of the 10 trajectories. There is an expanding gap between the single and multiple trajectories where the multiple short simulations outperform the long simulation both in terms of RMSD and GDT-HA.

Furthermore, in Table VI we compare the refinement performance by using structures either from a single 200 ns simulation or from multiple 10×20 ns simulations. We tested two selection protocols, using the lowest DFIRE score and subset selection and averaging followed by structure interpolation as described above. Selecting structures with the lowest DFIRE score performs poorly in both cases. However, applying our protocol improves the average RMSD with a similar level of accuracy ($-0.08 \text{ \AA}$), while the average improvement in GDT-HA is actually slightly higher in the case of the single 200 ns simulations (2.9 vs. 2.5 for 10×20 ns simulations). Given the increased sampling of refined structures with multiple short simulations, this is somewhat surprising and warrants further investigation. Assuming that the differences are statistically significant, it may be that much longer simulations generate a broader sampling that when averaged result in a structure that is closer to the experimentally averaged structures.

Final Stage of Refinement

As mentioned in the methods section, structure averaging and interpolation cause some unphysical conformation with bad bonds, angles, dihedrals and steric clashes. Therefore, an extra stage of refinement is required to generate stereochemically acceptable structures. Table VII shows the MolProbity measures for individual targets before (avg) and after final refinement stage (MD). This final refinement had only a small effect on RMSD from native and GDT-HA, as before this stage, the average change in RMSD was $-0.08 \text{ \AA}$ before the final stage and $-0.06 \text{ \AA}$ after the final stage while the average GDT-HA did not change during the final refinement stage.

Quality of refined models

The MolProbity score shown in Table VII indicates that the final refined structures are of high stereochemical quality. However, a remaining concern is that the averaging procedure applied here led to structure compression, which could improve GDT scores by itself, rather than actual refinement. In order to test this possibility we analyzed two quantities: C_{α} - C_{α} distances for subsequent residues i and $i+1$ and radii of gyration based on C_{α} atoms. Compression would result in reduced values for either of these measures. Table VIII compares the average values for all of the targets between the initial models provided by CASP, the structures sampled with MD, the averaged models, the final models after the final stage of refinement, and the reported experimental structures. It can be seen that MD sampling leads to increases in both C_{α} - C_{α} distances and radii of gyration from the initial models due to thermalization. Averaging then reduces both quantities but only to about the same average values as the initial values. The final refinement stage then again increases both C_{α} - C_{α} distances and radii of gyration to values that exceed the initial model while not affecting the GDT-HA scores. The C_{α} - C_{α} distances in the final models are quite close to the experimental structures while the radii of gyration are significantly larger than the

experimental structures. This suggests that after the initial averaging there might be slight structure compression but this is largely reversed during the final stage of refinement.

CONCLUSION

We applied a recently established molecular dynamics-based structure refinement protocol to CASP10 targets. Overall, we were able to reliably refine most of the targets both in terms of RMSD and GDT-HA relative to the experimental structures. The key components of our protocol are the use of restraints during MD simulations, the selection of trajectories based on a quality assessment score, and the generation of refined structures following structure subset selection and averaging.

We compared the results of using strong restraints on selected residues vs. weak restraints on all C_{α} s, and concluded that using strong restraints on selected C_{α} s leads to improved RMSD values, while weak restraints can improve GDT-HA measures better.

Another question in MD based refinement is the time scale of the simulation, and in this study we compared the sampling in multiple short MD simulations vs. one single long simulation, and we observed that multiple short MD simulations may have a higher sampling efficiency.

Although our protocol outperformed other refinement methods, overall, the improvements in RMSD and GDT-HA measures in refining protein structures are still relatively minor and it is clear that further progress is needed. It appears that the selection and averaging method used here actually performed very well, realizing about half in terms of refinement progress of what could have been achieved if the best MD-generated structure would have been selected for each target. This suggests that further improvements in sampling more native like structures is likely the key to further refinement progress. One possibility is to take advantage of the consistent and reliable refinement obtained here and extrapolate along the initial direction. Another direction is the further improvement of structure selection methods since for many targets significantly more refined structures were generated than what we submitted as predictions.

REFERENCES

1. Mariani V, Kiefer F, Schmidt T, Haas J, Schwede T. Assessment of template based protein structure predictions in CASP9. *Proteins*. 2011; 79:37–58. [PubMed: 22002823]
2. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*. 2010; 5:725–738. [PubMed: 20360767]
3. Kryzhanovych A, Venclovas C, Fidelis K, Moult J. Progress over the first decade of CASP experiments. *Proteins*. 2005; 61:225–236. [PubMed: 16187365]
4. Moult J. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr Opin Struct Biol*. 2005; 15:285–289. [PubMed: 15939584]
5. Moult J, Fidelis K, Rost B, Hubbard T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP) - Round 6. *Proteins*. 2005; 61:3–7. [PubMed: 16187341]
6. Zhang Y, Skolnick J. The protein structure prediction problem could be solved using the current PDB library. *Proc Natl Acad Sci USA*. 2005; 102:1029–1034. [PubMed: 15653774]
7. Henrick K, Feng ZK, Bluhm WF, Dimitropoulos D, Doreleijers JF, Dutta S, Flippen-Anderson JL, Ionides J, Kamada C, Krissinel E, Lawson CL, Markley JL, Nakamura H, Newman R, Shimizu Y,

- Swaminathan J, Velankar S, Ory J, Ulrich EL, Vranken W, Westbrook J, Yamashita R, Yang H, Young J, Yousufuddin M, Berman HM. Remediation of the protein data bank archive. *Nucleic Acids Res.* 2008; 36:D426–D433. [PubMed: 18073189]
8. Bazzoli A, Tettamanzi AGB, Zhang Y. Computational Protein Design and Large-Scale Assessment by I-TASSER Structure Assembly Simulations. *J Mol Biol.* 2011; 407:764–776. [PubMed: 21329699]
 9. Fiser, A.; Sali, A. MODELLER: Generation and refinement of homology-based protein structure models. In: Carter, CW.; Sweet, RM., editors. *Macromolecular Crystallography, Pt D. Volume 374, Methods in Enzymology.* Elsevier Academic Press Inc; San Diego: 2003. p. 461
 10. Xu D, Zhang Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins.* 2012; 80:1715–1735. [PubMed: 22411565]
 11. Zhang Y, Skolnick J. Tertiary structure predictions on a comprehensive benchmark of medium to large size proteins. *Biophys J.* 2004; 87:2647–2655. [PubMed: 15454459]
 12. Ko J, Park H, Seok C. GalaxyTBM: template-based modeling by building a reliable core and refining unreliable local regions. *BMC Bioinformatics.* 2012; 13
 13. Qu XT, Swanson R, Day R, Tsai J. A Guide to Template Based Structure Prediction. *Curr Protein Pept Sci.* 2009; 10:270–285. [PubMed: 19519455]
 14. MacCallum JL, Hua L, Schnieders MJ, Pande VS, Jacobson MP, Dill KA. Assessment of the protein-structure refinement category in CASP8. *Proteins.* 2009; 77:66–80. [PubMed: 19714776]
 15. MacCallum JL, Pérez A, Schnieders MJ, Hua L, Jacobson MP, Dill KA. Assessment of protein structure refinement in CASP9. *Proteins.* 2011; 79:74–90. [PubMed: 22069034]
 16. Zhang Y. Protein structure prediction: when is it useful? *Curr Opin Struct Biol.* 2009; 19:145–155. [PubMed: 19327982]
 17. Raval A, Piana S, Eastwood MP, Dror RO, Shaw DE. Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. *Proteins.* 2012; 80:2071–2079. [PubMed: 22513870]
 18. Zhang J, Liang Y, Zhang Y. Atomic-Level Protein Structure Refinement Using Fragment-Guided Molecular Dynamics Conformation Sampling. *Structure.* 2011; 19:1784–1795. [PubMed: 22153501]
 19. Chen JH, Brooks CL. Can molecular dynamics simulations provide high-resolution refinement of protein structure? *Proteins.* 2007; 67:922–930. [PubMed: 17373704]
 20. Lee MR, Tsai J, Baker D, Kollman PA. Molecular dynamics in the endgame of protein structure prediction. *J Mol Biol.* 2001; 313:417–430. [PubMed: 11800566]
 21. Chopra G, Summa CM, Levitt M. Solvent dramatically affects protein structure refinement. *Proc Natl Acad Sci USA.* 2008; 105:20239–20244. [PubMed: 19073921]
 22. Bhattacharya D, Cheng JL. 3Drefine: Consistent protein structure refinement by optimizing hydrogen bonding network and atomic-level energy minimization. *Proteins.* 2013; 81:119–131. [PubMed: 22927229]
 23. Rodrigues J, Levitt M, Chopra G. KoBaMIN: a knowledge-based minimization web server for protein structure refinement. *Nucleic Acids Res.* 2012; 40:W323–W328. [PubMed: 22564897]
 24. Zhu J, Xie L, Honig B. Structural refinement of protein segments containing secondary structure elements: Local sampling, knowledge-based potentials, and clustering. *Proteins.* 2006; 65:463–479. [PubMed: 16927337]
 25. Zhu J, Fan H, Periole X, Honig B, Mark AE. Refining homology models by combining replica-exchange molecular dynamics and statistical potentials. *Proteins.* 2008; 72:1171–1188. [PubMed: 18338384]
 26. Olson MA, Chaudhury S, Lee MS. Comparison Between Self-Guided Langevin Dynamics and Molecular Dynamics Simulations for Structure Refinement of Protein Loop Conformations. *J Comput Chem.* 2011; 32:3014–3022. [PubMed: 21793008]
 27. Olson MA, Lee MS. Structure refinement of protein model decoys requires accurate side-chain placement. *Proteins.* 2013; 81:469–478. [PubMed: 23070940]
 28. Wroblewska L, Jagielska A, Skolnick J. Development of a physics-based force field for the scoring and refinement of protein models. *Biophys J.* 2008; 94:3227–3240. [PubMed: 18178653]

29. Jagielska A, Wroblewska L, Skolnick J. Protein model refinement using an optimized physics-based all-atom force field. *Proc Natl Acad Sci USA*. 2008; 105:8268–8273. [PubMed: 18550813]
30. Ponder JW, Case DA. Force fields for protein simulations. *Protein Sim*. 2003; 66:27.
31. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles. *J Chem Theory Comput*. 2012; 8:3257–3273. [PubMed: 23341755]
32. Yang YD, Zhou YQ. Ab initio folding of terminal segments with secondary structures reveals the fine difference between two closely related all-atom statistical energy functions. *Protein Sci*. 2008; 17:1212–1219. [PubMed: 18469178]
33. Yang YD, Zhou YQ. Specific interactions for ab initio folding of protein terminal regions with secondary structures. *Proteins*. 2008; 72:793–803. [PubMed: 18260109]
34. Zhou HY, Skolnick J. GOAP: A Generalized Orientation-Dependent, All-Atom Statistical Potential for Protein Structure Prediction. *Biophys J*. 2011; 101:2043–2052. [PubMed: 22004759]
35. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci*. 2006; 15:2507–2524. [PubMed: 17075131]
36. Mirjalili V, Feig M. Protein Structure Refinement through Structure Selection and Averaging from Molecular Dynamics Ensembles. *J Chem Theory Comput*. 2012; 9:1294–1303. [PubMed: 23526422]
37. Fan H, Periole X, Mark AE. Mimicking the action of folding chaperones by Hamiltonian replica-exchange molecular dynamics simulations: Application in the refinement of de novo models. *Proteins*. 2012; 80:1744–1754. [PubMed: 22411697]
38. Gront D, Kmiecik S, Blaszczyk M, Ekonomiuk D, Kolinski A. Optimization of protein models. *Wiley Interdiscip Rev-Comput Mol Sci*. 2012; 2:479–493.
39. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoseck M, Im W, Kucsera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: The Biomolecular Simulation Program. *J Comput Chem*. 2009; 30:1545–1614. [PubMed: 19444816]
40. Olsson MHM, Sndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput*. 2011; 7:525–537.
41. Sndergaard CR, Olsson MHM, Rostkowski M, Jensen JH. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J Chem Theory Comput*. 2011; 7:2284–2295.
42. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys*. 1983; 79:926–935.
43. Caves LSD, Evanseck JD, Karplus M. Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Sci*. 1998; 7:649–666. [PubMed: 9541397]
44. Feig M, Karanicolas J, Brooks CL. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J Mol Graph*. 2004; 22:377–395.
45. DeLano WL. PyMOL molecular viewer: Updates and refinements. *Abstr Pap Am Chem Soc*. 2009; 238
46. DeLano WL, Lam JW. PyMOL: A communications tool for computational models. *Abstr Pap Am Chem Soc*. 2005; 230:U1371–U1372.

Figure 1.
Flowchart of the refinement protocol, from simulation to model selection

Figure 2. Subset selection based on normalized iRMSD and DFIRE scores from the replicas in set 1 with $\text{corr}(\text{iRMSD}, \text{DFIRE}) < 0.4$ for TR674. The structures shown in the lower left corner of the scatter plot as green triangles were selected for averaging.

Figure 3. Initial model (blue), refined (green) and native (magenta) for (a) TR662, (b) TR723, (c) TR738 and (d) TR674

Figure 4. Sampling efficiency toward native, showing the cumulative minimum RMSD (top) and cumulative maximum GDT-HA (bottom), comparing the single 200 ns MD simulations (red) vs. 10x20 ns MD simulations (green) averaged over 27 CASP10 targets.

Table I

Type of restraints applied on C_α atoms for the two simulation sets; strong (type 2, 1 kCal/mol/Å²), weak (type 1, 0.05 kCal/mol/Å²) and a combination of both. The strong force constants are only applied to the selected regions.

Target	Set 1: 20×20 ns	Set 2: 10×20 ns	Strongly Restrained Regions
TR644	Combined	Strong	53:56, 61:66, 71:75, 84:87, 115:119, 129:132, 140:142, 151:153
TR655	Strong	Weak	21:50, 65:90, 94:141, 164:180
TR661	Weak	-	-
TR662	Weak	Strong	5:16, 38:50, 66:79
TR663	Combined	Strong	79:140, 182:204
TR671	Strong	Weak	38:54, 77:80, 85:89, 96, 108:125
TR674	Weak	Strong	284:288, 300:305, 310:312, 318:320, 333:335
TR679	Strong	Weak	1:24, 46:145, 157:186, 198:223
TR681	Strong	Weak	21:40, 51:57, 65:87, 102:118, 128:144, 153:157, 171:172, 200:224
TR688	Combined	Strong	46:54, 67:76, 89:98, 113:122, 137:145, 160:167, 182:190
TR689	Strong	Weak	14:21, 33:39, 48:59, 64:72, 81:89, 116:118, 143:147, 156:160, 165:169, 181:190, 197:207, 211:218, 226:234
TR696	Weak	Strong	18:22, 27:35, 41:43, 50:51, 58:60, 69:73, 93:96, 101:105
TR698	Strong	Weak	1:16, 36:89, 101:119
TR699	Weak	Strong	8:11, 37:45, 53:55, 86:94, 103:135, 161, 205:206, 219:234
TR704	Weak	Strong	25:32, 40:42, 50:55, 61:64, 81:83, 100:102, 113:120, 128:132, 141:149, 161:166, 188:189, 193:200, 204:209, 217:226, 236:237, 242:246
TR705	Weak	Strong	40:42, 65:67, 82:85, 90:91, 110:114, 119:126
TR708	Weak	Strong	24:27, 45:60, 66:70, 99:101, 113:119, 125:129, 136:152, 172:183
TR710	Weak	Strong	27:50, 67:83, 100:117, 135:152, 168:185, 201:220
TR712	Strong	Weak	38:79, 90:115, 130:140, 156:223
TR720	Weak	Strong	27:29, 53:71, 80:86, 91:103, 108:114, 127:139, 144:147, 154:157, 162:176
TR723	Strong	Weak	39:73, 99:112
TR724	Weak	Strong	135:136, 152:157, 198:202, 210:216, 232:238
TR738	Strong	Weak	1:38, 88:90, 103:249
TR747	Weak	Strong	24:26, 46:49, 55:59, 68:71, 80:83, 92:94, 103:109, 114:121
TR750	Weak	Strong	1:6, 28:29, 48:57, 64:66, 78:93, 98:100, 121:137, 168:182
TR752	Strong	Combined	1:40, 51:99, 111:124, 129:156
TR754	Weak	Strong	25, 33, 63:76 (weak restraints are not applied to the zinc fingers)

Table II

Refinement results showing the best observed structures in the trajectories, the first submitted model, the best of five submitted models, and best model among models 3–5.

Target	Best in 30×20ns trajectories		First submitted model		Best of five models		Best of models 3–5	
	RMSD (Å)	GDT-HA	RMSD (Å)	GDT-HA	RMSD (Å)	GDT-HA	RMSD (Å)	GDT-HA
TR644	−0.94	11.0	−0.03	2.8	−0.55	5.3	0.04	−1.4
TR655	−0.26	2.4	0.04	0.3	0.00	0.3	0.20	−0.9
TR661	−0.25	6.1	−0.03	1.9	−0.03	1.9	0.17	−2.2
TR662	−0.54	13.0	−0.20	5.3	−0.25	6.7	−0.25	6.7
TR663	−0.41	4.6	−0.12	2.6	−0.15	3.6	−0.15	3.3
TR671	−0.62	5.4	−0.01	0.6	−0.25	2.8	0.09	2.8
TR674	−0.78	7.0	0.00	4.9	−0.06	4.9	−0.06	−3.4
TR679	−0.55	4.8	0.01	0.6	−0.03	3.3	0.12	1.0
TR681	−0.13	5.2	−0.04	1.1	−0.15	5.4	−0.15	5.4
TR688	−0.14	6.9	0.01	1.5	−0.02	2.2	0.02	−0.1
TR689	−0.25	2.3	−0.10	3.5	−0.13	4.9	−0.12	2.3
TR696	−0.81	11.0	−0.13	3.5	−0.33	4.8	−0.33	4.8
TR698	−0.32	3.6	−0.02	−0.4	−0.02	−0.4	0.09	−0.6
TR699	−0.33	4.1	−0.09	4.6	−0.09	4.6	−0.07	3.7
TR704	−0.57	7.8	−0.17	3.9	−0.23	5.6	−0.23	5.6
TR705	−0.51	10.7	−0.14	6.0	−0.24	7.3	−0.24	7.3
TR708	−0.84	1.3	0.09	2.7	0.09	2.9	0.10	−2.4
TR710	−0.20	11.1	−0.04	4.3	−0.06	4.3	−0.04	2.1
TR712	−0.54	3.1	−0.08	3.4	−0.14	5.0	−0.14	5.0
TR720	−1.85	5.1	0.02	2.7	−0.99	3.2	−0.99	1.1
TR723	−0.71	11.8	−0.13	6.5	−0.39	9.7	−0.39	9.7
TR724	−1.49	8.5	−0.01	2.6	−0.48	3.7	--	--
TR738	−0.37	10.6	−0.20	6.0	−0.30	9.5	−0.09	3.5
TR747	−0.44	13.1	−0.10	0.8	−0.10	0.8	0.10	−0.6
TR750	−0.43	11.8	−0.16	4.8	−0.16	4.8	−0.04	2.5
TR752	−0.30	3.1	−0.12	1.4	−0.12	1.4	−0.05	−0.7
TR754	−0.35	2.6	0.09	−6.3	0.09	−6.3	0.09	−7.4
Avg.	−0.55	7.0	−0.06	2.6	−0.19	3.8	−0.09	1.8

Table III

Fraction of improved trajectory frames in replicas classified by correlation between iRMSD and DFIRE score; Fraction of improved frames in trajectories with correlation <0.4 that are larger by 10% than fractions with correlation 0.4 are highlighted.

Target	Fraction (%) of traj. frames improved in RMSD		# replicas with Corr 0.4	Fraction (%) of traj. frames improved in GDT-HA	
	Corr < 0.4	Corr 0.4		Corr < 0.4	Corr 0.4
TR644	46.5	N.A.	0	14.1	N.A.
TR655	4.4	3.7	16	2.3	0.5
TR661	12.6	N.A.	0	8.5	N.A.
TR662	70.4	8.4	6	76.1	33.8
TR663	28.0	22.7	15	49.0	41.0
TR671	18.7	5.4	12	13.2	3.0
TR674	23.1	5.2	6	9.3	1.3
TR679	15.3	35.9	3	16.7	30.7
TR681	0.4	0.0	2	2.2	0.1
TR688	1.7	0.5	3	19.8	17.3
TR689	33.4	10.5	7	0.7	0.3
TR696	55.1	33.8	1	33.6	15.6
TR698	60.3	59.7	4	13.6	13.2
TR699	29.2	26.8	1	1.6	0.2
TR704	53.2	NA	0	37.9	N.A.
TR705	33.7	3.0	4	55.3	20.5
TR708	3.4	9.8	6	0.2	0.1
TR710	26.3	NA	0	79.4	N.A.
TR712	28.4	20.0	1	5.5	0.6
TR720	27.8	39.2	8	15.6	32.1
TR723	57.4	68.7	3	55.7	47.0
TR724	37.6	99.6	1	22.9	60.8
TR738	85.7	99.9	2	78.8	99.8
TR747	38.2	8.0	5	20.8	5.7
TR750	54.3	22.4	6	76.6	36.2
TR752	35.7	13.9	6	6.1	0.7
TR754	1.0	0.5	22	0.1	0.1
Avg.	32.7	26.0		26.5	20.0

Table IV

Refinement results of different restraints for the best observed structure in terms of RMSD (Å) and GDT-HA; comparing strong (1 kCal/mol/Å²) restraint on selected residues vs. weak (0.05 kCal/mol/Å²) restraint on all C_αs, and a combination of both. Cases associated with * indicate targets that had suggestions from CASP on which regions need refinement.

Target	Strong Restraint		Weak on all C _α		Strong + weak	
	RMSD (Å)	GDT-HA	RMSD (Å)	GDT-HA	RMSD (Å)	GDT-HA
TR644	-0.94	5.1	--	--	-0.49	4.4
TR655 *	-0.26	1.3	-0.14	-0.3		
TR661	--	--	-0.22	3.9		
TR662	-0.22	5.0	-0.54	10.7		
TR663 *	-0.22	3.6	--	--	-0.34	3.6
TR671	-0.21	1.7	-0.62	4.3		
TR674	-0.78	3.8	-0.40	4.7		
TR679 *	-0.55	2.9	-0.23	2.4		
TR681	-0.02	1.6	-0.13	3.5		
TR688	-0.09	4.1	--	--	-0.13	5.0
TR689	-0.15	-0.2	-0.24	0.9		
TR696	-0.81	6.0	-0.50	7.8		
TR698 *	-0.32	2.3	-0.16	-1.5		
TR699	-0.32	1.2	-0.33	1.7		
TR704	-0.37	4.5	-0.57	6.6		
TR705	-0.48	8.3	-0.45	5.7		
TR708	-0.84	0.0	-0.43	-0.1		
TR710	-0.16	6.7	-0.19	8.8		
TR712*	-0.48	2.3	-0.23	-1.9		
TR720	-1.85	2.8	-0.21	2.8		
TR723	-0.71	9.7	-0.35	7.8		
TR724	-1.49	7.4	-0.40	3.1		
TR738 *	-0.34	8.5	-0.37	6.3		
TR747	-0.19	3.3	-0.44	9.2		
TR750	-0.26	4.5	-0.42	10.2		
TR752 *	-0.30	2.5	--	--	-0.24	1.0
TR754	-0.21	-0.7	-0.35	2.57		
<i>Avg. of common rows:</i>	-0.50	3.8	-0.35	4.3	Not enough data	Not enough data
<i>Avg. * (CASP sugg.)</i>	-0.35	3.3	-0.23	1		
<i>Avg. (no sugg.)</i>	-0.53	3.9	-0.38	5.2		

Table V

Refinement results of different restraints using the established structure generation protocol; comparing strong (1 kCal/mol/Å²) on selected residues vs. weak (0.05 kCal/mol/Å²) restraint on all C_α, and a combination of both

Target	Strong restraint		Weak on all C _α		Strong + weak	
	RMSD (Å)	GDT-HA	RMSD (Å)	GDT-HA	RMSD (Å)	GDT-HA
TR644	− 0.54	3.0	--	--	− 0.36	3.2
TR655 *	− 0.05	0.0	− 0.01	−1.7		
TR661	--	--	− 0.03	2.3		
TR662	− 0.03	1.3	− 0.20	4.7		
TR663 *	0.54	3.1	--	--	− 0.11	2.8
TR671	0.02	1.1	− 0.05	1.4		
TR674	− 0.11	3.4	0.00	5.3		
TR679 *	0.03	0.1	− 0.05	3.0		
TR681	− 0.06	−0.1	− 0.04	1.3		
TR688	− 0.02	2.3	--	--	0.01	1.4
TR689	− 0.11	3.6	− 0.14	4.4		
TR696	− 0.24	2.8	− 0.17	4.0		
TR698 *	− 0.01	−0.4	0.03	−1.3		
TR699	− 0.14	4.0	− 0.04	4.2		
TR704	− 0.10	2.3	− 0.18	3.9		
TR705	− 0.15	5.2	− 0.14	4.4		
TR708	0.11	2.8	0.08	2.2		
TR710	− 0.05	2.6	− 0.05	4.4		
TR712*	− 0.08	3.5	− 0.07	4.3		
TR720	− 0.54	1.1	0.01	2.4		
TR723	− 0.22	4.6	− 0.13	6.3		
TR724	− 0.51	3.7	− 0.02	2.8		
TR738 *	− 0.20	6.1	− 0.23	7.5		
TR747	− 0.08	0.3	− 0.11	0.8		
TR750	− 0.11	2.9	− 0.16	4.0		
TR752 *	− 0.13	1.7	--	--	− 0.10	1.0
TR754	0.16	−7.4	0.07	−5.5		
<i>Avg. of common rows:</i>	− 0.11	2.0	− 0.07	2.9	<i>Not enough data</i>	<i>Not enough data</i>
<i>Avg. * (CASPsugg.)</i>	0.01	2.0	− 0.07	2.4		
<i>Avg. (no sugg.)</i>	− 0.14	2.1	− 0.08	3.5		

Table VI

Comparison of refinement results between 10×20 ns simulations and single 200 ns simulations using the same restraint conditions.

	Corr. iRMSD vs. DFIRE		Best Structure in Trajectory				Lowest DFIRE score				Subset Average + Structure Interpolation			
			RMSD		GDT-HA		RMSD		GDT-HA		RMSD		GDT-HA	
	200	10×20	200	10×20	200	10×20	200	10×20	200	10×20	200	10×20	200	10×20
TR644	0.14	-0.09	-0.37	-0.49	5.8	6.2	-0.09	-0.18	2.6	-0.5	-0.08	-0.36	3.7	3.2
TR655	0.50	0.52	-0.15	-0.26	1.0	2.4	0.61	0.25	-5.3	-3.7	-0.01	-0.05	0.3	0.0
TR661	-0.24	-0.41	-0.15	-0.22	2.3	6.1	0.31	0.31	-10.3	-4.6	0.07	-0.03	0.3	2.3
TR662	-0.16	-0.14	-0.48	-0.54	11.7	13.0	-0.08	-0.23	7.0	7.7	-0.22	-0.20	5.3	4.7
TR663	0.24	0.23	-0.37	-0.34	3.1	4.4	-0.12	0.07	0.8	3.0	-0.15	-0.11	2.5	2.8
TR671	0.78	0.15	-0.12	-0.21	0.3	2.8	0.17	-0.12	-4.0	0.6	0.06	0.02	-2.3	1.1
TR674	-0.01	0.42	-0.39	-0.40	7.0	6.1	-0.14	-0.14	-4.9	-1.5	-0.02	0.00	4.2	5.3
TR679	-0.17	0.17	-0.23	-0.55	2.3	3.7	0.38	0.28	1.4	-0.6	0.06	0.03	0.3	0.1
TR681	-0.01	0.30	0.02	-0.03	2.2	2.3	0.32	0.24	0.3	-2.3	-0.06	-0.06	2.5	-0.1
TR688	-0.31	0.27	-0.13	-0.13	3.8	6.9	0.30	0.15	-2.2	-0.8	-0.01	0.01	1.6	1.4
TR689	-0.13	0.13	-0.11	-0.15	0.6	1.2	0.19	0.07	-8.3	-2.8	-0.17	-0.11	3.2	3.6
TR696	0.24	0.14	-0.38	-0.50	9.0	11.0	-0.06	0.05	5.3	1.0	-0.15	-0.17	5.3	4.0
TR698	0.34	0.33	-0.32	-0.32	3.6	2.7	-0.01	0.09	0.2	-0.6	-0.03	-0.01	0.0	-0.4
TR699	-0.44	0.13	-0.29	-0.33	3.8	4.1	0.15	0.10	-0.5	-3.5	-0.12	-0.04	4.3	4.2
TR704	0.04	-0.08	-0.45	-0.55	11.8	9.3	0.17	-0.09	2.3	-1.9	-0.21	-0.18	4.9	3.9
TR705	0.09	0.15	-0.37	-0.48	10.9	10.7	-0.02	-0.06	4.4	4.2	-0.09	-0.14	5.7	4.4
TR708	0.35	0.23	-0.32	-0.36	0.0	1.3	0.24	0.17	-4.0	-4.5	0.07	0.08	1.3	2.2
TR710	0.14	0.11	-0.19	-0.19	11.1	11.1	-0.01	-0.09	3.6	7.6	-0.07	-0.05	5.3	4.4
TR712	-0.52	-0.16	-0.53	-0.58	2.5	3.0	-0.32	-0.02	-1.4	-1.4	-0.05	-0.08	3.1	3.5
TR720	-0.25	0.21	-0.13	-0.21	2.4	3.5	0.16	-0.05	-5.3	1.9	0.11	0.01	1.5	2.4
TR723	0.00	0.24	-0.56	-0.35	6.3	9.5	-0.04	-0.04	0.6	2.5	-0.23	-0.22	4.6	4.6
TR724	0.27	0.16	-0.32	-0.40	5.7	4.1	0.06	0.11	0.7	-1.5	-0.06	-0.02	4.4	2.8
TR738	0.16	0.41	-0.38	-0.34	11.5	10.1	-0.24	-0.09	7.1	3.5	-0.25	-0.20	6.4	6.1
TR747	-0.22	-0.01	-0.36	-0.33	10.5	10.7	0.00	-0.04	3.6	-2.6	-0.12	-0.11	2.5	0.8
TR750	0.12	0.34	-0.40	-0.42	11.1	11.8	-0.17	0.20	4.7	-1.0	-0.23	-0.16	5.9	4.0

	Corr. iRMSD vs. DFIRE		Best Structure in Trajectory				Lowest DFIRE score				Subset Average + Structure Interpolation			
			RMSD		GDT-HA		RMSD		GDT-HA		RMSD		GDT-HA	
			200	10x20	200	10x20	200	10x20	200	10x20	200	10x20	200	10x20
TR752	0.49	0.52	-0.34	-0.30	6.4	3.4	-0.13	-0.05	1.0	-1.0	-0.18	-0.13	2.2	1.7
TR754	0.81	0.63	0.04	-0.16	0.7	1.5	0.11	-0.10	0.7	1.5	0.04	0.07	-1.1	-5.5
Avg.	0.08	0.18	-0.29	-0.34	5.5	6.0	0.06	0.03	0.0	-0.1	-0.08	-0.08	2.9	2.5

Table VII

MolProbity results of the structure obtained from the averaging and structure interpolation

Target	MolProbity	
	Avg.	Final
TR644	4.16	1.42
TR655	4.83	2.33
TR661	4.00	1.03
TR662	4.48	1.49
TR663	4.80	2.51
TR671	4.69	2.52
TR674	4.34	1.99
TR679	3.54	1.74
TR681	4.45	1.74
TR688	4.09	1.64
TR689	4.26	2.00
TR696	4.82	2.14
TR698	3.88	1.47
TR699	4.34	2.10
TR704	4.72	1.32
TR705	5.07	2.36
TR708	4.11	1.44
TR710	4.06	1.14
TR712	3.79	1.82
TR720	4.60	1.28
TR723	4.44	1.80
TR724	4.87	1.90
TR738	3.88	0.88
TR747	3.91	1.12
TR750	4.12	1.61
TR752	3.02	1.05
TR754	5.15	2.59
Avg.	4.31	1.72

Table VIII

Average Ca-Ca distances between subsequent residues and radii of gyration for initial models (Init), MD samples (MD), averaged models (Avg.), models after final refinement step (Final), and experimental structures (Exp.)

Target	C α -C α Distance between i,i+1					Radius of Gyration based on C α				
	Init	MD	Avg	Final	Exp	Init	MD	Avg	Final	Exp
TR644	3.81	3.83	3.77	3.81	3.79	16.49	16.06	15.97	16.57	16.15
TR655	3.80	3.83	3.74	3.76	3.85	15.18	15.57	15.36	15.27	15.37
TR661	3.80	3.84	3.77	3.80	3.79	16.83	16.91	16.86	16.84	16.92
TR662	3.84	3.84	3.79	3.82	3.86	10.93	11.15	11.05	10.99	10.86
TR663	3.77	3.83	3.75	3.78	3.80	15.24	15.84	15.74	15.51	15.51
TR671	3.75	3.83	3.72	3.77	3.84	15.80	16.02	15.90	15.85	14.12
TR674	3.76	3.84	3.76	3.79	3.84	13.34	13.87	13.71	13.55	13.24
TR679	3.79	3.84	3.77	3.80	3.82	16.58	16.87	16.72	16.65	16.09
TR681	3.77	3.83	3.73	3.79	3.87	22.42	19.44	19.35	22.24	17.78
TR688	3.80	3.83	3.79	3.81	3.81	17.24	17.54	17.54	17.40	17.35
TR689	3.77	3.83	3.77	3.80	3.83	18.94	17.15	17.13	18.96	16.63
TR696	3.76	3.83	3.73	3.75	3.81	13.62	13.84	13.70	13.66	13.91
TR698	3.79	3.84	3.81	3.82	3.80	12.97	13.17	13.11	13.05	13.93
TR699	3.80	3.95	3.75	3.80	3.82	17.80	25.04	17.98	17.89	17.78
TR704	3.76	3.83	3.77	3.79	3.80	18.45	18.69	18.65	18.56	18.80
TR705	3.75	3.82	3.66	3.73	3.82	13.07	13.61	13.40	13.25	13.22
TR708	3.76	3.84	3.77	3.80	3.81	15.49	15.72	15.67	15.59	16.22
TR710	3.82	3.84	3.80	3.83	3.80	18.78	18.99	18.94	18.87	19.16
TR712	3.76	3.82	3.78	3.79	3.83	16.35	16.51	16.41	16.39	16.34
TR720	3.79	3.84	3.75	3.80	3.79	18.09	18.52	18.38	18.25	16.98
TR723	3.80	3.84	3.79	3.81	3.82	13.95	14.14	14.07	14.02	13.76
TR724	3.73	3.83	3.69	3.75	3.83	13.97	14.65	14.46	14.25	13.43
TR738	3.77	3.84	3.81	3.81	3.81	17.14	17.33	17.28	17.22	17.19
TR747	3.75	3.83	3.77	3.79	3.83	12.61	12.96	12.84	12.74	12.26
TR750	3.80	3.84	3.78	3.81	3.81	14.99	15.27	15.19	15.10	15.17

Target	C_{α} - C_{α} Distance between $i,i+1$					Radius of Gyration based on C_{α}				
	Init	MD	Avg	Final	Exp	Init	MD	Avg	Final	Exp
TR752	3.79	3.83	3.79	3.81	3.79	15.45	15.70	15.59	15.53	15.44
TR754	3.82	3.84	3.70	3.78	3.86	11.29	12.01	11.67	11.50	11.21
<i>Avg:</i>	3.78	3.84	3.76	3.79	3.82	15.67	16.02	15.65	15.77	15.36