

Nat Methods. Author manuscript; available in PMC 2014 November 01

Published in final edited form as:

Nat Methods. 2013 November; 10(11): 1102-1104. doi:10.1038/nmeth.2648.

Improved low-resolution crystallographic refinement with Phenix and Rosetta

Frank DiMaio^{1,6}, Nathaniel Echols^{2,6}, Jeffrey J Headd², Thomas C Terwilliger³, Paul D Adams^{2,4}, and David Baker^{1,5}

¹Department of Biochemistry, University of Washington, Seattle, Washington, USA

²Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA

³Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA

⁴Department of Bioengineering, University of California Berkeley, Berkeley, California, USA

⁵Howard Hughes Medical Institute, University of Washington, Seattle, USA

Abstract

Refinement of macromolecular structures against low-resolution crystallographic data is limited by the ability of current methods to converge on a structure with realistic geometry. We developed a low-resolution crystallographic refinement method that combines the Rosetta sampling methodology and energy function with reciprocal-space X-ray refinement in Phenix. On a set of difficult low-resolution cases, the method yielded improved model geometry and lower free *R* factors than alternate refinement methods.

While determination of X-ray crystal structures at moderate to high resolutions has recently accelerated, structure determination and refinement at lower resolutions remains problematic¹ despite considerable recent work^{2–8}. We reasoned that combining the strengths of the Rosetta structure modeling methodology and the Phenix X-ray refinement software could yield improved refinement at low resolution. Rosetta utilizes a detailed all-atom force field that could partially compensate for the lack of high-resolution data, as well as search procedures combining backbone minimization with discrete side-chain optimization that more effectively explore alternative side-chain arrangements than does simulated annealing. Phenix is a state-of-the-art X-ray refinement package that can be readily integrated with

AUTHOR CONTRIBUTIONS

F.D., N.E., T.C.T., P.D.A. and D.B. designed the research; F.D., N.E. and J.J.H. performed the experiments; F.D. and N.E. wrote the manuscript, and all authors edited and read the final manuscript; P.D.A. and D.B. supervised the project.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

^{© 2013} Nature America, Inc. All rights reserved.

Correspondence should be addressed to F.D. (dimaio@u.washington.edu) or N.E. (nechols@lbl.gov). $^6\mathrm{These}$ authors contributed equally to this work.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

other computational methods. We therefore incorporated the maximum-likelihood reciprocal-space X-ray target function from phenix.refine⁹ into Rosetta.

To enable refinement in Rosetta, we use Phenix routines (called through Python bindings) to calculate the crystallographic refinement target function. Rosetta energy is weighted against the crystallographic likelihood function by normalizing the gradients of each before each minimization cycle¹⁰. Unlike in standard Rosetta structure prediction, non-ideal bond geometry is allowed throughout refinement; Rosetta symmetry¹¹ optimizes the energy of the protein in the crystal lattice. The combined approach (phenix.rosetta_refine) uses Phenix¹² to perform bulk solvent correction, calculate electron-density maps and refine B factors, while the Rosetta force field, minimizer and sampling methods optimize model geometry. Modular interfaces to the new refinement protocols provided by RosettaScripts¹³ and Phenix allow refinement protocols to be customized (Supplementary Figs. 1–3).

Using this framework, we developed a protocol to optimize poor starting models against low-resolution data. The method alternates real- and reciprocal-space refinement: the Rosetta force field constrains reciprocal-space refinement to physically plausible conformations; density maps restrain Rosetta side-chain and backbone sampling in real space. Significant backbone movement occurs during internal coordinate minimization, while side-chain optimization allows traversal of large energy barriers that impede traditional continuous refinement.

To assess the performance of our approach in realistic difficult cases, we assembled a collection of 26 starting models for 15 data sets at 3.0- to 4.5-Å resolution (Supplementary Table 1) by performing molecular replacement with low-resolution data sets using templates with nontrivial conformational changes. These cases are within the radius of convergence of molecular replacement but are far enough from the final structure that large errors were expected unless extensive manual rebuilding was applied.

We compared Rosetta-Phenix refinement (phenix.rosetta_refine) to three different low-resolution refinement strategies. As a control, structures were refined for 20 cycles in *phenix.refine*, optimizing X-ray weight at each cycle. We also refined structures in CNS¹⁴, using the DEN methodology (with full weight optimization)², and in REFMAC5 (ref. 15) with jelly-body refinement. We compared the generated structures using free R factor ($R_{\rm free}$), MolProbity score¹⁶ and r.m.s. deviation to the re-refined published structure. Running times of DEN and Rosetta-Phenix were similar (about 4 h per model for a 1,000-residue protein), with Phenix roughly 4 times faster and REFMAC5 about 30 times faster (not accounting for parallelization).

The results (Fig. 1, Supplementary Tables 2 and 3) show clear improvement for CNS-DEN and REFMAC5 compared to conventional refinement in Phenix, and further improvement with Rosetta-Phenix refinement. Although DEN and REFMAC5 refinement consistently showed a large radius of convergence, model quality (using MolProbity) was worse than that of Rosetta-Phenix in all but two cases. We discuss several illustrative examples below.

The starting molecular replacement models for a Ca^{2+} ATPase structure (PDB 3FPS) require significant conformational changes to match the published structure; initial R_{free} factors are

>0.5. Neither phenix.refine, DEN nor REFMAC5 was able to improve the models beyond an $R_{\rm free}$ of 0.43 or r.m.s. deviation of 6.0 Å. However, Rosetta-Phenix refinement starting from the conformation in the absence of calcium (PDB 2ZBG) resulted in a greatly improved fit to the data: $R_{\rm free} = 0.28$ with an r.m.s. deviation of 1.68 Å to the published structure (Fig. 2).

The starting models for a glutamate receptor (PDB 1ISR) differ from the native structure by large hinge motions. Both CNS-DEN and Rosetta-Phenix outperformed conventional refinement; the Rosetta-Phenix model is nearly identical to the published structure (r.m.s. deviation = 0.5 Å), but with superior geometry. The final Rosetta-Phenix model starting from the ligand-free conformation (PDB 1EWT) has an R_{free} below that of the published structure (see Supplementary Tables 2 and 3), despite a starting-model r.m.s. deviation of 4.5 Å.

For calcium-calmodulin-dependent protein kinase II (PDB 3KK9), CNS-DEN and Rosetta-Phenix yielded similar R factors, with the DEN model closer to the published structure. To evaluate whether a combined protocol would yield further improvement, we ran Rosetta-Phenix refinement on the DEN model; this reduced $R_{\rm free}$ to 0.31 (below the published structure) and corrected most of the geometry errors (Supplementary Table 4). This suggests that a combination of strategies may be most appropriate in some difficult cases, as the restraint data used by the methods—DEN drawing off homologous structures and Rosetta using detailed physical chemistry—are largely orthogonal.

We had used a preliminary version of our Rosetta-Phenix refinement to solve the structure of the TRIP8b channel subunit¹⁷. Standard refinement reduced the $R_{\rm free}$ slightly, but model geometry was poor. Rosetta-Phenix refinement followed by rebuilding in Coot¹⁸ improved $R_{\rm free}$ from 0.46 to 0.28 while also maintaining reasonable model geometry¹⁷ (Supplementary Fig. 4).

The refinement protocol described here differs from previously described refinement methods^{2,6–8} in several key aspects. First, the Rosetta energy function ensures that the protein conformation remains physically plausible even when models undergo large conformational rearrangement during refinement¹⁹. Second, Rosetta refinement makes use of discrete side-chain optimization, which allows for large-scale reorganization of side-chain geometry not achievable through minimization alone.

Our results using the DEN method in CNS and jelly-body restraints in REFMAC5 show that reference model—derived restraints are quite powerful in low-resolution refinement. Although these methods extend the radius of convergence², they are less effective at improving model geometry²⁰. In contrast to reference structure—based restraints, the Rosetta force field makes it possible to discover new energetically favorable interactions—not present in the reference model—during a refinement trajectory. As suggested by the results for PDB 3KK9, the combination of these approaches can be quite powerful.

We found that the combined Rosetta-Phenix method outperforms conventional refinement methods against low-resolution crystallographic data. Importantly, the method also does not degrade high-resolution structures: a high-resolution benchmark (Supplementary Table 5 and Supplementary Fig. 5) showed that Rosetta-Phenix refinement produced little change in

 $R_{\rm free}$ but improved MolProbity score. Further low-resolution improvements may be achievable by incorporating the fragment-based backbone rebuilding essential to Rosetta *de novo* structure prediction and to MR-Rosetta²¹.

The combined Rosetta-Phenix method requires installation of both Rosetta (version 3.6 or newer, available at https://www.rosettacommons.org) and Phenix (version 1.8.3 or newer, available at http://www.phenix-online.org), which are freely available to academics. The program *phenix.rosetta_refine* (distributed with Phenix) automates the process. All structures used for testing may be downloaded at http://www.phenix-online.org/phenix_data.

METHODS

Methods and any associated references are available in the online version of the paper.

ONLINE METHODS

Generation of test cases

Structures and crystallographic data were selected based on resolution, the absence of noncrystallographic symmetry and the availability of related structures to be used as search models for molecular replacement. Suitable MR models were identified with an r.m.s. deviation between 1.5 Å and 4.5 Å to the final deposited structure, and starting models for refinement were generated by molecular replacement in Phaser²² after minimal trimming of loops and terminal regions not found in the target structure. The search models for PDB 3IDQ, 3A8N and 3SNH were processed by the Sculptor program²³ to match the sequence of the target. Where necessary, crystallographic symmetry operators and origin shifts were applied using phenix.find_alt_orig_sym_mate²⁴ to place the MR solution in the same frame of reference as the published structure.

Rosetta flexible bond geometry and Cartesian refinement

During internal coordinate refinement, both bond angles and bond lengths were allowed to deviate from their ideal values. Harmonic potentials were added to restrain both bond angles and bond lengths; planarity was enforced using pseudotorsional restraints. Cartesian-space refinement was implemented using the same harmonic restraints and was found to produce significant improvement in structure optimization against the data while simultaneously improving model geometry. Both minimization strategies can be used and modified via RosettaScripts¹³.

Rosetta refinement strategy

Rosetta was compiled with a Python interpreter that directly calls Phenix functions using C+ + Python bindings²⁵. Phenix functionality was used to perform bulk solvent correction and anisotropic scaling²⁶ and to calculate the X-ray energy and gradients using the ML^{27} or $MLHL^{28}$ target functions. Phenix was also used to compute $2mF_0 - DF_c$ (2mFo–DFc) maps for real-space modeling and refinement in Rosetta; this can optionally include density modification in RESOLVE²⁹. *B*-factor refinement was performed by *phenix.refine*⁹.

The refinement script used in Rosetta consists of the following steps:

 Three cycles of side-chain optimization, followed by reciprocal-space torsion-angle minimization;

- Five cycles of side-chain optimization, followed by both real-space and reciprocalspace torsion-angle minimization;
- Two cycles of side-chain optimization, followed by Cartesian reciprocal-space minimization.

Each step also starts with the calculation of a new σ_A -weighted 30 $2mF_o-DF_c$ electron density map; this map—together with the Rosetta energy function—is used for side-chain optimization and (in some cycles) real-space refinement. Reflections flagged for cross validation 31 are replaced by DF_c to avoid bias 9 . By default, the grid spacing for the CCP4-format maps 32 was $d_{\rm min}/2$. A single macrocycle of individual B-factor refinement was performed by phenix.refine at the end of each step. To ensure consistent calculation of the scattering contribution from bulk solvent and hydrogen atoms, the Rosetta models were run in phenix.refine for one cycle with a null strategy to recalculate R factors after refinement. The complete XML script is provided in Supplementary Figure 1.

To counter refined model variation in Rosetta-Phenix (typically 3–6% $R_{\rm free}$ variation between trajectories, but over 25% in extreme cases; see Supplementary Fig. 6), we ran multiple independent refinements in parallel and selected the result with the lowest $R_{\rm free}$ —a process similar to the protocol used for DEN refinement². While this potentially biases an important validation statistic, in practice we found that for the test cases described here, only five runs were required to obtain adequate sampling, with an overall runtime comparable to that of DEN refinement in CNS. If a fully unbiased $R_{\rm free}$ statistic is desired, then the procedure could be carried out using a second fully free test set³³.

Refinement in Phenix

Coordinate and individual *B*-factor refinement were performed for 20 macrocycles, with automatic optimization of restraint weights using a grid search³⁴. Only the default stereochemistry restraints were applied^{35,36}. An alternate 'optimized' protocol, shown only in Supplementary Tables 2 and 3, was also run with 20 cycles of simulated annealing, using reference restraints derived from the starting structure.

Refinement in CNS

DEN refinements were performed as described², using a grid search to optimize the *gamma* and w_den parameters. Because the DEN protocol in CNS uses grouped B-factor refinement, additional individual B-factor refinement and bulk-solvent correction was performed in Phenix to obtain R factors comparable to the other methods.

Refinement in REFMAC

REFMAC5 refinements were run for 100 cycles, using automatic weighting and jelly-body restraints with σ = 0.02. R factors were then recalculated using phenix.refine to ensure consistent bulk solvent handling.

Evaluation of results

All refined models were validated using MolProbity 16 as implemented in the Phenix suite. R.m.s. deviations to the published structures were calculated using *phenix.superpose_pdbs*, including all non-hydrogen backbone atoms and side chains out to the C γ atom where present.

High-resolution data set

The HiQ54 data set³⁷ consists of 54 non-redundant, monomeric, atomic-resolution structures (0.8–1.4 Å) with MolProbity score 1.4, near-zero geometry outliers and no large, tightly bound ligands. Structures were prepared for Rosetta refinement by removing all ligands except waters and all alternate conformations and then refining in *phenix.refine* for three macrocycles with all default parameters. Rosetta-Phenix refinements were performed with explicit hydrogens using a streamlined protocol consisting of only the final two Cartesian-minimization cycles. The refined structures were compared to the starting models by removing hydrogen atoms and re-running *phenix.refine* on both input and output models with a null strategy as above.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank P. Afonine and R. Grosse-Kunstleve for technical advice, J. Richardson for the HiQ54 test structures, and P. Afonine, J. Fraser, R. Read and J. Richardson for comments on the manuscript. Funding was provided by the US National Institutes of Health (grant nos. GM063210 and GM092802). This work was supported in part by the US Department of Energy under contract no. DE-AC02-05CH11231.

References

- Brünger AT, DeLaBarre B, Davies JM, Weis WI. Acta Crystallogr D Biol Crystallogr. 2009; 65:128–133. [PubMed: 19171967]
- 2. Schröder GF, Levitt M, Brünger AT. Nature. 2010; 464:1218–1222. [PubMed: 20376006]
- 3. Fenn TD, Schnieders MJ. Acta Crystallogr D Biol Crystallogr. 2011; 67:957–965. [PubMed: 22101822]
- 4. Haddadian EJ, et al. Biophys J. 2011; 101:899–909. [PubMed: 21843481]
- 5. Koparde VN, Scarsdale JN, Kellogg GE. PLoS ONE. 2011; 6:e15920. [PubMed: 21246043]
- 6. Headd JJ, et al. Acta Crystallogr D Biol Crystallogr. 2012; 68:381–390. [PubMed: 22505258]
- 7. Nicholls RA, Long F, Murshudov GN. Acta Crystallogr D Biol Crystallogr. 2012; 68:404–417. [PubMed: 22505260]
- 8. Smart OS, et al. Acta Crystallogr D Biol Crystallogr. 2012; 68:368–380. [PubMed: 22505257]
- 9. Afonine PV, et al. Acta Crystallogr D Biol Crystallogr. 2012; 68:352-367. [PubMed: 22505256]
- Adams PD, Pannu NS, Read RJ, Brünger AT. Proc Natl Acad Sci USA. 1997; 94:5018–5023.
 [PubMed: 9144182]
- DiMaio F, Leaver-Fay A, Bradley P, Baker D, Andre I. PLoS ONE. 2011; 6:e20450. [PubMed: 21731614]
- 12. Adams PD, et al. Acta Crystallogr D Biol Crystallogr. 2010; 66:213-221. [PubMed: 20124702]
- 13. Fleishman SJ, et al. PLoS ONE. 2011; 6:e20161. [PubMed: 21731610]
- 14. Brünger AT, et al. Acta Crystallogr D Biol Crystallogr. 1998; 54:905–921. [PubMed: 9757107]

15. Murshudov GN, et al. Acta Crystallogr D Biol Crystallogr. 2011; 67:355–367. [PubMed: 21460454]

- 16. Chen VB, et al. Acta Crystallogr D Biol Crystallogr. 2010; 66:12-21. [PubMed: 20057044]
- 17. Bankston JR, et al. Proc Natl Acad Sci USA. 2012; 109:7899–7904. [PubMed: 22550182]
- 18. Emsley P, Lohkamp B, Scott WG, Cowtan K. Acta Crystallogr D Biol Crystallogr. 2010; 66:486–501. [PubMed: 20383002]
- 19. Das R, Baker D. Annu Rev Biochem. 2008; 77:363-382. [PubMed: 18410248]
- 20. Brünger AT, et al. Acta Crystallogr D Biol Crystallogr. 2012; 68:391–403. [PubMed: 22505259]
- 21. DiMaio F, et al. Nature. 2011; 473:540-543. [PubMed: 21532589]
- 22. McCoy AJ, et al. J Appl Crystallogr. 2007; 40:658–674. [PubMed: 19461840]
- Bunkóczi G, Read RJ. Acta Crystallogr D Biol Crystallogr. 2011; 67:303–312. [PubMed: 21460448]
- 24. Oeffner R, Bunkóczi G, Read RJ. Computat Crystallogr Newslett. 2012; 3:5–10.
- 25. Abrahams D, Grosse-Kunstleve RW. C/C++ Users J. 2003; 21:29–36.
- Afonine PV, Grosse-Kunstleve RW, Adams PD. Acta Crystallogr D Biol Crystallogr. 2005;
 61:850–855. [PubMed: 15983406]
- 27. Lunin VY, Afonine PV, Urzhumtsev AG. Acta Crystallogr A. 2002; 58:270–282. [PubMed: 11961289]
- 28. Pannu NS, Murshudov GN, Dodson EJ, Read RJ. Acta Crystallogr D Biol Crystallogr. 1998; 54:1285–1294. [PubMed: 10089505]
- 29. Terwilliger TC. Acta Crystallogr D Biol Crystallogr. 2000; 56:965–972. [PubMed: 10944333]
- 30. Read RJ. Acta Crystallogr A. 1986; 42:140-149.
- 31. Brünger AT. Nature. 1992; 355:472–475. [PubMed: 18481394]
- 32. Winn MD, et al. Acta Crystallogr D Biol Crystallogr. 2011; 67:235–242. [PubMed: 21460441]
- 33. Kleywegt GJ. Acta Crystallogr D Biol Crystallogr. 2007; 63:939–940. [PubMed: 17704561]
- 34. Afonine PV, Echols N, Grosse Kunstleve RW, Moriarty NW, Adams PD. Computat Crystallogr Newslett. 2011; 2:99–103.
- 35. Engh RA, Huber R. Acta Crystallogr A. 1991; 47:392–400.
- 36. Vagin AA, et al. Acta Crystallogr D Biol Crystallogr. 2004; 60:2184–2195. [PubMed: 15572771]
- 37. Leaver-Fay A, et al. Methods Enzymol. 2013; 523:109-143. [PubMed: 23422428]

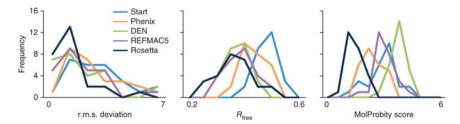


Figure 1. Comparison of Phenix, CNS-DEN, REFMAC5 and Rosetta-Phenix refinements on a realistic low-resolution test set of 15 proteins (Supplementary Table 1). Histograms of the r.m.s. deviation to the deposited structure (left), $R_{\rm free}$ (middle), and MolProbity score, indicating stereochemical quality (right), after refinement of low-resolution test models. Starting model distributions are in blue. Source data for this figure are provided in Supplementary Tables 2 and 3.

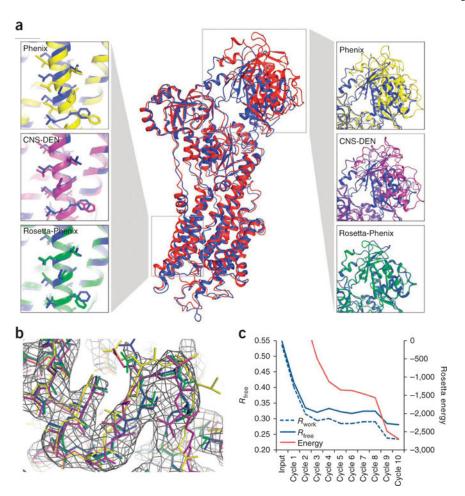


Figure 2. Refinement of the Ca²⁺ ATPase (PDB 3FPS, PDB 2ZBG) using Rosetta-Phenix. (a) Alphacarbon traces comparing the results of refinement of a molecular replacement (MR) model using rigid-body refinement (red), Phenix (yellow), CNS-DEN (purple) and Rosetta-Phenix (green); the published structure is in blue. (b) Close-up showing all non-hydrogen atoms (colored as above), with $2mF_0 - DF_c$ electron density for the final structure contoured at 1σ . (c) Decrease in Rosetta energy and $R_{\rm work}$ and $R_{\rm free}$ to the deposited structure during Rosetta-Phenix refinement.