HHS Public Access

Author manuscript

Nat Protoc. Author manuscript; available in PMC 2016 March 01.

Published in final edited form as:

Nat Protoc. 2015 September ; 10(9): 1275–1284. doi:10.1038/nprot.2015.069.

X-ray structure determination using low-resolution electron microscopy maps for molecular replacement

Ryan N. Jackson[1], **Airlie J. McCoy**[3], **Thomas C. Terwilliger**[2], **Randy J. Read**[3], and **Blake Wiedenheft**[1,*]

[1]Department of Microbiology and Immunology, Montana State University, Bozeman MT 59717, USA

[2]Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545

[3]Department of Haematology, University of Cambridge, Cambridge Institute for Medical Research, Cambridge CB2 0XY, U.K.

Abstract

Structures of multi-subunit macromolecular machines are primarily determined by either electron microscopy (EM) or X-ray crystallography. In many cases, a structure for a complex can be obtained at low resolution (at a coarse level of detail) with EM and at higher resolution (with finer detail) by X-ray crystallography. The integration of these two structural techniques is becoming increasingly important for generating atomic models of macromolecular complexes. A low-resolution EM image can be a powerful tool for obtaining the "phase" information that is missing from an X-ray crystallography experiment, however integration of EM and X-ray diffraction data has been technically challenging. Here we present a step-by-step protocol that explains how low-resolution EM maps can be placed in the crystallographic unit cell by molecular replacement, and how initial phases computed from the placed EM density are extended to high resolution by averaging maps over non-crystallographic symmetry. As the resolution gap between EM and X-ray crystallography continues to narrow, the use of EM maps to help with X-ray crystal structure determination, as described in this protocol, will become increasingly effective.

INTRODUCTION

Overview

Structure determination of macromolecules is one of the most effective tools for understanding biological function and mechanisms of action. Traditionally, electron microscopy (EM) has been used to determine structures of large macromolecular assemblies

*Corresponding author. bwiedenheft@gmail.com.

Supplementary Information

Description of Supplementary Data Files S1-S13

Author contributions statements

All authors contributed to writing the protocol. AJM developed the EM refinement parameter in *Phaser-MR* necessary to account for uncertainty in the EM magnification error.

Competing financial interests

The authors declare no competing financial interests

(> 350 kDa); however, recent advances in high-throughput crystal screening, X-ray detectors, data processing software, and synchrotron radiation sources are making structure determination of large macromolecular assemblies by X-ray crystallography increasingly more common [1]. At the same time, advances in direct electron detector technology and cryo-EM image reconstruction algorithms are dramatically improving the quality and resolution of EM densities [2,3]. Collectively, these advances present new opportunities for structural biologists capable of integrating data from X-ray crystallography and EM [4].

The amplitudes and phases of diffracted X-rays are required to determine three-dimensional structures from macromolecular crystals. Amplitudes are measured using an X-ray detector, but phase information is lost in native diffraction data. Phases can be determined experimentally from anomalous and/or isomorphous differences in heavy atom diffraction, or phases can be approximated from a structurally similar model using a technique called Molecular Replacement (MR) [5]. MR was first implemented in the 1960s, and has now been used in approximately 65% of X-ray crystal structures in the Protein Data Bank (PDB). While most of these structures were determined using phases from homologous X-ray structures, phases can also be derived from low-resolution EM densities [6,7]. In this case, the phase information will normally only be applicable to the X-ray diffraction data that corresponds to low-resolution features in the structure. Consequently some means of using this low-resolution phase information as a starting point and obtaining high-resolution phase information is necessary. This remains technically challenging and few structures in the PDB have been determined using phases from EM densities. The technical challenges associated with using EM density maps to phase X-ray data can be broken into four general steps, (i) the EM map must be placed into a large unit cell from which structure factors (amplitudes and phases) are generated by Fourier methods, (ii) the MR method must be able to account for uncertainty in the magnification factor of an EM map, (iii) non-crystallographic-symmetry (NCS) must be identified in the electron density, and (iv) density related by NCS must be averaged during iterative rounds of density modification to extend the phases to high-resolution data (Figure 1). Each of these steps is achievable using a combination of data processing and analysis software [7,8], however a detailed protocol that guides users though each of these steps is currently unavailable. Here we present a step-by-step protocol that uses CCP4 and Phenix software tools to determine X-ray structure solutions with an EM map as an MR search ensemble.

Development of the protocol

The protocol presented here was used to determine the 3.24 Å resolution X-ray crystal structure of a 405kDa RNA-guided surveillance complex from Escherichia coli, called Cascade (CRISPR-associated complex for antiviral defense) [9]. The 8 Å resolution cryo-EM reconstruction of Cascade was prepared as an MR search ensemble for use in Phaser [10,11]. After modifying the bulk solvent parameter Siga fsol and refinement of the EM scale factor, we were able to determine an initial low-resolution structure. Initial low-resolution phases were extended to high-resolution by iterative rounds of density modification that averaged NCS-related electron density. In this protocol we use the Cascade example as a tutorial. We provide supplemental material necessary to solve this structure, and intermediate files

generated while determining the structure of Cascade. These files are intended to help guide new users through the process and serve as positive controls for users at every skill level.

To test the general applicability of this protocol, we also used this protocol to determine phases of the archaeal 20S proteasome core particle from *Thermoplasma acidophilum* (PDB ID: 1pma) using a 6.6 Å resolution EM map (EMDB: 1733) [12,13]. Seven fold NCS identified in the low-resolution proteasome solution was used to extend phases. For both the 20S proteasome and Cascade, phase extension to high-resolution produced continuous electron density maps with obvious secondary structure elements and density for side chains (Figure 2).

In addition to the proteasome, we also determined phases for the 70S ribosome from *Thermus thermophilus* (PDB ID: 4v51) using an 11 Å (EMDB 1008) resolution EM map as a search ensemble [14,15]. To determine the resolution limits of cryo-EM maps that can be used for MR, we sequentially reduced the resolution of each search ensemble until the protocol failed. Phasing by MR failed at 10 Å for Cascade, 15 Å for the Proteasome, and 19 Å for the Ribosome. These results indicate that this protocol will work with search ensembles of approximately 20 Å resolution or higher, however the limits of this protocol will vary to some degree depending on the size, resolution, and symmetry of the macromolecule.

Applications and Limitations

Extending phases to high-resolution data with this protocol is dependent on averaging related density within the asymmetric unit of a single crystal. It is possible that similar approaches could be used without NCS if multiple crystal forms were available or if an exceptionally high fraction of the crystal were disordered solvent, a feature that can be used in density modification.

If NCS, high solvent content, and multiple crystal forms are not available, then initial low-resolution phases can be used to help identify heavy atom positions using isomorphous or anomalous dispersion methods. Once heavy atom positions are found, low-resolution phases can be estimated, and extended to high-resolution using density modification strategies.

Procedure Overview—Prepare the EM map for molecular replacement (Supplemental movie 1 https://www.youtube.com/watch?v=NH0E2suSiA4)

- Steps 1-3 Generate structure factors and phases from an EM map
- Steps 4-6 Determine the center and extents of the EM density
- Steps 7-17 Place the EM density into a large P1 unit cell using *Phenix*

Molecular Replacement using Phaser-MR (Supplemental movie 2 https://www.youtube.com/watch?v=9GMutfmhRZY)

- Steps 18-27 Modify *Phaser-MR* parameters to account for EM magnification error
- Steps 28-38 Set up and run *Phaser-MR*

Find NCS-related density (Supplemental movie 3 https://www.youtube.com/watch?v=oqGbAL3FqIc)

- Steps 39-43 Find NCS from density using *Phenix*.

Phase extension (Supplemental movie 4 https://www.youtube.com/watch?v=DyG3jULPGEM)

- Steps 44-52 Prepare the MR solution for phase extension
- Steps 53-65 Merge high-resolution X-ray diffraction data with low-resolution phases
- Steps 66-77 Extend phases to high-resolution using density modification

MATERIALS

Equipment setup

Download Chimera, *Phenix*, Coot, and CCP4 software packages. This protocol was developed using Chimera version 1.10, Phenix unofficial release dev-1930, Coot version 0.8, and CCP4 version 6.4.0. These programs are free for academic users and are supported on Linux, MacOS-X, and Windows platforms. Download instructions, tutorials and list servers are available [16-19].

The computational tools used in this protocol are continually being updated to meet the demands of structural biologists. As the fields of X-ray crystallography and cryo-EM continue to converge, we expect that updates to the software will make this approach to structure determination even more accessible. Developments that will improve this protocol include GUI tools that eliminate the need for the command line, improvements in algorithms aimed at finding NCS related density in low-resolution density, and tools that will quickly merge phases from a low-resolution MR-solution with collected X-ray amplitudes.

Data

Structure determination performed using this procedure requires a scaled set of merged structure factors and associated errors derived from X-ray diffraction (Supplementary Data S1), electron density map coordinates (Supplementary Data S2), and sequence information associated with the macromolecules in the crystal (Supplementary Data S3). The estimated solvent fraction of the crystal must be known, and can be determined using phenix.xtriage or the Matthews Probability Calculator (www.ruppweb.org/mattprob/default.html) [20-23]. All formats of scaled X-ray data should work with this procedure, but EM maps must be in CCP4 format (i.e., .map or .ccp4). It is critical that low-resolution amplitudes are included in the X-ray data; otherwise MR experiments performed using low-resolution search ensembles are likely to fail.

PROCEDURE

Prepare the EM map for molecular replacement

1. Generate structure factors and phases from an EM map (Steps 1-3). **1-5 minutes.** Download published EM maps from the Electron Microscopy Data Bank (EMDB) (http://www.ebi.ac.uk/pdbe/emdb/), or obtain a map directly from EM experimentalists (Supplementary Data S2).

2. Set the origin of the EM map to zero. Open the .map file in Chimera. Under the *Tools* menu select *Volume Data*, *Volume Viewer*, *Features*, *Coordinates*, change *Origin index* to 0 (hit return). Go to *File* and save map as map_shifted.mrc (Supplementary Data S4).

3. Convert map_shifted.mrc to structure factors and phases (i.e. map_shifted.mtz). Open a new terminal and set up the *Phenix* command-line environment as described in the *Phenix* documentation: http://www.phenix-online.org/documentation/install-setup-run.html. Source *Phenix* and type “phenix.map_to_structure_factors d_min=8.0 output_file_name=map_coeff.mtz map_shifted.mrc”. The value of “d_min” is the resolution of the EM map (e.g. 8Å), and map_shifted.mrc is the path to the newly created .mrc file (e.g. */Users/username/directory/map_shifted.mrc*). See Supplemental file S5.

<CRITICAL STEP> The Phenix GUI can be used to execute most Phenix functions, but phenix.map_to_structure_factors must be run from the command-line.

4. Determine the center and extent of the EM density. (Steps 4-6). <TIMING> 5-10 minutes. Molecular replacement calculations performed later in the protocol will require extent values and x, y, z coordinates of the center of the EM density. To determine these values open the .mtz file in Coot and scroll up or down to adjust the contour. To display the entire map, you may need to adjust the map radius by selecting *Edit*, *Map parameters*. To display the unit cell select *Draw*, *Cell and Symmetry*, and choose “yes” under *Show unit cells.*

5. To identify the center of the EM density move the Coot center pointer to the center of the EM density, and place a water molecule with the *Place Atom at Pointer Tool.* To display the x, y, z coordinates of the EM density, middle click on the water molecule. The x, y, z coordinates will be displayed in the information bar at the bottom of the Coot window.

6. Determine the x, y, z extents of the EM density by placing water molecules at opposite edges of the density along the x-, y-, z-axes. Select the *Measure* tab, and use the *Distance* tool to find the distance in ångstroms between the water atoms along each axis. The locations of the EM center and volume extents are important for placing the density into a large P1 unit cell (steps 13-14).

7. Place the EM density into a large P1 cell using *Phenix*. (Steps 7-17). <TIMING> 10-15 minutes. Molecular replacement calculations require that structure factors be finely sampled. To obtain fine sampling of structure factors, the EM density must be placed in a unit cell that is three to four times larger in volume. Using the *Phenix GUI*, open *Maps*, and select the *Cut out density* tool.

8. In the *Map coefficients (MTZ)* field, load the .mtz file generated in step 3 (Supplementary Data S5).

9. Leave the *PDB file*, and *Atom selection* fields empty. Define the *Output directory* and *Job title*.

10. Select "*box*" as the *Cutout type*.

11. Fill in the *Output resolution* field with a value smaller than the resolution limit of the original EM map (e.g. if the resolution of the EM map is 8 Å, then use 7.8).

12. To expand the volume of the unit cell, define the *Padding around cutout density* parameter as the largest extent in ångstroms of the EM density determined in step 6.

13. Define the *Cutout center*. This will be 0, 0, 0 if the electron density is positioned at the origin (i.e. corner of the unit cell). If the density of the map you download is not located at the origin, then use the x, y, z coordinates that define the center of the EM density (see step 5).

14. In the field *Cutout dimensions*, define x, y, z lengths in ångstroms that span the full volume of the density. These values were determined using the measure distance tool in Coot (see step 6).

15. Do not change the default parameters for *Cutout sphere radius* and *Cutout model radius*. These are ignored when the *Cutout type* is "*box*". Do not select *subtract mean density from map*.

16. Run the *Phenix Cut out density* tool. This will position the density in a new .mtz file at the origin of a large P1 unit cell (Supplementary Data S6).

17. Evaluate the .mtz file in Coot to ensure that the entire density of the EM map is positioned at the origin of a large P1 unit cell. If the density is cut off on one edge, then increase the cutout dimension along the corresponding axis (step 14) and repeat the cut out density procedure.

Molecular replacement using *Phaser-MR*

18. Modify *Phaser-MR* parameters to account for EM magnification error. (Steps 18-27). <TIMING> 5 minutes. Open *Phaser-MR* from the *Molecular Replacement* tab in the *Phenix GUI*.

19. Under *Phaser mode*, select *Automated molecular replacement*.

20. In the *Data file* field, load the high-resolution X-ray data (Supplementary Data S1). The *Unit cell*, *Space group*, and *Data labels* fields should automatically populate.

21. Fill the *High resolution*, and *High resolution for refinement* fields with the original resolution limit of the EM map.

22. Define the path for the *Output directory* and *Title* of the run.

23. Select the *other setting* tab located at the top center of the *Input and general options* panel.

24. At the bottom of the *Phaser parameters* window change the *User level* to *Developer*.

25. Find the parameters section called *MR refinement*. Change the *MR refinement* protocol setting to *custom*. Check all the boxes, including *Refine EM scale factor*, which accounts for magnification errors. Set the Maximum number of cycles to 50.

26. Find the *Bulk solvent parameters* section, and place a number less than 1 (empirically, 0.3 has proven to be a good starting point) in the *Siga fsol* parameter field. The default is 1.05. This is appropriate for atomic models, where bulk solvent is omitted, but EM data includes the average contribution of vitreous ice.

27. Exit the *Other parameters* window by clicking OK.

28. Set up and run *Phaser-MR.* (Steps 28-38). <TIMING> 10-45 minutes. Open the *Ensembles* tab, and select the *Add map ensemble* button at the bottom of the panel.

29. Open the newly created map *Ensemble* tab, and define a name for the Ensemble in the *Model ID* field.

30. In the *Reflections file* field, load the .mtz file generated in step 16 (Supplemental File S6). The *Column labels* should automatically fill with FWT and PHWT.

31. Set the RMSD of the ensemble. One-half the value of the resolution of the EM map is a good rule of thumb. For example if the resolution limit of the EM map is 8 Å, then set the RMSD to 4.

32. In the *Centre* field, enter the values 0, 0, 0. The EM map was placed at the unit cell origin in step 16 using the phenix.cut_out_density tool.

33. In the *Map extent* field, enter the x, y, z extent values in ångstroms. These were determined in step 6.

34. Define the protein and nucleic acid mass contributions of the ensemble. Leave the *Cell scale factor* value as 1.0.

35. In the *Composition* tab, define the *Chain type*, *Number of copies*, and molecular weight or sequence of each macromolecule expected in the asymmetric unit of the crystal (Supplementary Data S3).

<CRITICAL STEP> All subunits present in the asymmetric unit of the crystal must be accounted for in the composition. For example, if two hexamers are expected in the asymmetric unit, then twelve copies of the chain should be indicated.

36. In the *Search procedure* tab, check the box next to the search ensemble and define how many copies are expected in the asymmetric unit of the crystal. For example, two would be selected if two copies of the prepared EM map ensemble were expected to be present in the asymmetric unit of the crystal.

37. Keep the defaults of other *Search options* and run *Phaser-MR* (Supplementary Data S7).

38. Evaluate if *Phaser* solved the structure. Usually a final TFZ score above 8 indicates a correct solution. In some cases, multiple solutions may result from internal symmetry in the search ensemble. For example with the 20S archaeal proteasome 14 solutions are found due to 7- and 2-fold internal symmetry of the complex. However, if more solutions are found than expected it is likely a consequence of a low signal to noise ratio and may require more low-resolution reflections in the X-ray data set, a more complete X-ray data set, or a higher-resolution search ensemble.

For more information about evaluating an MR solution see the Phaser user manual: http://www.phaser.cimr.cam.ac.uk/index.php/Molecular_Replacement#Has_Phaser_Solved_It.3F.

<CRITICAL STEP> A solution should be found by Phaser within minutes. If the run takes more than 0.5 hrs, abort the run and refer to the advice in the Troubleshooting section.

Find NCS-related density

39. Find NCS-related density using *Phenix*. (Steps 39-43). <TIMING> 40-60 minutes. Open *Find NCS operators* from the *Model tools* tab in the *Phenix GUI*.

40. Leave the *Input file type* option as *Automatic* and the *PDB file* field blank. In the *Map coefficients* field, load the .mtz file generated in step 37 by Phaser-MR (Supplementary Data S7). Define the *Output directory* and *job title*. The *Map labels* should be automatically recognized. Insert the resolution of the cryo-EM map into the *High-resolution limit for map* field.

41. Open the *Center-finding parameters* window. Fill in the *Smoothing radius* field with the number ten and use twenty in the *Maximum number of centers to try*. Both of these parameters can be modified to expand or limit the search procedure (see Troubleshooting section below). Select OK to exit the window.

42. Open the *Density search parameters* window. We recommend the following as starting parameter defaults: Density radius = 20, Peak separation = 30, Density peaks = 20, Delta phi for search = 10, Max NCS copies = none, Min ratio to top CC = 0.75, Min overall NCS CC = 0.4, Min fraction NCS = 0.01. Select OK to leave the window.

43. The NCS operators and a correlation coefficient (CC) of related density are reported as rotation and translation matrices for each search center (e.g. twenty different matrices will be produced with the default *n_center_use*=20). The NCS group with the overall best score is saved as the file *find_ncs.ncs_spec* (Supplementary Data S8). Open the *find_ncs.ncs_spec* file in a text editor to evaluate the results.

It is useful to note that the matrices generated by all searched centers are listed in the *Run status* window and can be copied and saved as a new .ncs_spec file. A group of NCS matrices with high CC, but few operators might work better than a group with more NCS operators and low CC.

Phase extension

44. Prepare the MR solution for phase extension. **(Steps 44-52). <TIMING> 5 minutes.** Generate Hendrickson-Lattman (HL) coefficients from the Figure of Merit (FOM) generated by the *Phaser-MR* solution.

45. Open the *Reflection Data Utilities* in CCP4i and select the *Convert FoM to/from HL* tool.

46. In the *Convert* pull down menu select from *phi/fom to HL coeffs*.

47. Load the .mtz file generated by *Phaser-MR* in step 37 (Supplementary Data S7), into the *MTZ in* field.

48. After loading the .mtz file, the *PHI* field should automatically populate with *PHIC*. If this field remains empty, or contains a different parameter, then use the pull-down menu to select PHIC.

49. Select *FOM* from the *FOM* pull-down menu.

50. Assign a location and file name for the output .mtz file in the *Work MTZ out* field.

51. Change the *Output column label prefix* to HL.

52. Run the *Convert between HL coeffs and phi/fom* tool to generate an .mtz file that contains HL coefficients from the FOM of the *Phaser-MR* solution (Supplementary Data S9).

53. Merge high-resolution X-ray diffraction data with low-resolution phases. **(Steps 53-65). 5-10 minutes.** Open the *Reflection tools* tab in the *Phenix* GUI *Actions* menu, and select the *Reflections file editor*.

54. Add the .mtz file containing high-resolution X-ray data to the *Input file selection* panel. This should be the same data file used in *Phaser-MR* (step 20, Supplementary Data S1).

55. Add the .mtz file with HL coefficients generated in step 52 to the *input file* box (Supplementary Data S9).

56. Once loaded as input files, the data arrays will be listed automatically in the *All input arrays* box located at the bottom left.

57. Select the array from the X-ray data .mtz file containing structure factors and structure factor errors/sigmas (F and SIGF) or intensities and associated errors (I and SIGI). For example, the array containing F_XDS_dataset, and SIGF_XDS_dataset would be selected from an .mtz file generated by XDS (Supplementary Data S1).

58. Move the selected array to the *Output arrays* box by clicking the “+” at the bottom of the window.

59. Select all arrays associated with the .mtz file containing the HL coefficients in the *All input arrays* box. Move the selected arrays to the *Output* file box.

60. Edit the *Output array* names by selecting the *Edit arrays* button.

61. Array names are usually simplified in the *Output column labels* field (e.g. F_XDSdataset = F, SIGF_XDSdataset = SIGF). In some cases it may be necessary to manually simplify names in the *Output column labels* field.

62. Change the *Output column labels* of the HL array to HLA, HLB, HLC, and HLD.

63. In the *Output options* tab of the *Reflection file editor* tool, ensure that Extend existing R-free array(s) to full resolution range and *Generate R-free flags if not already present* are selected. Leave all other parameters in the *Output options* tab as default.

64. Define the path and name of the new output .mtz. This file will be used for phase extension.

65. Run the *Reflection file editor* to merge the high-resolution amplitudes/intensities from the X-ray data with the low-resolution phases in the *Phaser-MR* solution (Supplementary Data S10).

66. Extend phases to high-resolution using density modification. (Steps 66-77). Usually 10-30 minutes per macrocycle. Open *RESOLVE,* found under the *Maps* tab in the *Phenix* GUI *Actions* menu.

67. Add the merged .mtz file generated in step 65 (Supplementary Data S10) to the *Input files* box. *RESOLVE* will recognize the *data type* of this file as an *Initial map* (shown in the top right of the input files box).

68. Highlight the file and click the *Modify file data type* button. Change the data type so the file is listed as *experimental data* and *initial map*. Select *Input file options*.

69. *RESOLVE* should automatically identify the data labels it needs to run in the merged .mtz file. However, fill the *High-resolution data labels* field with the labels associated with the collected X-ray data (e.g. IOBS and SIGIOBS or FOBS and SIGFOBS) to ensure that the correct array is used for phase extension.

70. Select *Crystal Info* under the *All parameters* pull-down menu, complete the estimated solvent fraction of the crystal, and select *Chain type*. If your structure is a nucleoprotein complex select the chain type that is most abundant.

71. Set the *High-resolution* limit to the highest-resolution shell of the diffraction data. Set the *Low-resolution* limit to 500. Input protein and nucleic acid sequence. The program needs some sequence information to begin the run, but the sequence does not have to be complete. For instance a string of “A’s” is sufficient. Exit the *Crystal info* window by clicking OK.

72. Select NCS from the *All parameters* pull-down menu. In the *NCS parameter* window, load the find_ncs.ncs_spec file, generated in step 43 (Supplementary Data S8). Exit the window by clicking OK.

73. Define the *High-resolution limit* as the highest resolution of the X-ray data (e.g. 3.2).

74. Open the *Density modification* window. Define the *Start resolution for macro-cycle extension* as 0.25 Å less than the high-resolution limit of the EM map. For example, if

the resolution limit of the EM map used in *Phaser-MR* was 8 Å, then place 7.75 Å in the field. Leave the *Step size macro-cycle extension* as 0.25. Select OK to exit the window.

75. In the *Output file* box, define a unique name for the run title. It is helpful for bookkeeping purposes to use the resolution limit in the *Run title*.

76. Define the location of the *Output* directory and run *RESOLVE*. Density modification by *RESOLVE* will run in iterative macro-cycles that extend phases in 0.25 Å increments. After each macro-cycle the .ncs_spec file is refined and applied to the next macro-cycle.

77. Evaluate the phase extension process while running *RESOLVE* by opening the *TEMP0* directory in the *Autobuild_run* folder. After each 0.25 Å macro-cycle a resolve_work.mtz file is deposited in the *TEMP0* directory that can be opened in Coot. Once *RESOLVE* has finished, the final phase-extended map will be deposited as *overall_best_denmod_map_coeffs.mtz* in the *Autobuild_run* folder and the files in the *TEMP0* directory will be automatically deleted. The density should be continuous and secondary structure elements should be recognizable once phases have extended to sub-nanometer resolution (Figure 2). Side chains should begin to appear around resolutions of 4-5 Å (see Supplemental Files S11-S13).

TIMING

Depending on the starting resolution of the EM search ensemble the protocol presented here usually can be completed in an 8-hour period, with phase extension into high-resolution taking up the majority of the time. However, the protocol can be stopped after each section.

Steps 1-17: It takes 5 to 15 minutes to prepare an EM map for *Phaser-MR*, using a combination of *Chimera*, *Phenix* and *Coot*.

Steps 18-38: The set up for *Phaser-MR* usually takes 30 minutes and *Phaser-MR* usually finishes in less than 30 minutes. Longer runs indicate that *Phaser* has not found a solution and the run should be aborted.

Steps 39-43: A find_ncs_from_density search with the default parameters of 20 centers and a delta phi of 10 takes approximately 40-60 minutes. The procedure time can be reduced if the number of search centers is reduced or the delta phi angle is increased.

Steps 44-65: Generating Hendrickson-Lattman coefficients from the MR solution and merging high-resolution diffraction data with low-resolution phases usually takes 5 to 30 minutes.

Steps 66-77: Each macro-cycle of 0.25 Å take 10 to 30 minutes. Depending on the starting and final resolution, a full phase extension may take several hours. For example, a *RESOLVE* run that extends phases from 8 to 3.2 Å with the Cascade tutorial data generally takes 4 to 5 hours.

TROUBLESHOOTING

Molecular phasing using Phaser-MR (Steps 18-38)

Phaser-MR accounts for uncertainty in the magnification error of the original EM map (step 25). However, if *Phaser* fails it may be advantageous to try varying the original scale of the map before the *Phaser* run by modifying the *Cell scale factor* value in the *Ensembles* tab (step 34). Start by changing the *Cell scale factor* in increments of plus or minus 0.03. Additionally, crystal contacts or conformational dynamics may result in conformational differences between molecules in the crystal and cryo-EM structure. Thus, in some cases it might be advantageous to cut out two or three separate pieces of the density by modifying the center and extent values during the cut-out-density procedure (see steps 13-16). Each piece can then be used as an independent search ensemble.

Find NCS-related density (Steps 39-43)

Search parameters can be modified if initial efforts to find_ncs_from_density fail.. The *Smoothing radius* can be lowered to more finely sample the electron density for molecular centers, and the *Maximum number of centers to try* can be increased to add molecular centers to the NCS search (step 41). The *density_radius* parameter default of 20 can be varied to change the size of the sphere of density used in the search procedure. Generally if the radius is higher, delta phi should be lowered. Delta phi can be lowered to search more finely, however this increases search time. Minimum_ncs_cc_from_density can be lowered to find weak NCS. An explanation of the other parameters is available at http://www.phenix-online.org/documentation/reference/find_ncs_from_density.html

The matrices generated by all searched centers are listed in the *Run status* window and can be copied and saved as a new .ncs_spec file. A group of NCS matrices with high CC, but few operators might work better during phase extension than a group with more NCS operators and low CC.

Phase extension (Steps 66-77)

Failure to produce continuous maps with identifiable secondary structure may be caused by a lack of NCS or poor CC between NCS-related densities. To identify additional NCS operators or to improve the CC of NCS-related density, search for NCS in a density modified map. This can be done by repeating steps 39-43 with a density modified .mtz file (e.g. resolve_work.mtz).

If all available NCS-related density is found and maps are still not interpretable, then other methodologies such as multi-crystal averaging, isomorphous replacement, and anomalous dispersion may be required to extend the phases into high-resolution X-ray diffraction data.

Additional troubleshooting advice can be found in Table 1.

ANTICIPATED RESULTS

This protocol is designed to produce electron density maps to the resolution limit of collected X-ray data. The electron density should be continuous and secondary structure

features (i.e. beta-strands and alpha-helices) should be identifiable (Figure 2). To facilitate implementation of this protocol, we have included Supplementary Data files containing 3.2 Å resolution X-ray data (Supplementary Data S1), the 8 Å cryo-EM map of Cascade (Supplementary Data S2), Cascade subunit sequence information (Supplementary Data S3), and intermediate output files (Supplemental Files S4-S13).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

RNJ is supported by the NRSA postdoctoral fellowship (F32 GM108436) from the National Institutes of Health (NIH). RJR and TCT are supported by a grant (GM063210) from the NIH. RJR is supported by a Principal Research Fellowship from the Wellcome Trust (grant no. 082961/Z/07/Z). Research in the Wiedenheft lab is supported by the NIH IDeA Program COBRE (GM110732), an R01 to BW (GM108888), the National Science Foundation EPSCoR (EPS-110134), the M.J. Murdock Charitable Trust, and the Montana State University Agricultural Experimental Station.

REFERENCES

1. Garman EF. Developments in x-ray crystallographic structure determination of biological macromolecules. Science. 2014; 343:1102–1108. [PubMed: 24604194]
2. Rossmann MG, Morais MC, Leiman PG, Zhang W. Combining X-ray crystallography and electron microscopy. Structure. 2005; 13:355–362. [PubMed: 15766536]
3. Bai, X.-c.; McMullan, G.; Scheres, SHW. How cryo-EM is revolutionizing structural biology. Trends Biochem Sci. 2014; 40:49–57. [PubMed: 25544475]
4. Lengyel J, Hnath E, Storms M, Wohlfarth T. Towards an integrative structural biology approach: combining Cryo-TEM, X-ray crystallography, and NMR. J Struct Funct Genomics. 2014; 15:117–124. [PubMed: 24748171]
5. Rossmann MG. Molecular replacement - historical background. Acta Crystallogr D. 2001; 57:1360–1366. [PubMed: 11567146]
6. Stuart DI, Abrescia NG. From lows to highs: using low-resolution models to phase X-ray data. Acta crystallographica. Section D, Biological crystallography. 2013; 69:2257–2265.
7. Xiong Y. From electron microscopy to X-ray crystallography: molecular-replacement case studies. Acta crystallographica. Section D, Biological crystallography. 2008; 64:76–82.
8. Dodson EJ. Using electron-microscopy images as a model for molecular replacement. Acta crystallographica. Section D, Biological crystallography. 2001; 57:1405–1409.
9. Jackson RN, et al. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. Science. 2014; 345:1473–1479. [PubMed: 25103409]
10. Mccoy AJ, et al. Phaser crystallographic software. J Appl Crystallogr. 2007; 40:658–674. [PubMed: 19461840]
11. Wiedenheft B, et al. Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature. 2011; 477:486–489. [PubMed: 21938068]
12. Lowe J, et al. Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution. Science. 1995; 268:533–539. [PubMed: 7725097]
13. Rabl J, et al. Mechanism of gate opening in the 20S proteasome by the proteasomal ATPases. Mol Cell. 2008; 30:360–368. [PubMed: 18471981]
14. Rawat UB, et al. A cryo-electron microscopic study of ribosome-bound termination factor RF2. Nature. 2003; 421:87–90. [PubMed: 12511960]
15. Selmer M, et al. Structure of the 70S ribosome complexed with mRNA and tRNA. Science. 2006; 313:1935–1942. [PubMed: 16959973]

16. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica. Section D, Biological crystallography. 2010; 66:213–221.
17. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D. 2004; 60:2126–2132. [PubMed: 15572765]
18. Winn MD, et al. Overview of the CCP4 suite and current developments. Acta crystallographica. Section D, Biological crystallography. 2011; 67:235–242.
19. Goddard TD, Huang CC, Ferrin TE. Software extensions to UCSF chimera for interactive visualization of large molecular assemblies. Structure. 2005; 13:473–482. [PubMed: 15766548]
20. Kantardjieff KA, Rupp B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. Protein science : a publication of the Protein Society. 2003; 12:1865–1871. [PubMed: 12930986]
21. Weichenberger CX, Rupp B. Ten years of probabilistic estimates of biocrystal solvent content: new insights via nonparametric kernel density estimate. Acta crystallographica. Section D, Biological crystallography. 2014; 70:1579–1588.
22. Zwart P, Grosse-Kunstleve R, Adams P. Xtriage and Fest: Automatic assessment of X-ray data and substructure structure factor estimation. CCP4 News Letter on Protein Crystallography. 2005
23. Zwart PH, Grosse-Kunsteleve RW, Adams PD. Characterization of X-ray data sets. CCP4 newsletter. 2005

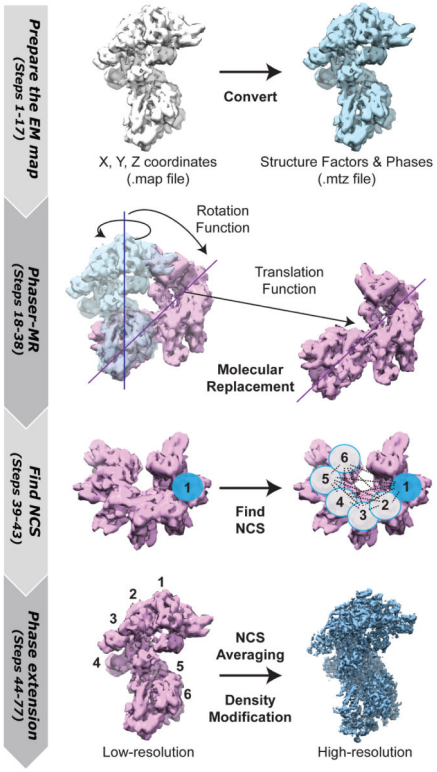

Figure 1. Using low-resolution EM maps to determine high-resolution X-ray structures

The protocol proceeds in four consecutive parts (left arrows). Map coordinates are converted to structure factors and phases (top), Phaser-MR is used to determine low-resolution phases of the crystal structure (second panel), NCS is identified (third panel), and phases are extended to high-resolution with iterative rounds of NCS averaging and density modification (bottom).

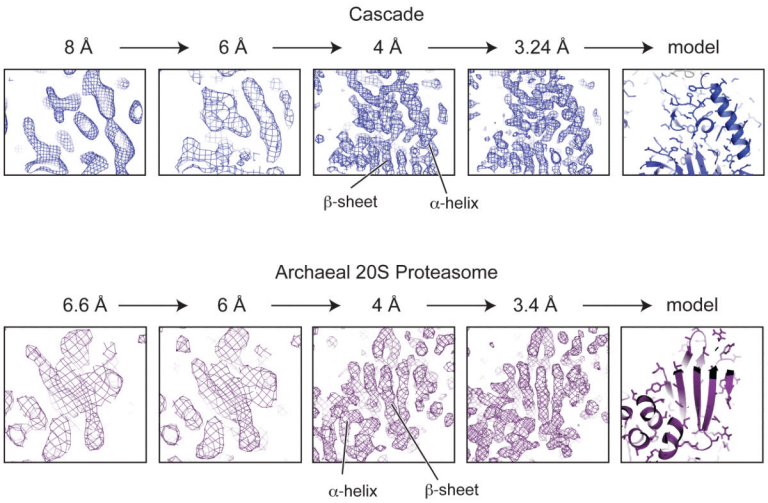

Figure 2. Extending low-resolution phases to high-resolution data

A section of density from the *E. coli* Cascade complex (top) and the *T. acidophilum* 20S proteasome (bottom) is shown at different intervals of phase extension. Initial maps were generated using low-resolution phases from 8 Å and 6.6 Å resolution cryo-EM reconstructions, respectively (left). Phases were extended to 3.24 Å for Cascade and 3.4 Å for the proteasome (right), using NCS averaging and iterative rounds of density modification.

Table 1

Troubleshooting

Step	Problem	Possible Reason	Solution
38	No clear Phaser solution (i.e. TFZ score < 8)	Magnification error.	Change the *Cell scale factor* located in the *Ensembles tab* in increments of 0.03 before running *Phaser* (step 34).
		Conformational differences between the EM and crystal structures.	Use the *Cut out density* tool to cut the cryo-EM map into multiple search ensembles and search for each piece independently.
42	No NCS is found	The density search parameters limit the NCS search.	Modify the find_ncs search parameters. Start by modifying the search radius, delta phi, or number of centers. Also, the *minimum_ncs_cc_from_density* can be lowered to find weak NCS.
77	Continuous high-resolution maps are not produced.	Not all NCS related copies are used in phase extension.	Run steps 39-42 with a density modified map at higher-resolution to identify all NCS-related subunits. Combine NCS groups of non-related subunits into a single .ncs_spec file.