Title

Cryo_fit: Democratization of flexible fitting for cryo-EM

Journal

Journal of Structural Biology, 208(1)

ISSN

1047-8477

Authors

Kim, Doo Nam

Moriarty, Nigel W

Kirmizialtin, Serdal

et al.

Publication Date

2019-10-01

DOI

10.1016/j.jsb.2019.05.012

Peer reviewed

HHS Public Access

Author manuscript

J Struct Biol. Author manuscript; available in PMC 2020 April 01.

Published in final edited form as:

J Struct Biol. 2019 October 01; 208(1): 1–6. doi:10.1016/j.jsb.2019.05.012.

cryo_fit: Democratization of Flexible Fitting for Cryo-EM

Doo Nam Kim[‡], **Nigel W. Moriarty**[†], **Serdal Kirmizialtin**[§], **Pavel V. Afonine**[†], **Billy Poon**[†], **Oleg V. Sobolev**[†], **Paul D. Adams**[†,a], **Karissa Sanbonmatsu**[‡,*]

[‡]Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM, USA

[†]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA, USA

[§]Chemistry Program, Science Division, New York University, Abu Dhabi, United Arab Emirates

[a]Department of Bioengineering, University of California Berkeley, Berkeley, CA, USA

[*]New Mexico Consortium, Los Alamos, NM, USA

Abstract

Cryo-electron microscopy (cryo-EM) is becoming a method of choice for describing native conformations of biomolecular complexes at high resolution. The rapid growth of cryo-EM in recent years has created a high demand for automated solutions, both in hardware and software. Flexible fitting of atomic models to three-dimensional (3D) cryo-EM reconstructions by molecular dynamics (MD) simulation is a popular technique but often requires technical expertise in computer simulation. This work introduces *cryo_fit*, a package for the automatic flexible fitting of atomic models in cryo-EM maps using MD simulation. The package is integrated with the *Phenix* software suite. The module was designed to automate the multiple steps of MD simulation in a reproducible manner, as well as facilitate refinement and validation through *Phenix*. Through the use of *cryo_fit*, scientists with little experience in MD simulation can produce high quality atomic models automatically and better exploit the potential of cryo-EM.

Keywords

Cryo-EM; molecular dynamics; flexible fitting; correlation coefficient; *Phenix*

Corresponding author: Sanbonmatsu, Karissa (kys@lanl.gov).

Author Contributions

K.S., P.D.A., D.K., N.W.M., and S.K. conceived and designed the project. D.K., N.W.M., S.K., B.P. and O.V.S. developed the program. D.K., P.V.A, K.S. and S.K. wrote the manuscript.

Program Development and Availability

Other than the *Phenix* GUI interface and simple command line execution with *cctbx* libraries, all *cryo_fit* code is either embedded in C language based *gromacs* 4.5.5 extensions or as python 2.7 based scripts to automate the input, output and execution of *cryo_fit*. *cryo_fit* can be installed into macOS and linux (CentOS and Ubuntu), and all codes are available at https://github.com/cryoFIT/cryo_fit. The *gromacs_cryo_fit* installation file is located at https://github.com/cryoFIT/cryo_fit_install. Documentation is available at www.phenix-online.org[62].

Introduction

Cryo-electron microscopy (cryo-EM) has been changing the landscape of structural biology. Cryo-EM was selected as the method of the year 2015[1] and its major contributors received the Nobel prize in 2017[2]. Indeed, it has been used to determine biomolecular structures of a wide range of molecular weights (ranging from 64 kDa[3,4] to 150 MDa[5]), achieving up to 1.8 Å resolution[6,7]. Recent discoveries of improved focusing waves[8] and the ability to record super-resolution (beyond the Nyquist frequency)[9] are expected to further improve resolution. Among the four methods of cryo-EM (single particle analysis, cryo-tomography, electron crystallography, and micro-electron diffraction), single particle analysis (SPA) is most popular (78% among all cryo-EM methods)[10].

For SPA, flexible fitting has been an essential tool for better elucidating dynamical aspects of biological complexes[5,11]. These dynamical aspects produce conformational heterogeneity, one of the important causes of lower resolution in cryo-EM reconstructions. If elucidated, this can help reveal crucial structure-function relationships involving dynamics[12]. Flexible fitting has been used to capture dynamics[11] with a larger radius of convergence than standard refinement,[13] as large as 34 Å[14]. Additionally, flexible fitting can describe multiple conformational changes (supporting Fig. S1). One popular technique for flexible fitting employs molecular dynamics simulations based on classical molecular mechanics force fields, which, among many other examples, have been used to simulate tRNA movement into the ribosome[15]. We initially developed a hybrid molecular dynamics method that uses a physics-based all-atom force field together with secondary structure restraints from an initial model to fit to cryo-EM maps. This method[16] is implemented in the *gromacs*[17] suite of programs that provides access to all *gromacs* functionalities such as enhanced sampling. In principle, our method allows one to choose the complexity of the physical model, from a reduced representation to an all-atom representation, while fitting a cryo-EM map.

However, molecular dynamics simulation is a highly technical field in its own right and, like other molecular dynamics based cryo-EM map fitting programs[18,19,20,21,22], the technique has been challenging to use by those who are not experts in molecular dynamics. Therefore, we automated all procedures of our hybrid molecular dynamics method and integrated them into *Phenix*[23], a widely used structural biology software suite with a graphical user interface (GUI). Our new method for streamlined automatic flexible fitting of atomic models into a cryo-EM map is named *cryo_fit*. As a part of *Phenix*, *cryo_fit* makes use of all pre- and post-processing functionalities of *Phenix* providing the user a unified platform for cryo-EM structure construction and interpretation. *Cryo_fit* is particularly useful when attempting to fit novel atomic structures into cryo-EM maps, *e.g.* cases where the starting atomic models are significantly different from the native conformation (Fig. 1, 4). When the initial model does not require large changes to fit the map, refinement can be used. However, if larger changes are required (that are beyond convergence radius of gradient-driven minimization) then *cryo_fit* is recommended (Fig. S2). The *cryo_fit* module is expected to be most useful when an atomic model is approximately placed into the map, and large conformational changes are required for fitting (*i.e.*, model-to-map fit is not sufficient for gradient-based refinement to be successful). Importantly, the current implementation of *cryo_fit* can only handle standard amino- and nucleic-acid based molecules (proteins, RNA, DNA), while non-

standard entities or ligands are excluded from fitting. Excluded entities can be modeled at later stages after the macromolecule has been fitted into the map by *cryo_fit*.

Procedures

The algorithms used in *cryo_fit* are explained in detail in Kirmizialtin et al.[16] Briefly, while *phenix.real_space_refine* sums data-based and restraints-based components in a target function[13], *cryo_fit* uses a molecular dynamics potential function including a term proportional to the correlation between the experimentally determined cryo-EM map and a map calculated from the current model, an *ab initio* molecular dynamics potential, and an all-atom structure-based potential to preserve the stereochemistry of initial models while exploring alternative conformations[16]. Several different measures of the correlation between map and model can be used, as described recently[24]. The correlation coefficient in *cryo_fit* is based on those originally used for cryo-EM fitting by Tama and co-workers and is similar to previous implementations[14,16,19,25]:

$$CC_{cryo\_fit} = \frac{\sum_{i,j,k} \rho_{EM}(i,j,k)\rho_{sim}(i,j,k)}{\sqrt{\sum_{i,j,k} \rho_{EM}(i)^2 \sqrt{\sum_{i,j,k} \rho_{sim}(i)^2}}} \quad (1)$$

$$\rho_{sim}(i,j,k,\mathbf{r}_l) = \exp\left[-\frac{1}{2}\left(\frac{\mathbf{r}_{i,j,k} - \mathbf{r}_l}{\sigma}\right)^2\right] \quad (2)$$

where $\rho_{sim}(i,j,k)$ is a simulated map computed from atomic coordinates using the same grid spacing as the cryo-EM map assuming a Gaussian distribution for each atom defined by:

Here, $\rho_{sim}(i,j,k, \mathbf{r}_l)$ is proportional to the electron density (which is directly related to the electrostatic potential measured in cryo-EM) at grid site (i,j,k) due to atom l at $\mathbf{r}_l$ and 2σ is the resolution of the cryo-EM map. This technique of determining $\rho_{sim}(i,j,k, \mathbf{r}_l)$ is used in other flexible fitting methods and was chosen for ease of calculation since it is used in the MD force field potential[26]; however, we note that more accurate FFT (fast fourier transform)-based methods may be used, such as those implemented elsewhere in *Phenix*[23].

This article is focused on the implementation of *cryo_fit* in *Phenix* and its utility in biomolecular complexes comprised of RNA, DNA and proteins. Similar to other *Phenix* applications, in developing *cryo_fit* we prioritized ease of installation and use. *Cryo_fit* requires an atomic model of a structure (PDB or mmCIF formatted file) approximately placed into a cryo-EM map (ccp4 formatted binary map or a map in *Situs*[27] format).

There are several approaches to build a starting atomic model before using *cryo_fit*. For example, an initial atomic model can be built from a map and sequence information using *phenix.map_to_model*[28]. Alternative approaches such as a homology modeling or a *pathwalker*[29] (among other programs) derived model can be used to provide *cryo_fit* a starting atomic model.

If an initial atomic model is available but is not placed into the map yet, then tools such as *phenix.dock_in_map*[30] can be used to automatically dock a model. Alternatively, *UCSF*

Chimera[31] provides manual translation/rotation of the model and a '*Fit in Map*' function. To allow for easier interpretation, we recommend automatically sharpening or blurring the cryo-EM map with *phenix.auto_sharpen*[32].

Once the initial atomic model and map files are prepared, *cryo_fit* can be executed on the command line,

$$phenix.cryo\_fit\ \langle initial\ atomic\ model\rangle\ \langle cryo\text{-}EM\ map\rangle$$

or in the *Phenix* GUI (supporting Fig. S3).

Cryo_fit automatically performs a number of steps, from processing input files and generating various geometrical restraints to performing MD simulations in order to optimize the model to map fit. The model to map fit is quantified by the correlation coefficient between the cryo-EM map and the simulated map calculated from the atomic model (see Discussion section). *Cryo_fit* runs molecular dynamics in the presence of the cryo-EM term (proportional to the correlation between the experimental electric potential map and a density map calculated from the atomic model) with neutralized charges as a default option to improve speed.

Domain decomposition[33] and molecular dynamics time steps are automatically assigned[34] to ensure numerical stability over a large numbers of time steps. Depending on the computing hardware used, *cryo_fit* automatically utilizes multiple cores for optimal performance. Automatic choice of the relative weighting between the map and geometry restraints is used while maintaining acceptable stereochemistry. By default, *cryo_fit* uses simulated annealing in combination with molecular dynamics to escape local energy minima effectively[16]. The benefit of this was verified by our benchmark with five pairs of deposited models and maps (unpublished). *Cryo_fit* outputs the top three fitted models that have the highest model-to-map correlation, as well as molecular dynamics trajectory files, which can be visualized using *PyMOL*[35], *UCSF Chimera*[31], *UCSF ChimeraX*[36], and *VMD*[37] (an example trajectory is given in supporting information movie S1). Analyzing MD trajectories may be particularly useful to reveal possible conformational changes. The resulting fitted model may require some further refinement using *phenix.real_space_refine*[13] to improve model geometry (*e.g.*, eliminate residue rotamer or Ramachandran plot outliers). Continuous model and model-to-map fit validation[24] is also desirable.

Results

Cryo_fit has been applied to a variety of systems, as small as the GTPase associated center (GAC) of the ribosome (~6,000 atoms) and as large as the intact ribosome (~200,000 atoms). *Cryo_fit* can fit both protein and nucleic acid atomic models into cryo-EM maps (Fig. 2–3). Running *cryo_fit* with default parameters was sufficient to improve cross-correlation values relative to a *UCSF Chimera* ('Fit in Map') fitted model of a protein and DNA complex (H1.0b-bound 197 bp nucleosome; PDB code 5nl0) from the Protein Data Bank (PDB)[38] (Fig. 3).

For fittings that require large conformational changes (RMSD by *PyMOL* ≥ 7.0 Å), we calculated CC_{cryo_fit} and CC_{mask} with 4 benchmark pairs of starting models and maps (Fig. 4). The value of 7.0 Å was chosen by clustering the data according to RMSD. For *phenix.real_space_refine*, we turned on secondary structure restraints that are determined by default. The workflow that resulted in the highest correlation, on average, was the application of *phenix.cryo_fit* followed by the application of *phenix.real_space_refine*. This was true when measuring correlation with CC_{cryo_fit} and when measuring the correlation with CC_{mask}.

For fittings that require smaller conformational changes (RMSD by *PyMOL* < 7.0 Å), we calculated CC_{cryo_fit} and CC_{mask} with 8 benchmark pairs of starting models and maps (Fig. 5). Other than one pair (1gru/1046) whose cryo-EM map has low resolution (24.0 Å), *cryo_fit* produced either similar or improved fits. Thus, independent of the method that the original authors used (manual/semi-automatic/automatic), *cryo_fit* produces either comparable or better results in terms of fitting the cryo-EM map. Note that most of these fittings are achieved with default settings that can be changed to improve the cross-correlation scores based on the cryo-EM map properties. When measuring correlation with CC_{cryo_fit}, the workflow of *phenix.cryo_fit* followed by *phenix.real_space_refine* resulted in the highest correlation values on average (0.60). In these same cases, when measuring by CC_{mask}, *phenix.real_space_refine* alone resulted in the highest fit (0.70). This means that as we expected, gradient-based refinement only is sufficient for local fittings that require smaller conformational changes (Supporting Fig. S2). It should be noted that closer comparisons can be made by using *phenix.real_space_refine* with simulated annealing enabled.

Discussion

In our test cases, *cryo_fit* successfully fit most protein, DNA and RNA systems into cryo-EM maps. Here, we note that *cryo_fit* can be extended to lipids and carbohydrates when the parameters are added to the force field, which will be implemented in the future. Currently, unusual/unknown residues (such as *UNK*) or ligands are excluded from fitting[39].

As previously shown (Fig. 5), *cryo_fit* does not always find a better cross-correlation (CC) value than other methods. During the process of development and implementation of cryo-EM map fitting software, we often found that large improvements of the correlation (both CC_{cryo_fit} and CC_{mask}) can degrade the model geometry, while severe restrictions in model geometry may also degrade the correlation. This necessitates a balance between the weight between the CC_{cryo_fit} term and model geometry in the MD potential. However, as we assume that model geometry is adequate (as we have seen in supporting Table S1–2), these particular cases (*i.e.*, when the CC value is not significantly increased even by *cryo_fit*) can be understood as follows: 1) the initial model is already well fit to the given cryo-EM map; 2) the cryo-EM map has low resolution; 3) the cryo-EM map needs to be filtered to remove unresolved regions; 4) the system is trapped in local energy/configuration minima and requires more time steps than those used in the default scheme for fitting, or 5) *cryo_fit* reduces overfitting, very much like crystallography modeling increases R_{work} to reduce R_{free}-R_{work} gap.

One can forcefully increase cross-correlation by using a very high map weight. However, this may result in overfitting, distorting the stereochemistry of the model. The user can specify (*e.g.*, in the case of an intrinsically disordered domain) high values of *emweight_multiply_by*. Then, *cryo_fit* will start with very high bias toward the cryo-EM map. In the case of low-resolution maps, it is reasonable to rely on the physically realistic force field more than forcing the model to fit into a less informative map. In certain cases, preceding the final fitting of the actual cryo-EM map by a preliminary step, fitting to a filtered map can be used to further improve cross-correlation values[1619]. By filtering out less-resolved/dynamic region/map, the fitting procedure helps to produce large conformational changes (when necessary) by decreasing the number of local minima; the output of this step can then be used to fit the unfiltered map. When the system is trapped in local minima, longer simulation times can also be helpful (*e.g.*, a larger value for *number_of_steps_for_cryo_fit*). In addition, one can perform local refinement on problematic regions using software packages such as *ISOLDE*[40], a *ChimeraX*[36] bundled package that allows a user to directly tug atoms with a mouse interface or applying/removing restraints for manual inspection and fixing local errors. *Rosetta* based[4142] or *rosetta*+MD simulation based[43] approaches may also be useful in this case.

Interestingly, the various measures of correlation between the modeled atomic structure and the cryo-EM map report on different aspects of agreement and disagreement between model and map. A comprehensive benchmark study examined correlations related to (i) highest value points the model-calculated map (CC_{volume} and CC_{peaks}), (ii) masked regions of the map with the capability of more localized fits (CC_{mask}), and (iii) a measure covering the entire map (CC_{box}), in an analogous fashion to CC_{cryo_fit}[24]. CC_{mask} provides an excellent method for more localized fits since CC_{mask} can use map values inside a mask calculated around the macromolecule only. CC_{box} and CC_{cryo_fit} can be useful in the cases of large-scale conformational changes, where the trajectory of the molecule is not known *a priori* (*e.g.*, > 50Å motions of tRNAs required to fit intermediate states of tRNAs from the classical states of tRNAs within the ribosome). Generally speaking, CC_{box} is useful for determining how well a given atomic model matches the entire volume of the map. For example, if only one chain is modeled, while the full system consists of many chains, then CC_{box} will depict this. Just as *cryo_fit* uses the entire map for the CC calculation (*e.g.* $CC_{cryo\_fit,}$ Kirmizialtin et al.[16] eq. 4), most MD simulation based cryo-EM map fitting programs use correlation measures similar to Eq. 1[192125] or related equations (*e.g.*, similar to Eq. 1, but subtracting the average from ρ_{sim} and ρ_{EM})[4445]. It should be emphasized the CC_{mask} is the standard measure of model-to-map fit that is a more focused, localized and is a more accurate method of measuring the correlation.

To accelerate scientific discoveries of biological molecular structures and dynamics, all aspects of cryo-EM need to be readily accessible to structural biologists. This will require consistent and reproducible outcomes, and procedural automations, such as automated blotting[46], direct ice thickness measurement in the single particle collection workflow[47], prediction of image-quality of particles[48], automated data acquisition[495051], beam induced motion correction[52], contrast transfer function correction[53,54], map reconstruction and refinement[125556], determination of map local resolution[57], map segmentation[58], map sharpening[32], atomic model building[2829], comparison of time-averaged density maps from

molecular dynamics simulations[44], atomic model refinement[1359], and atomic model validation[6061]. It is hoped that *cryo_fit* helps democratize global and local flexible map fitting procedures in these automation endeavors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Authors appreciate beta-testers of *cryo_fit* and their valuable feedbacks. This work was supported by NIH grants: NIGMS R01-GM072686 (to KS), and NIGMS P01-GM063210 (to PDA). This work was supported in part by US Department of Energy under Contract No. DE-AC02-05CH11231 (Lawrence Berkeley National Laboratory). KS and DK were supported in part by LANL LDRD for portions of *cryo_fit*. SK was supported by NYU Abu Dhabi Faculty Research Fund (ST181).

References

1. Method of the Year 2015. Nat. Methods 13, 1 (2015).
2. Shen PS The 2017 Nobel Prize in Chemistry: cryo-EM comes of age. Anal. Bioanal. Chem 410, 2053–2057 (2018). [PubMed: 29423601]
3. Khoshouei M, Radjainia M, Baumeister W & Danev R Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. Nat. Commun 8, 1–6 (2017). [PubMed: 28232747]
4. Danev R & Baumeister W Expanding the boundaries of cryo-EM with phase plates. Curr. Opin. Struct. Biol 46, 87–94 (2017). [PubMed: 28675816]
5. Murata K & Wolf M Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. Biochim. Biophys. Acta - Gen. Subj 1862, 324–334 (2018). [PubMed: 28756276]
6. Chiu W & Downing KH Editorial overview : Cryo Electron Microscopy : Exciting advances in CryoEM Herald a new era in structural biology. Curr. Opin. Struct. Biol 46, 1–5 (2017). [PubMed: 28342396]
7. Earl LA, Falconieri V, Milne JLS & Subramaniam S Cryo-EM : beyond the microscope. Curr. Opin. Struct. Biol 46, 71–78 (2017). [PubMed: 28646653]
8. Lou K, Granick S & Amblard F How to better focus waves by considering symmetry and information loss. Proc. Natl. Acad. Sci 115, 6554–6559 (2018). [PubMed: 29899145]
9. Afanasyev P et al. Single-particle cryo-EM using alignment by classification (ABC): The structure of Lumbricus terrestris haemoglobin. IUCrJ 4, 678–694 (2017).
10. Lawson CL et al. EMDataBank unified data resource for 3DEM. Nucleic Acids Res. 44, D396–D403 (2016). [PubMed: 26578576]
11. Kim DN & Sanbonmatsu K Tools for the Cryo-EM Gold Rush: Going from the cryo-EM map to the atomistic model. Biosci. Rep 37, BSR20170072 (2017).
12. Ludtke SJ Single-Particle Refinement and Variability Analysis in EMAN2.1. Methods Enzymol. 579, 159–189 (2016). [PubMed: 27572727]
13. Afonine PV et al. Real-space refinement in Phenix for cryo-EM and crystallography. Acta Crystallogr. Sect. D Struct. Biol D74, 531–544 (2018).
14. Ahmed A, Whitford PC, Sanbonmatsu KY & Tama F Consensus among flexible fitting approaches improves the interpretation of cryo-EM data. J. Struct. Biol 177, 561–570 (2012). [PubMed: 22019767]
15. Sanbonmatsu KY, Joseph S & Tung C-S Simulating movement of tRNA into the ribosome during decoding. Proc. Natl. Acad. Sci 102, 15854–15859 (2005). [PubMed: 16249344]
16. Kirmizialtin S, Loerke J, Behrmann E, Spahn CMT & Sanbonmatsu KY Using molecular simulation to model high-resolution cryo-EM reconstructions. Methods Enzymol. 558, 497–514 (2015). [PubMed: 26068751]

17. Hess B, Uppsala S- & Lindahl E GROMACS 4 : Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J. Chem. Theory Comput 4, 435–447 (2008). [PubMed: 26620784]
18. Singharoy A et al. Molecular dynamics-based refinement and validation for sub-5 Å cryo-electron microscopy maps. Elife 5, 1–33 (2016).
19. Ratje AH et al. Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. Nature 468, 713–716 (2010). [PubMed: 21124459]
20. Blau C densfit. Available at: https://www.mpibpc.mpg.de/grubmueller/densityfitting.
21. Mori T et al. Acceleration of cryo-EM Flexible Fitting for Large Biomolecular Systems by Efficient Space Partitioning. Structure 27, 1–14 (2019). [PubMed: 30605659]
22. Igaev M, Kutzner C, Bock LV, Vaiana AC & Grubmüller H Automated cryo-EM structure refinement using correlation-driven molecular dynamics. Elife 8, (2019).
23. Adams PD et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D Biol. Crystallogr 66, 213–221 (2010). [PubMed: 20124702]
24. Afonine PV et al. New tools for the analysis and validation of Cryo-EM maps and atomic models. Acta Crystallogr. Sect. D Struct. Biol D74, 814–840 (2018).
25. Orzechowski M & Tama F Flexible fitting of high-resolution x-ray structures into cryoelectron microscopy maps using biased molecular dynamics simulations. Biophys. J 95, 5692–5705 (2008). [PubMed: 18849406]
26. Zheng W Accurate flexible fitting of high-resolution protein structures into cryo-electron microscopy maps using coarse-grained pseudo-energy minimization. Biophys. J 100, 478–488 (2011). [PubMed: 21244844]
27. Wriggers W Using Situs for the integration of multi-resolution structures. Biophys. Rev 2, 21–27 (2010). [PubMed: 20174447]
28. Terwilliger TC, Adams PD, Afonine PV & Sobolev OV A fully automatic method yielding initial models from high-resolution cryo-electron microscopy maps. Nat. Methods 15, 905–908 (2018). [PubMed: 30377346]
29. Chen M, Baldwin PR, Ludtke SJ & Baker ML De Novo modeling in cryo-EM density maps with Pathwalking. J. Struct. Biol 196, 289–298 (2016). [PubMed: 27436409]
30. Terwilliger Tom. phenix.dock_in_map. Available at: https://www.phenix-online.org/documentation/reference/dock_in_map.html.
31. Pettersen EF et al. UCSF Chimera - A visualization system for exploratory research and analysis. J. Comput. Chem 25, 1605–1612 (2004). [PubMed: 15264254]
32. Terwilliger TC, Sobolev OV, Afonine PV, Adams PD & IUCr. Automated map sharpening by maximization of detail and connectivity. Acta Crystallogr. Sect. D Struct. Biol 74, 545–559 (2018). [PubMed: 29872005]
33. Domain decomposition decomposes the component of the non-bonded interactions into domains that share spatial locality. This is useful for multi-core runs. Available at: http://ftp.gromacs.org/pub/manual/manual-4.5.4.pdf.
34. Molecular dynamics time step which is denoted as ‘dt’ in gromacs and ‘time_step_for_cryo_fit’ in cryo_fit often allows more stable running when it is smaller.
35. The PyMOL Molecular Graphics System, Schrödinger, LLC.
36. Goddard TD et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci. 27, 14–25 (2018). [PubMed: 28710774]
37. Humphrey W, Dalke A & Schulten K VMD: Visual Molecular Dynamics. J. Mol. Graph 14, 33–38 (1996). [PubMed: 8744570]
38. Bednar J et al. Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. Mol. Cell 66, 384–397 (2017). [PubMed: 28475873]
39. If needed for their biological targets, we would encourage any amber03 forcefield experts to update current cryo_fit forcefield and share with the community.
40. Croll TI ISOLDE : a physically realistic environment for model building into low-resolution electron-density maps. Acta Crystallogr. Sect. D Struct. Biol 74, 519–530 (2018). [PubMed: 29872003]

41. Rosetta modeling of cryo-EM data on the cloud. Available at: http://cryoem-tools.cloud/rosetta-aws/.
42. Meilar lab, Fitting molecules in low resolution electron density maps. Available at: http://www.meilerlab.org/research/show/w_text_id/19.
43. Leelananda SP & Lindert S Iterative Molecular Dynamics-Rosetta Membrane Protein Structure Refinement Guided by Cryo-EM Densities. J. Chem. Theory Comput 13, 5131–5145 (2017). [PubMed: 28949136]
44. Briones R, Blau C, Kutzner C, de Groot BL & Aponte-Santamaría C GROmaps: a GROMACS-based toolset to analyse density maps derived from molecular dynamics simulations. Biophys. J 116, 1–8 (2019). [PubMed: 30558887]
45. Trabuco LG, Villa E, Mitra K, Frank J & Schulten K Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics. Structure 16, 673–683 (2008). [PubMed: 18462672]
46. Dandey VP et al. Spotiton: New features and applications. J. Struct. Biol 202, 161–169 (2018). [PubMed: 29366716]
47. Rice WJ et al. Routine determination of ice thickness for cryo-EM grids. J. Struct. Biol (2018).
48. Han B-G, Watson Z, Cate JHD & Glaeser RM Monolayer-crystal streptavidin support films provide an internal standard of cryo-EM image quality. J. Struct. Biol 200, 307–313 (2017). [PubMed: 28259651]
49. Suloway C et al. Automated molecular microscopy: The new Leginon system. J. Struct. Biol 151, 41–60 (2005). [PubMed: 15890530]
50. Mastronarde DN Automated electron microscope tomography using robust prediction of specimen movements. J. Struct. Biol 152, 36–51 (2005). [PubMed: 16182563]
51. Li X, Zheng S, Agard DA & Cheng Y Asynchronous data acquisition and on-the-fly analysis of dose fractionated cryoEM images by UCSFImage. J. Struct. Biol 192, 174–178 (2015). [PubMed: 26370395]
52. Li X et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat. Methods 10, 584–590 (2013). [PubMed: 23644547]
53. Mindell JA & Grigorieff N Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol 142, 334–347 (2003). [PubMed: 12781660]
54. Zhang K Gctf: Real-time CTF determination and correction. J. Struct. Biol 193, 1–12 (2016). [PubMed: 26592709]
55. Nakane T, Kimanius D, Lindahl E & Scheres Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. bioRxiv (2018).
56. Punjani A, Rubinstein JL, Fleet DJ & Brubaker MA cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods 14, 290–296 (2017). [PubMed: 28165473]
57. Vilas JL et al. MonoRes: Automatic and Accurate Estimation of Local Resolution for Electron Microscopy Maps. Structure 26, 337–344 (2018). [PubMed: 29395788]
58. Terwilliger TC, Adams PD, Afonine PV, S. O Map segmentation, automated model-building and their application to the Cryo-EM Model Challenge Thomas. J Struct Biol. 18, 30193 (2018).
59. Wang RYR et al. Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. Elife 5, 1–22 (2016).
60. Williams CJ et al. MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. 27, 293–315 (2018). [PubMed: 29067766]
61. Barad BA et al. EMRinger: Side chain-directed model and map validation for 3D cryo-electron microscopy. Nat. Methods 12, 943–946 (2015). [PubMed: 26280328]
62. Kim DN phenix.cryo_fit. Available at: https://www.phenix-online.org/documentation/reference/cryo_fit.html.
63. Moriarty NW Editor's Note. Comput. Crystallogr. Newsl 6, 26 (2015).
64. Müller CW & Schulz GE Structure of the complex between adenylate kinase from Escherichia coli and the inhibitor Ap5A refined at 1.9 Å resolution. A model for a catalytic transition state. J. Mol. Biol 224, 159–177 (1992). [PubMed: 1548697]

65. Müller CW, Schlauderer GJ, Reinstein J & Schulz GE Adenylate kinase motions during catalysis: An energetic counterweight balancing substrate binding. Structure 4, 147–156 (1996). [PubMed: 8805521]

66. Smog. Available at: http://smog-server.org/extension/MDfit.html.

Highlights

- Molecular dynamics based atomic model fitting into cryo-EM map is automated.
- This automated program, *cryo_fit*, is integrated into the *Phenix* suite.
- *cryo_fit* enables larger conformational changes relative to rigid fitting and refinement.

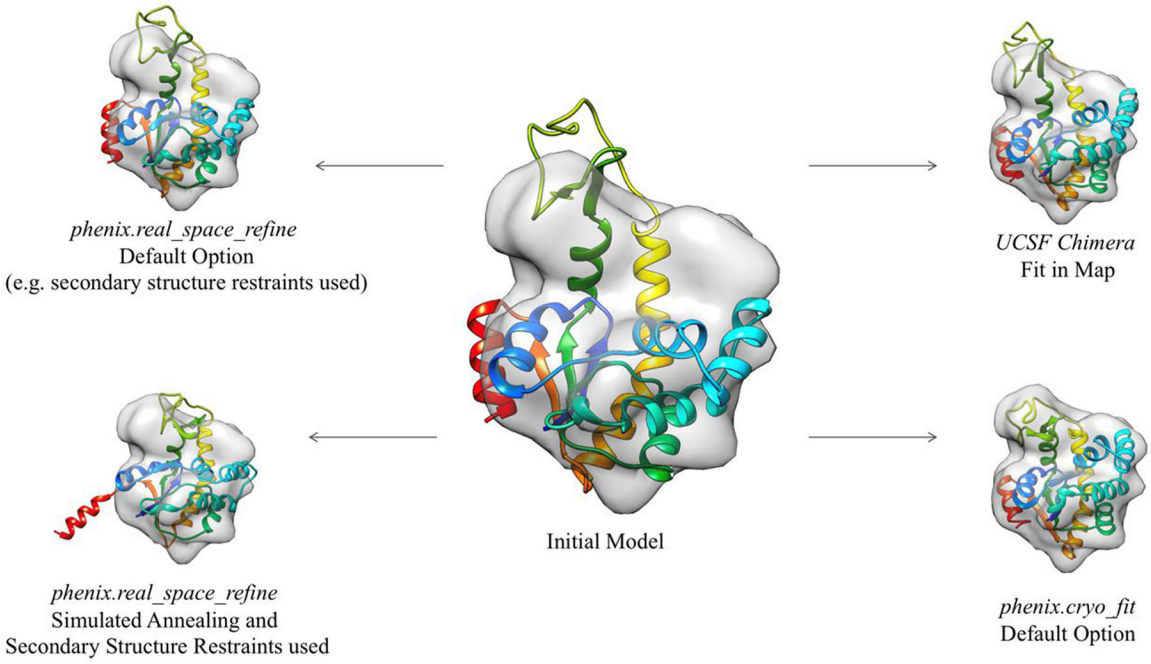

Figure 1.

Comparison of different programs for adenylate kinase (PDB code 1ake, 4ake) atomic model fitting.

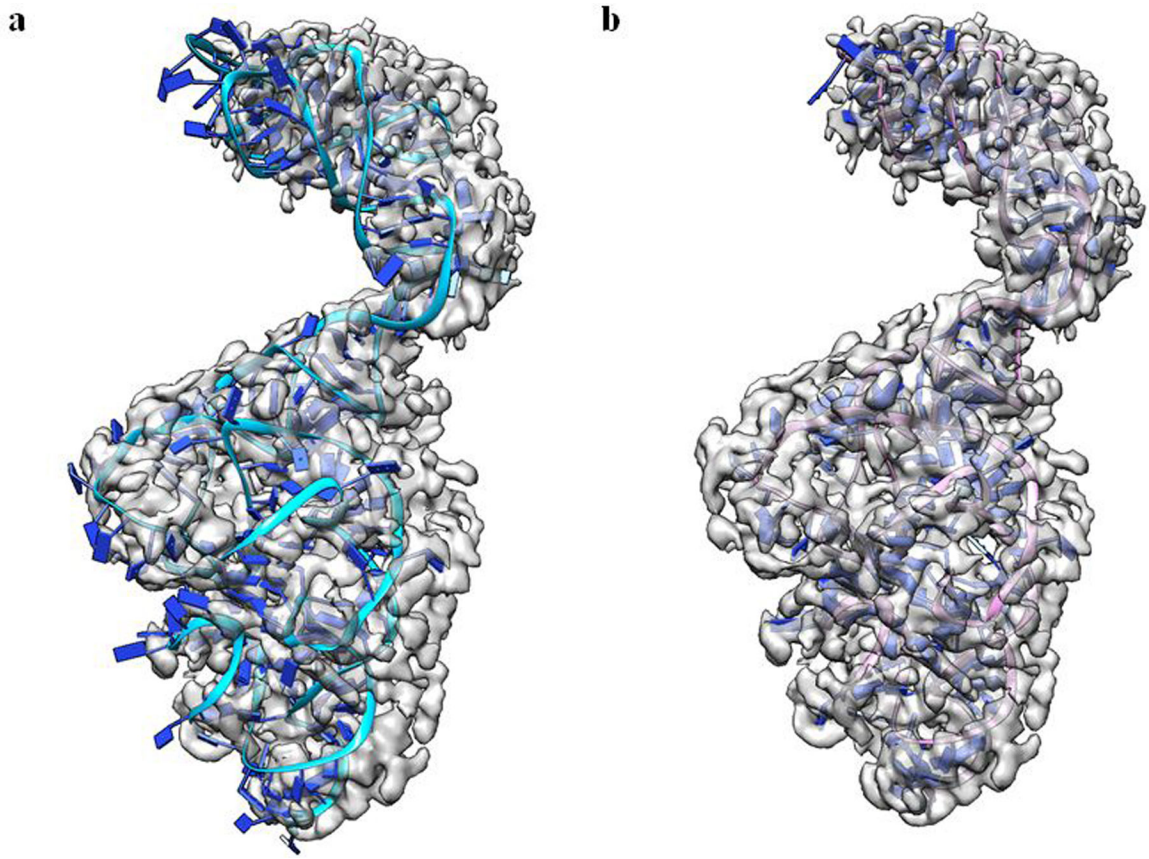

Figure 2.

Atomic models of a H40–44 region (GTPase activation center) from human ribosome. Map resolution = 3.2 Å. (a) Model before flexible fitting by *cryo_fit* (CC_{cryo_fit} = 0.53) and (b) after flexible fitting by *cryo_fit* (CC_{cryo_fit}=0.69). Atomic model and map data are courtesy of Christian Spahn group which were shown in Kirmizialtin et al. (2015).[16]

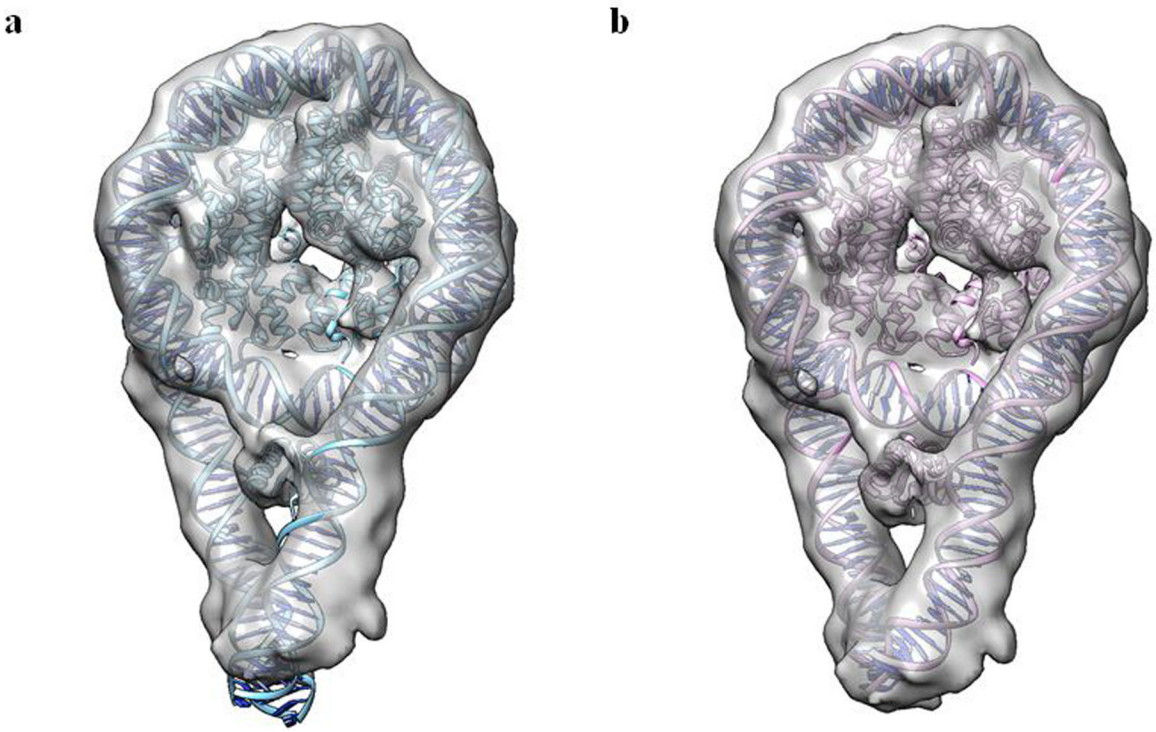

Figure 3.

Atomic models of a nucleosome[38] (PDB code 5nL0, EMDB code: 3659). (a) *Chimera* fitted model of PDB deposited model. (b) Same model after flexible fitting by *cryo_fit*. CC_{cryo_fit} values are 0.735 and 0.747, before and after fitting, respectively. *Cryo_fit* successfully fits DNA region (bottom region) into cryo-EM map.

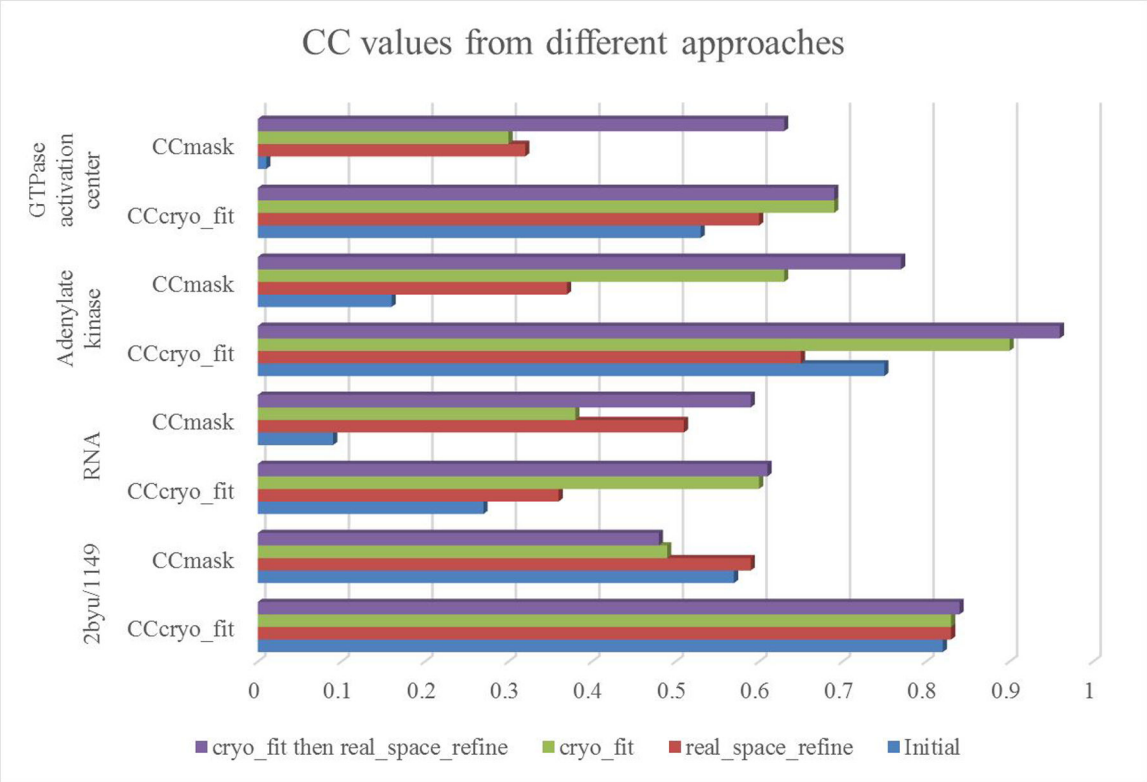

Figure 4.

Fittings for maps that require large conformational changes

Refinement statistics are in supporting Table S1.

The maps used for these cases have 3.4, 10.5, 13.9 and 17.0 Å resolutions respectively (calculated by *phenix.mtriage*). *phenix.real_space_refine* significantly improves CC_{mask} both for all cases. While *phenix.real_space_refine* slightly decreases CC_{cryo_fit} for adenylate kinase, this may be due to a short perturbation at the initial steps of *cryo_fit* required to calculate CC_{cryo_fit} may have slightly decreased CC_{cryo_fit}. We note that this was the only case which resulted in a decrease of CC_{cryo_fit} more than 0.01.

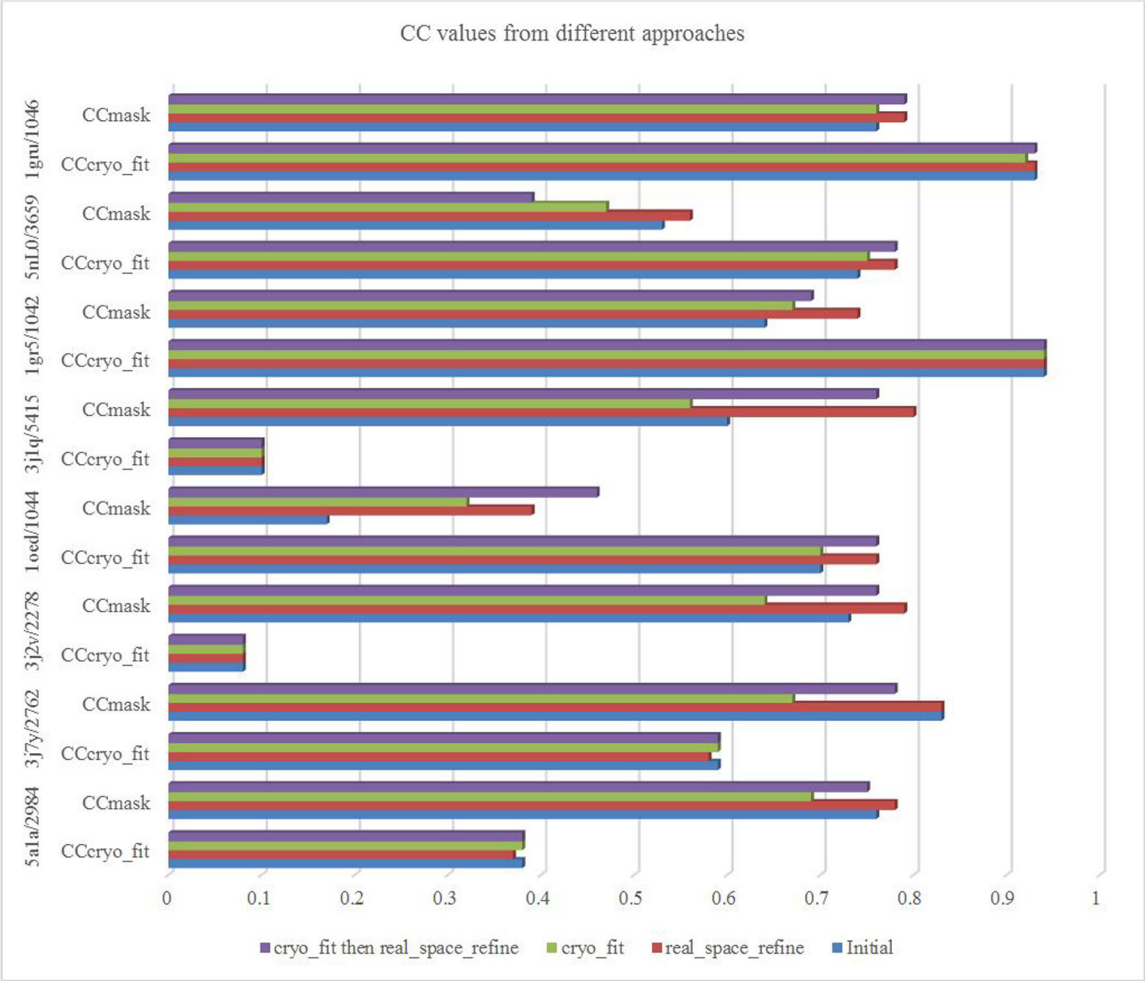

Figure 5.

Fittings for maps that require small conformational changes

On the y axis, we show PDB code/EMDB id. We followed human readable convention for PDB code throughout this paper[63]. Maps used for these cases have 2.2 ~ 24.0 Å resolutions (reported in EMDB). Refinement statistics are in supporting Table S2. For 3j1q/5415 and 3j2v/2278 pairs, the maps are much larger than the atomic model, therefore CC_{cryo_fit} value are calculated as quite low. For 5nL0/3659 pair, we applied *UCSF Chimera* 'fit in map' to the deposited model.