



## Deriving and refining atomic models in crystallography and cryo-EM: the latest *Phenix* tools to facilitate structure analysis

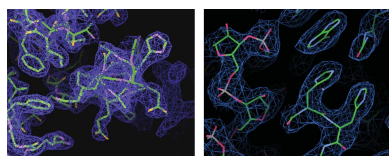
Bruno P. Klaholz\*

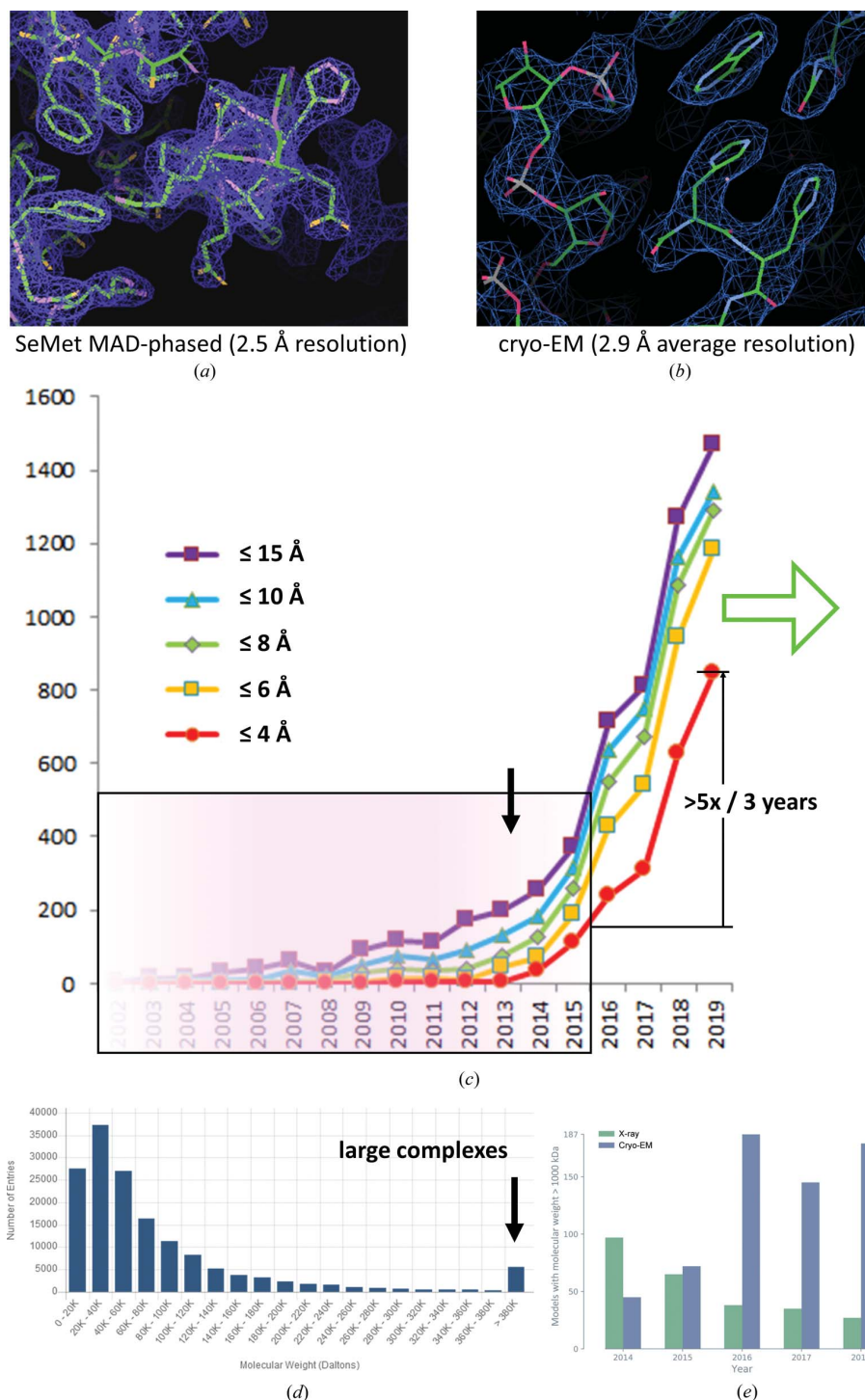
Centre for Integrative Biology (CBI), Department of Integrated Structural Biology, IGBMC, CNRS, Inserm, Université de Strasbourg, 1 rue Laurent Fries, Illkirch 67404, France, Institute of Genetics and of Molecular and Cellular Biology (IGBMC), 1 rue Laurent Fries, Illkirch, France, Centre National de la Recherche Scientifique (CNRS), UMR 7104, Illkirch, France, Institut National de la Santé et de la Recherche Médicale (Inserm), U964, Illkirch, France, and Université de Strasbourg, Illkirch, France. \*Correspondence e-mail: klaholz@igbmc.fr

**Keywords:** structural biology; atomic model refinement; crystallography; cryo electron microscopy; *Phenix*.

Over the last few years, the technical breakthrough of single-particle cryo electron microscopy (cryo-EM) thanks to the recent developments of direct electron detectors and advanced image processing software (for reviews see, for example, Orlova & Saibil, 2011; Stark & Chari, 2016; Orlov *et al.*, 2017; Kim *et al.*, 2018; Ognjenović *et al.*, 2019), has in many cases allowed the 3 Å resolution range to be reached (Fig. 1). In X-ray crystallography, this resolution range is generally considered to be required for deriving atomic models with properly defined geometrical properties, such as the peptide backbone and side-chain geometry of amino acids and nucleotides in macromolecular complexes. The reason why deriving reliable atomic models is so important is that it provides the basis for the detailed analysis of the three-dimensional structures of interest, such as nucleoprotein complexes in the cell nucleus, membrane proteins or viruses and various drug targets. An atomic model containing flaws because of incorrect or insufficient refinement may lead to incorrect conclusions on the detailed interpretation of the structure, with direct implications on the analysis of interactions between residues. For example, the accurate description of hydrogen bonds (which relies both on proper distances and angular orientations of the acceptor and donor, *e.g.* Klaholz & Moras, 2002; Coulocheri *et al.*, 2007) are directly relevant for the analysis of molecular recognition or catalysis events, specificity of drug interactions, effects of point mutants *etc.*, and the precise depiction of base-pairing and stacking interactions between nucleotide bases is essential for the analysis of RNA and DNA complexes (Leontis & Westhof, 2001).

Three-dimensional maps obtained from X-ray diffraction describe the spatial electron-density distribution, while cryo-EM maps are electrostatic potential (ESP) maps as a result of the charged nature of the electrons used to image the biological sample. This difference can have important implications for the analysis of charged residues and ions (Wang & Moore, 2017; Hryc *et al.*, 2017; Wang *et al.*, 2019), but the general properties of these maps are similar such that analogous tools can be used to build and refine atomic models into them (Fig. 1). In the X-ray crystallography field, a series of software packages are available (see, for example, [https://www.rcsb.org/pdb/static.do?p=software/software\\_links/crystallography.html](https://www.rcsb.org/pdb/static.do?p=software/software_links/crystallography.html)), some of which have been more specifically adapted to allow them to be used on cryo-EM maps [*e.g.* *REFMAC* (Brown *et al.*, 2015), *Buster* (Smart *et al.*, 2012) or *Phenix* (Afonine, Poon *et al.*, 2018), see also overall procedures described in Natchiar *et al.* (2017b)], while others originated more from the cryo-EM field [see review by Malhotra *et al.* (2019), and references therein; Pintilie & Chiu (2018)]. This commonly involves real-space refinement because the cryo-EM maps can be used to refine the atomic model directly, without modifying the map (*i.e.* they intrinsically comprise experimental phases). ‘Structure’ refinement in cryo-EM means a map refinement that primarily comprises particle centring (translation and rotation) and Euler angle assignment to iteratively improve the 3D reconstruction (technically, this is a back-projection obtained from individual 2D particle views), but once the map is fully refined, it is not modified further apart from applying a high-pass filter to better visualize high-resolution features (if present, otherwise this only increases the noise level). By contrast, for diffraction data the map is iteratively modified and refined using phase information derived from the atomic model that is under refinement (phase information can also come from experimental phasing, such as native sulfur phasing, SAD, MIR *etc.*, but it




**Figure 1**

(a) X-ray crystallography electron-density map: SeMet (selenomethionine) MAD (multiple-wavelength anomalous dispersion), phased at 2.5 Å resolution (translation initiation factor 2, IF2; PDB ID 4b3x; Simonetti *et al.*, 2013). Note the histidine residue defined by the map. (b) Cryo-EM map: obtained by 3D reconstruction from individual 2D particle images (60S ribosomal subunit of the 80S human ribosome, 2.9 Å average resolution with local resolution extending this; Natchiar *et al.*, 2017a). Note the histidine residues defined by the map (to be compared with panel a) and the nucleotides in the vicinity. (c) Increasing number of cryo-EM maps deposited in the EMDB and achieving the specified resolution levels. The data are taken from the <http://www.ebi.ac.uk/pdbe/emdb/> and <http://www.rcsb.org/pdb/> websites [updated graph compared with the one shown in Orlov *et al.* (2017) as of 26 September 2016, within the shaded light pink box; and as of 20 September 2019], illustrating an over fivefold increase in cryo-EM structures at high resolution (4 Å or better, red curve) within the last three years. The black arrow marks the year 2013 where high-sensitivity detectors entered the cryo-EM field. (d) Graph showing the PDB data distribution by molecular weight. While most structures lie below 60 kDa and are determined by X-ray crystallography, those at high molecular weight (right-hand end) are more amenable to cryo-EM, although complexes in the 50–150 kDa range can now be targeted by cryo-EM as well. (e) Graph from Liebschner *et al.* (2019) in this issue, illustrating that since 2015, cryo-EM depositions have accounted for the majority of large macromolecular structures currently in the PDB.

rarely extends to high resolution and it often needs to be combined with model-derived phases). In the cryo-EM field, various software has been developed in the past for the analysis of low to medium resolution maps (including flexible fitting *etc.*), which is not the focus of this commentary because specifically for high-resolution maps (regardless of whether they come from crystallography or cryo-EM), it is essential to refine the detailed geometry of the model and validate it for data deposition into the appropriate databases (PDB, EMDB and associated databases) to which cryo-EM is increasingly contributing (Fig. 1).

Here, we comment on the article by Liebschner *et al.* in this issue of *Acta Cryst. D* (Liebschner *et al.*, 2019), which describes the various tools available in the software suite *Phenix*, including the most recent developments, thus providing a comprehensive and extensive description of the latest version. The major aim is to provide any user with informatics tools, including robust default settings, which allow a high level of automation. This is not only to simplify the work, but also contributes to reducing errors in refinements because iteration between automatic refinement and manual model building/validation are often required. The article addresses the challenge of listing and briefly describing all the main parts of the program suite, which can handle X-ray and neutron diffraction data, and cryo-EM data. The article will be useful for any reader, specialist or newcomer: it gives an overview of the structure determination steps, the specifics of structure determination from X-ray and neutron diffraction data such as crystal twinning analysis, native Patterson functions, SAD phasing and related methods based on the presence of an anomalous signal, molecular replacement *etc.*, refinement of atomic models into maps obtained using diffraction or cryo-EM methods, and new specific tools for cryo-EM map interpretation. As a software package, *Phenix* integrates all of these aspects, which is a major achievement and is very helpful for the community. As a suggestion to both the crystallography and cryo-EM fields, and considering that the resolution levels reached in cryo-EM nowadays allow the derivation of detailed atomic models, one should probably be more cautious with regards to the confusing usage of the term 'model', which implies the *atomic* model (the model being built into the map), while in the cryo-EM field the term is often used to mean the cryo-EM map itself. A suggestion would be to specify 'atomic' when we speak about atomic models in general, and in cryo-EM avoid the term 'model', instead using the terms cryo-EM map or 3D reconstruction, for example in the name of software subroutines (this can be an initial map or a refined map depending on the refinement stage during the structure determination process; this involves no atomic models unless they are used as the initial reference in the form of a calculated map that is low-pass filtered).

Several new tools, some coming from external developments, are integrated or interfaced with *Phenix*, for example *phenix.dock\_in\_map*, *CryoFit* and *ISOLDE* for flexible fitting, and *Pathwalker* to trace the backbone, which all help in building, refining and validating atomic models, *e.g.* with

*phenix.mtriage* there are tools to estimate resolution ( $d_{09}$ ) or to calculate map-model Fourier correlation curves (Afonine, Klaholz *et al.*, 2018). However, for disordered regions it can be difficult to build a reliable atomic model, in which case the presence of flexible structures needs to be addressed, *e.g.* by ensemble refinement (Burnley *et al.*, 2012). In cryo-EM, various particle-sorting methods [based on 2D or 3D classification methods using multi-variate statistical analysis or maximum-likelihood approaches (Klaholz *et al.*, 2004; White *et al.*, 2004; Penczek *et al.*, 2006; Orlova & Saibil, 2010; Scheres, 2010; Lyumkis *et al.*, 2013; Klaholz, 2015; Serna, 2019)] have been developed to separate different structures, describe several conformational states and address the dynamics of macromolecular complexes. The maps of the particle subpopulations that describe a similar conformation can then be further refined using focused classifications and specific refinements to reach a high-resolution for the entire complex (Ilca *et al.*, 2015; von Loeffelholz *et al.*, 2017; Nakane *et al.*, 2018) for which *Phenix* provides a tool for assembling a weighted composite map from the refined sub-regions. As the resolution is often not constant throughout a cryo-EM map (the concept of local resolution; Cardone *et al.*, 2013; Kucukelbir *et al.*, 2014) there is a tool for local filtering (*phenix.auto\_sharpen*), which uses the current atomic model taking into account the atomic displacement ( $B$ ) factors, similarly to *LocScale* (Jakobi *et al.*, 2017); however, the recent software *LocalDeblur* does not use an atomic model (Ramírez-Aportela *et al.*, 2019). To reflect a certain degree of flexibility, it is important to also refine temperature factors for cryo-EM derived atomic models (Wlodawer *et al.*, 2017; usually a restrained  $B$ -factor refinement of all the atoms in an amino acid, to reduce the number of parameters to be refined). Moreover, including hydrogen atoms in the final atomic model refinements can also improve the clash score for cryo-EM data (Orlov *et al.*, 2019). The *Phenix* graphical user interface (GUI) is interfaced with the graphics programs *Coot* (Emsley *et al.*, 2010) and *Pymol* (DeLano, 2002) to facilitate switching between automatic and manual refinement modes (*e.g.* for checking backbone  $C\alpha$  atom positions, flipping backbone peptides to cure Ramachandran plot outliers, correcting side-chain conformations, validating the entire structure *etc.*) and for performing detailed structure analysis, which is the original aim of a structural biology project. Finally, a convenient feature is also the possibility to prepare a table summarizing statistics for the structure determination and the geometrical parameters of the atomic model in crystallography or cryo-EM, together with the validation report linked with the wwPDB (<https://www.wwpdb.org/validation/validation-reports>). As for other tools, there is also a specific 'bulletin board' mailing list and an online tutorial (see <http://www.phenix-online.org/mailman/listinfo/phenixbb> and <https://www.youtube.com/c/phenixutorials>). Taken together, the latest features of *Phenix* are not only convenient for full workflows but also respond to specific needs, depending on the applications and user expertise.

Clearly, the next challenge will be to integrate atomic model building into large-scale approaches, particularly in

cryo electron tomography, which when combined with sub-tomogram averaging can provide maps in the 30–10 Å resolution range and in exceptional cases that comprise internal symmetry even up to the 3–4 Å resolution range (Schur *et al.*, 2016). For this, various medium-resolution tools exist (including those in *Phenix*) and will need to be developed further, illustrating the ongoing move of the field towards multi-scale and multi-resolution, and correlative approaches to *in situ* macromolecular complexes (Orlov *et al.*, 2017; Jun *et al.*, 2019; Schaffer *et al.*, 2019). This includes super-resolution fluorescence imaging (nowadays single-molecule localization microscopy, SMLM, is also feasible in 3D, see for example Andronov *et al.*, 2018, 2019) to integrate all scales and achieve cellular structural biology in the future.

### Funding information

Support was given by CNRS, Inserm, University of Strasbourg, Institut National du Cancer (INCa), Ligue nationale contre le cancer, Association pour la Recherche sur le Cancer (ARC), Agence National pour la Recherche (ANR), Fondation pour la Recherche Médicale (FRM) and USIAS (USIAS-2018–012) of the University of Strasbourg, and by the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05–01, by Instruct-ULTRA and Instruct-ERIC.

### References

- Afonine, P. V., Klaholz, B. P., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., Adams, P. D. & Urzhumtsev, A. (2018). *Acta Cryst. D* **74**, 814–840.
- Afonine, P. V., Poon, B. K., Read, R. J., Sobolev, O. V., Terwilliger, T. C., Urzhumtsev, A. & Adams, P. D. (2018). *Acta Cryst. D* **74**, 531–544.
- Andronov, L., Michalon, J., Ouararhni, K., Orlov, I., Hamiche, A., Vonesch, J.-L. & Klaholz, B. P. (2018). *Bioinformatics*, **34**, 3004–3012.
- Andronov, L., Ouararhni, K., Stoll, I., Klaholz, B. P. & Hamiche, A. (2019). *Nat. Commun.* **10**, 4436.
- Brown, A., Long, F., Nicholls, R. A., Toots, J., Emsley, P. & Murshudov, G. (2015). *Acta Cryst. D* **71**, 136–153.
- Burnley, B. T., Afonine, P. V., Adams, P. D. & Gros, P. (2012). *eLife*, **1**, e00311.
- Cardone, G., Heymann, J. B. & Steven, A. C. (2013). *J. Struct. Biol.* **184**, 226–236.
- Coulocheri, S. A., Pigis, D. G., Papavassiliou, K. A. & Papavassiliou, A. G. (2007). *Biochimie*, **89**, 1291–1303.
- DeLano, W. (2002). *PyMOL*. <http://www.pymol.org>.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst. D* **66**, 486–501.
- Hryc, C. F., Chen, D.-H., Afonine, P. V., Jakana, J., Wang, Z., Haase-Pettingell, C., Jiang, W., Adams, P. D., King, J. A., Schmid, M. F. & Chiu, W. (2017). *Proc. Natl Acad. Sci. USA*, **114**, 3103–3108.
- Iica, S. L., Kotecha, A., Sun, X., Poranen, M. M., Stuart, D. I. & Huiskonen, J. T. (2015). *Nat. Commun.* **6**, 8843.
- Jakobi, A. J., Wilmanns, M. & Sachse, C. (2017). *eLife*, **6**, e27131.
- Jun, S., Ro, H.-J., Bharda, A., Kim, S. I., Jeoung, D. & Jung, H. S. (2019). *Protein J.*, <https://dx.doi.org/10.1007/s10930-019-09856-1>.
- Kim, L. Y., Rice, W. J., Eng, E. T., Kopylov, M., Cheng, A., Raczkowski, A. M., Jordan, K. D., Bobe, D., Potter, C. S. & Carragher, B. (2018). *Front. Mol. Biosci.* **5**, 50.
- Klaholz, B. & Moras, D. (2002). *Structure*, **10**, 1197–1204.
- Klaholz, B. P. (2015). *Open J. Stat.* **5**, 820–836.
- Klaholz, B. P., Myasnikov, A. G. & van Heel, M. (2004). *Nature*, **427**, 862–865.
- Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. (2014). *Nat. Methods*, **11**, 63–65.
- Leontis, N. B. & Westhof, E. (2001). *RNA*, **7**, 499–512.
- Loeffelholz, O. von, Natchiar, S. K., Djabeur, N., Myasnikov, A. G., Kratzat, H., Ménétret, J.-F., Hazemann, I. & Klaholz, B. P. (2017). *Curr. Opin. Struct. Biol.* **46**, 140–148.
- Lyumkis, D., Brilot, A. F., Theobald, D. L. & Grigorieff, N. (2013). *J. Struct. Biol.* **183**, 377–388.
- Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T., Hintze, I., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst. D* **75**, 861–877.
- Malhotra, S., Träger, S., Dal Peraro, M. & Topf, M. (2019). *Curr. Opin. Struct. Biol.* **58**, 105–114.
- Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. (2018). *eLife*, **7**, e36861.
- Natchiar, S. K., Myasnikov, A. G., Kratzat, H., Hazemann, I. & Klaholz, B. P. (2017a). *Nature*, **551**, 472–477.
- Natchiar, S. K., Myasnikov, A., Kratzat, H., Hazemann, I. & Klaholz, B. P. (2017b). *Protoc. Exch.*, <https://dx.doi.org/10.1038/protex.2017.122>.
- Ognjenović, J., Grisshammer, R. & Subramaniam, S. (2019). *Annu. Rev. Biomed. Eng.* **21**, 395–415.
- Orlova, E. V. & Saibil, H. R. (2010). *Methods Enzymol.* **482**, 321–341.
- Orlova, E. V. & Saibil, H. R. (2011). *Chem. Rev.* **111**, 7710–7748.
- Orlov, I., Hemmer, C., Ackerer, L., Lorber, B., Ghannam, A., Poignavent, V., Hleibieh, K., Sauter, C., Schmitt-Keichinger, C., Belval, L., Hily, J., Marmonier, A., Komar, V., Gersch, S., Schellenberger, P., Bron, P., Vigne, E., Muyldermans, S., Lemaire, O., Demangeat, G., Ritzenthaler, C. & Klaholz, B. P. (2019). *bioRxiv*, <https://biorxiv.org/cgi/content/short/728907v1>.
- Orlov, I., Myasnikov, A. G., Andronov, L., Natchiar, S. K., Khatter, H., Beinsteiner, B., Ménétret, J.-F., Hazemann, I., Mohideen, K., Tazibt, K., Tabaroni, R., Kratzat, H., Djabeur, N., Bruxelles, T., Raivoniaina, F., Pompeo, L., Torchy, M., Billas, I., Urzhumtsev, A. & Klaholz, B. P. (2017). *Biol. Cell*, **109**, 81–93.
- Penczek, P. A., Frank, J. & Spahn, C. M. T. (2006). *J. Struct. Biol.* **154**, 184–194.
- Pintilie, G. & Chiu, W. (2018). *J. Struct. Biol.* **204**, 564–571.
- Ramírez-Aportela, E., Vilas, J. L., Glukhova, A., Melero, R., Conesa, P., Martínez, M., Maluenda, D., Mota, J., Jiménez, A., Vargas, J., Marabini, R., Sexton, P. M., Carazo, J. M. & Sorzano, C. O. S. (2019). *Bioinformatics*, <https://dx.doi.org/10.1093/bioinformatics/btz671>.
- Schaffer, M., Pfeffer, S., Mahamid, J., Kleindiek, S., Laugks, T., Albert, S., Engel, B. D., Rummel, A., Smith, A. J., Baumeister, W. & Plitzko, J. M. (2019). *Nat. Methods*, **16**, 757–762.
- Scheres, S. H. W. (2010). *Methods Enzymol.* **482**, 295–320.
- Schur, F. K. M., Obr, M., Hagen, W. J. H., Wan, W., Jakobi, A. J., Kirkpatrick, J. M., Sachse, C., Kräusslich, H.-G. & Briggs, J. A. G. (2016). *Science*, **353**, 506–508.
- Serna, M. (2019). *Front. Mol. Biosci.* **6**, 33.
- Simonetti, A., Marzi, S., Fabbretti, A., Hazemann, I., Jenner, L., Urzhumtsev, A., Gualerzi, C. O. & Klaholz, B. P. (2013). *Acta Cryst. D* **69**, 925–933.
- Smart, O. S., Womack, T. O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vornrhein, C. & Bricogne, G. (2012). *Acta Cryst. D* **68**, 368–380.
- Stark, H. & Chari, A. (2016). *Microscopy (Tokyo)*, **65**, 23–34.
- Wang, J. & Moore, P. B. (2017). *Protein Sci.* **26**, 122–129.
- Wang, J., Natchiar, S. K., Myasnikov, A. G., Hazemann, I., Moore, P. B. & Klaholz, B. P. (2019). Submitted.
- White, H. E., Saibil, H. R., Ignatiou, A. & Orlova, E. V. (2004). *J. Mol. Biol.* **336**, 453–460.
- Wlodawer, A., Li, M. & Dauter, Z. (2017). *Structure*, **25**, 1–9.