









Image processing tools for the validation of CryoEM maps†

C. O. S. Sorzano, ^{id}*^a J. L. Vilas,^a E. Ramírez-Aportela,^a J. Krieger, ^{id}^a D. del Hoyo,^a D. Herreros,^a E. Fernandez-Giménez,^a D. Marchán, ^{id}^a J. R. Macías, ^{id}^a I. Sánchez, ^{id}^a L. del Caño,^a Y. Fonseca-Reyna,^a P. Conesa,^a A. García-Mena,^a J. Burguet, ^{id}^b J. García Condado, ^{id}^c J. Méndez García,^d M. Martínez, ^{id}^a A. Muñoz-Barrutia,^e R. Marabini,^f J. Vargas^b and J. M. Carazo^a

Received 7th March 2022, Accepted 4th April 2022

DOI: 10.1039/d2fd00059h

The number of maps deposited in public databases (Electron Microscopy Data Bank, EMDB) determined by cryo-electron microscopy has quickly grown in recent years. With this rapid growth, it is critical to guarantee their quality. So far, map validation has primarily focused on the agreement between maps and models. From the image processing perspective, the validation has been mostly restricted to using two half-maps and the measurement of their internal consistency. In this article, we suggest that map validation can be taken much further from the point of view of image processing if 2D classes, particles, angles, coordinates, defoci, and micrographs are also provided. We present a progressive validation scheme that qualifies a result validation status from 0 to 5 and offers three optional qualifiers (A, W, and O) that can be added. The simplest validation state is 0, while the most complete would be 5AWO. This scheme has been implemented in a website <https://biocomp.cnb.csic.es/EMValidationService/> to which reconstructed maps and their ESI can be uploaded.

Cryo-electron microscopy is currently one of the most active techniques in structural biology. The number of maps deposited at the Electron Microscopy Data Bank is rapidly growing every year^{1,2} and keeping the quality of the submitted maps is essential to maintain the scientific quality of the field. Additionally, all

^aNatl. Center of Biotechnology, CSIC, c/Darwin, 3, 28049, Madrid, Spain. E-mail: coss@cnb.csic.es; Fax: +34 91 585 4506; Tel: +34 91 585 45 10

^bDepto. de Óptica, Univ. Complutense de Madrid, Pl. Ciencias, 1, 28040, Madrid, Spain

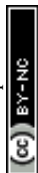
^cBiocruces Bizkaia Instituto Investigación Sanitaria, Cruces Plaza, 48903, Barakaldo, Bizkaia, Spain

^dUniv. Tecnológica de Pereira, 660003, Pereira, Colombia

^eUniv. Carlos III de Madrid, Avda. de la Universidad 30, 28911Leganés, Madrid, Spain

^fEscuela Politécnica Superior, Univ. Autónoma de Madrid, CSIC, C. Francisco Tomás y Valiente, 11, 28049, Madrid, Spain

† Electronic supplementary information (ESI) available. See <https://doi.org/10.1039/d2fd00059h>

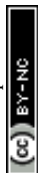


scientific domains have recognized open science as a way to accelerate research.³ In this way, disclosing sufficient information in order to understand the limitations and strengths of a CryoEM map is crucial for a better use of the map. Ideally, reproducibility of the results should be achieved, and the possibility of depositing the raw data at the EMPIAR (Electron Microscopy Public Image Archive, Iudin *et al.*⁴) has certainly been a huge step forward. However, as a community we are still far from having generally adopted this reproducibility goal. For instance, from the 950 EMDB entries labelled as SARS-CoV-2, only 23 (less than 2.5%) are deposited at EMPIAR (as of March 7th, 2022). The availability of the raw data could be complemented with the availability of the image processing workflow and decisions taken to go from the raw acquisition to the final map. Despite the fact that scientific articles are generally trusted by other scientists and the general public, severe concerns about a reproducibility crisis in science have been raised.⁵

The ultimate quality measure is the consistency of the map and an atomic model.⁶ However, this is only possible for high-resolution maps. Alternatively, the standard map validation practice has mostly been restricted to the internal consistency of two half-maps calculated from independent halves of the whole dataset.⁷ This internal consistency is essential, and it is a good measure of the presence of random fluctuations. Its main drawback is that it is not immune to systematic biases,⁸ that is, systematic mistakes committed in both halves would be rewarded in terms of the Fourier shell correlation.

Over the years, there have been many suggestions about validation measures of CryoEM maps.⁹ Unfortunately, most of these measures are not currently used due to their spread across multiple software tools and the associated difficulty accessing them. To alleviate this problem, we present a validation grading system and its public availability through a web server that qualifies the CryoEM map depending on the information available to assess it. This system grades a structure map at six different levels. In this way, a map could be validated at level 0 (the deposited map), 1 (two half maps), 2 (2D classes), 3 (particles), 4 (...+angular assignment), 5 (...+micrographs and coordinates). For each of these levels, we list algorithms that can be employed. In addition to this grading system, we have three optional qualifiers: A (...+atomic model), W (...+image processing workflow), and O (...+other techniques). Those depositions wanting to achieve the highest level of validation should deposit a relatively small number of particles (in the order of 10k randomly chosen from the final set of particles) along with their micrographs and all alignment parameters. They should also include a detailed description of the image processing workflow so that the final result can be fully understood (and ideally reproduced) and some extra validation by other experimental data. This high-quality standard is ideal and will not be accomplished shortly for all deposited structures. Nevertheless, it is essential to have it as a compass to direct our community efforts.

The web server is publicly available at <https://biocomp.cnb.csic.es/EMValidationService/>. It returns a PDF report that evaluates the correctness of the submitted map from multiple figures of merit. Obviously, map submission to a public database cannot require a compulsory deposition of all these elements. Still, a validation grading system could be adopted in which, if they are given, the consistency of the map with the different aspects that give rise to it can be assessed.



The server is aimed at structures determined by single particle analysis (SPA). Maps determined by subtomogram averaging (STA) in cryo-electron tomography can share the levels 0, 1, A, and W with the maps coming from SPA. However, more specific analyses could be developed for STA maps.

We hope that the validation server proposed in this paper will help improve the understanding of CryoEM maps and reduce the reproducibility crisis, especially if the image processing workflow is also disclosed.

1 Validation methods

The following sections describe the different methods at one's disposal to validate a CryoEM map. The availability of these methods usually depends on the accessibility of extra information, like the set of particles supporting it, their angular assignment, *etc.* The level of map validation is defined as the highest consecutive number up to which there is information available. The three extra qualifiers can be added to any of the levels. The highest degree of validation would be 5AWO. Still, we could have, for instance, a map whose validation is 4W, meaning that it has reached level 4 and a detailed description of the image processing workflow is available. At present, the highest validation that a typical map from EMDB can achieve is 1A (map + half maps + atomic model). Interestingly, since February 2022, the deposition of half maps at the EMDB has been compulsory, meaning that we have moved from validation level 0 to level 1.

1.0 Level 0: map

The first level of validation is performed when just the reconstructed map is available along with a visualization threshold. At this point, several methods can evaluate the local resolution of the map and its hand.

0.a Center analysis. Centering of the mass and extra space available to correct for the contrast transfer function (CTF). There should be at least 30–40 Å on each side for a proper correction.

0.b Mask analysis. At the threshold value specified by the user, most of the mass should be collected in a single connected component.

0.c Background analysis. If we analyze the gray values outside the mask, they should not have too negative values (*e.g.*, values below five times the standard deviation of the background noise).

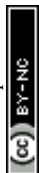
0.d B-factor analysis. The *B*-factor line,¹⁰ fitted between 15 Å and the resolution reported, should have a slope that is between 0 and 300 Å².

0.e DeepRes.¹¹ This method is based on a deep learning algorithm that assesses the similarity of the texture features present in the map to the texture features observed in atomic structures.

0.f LocBfactor.¹² This method estimates a local resolution *B*-factor by decomposing the input map into a local magnitude and phase term using the spiral transform.

0.g LocOccupancy.¹² This method estimates the occupancy of a voxel by the macromolecule.

0.h DeepHand.¹³ This method determines for maps whose resolution is higher than 5 Å whether the map has the right hand or, on the contrary, it is the mirrored version of the correct map.



Although not yet implemented, it would be possible to detect preferential orientations or artifacts by analyzing the macromolecule's local texture and noise.

1.1 Level 1: ...+half maps

If independent half maps are available, then we can further assess the local resolution with different means:

1.a Global resolution.¹⁴ The Fourier shell correlation (FSC) between the two half maps is the most standard method to determine the global resolution of a map. However, other measures exist, such as the spectral signal-to-noise ratio and the differential phase residual. There is a long debate about the correct thresholds for these measures. Probably the clearest threshold is the one of the SSNR (SSNR = 1). For the DPR, we have chosen 103.9° (ref. 14) and for the FSC, the standard 0.143.

1.b Permutation test FSC.¹⁵ This method calculates a global resolution by formulating a hypothesis test in which the distribution of the FSC of noise is calculated from the two maps.

1.c BlocRes.¹⁶ This method computes a local Fourier shell correlation (FSC) between the two half maps.

1.d Resmap.¹⁷ This method is based on a test hypothesis testing the superiority of signal over noise at different frequencies.

1.e MonoRes.¹⁸ This method evaluates the local energy of a point to the distribution of energy in the noise. This comparison is performed at multiple frequencies, and for each one, the monogenic transformation separates the amplitude and phase of the input map.

1.f MonoDir.¹⁹ This method extends the concept of local resolution to local and directional resolution by changing the shape of the filter applied to the input map. The directional analysis can reveal image alignment problems.

1.g FSO. This method calculates the anisotropy of the energy distribution in Fourier shells. It is an indirect measure of the anisotropy of the angular distribution or the presence of heterogeneity.

1.h FSC Directional.²⁰ This method analyzes the FSC in different directions and evaluates its homogeneity through the sphericity of the FSC surface.

1.2 Level 2: ...+2D classes

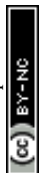
If 2D classes of the particles used for the reconstruction are available, then the following method can be applied:

2.a Reprojection consistency. The 2D classes can be aligned against the reconstructed map, then the correlation between rejections of the map and the 2D classes can be analyzed. Also, analyzing the residuals (2D class minus the corresponding reprojction) can reveal systematic differences.

1.3 Level 3: ...+particles

If a random subset of the particles is provided, the following actions can be performed:

3.a Outlier detection. The set of particles is classified into the input set of 2D classes of level 2. The number of particles considered to be outliers in those classes is reported. A particle is an outlier if its Mahalanobis distance to the centroid of the class is larger than 3.²¹



3.b 2D classification internal consistency. The input particles are classified in 2D clusters. The quality of the 2D clusters is assessed through Fourier ring correlation.

3.c 2D classification external consistency. We measure the overlap between the subspace spanned by the classes in level 2 and the classes of level 3.

1.4 Level 4: ...+angular assignment

If the angular assignment of the particles in level 3 is available, then the following methods can be applied:

4.a Similarity criteria. Analysis of the distribution of the similarity between the input particles and the reprojection from the same angular orientation by different scores.

4.b Alignability smoothness.²² This algorithm analyzes the smoothness of the correlation function over the projection sphere and the stability of its maximum.

4.c Alignability precision and accuracy. The precision²³ analyzes the orientation distribution of the best matching reprojections from the reference volume. If the high values are clustered around the same orientation, the precision is close to 1. Otherwise, it is closer to -1 . Below 0.5, the best directions tend to be scattered. The alignability accuracy²⁴ compares the final angular assignment with the result of a new angular assignment. The similarity between both is again encoded between -1 and 1.

4.d Angular error distribution. Angular error distribution between the provided angles and an independent angular assignment performed with state-of-the-art algorithms.

4.e Classification without alignment. 3D classification of the input particles without angular refinement.

4.f Detection of overfitting.²⁵ This method compares the resolution achieved by subsets of images of increasing size and by subsets of noise images of the same size.

4.g Angular distribution efficiency.²⁶ This method evaluates the ability of the angular distribution to fill the Fourier space.

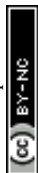
4.h Sampling compensation factor.²⁷ This method is another way of measuring the ability of the angular distribution to fill the Fourier space.

4.i CFT stability. Analysis of the stability of the defocus parameters. For this purpose, the defocus, *B*-factor, astigmatism, and phase shift can be estimated from the given particles, and these refined parameters' deviations are reported. Ideally, the differences in defoci cannot be larger than the ice thickness. The same can be done with local magnification offsets (which should be around 0) and the *B*-factor.

1.5 Level 5: ...+micrographs and coordinates

If a random subset of micrographs and their corresponding coordinates are available, then the following measure can be taken:

5.a Micrograph cleaner.²⁸ This method assigns a score between 0 and 1, reflecting the probability that the coordinate is outside a region with aggregations, ice crystals, carbon edges, *etc.*



1.6 Level A: ...+atomic model

If a fitted atomic model is available, then we may apply the following validation methods:

A.a Map-Q.²⁹ This method computes the local correlation between the map and each one of its atoms assumed to have a Gaussian shape.

A.b FSC-Q.³⁰ This method compares the local FSC between the map and the atomic model to the local FSC of the two half maps.

A.c Model ambiguity by molecular dynamics.³¹ This method estimates the ambiguity of the atomic model in each region of the CryoEM map due to the different local resolutions or local heterogeneity.

A.d Guinier plot of model and map.³² This method compares the falloff in Fourier space between the map and atomic model.

A.e Phenix CryoEM validation tools.³³ Phenix provides several tools to assess the agreement between the experimental map and its atomic model. Two large clusters of these measurements are: (1) different ways of measuring the cross-correlation between the map and model, and (2) different ways of measuring the resolution between the map and model.

A.f EMRinger.³⁴ This algorithm compares the side chains of the atomic model to the CryoEM map.

A.g DAQ.³⁵ This algorithm uses deep learning that can estimate the residue-wise local quality for protein models from cryo-electron microscopy (EM) maps. The method calculates the likelihood that a given density feature corresponds to an amino acid, atom, and secondary structure. These likelihoods are combined into a score that ranges from -1 (bad quality) to 1 (good quality).

We are aware that the report generated at the submission to EMDB/PDB is relatively rich in this area, including methods such as Molprobity³⁶ or TEMPy2.³⁷ Our goal is to complement this analysis with alternative tools. Some of these analyses (0.a, 0.b, 0.c, 2.a, 3.b, 3.c, 4.a, 4.d, 4.e, and 4.i) have been newly developed for this server.

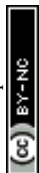
1.7 Qualifier W: ...+workflow

Another level of validation purely based on image processing is a detailed description of each of the image processing steps used to produce the final map so that an external user can understand the results of each step and has all the information to reproduce the whole pipeline. This information is produced, for example, by Scipion³⁸ in the form of a JSON that can be submitted to EMPIAR¹ or Scipion's workflow repository (<https://nolan.cnb.csic.es/cryoemworkflowviewer>) and visualized with a JavaScript viewer integrated into these databases. Alternative methods would also be valid, specifically tailored to each one of the different image processing packages.

1.8 Qualifier O: ...+other techniques

Finally, if extra experimental work is available, then we may apply the following techniques.

O.a Mass spectroscopy.³⁹ This method uses information from cross- and mono-links to validate the atomic model.



O.b SAXS.⁴⁰ This method compares the expected energy profile from the reconstructed map to the one obtained by a SAXS experiment.

O.c Tilt pair validation.⁴¹ This method is capable of experimentally validating the hand of the reconstructed map by comparing the angular assignment of two sets of particles related by a single-axis tilt.

1.9 Availability

We have created a web server located at <https://biocomp.cnb.csic.es/EMValidationService/>. The server has a web interface that guides the user through the different steps. We have not made the levels to be compulsorily progressive. For instance, one could have validation levels 0 and 2 without having the information for level 1. Although possible, this option is discouraged. Method A.c is sensible, but this method takes much time to execute due to the molecular dynamics underneath (about 6 hours per constructed model). For this reason, we recommend not to run it regularly to not saturate the server.

Once a job is submitted to the server, the execution time varies from 20 minutes to 16 hours if A.c is not executed or 3 days if it is. The execution time depends on the number of validations to perform, the size of the reconstructed map, and the number of images provided for validation. During its execution, the user gets a token so that they can check whether the job has already been finished.

The server asks for a URL where the image processing workflow can be visualized for validation. If Scipion has created the workflow, it is included in the report. However, the validation information truly lies on the given URL. For submitting the Scipion workflow, the user may use the `scipion-em-datamanager` plugin that submits the current project to the workflow repository at <https://nolan.cnb.csic.es/cryoemworkflowviewer/entries>. An example of such a workflow can be seen in Fig. 1.

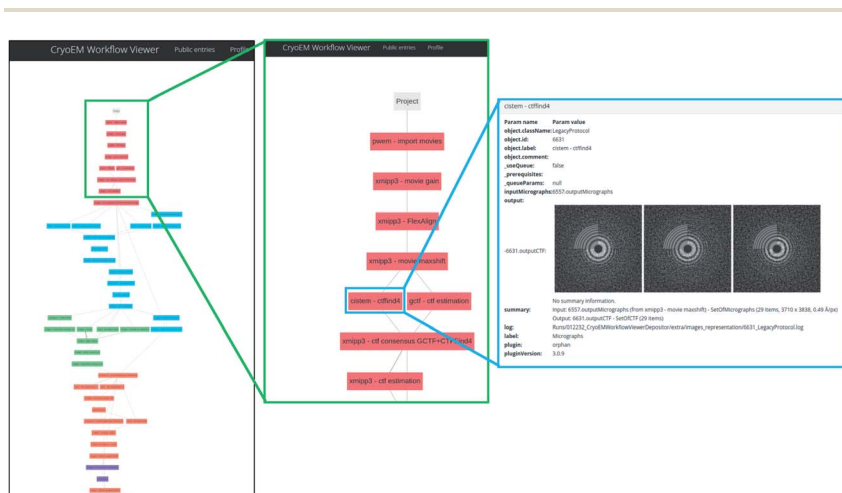
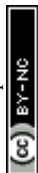


Fig. 1 An example of workflow visualization at Scipion's workflow repository. Each of the colored boxes represent a protocol executed within Scipion. The last panel on the right shows how the input parameters, data and results can be inspected using a web browser.



The server automatically constructs a Scipion workflow³⁸ based on the input data. Depending on the amount of data available (levels 0, 1, 2, ...), the appropriate protocols are instantiated and interconnected. The source code of the server is available at <https://github.com/I2PC/scipion-em-validation>. The program creates the Scipion workflow, automatically analyzes the results, and constructs a Latex document which is later compiled into a PDF. This report is handed to the user.

2 Results

To show the usefulness of the validation reports, we have applied this methodology to multiple datasets. The first one comes from the tutorial of Scipion,³⁸ while all others are examples from the EMDB. In the first one, we have information for evaluating all validation levels (5AWO), while for the EMDB, the validation reports are of levels 0A or 1A.

2.1 Full report

ESI 1† shows the validation report for Scipion's tutorial. The data corresponds to 30 micrographs of the apoferritin EMPIAR 10248 dataset.⁴² The tutorial used 1457

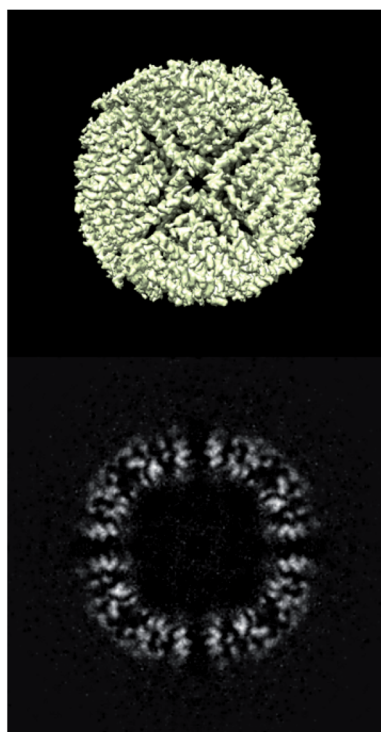
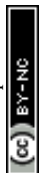


Fig. 2 Isosurface and central slice of the Scipion's tutorial dataset, apoferritin.



Abstract

The map seems to be well centered. There is no problem with the suggested threshold. There seems to be a problem with the map's background (see Sec. 2.3). The resolution does not seem to be uniform in all directions (see Sec. 4.6). The 2D classes provided by the user do not seem to correlate well with the reprojections of the map (see Sec. 6.1). It seems that the input particles cannot be easily aligned (see Sec. 9.2). It seems that the angular assignment given by the user does not match with the one produced by CryoSparc (see Sec. 9.5). It seems that the angular assignment produced by Relion does not match with the one produced by Cryosparc (see Sec. 9.6). This is probably a sign of the difficulty to align these particles. It seems that there is some problem with the CTF (see Sec. 9.11). According to phenix, it seems that there might be some mismatch between the map and its model (see Sec. 13.5). The EMRinger score is negative, it seems that the model side chains do not match the map (see Sec. 13.6). DAQ detects some mismatch between the map and its model (see Sec. 13.7).

The average resolution of the map estimated by various methods goes from 2.0 Å to 4.6 Å with an average of 3.3 Å. The resolution provided by the user was 2.6 Å. The resolution reported by the user may be overestimated.

The overall score (passing tests) of this report is 26 out of 36 evaluable items.

Fig. 3 Example of the abstract of the full report generated for Scipion's tutorial data.

experimental projections to construct a map whose nominal resolution is 2.6 Å. Although generally, the map is good (see Fig. 2) and reasonably agrees with its atomic model by visual inspection, the validation report shows that it has several problems:

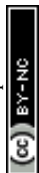
- (1) The background outside the map does not have zero mean, and it contains a significant amount of intensity outliers.
- (2) The reported resolution seems to be overestimated according to several local resolution algorithms (DeepRes, Resmap, and MonoDir) and Phenix's comparison to the atomic model. The average resolution seems to be more around 4.4 Å, rather than 2.6 Å.
- (3) Images are difficult to align, as reported by the alignability smoothness, CryoSparc alignment, and the comparison between CryoSparc and Relion alignments. This uncertainty in the alignment was also shown in MonoDir radial plots.
- (4) Relion could not reliably determine the scaling factor of the images.
- (5) The resolvability of the side chains is not good, as reported by EMRinger and DAQ.

None of these errors is terrible, and it is the expected result for a tutorial reconstruction from only 30 micrographs. However, this contrasts with the resolution reported by the FSC, which gives the false impression of having achieved a better map than the one obtained.

Fig. 3–5 show the summary of the report generated for this dataset. In a single glimpse, the main problems can be easily identified.

2.2 A comparison of three SARS-CoV 2 spike structures

With the Covid-19 pandemic, many structural studies have addressed all the proteins amenable to CryoEM. We have chosen three reconstructions of the spike (EMDB entries 11 337,⁴³ 22 301,⁴⁴ and 22 838⁴⁵) which are supposed to be



0.a Mass analysis	Sec. 2.1	OK
0.b Mask analysis	Sec. 2.2	OK
0.c Background analysis	Sec. 2.3	2 warnings
0.d B-factor analysis	Sec. 2.4	OK
0.e DeepRes	Sec. 2.5	1 warnings
0.f LocBfactor	Sec. 2.6	OK
0.g LocOccupancy	Sec. 2.7	OK
0.h DeepHand	Sec. 2.8	OK
1.a Global resolution	Sec. 4.1	OK
1.b FSC permutation	Sec. 4.2	OK
1.c Blocres	Sec. 4.3	OK
1.d Resmap	Sec. 4.4	1 warnings
1.e MonoRes	Sec. 4.5	OK
1.f MonoDir	Sec. 4.6	1 warnings
1.g FSO	Sec. 4.7	OK
1.h FSC3D	Sec. 6.1	OK
2.a Reprojection consistency	Sec. 6.1	OK
3.a Outlier detection	Sec. 9.1	OK
3.b 2D Classification internal consistency	Sec. 8.2	Cannot be automated
3.c 2D Classification external consistency	Sec. 8.3	OK
4.a Similarity criteria	Sec. 9.1	Cannot be automated
4.b Alignability smoothness	Sec. 9.2	1 warnings
4.c Alignability precision and accuracy	Sec. 9.3	OK
4.d1 Relion alignment	Sec. 9.4	OK
4.d2 CryoSparc alignment	Sec. 9.5	1 warnings
4.d3 Relion/CryoSparc alignments	Sec. 9.6	1 warnings
4.e Classification without alignment	Sec. 9.8	OK
4.f Overfitting detection	Sec. 9.8	OK
4.g Angular distribution efficiency	Sec. 9.9	OK
4.h SCF	Sec. 9.10	OK
4.i CTF stability	Sec. 9.11	1 warnings
5.a Micrograph cleaner	Sec. 11.1	OK
A.a MapQ	Sec. 13.1	OK
A.b FSC-Q	Sec. 13.2	OK
A.c Multimodel	Sec. 13.3	OK
A.d Map-Model Guinier	Sec. 13.4	OK
A.e Phenix validation	Sec. 13.5	1 warnings
A.f EMRinger	Sec. 13.6	1 warnings
A.g DAQ	Sec. 13.7	1 warnings
W Workflow	Sec. 14	Cannot be automated
O.b SAXS	Sec. 15.1	Cannot be automated

3

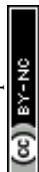
Fig. 4 Example of the index of the full report generated for Scipion's tutorial data.

reconstructed at similar resolutions (3.3, 3.7 and 3.84 Å, respectively; see Fig. 6). The first entry could be validated at level 1A, while the other two could only be validated at level 0A. The validation reports for these structures can be seen in ESI 2.†

The average resolutions measured by several methods for the different structures were 4.9, 8.2, and 3.8 Å, respectively. This points out that the internal resolution variability is much higher than the one reported by the FSC. We have empirically verified that very often, the resolution based on the FSC is at the lower extreme of the distribution of resolutions reported by many local resolution methods.¹¹ For this reason, this FSC resolution must be understood as “*there is a region in the map whose local resolution is this number*”.

In this analysis, we can identify the following problems:

(1) The backgrounds of the three structures have problems as they are not equal to 0 and contain significant outliers.



Summary of the warnings across sections.

If it is empty below this point, it means that there are no warnings.

Section 2.3 (0.c Background analysis)

1. **The null hypothesis that the background mean is 0 has been rejected because the p-value of the comparison is smaller than 0.001**
2. **There is a significant proportion of outlier values in the background (cdf5 ratio=2031.06)**

Section 2.5 (0.e DeepRes)

1. **The reported resolution, 2.60 Å, is particularly with respect to the local resolution distribution. It occupies the 0.00 percentile**

Section 4.4 (1.d Resmap)

1. **The reported resolution, 2.60 Å, is particularly with respect to the local resolution distribution. It occupies the 0.00 percentile**

Section 4.6 (1.f MonoDir)

1. **The distribution of best resolution is not uniform in all directions. The associated p-value is 0.000000.**

Section 9.2 (4.b Alignability smoothness)

1. **The percentage of images whose angular assignment is significantly away from the smoothed maximum is too high, 50.2%**

Section 9.5 (4.d2 CryoSparc alignment)

1. **The percentage of images with uncertain shift is larger than 20%**

Section 9.6 (4.d3 Relion/CryoSparc alignments)

1. **The percentage of images with uncertain shift is larger than 20%**

Section 9.11 (4.i CTF stability)

1. **The 95% confidence interval of scale factor is not centered.**

Section 13.5 (A.e Phenix validation)

1. **The resolution reported by the user, 2.6 Å, is significantly smaller than the resolution estimated between map and model (FSC=0.5), 4.4 Å**

Section 13.6 (A.f EMRinger)

1. **The EMRinger score is smaller than 1, it is 0.892.**

Section 13.7 (A.g DAQ)

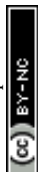
5

Fig. 5 Example of the warnings abstract of the full report generated for Scipion's tutorial data.

(2) Two of the maps, EMDB11337 and EMDB22838, have a reported resolution that is particularly good compared to the one computed by DeepRes (level 0) or other local resolution methods (level 1).

(3) EMDB22838 and EMDB22301 have either too low a resolution or a severe problem of anisotropic resolution, probably caused by the attraction of particles towards specific directions. Both issues are translated into the uncertainty of DeepHand to determine the handedness of the structure.

(4) EMDB11337 has problems fitting to the atomic map according to FSC-Q and Phenix. EMDB22301 has issues in this fitting according to MapQ, Phenix, and DAQ. Finally, EMDB 22838 has problems according to MapQ, EMRinger, and DAQ.



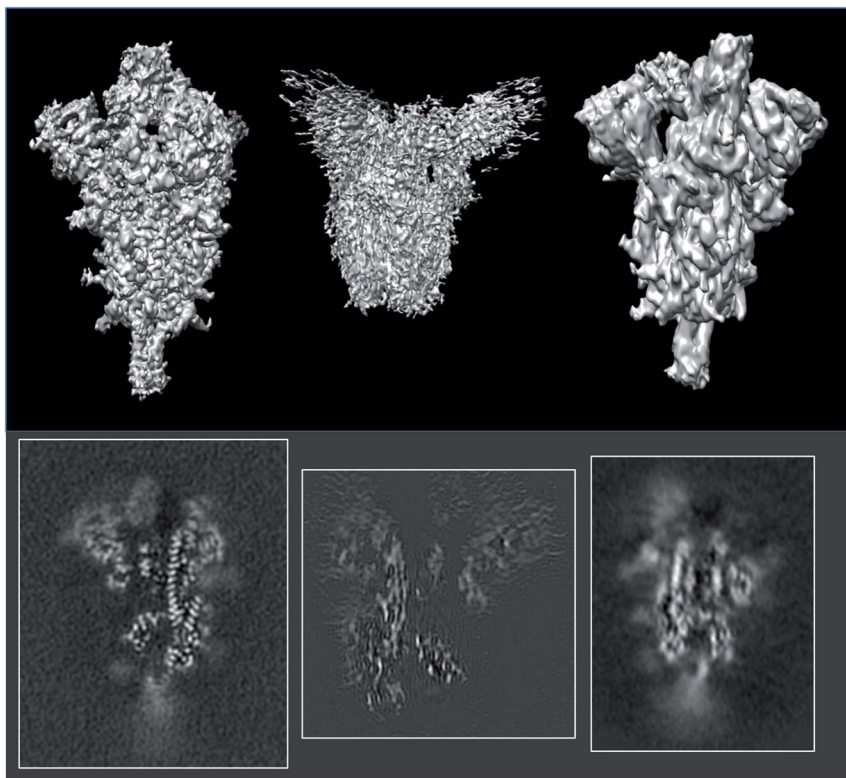


Fig. 6 Isosurfaces and central slices of EMDB 11337 (left), EMDB 22301 (middle), and EMDB 22838 (right), all of which are SARS-CoV2 spikes.

2.3 An analysis of a 1.15 Å resolution apoferritin

EMDB11668 is currently the map with the best resolution in the EMDB reconstructed by single particle analysis (see Fig. 7). It is the reconstruction of human apoferritin in a Titan Krios with a second-generation spherical aberration correction.⁴⁶ An analysis of the deposited structure reveals that the recommended threshold, 0.15, is too high and causes many mass and mask problems (see ESI 3 Report 11668_015†). In the following, we will analyze the map at a threshold of 0.05 (see ESI 3 Report 11668_005†).

In this analysis, we can identify the following problems:

- (1) The mask of the map is very fragmented, with 478 289 connected components. The largest component only takes 42% of the mass. This is due to the boosting of the high frequencies caused by the *B*-factor correction.
- (2) The mean of the background is not 0, and there is a significant amount of outlier values.
- (3) The map has been *B*-factor corrected, resulting in a boost of noise at high frequency and a disagreement between the falloffs of the map and the model.
- (4) There is a significant disagreement between the map and its model according to MapQ, Phenix, and DAQ in some regions. In Fig. 8, we show a region of the map and model in Coot. Although the model is excellent in many regions, there are portions in which the map and model do not match, as highlighted in the figure.



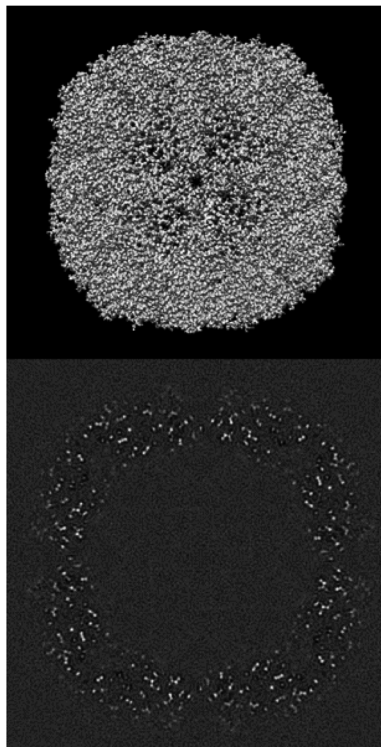


Fig. 7 Isosurface and central slice of EMDB 11668, apoferritin.

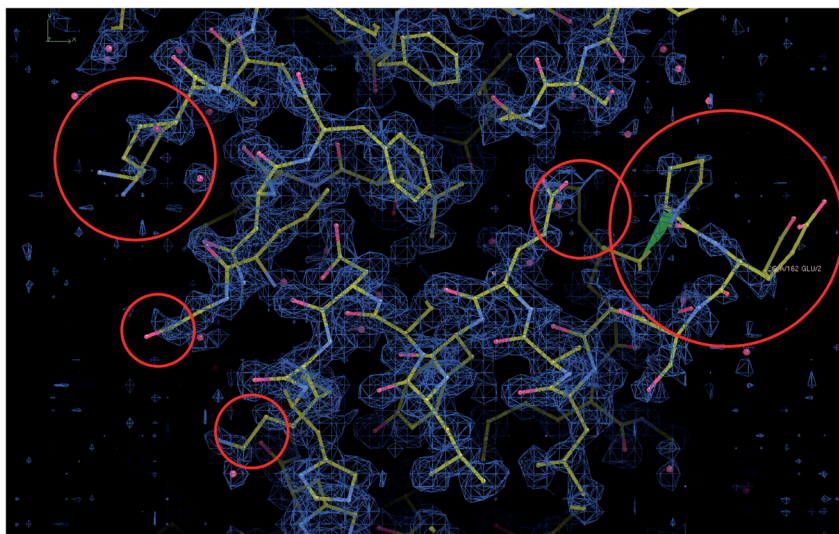
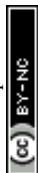


Fig. 8 Map and model representations of apoferritin with some mismatching areas highlighted.



3 Discussion

The determination of a three-dimensional map compatible with the CryoEM measurements of a macromolecule is full of decisions along the image processing pipeline (whether to keep or not this micrograph according to its contrast transfer function or beam-induced alignment, whether a region of that micrograph represents a centered projection of the macromolecule, which is the orientation of that projection with respect to a reference map and which is the exact conformation of that macromolecule, *etc.*). All these decisions involve parameter estimates. Some of these parameters are continuous, such as the micrograph defocus or the in-plane shifts of a given projection with respect to a reference map. Some others are discrete, such as whether a given projection is in one conformation or another. From this point of view, the reconstructed map is a signal (another parameter) to be estimated in a N^3 -dimensional space, with N^3 being the number of map voxels.

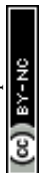
All algorithms that have to estimate a parameter can commit an error (false positives or negatives if the parameter is binary, or a residual if the parameter is continuous). Our final map is a mixture of our good and bad decisions for all the particles involved in the reconstruction. To simplify the analysis, let us consider that we perfectly estimate all parameters of a given projection with probability p , and we do not, with probability $1 - p$. The fraction of well-estimated projections will provide us with a perfect reconstruction, while the fraction of incorrectly estimated parameters will result in an imperfect reconstruction:

$$\begin{aligned} V_{\text{reconstruction}} &= pV_{\text{perfect}} + (1 - p)V_{\text{imperfect}} \\ &= V_{\text{perfect}} + (1 - p)(V_{\text{imperfect}} - V_{\text{perfect}}) \end{aligned}$$

In this way, we see that our reconstruction is a perfect map plus bias that depends on the fraction of incorrectly estimated parameters, $1 - p$, and the difference between the imperfect and perfect reconstructions (if we make a mistake in the estimation of the projection direction of an experimental projection, but the map from the wrong direction is very similar to the map from the correct direction, the mistake will have a negligible effect). It is this bias that we call overfitting in our field. Hence, overfitting is directly linked to incorrect parameter estimates. In Sorzano *et al.*,⁸ we discuss the different sources of map bias and how we can design image processing workflows that try to identify the incorrectly estimated parameters. Ultimately, this identification requires multiple estimates of the same parameter by, ideally, different algorithms. In this way, we can determine the reliability of any particular value.

The use of multiple algorithms to discard incorrectly estimated parameters or average those that agree more is seldom seen in our applied papers. However, this practice of estimating parameters only once leaves us in a position in which we cannot know whether the estimated parameter is stable or not.

Map and model validations have become a relevant research line in the field, responding to the need to assess the reliability of the reconstructed maps. Despite the multiple methods available to perform this validation, the most common is the reconstruction of two half-maps and their subsequent comparison through the Fourier shell correlation. This practice is known in the field as the gold standard. It is a cheap substitute for the multiple estimations of the same



parameter: we use the estimate of the N^3 -dimensional signal as a proxy of the various estimations at a cost that involves estimating the parameters only once for each of the experimental images. If the parameter estimation errors are independent, zero-mean processes, this is a good practice. However, it does not protect us if the estimation errors are dependent or do not have a zero mean. For instance, a wrong estimate of the defocus of a micrograph will systematically affect the CTF correction of many particles. Suppose we have an attraction problem⁸ (a projection direction has a larger signal-to-noise ratio, SNR, than its surrounding or a 3D class has a larger SNR than the alternatives). In that case, the parameters would also be systematically misestimated.

In Sorzano *et al.*,⁸ we show that splitting the data into two halves is not a sufficient nor necessary condition for avoiding overfitting. It is not sufficient because we can make systematic mistakes in estimating the parameters in both halves. It is not necessary because not splitting the data into two halves does not necessarily lead to an incorrect estimation of the underlying parameters. The most common measure to assess the correction of a reconstructed map is the Fourier shell correlation between the two half maps. This measure is excellent when the difference between the two maps is independent and randomly distributed with zero mean. However, it is misled by systematic errors appearing in both halves.⁸

Despite the plethora of alternative validation methods, these are seldom used due to the difficulty of accessing them conveniently and smoothly. Scipion³⁸ is an appropriate platform for this evaluation as it integrates all the methods described in this paper, a total of 37, and provides effective ways of allowing them to interoperate. The server is open to new approaches, and any new validation tool is welcome to be incorporated into the validation report.

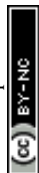
4 Conclusions

In this work, we have presented a validation server that assesses CryoEM maps' correction from multiple perspectives. The validation can be more profound as more data is available about the structure, the experimental images supporting it, and the image processing workflow followed to achieve it. In this way, we provide a progressive validation level that evaluates the map from different perspectives: (0) the map itself, (1) its two-halves, (2) the 2D classes of the particles, (3) the particles themselves, (4) the angular assignment of those particles, (5) the coordinates of those particles in the micrographs, (A) an atomic model of the structure, (W) the image processing workflow leading to the final result, and (O) validation through alternative experiments.

We believe that disclosing as much information as possible about the supporting experimental evidence leading to a CryoEM map will help the user of that structure better understand the strengths and weaknesses of the map at hand. And finally, it will ultimately contribute to more open and reproducible science.

Author contributions

COSS conceptualized the project, implemented its calculation engine, coordinated the integration of the different methods and performed the experiments. COSS and JMC analyzed the application and wrote the article. IS and JRM



implemented the webserver. All others helped to make the methods available through Scipion plugins. All authors have revised the manuscript.

Conflicts of interest

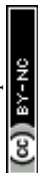
There are no conflicts to declare.

Acknowledgements

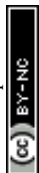
We are thankful to Philip Baldwin, Dmitry Lyumkis, and Gabriel Lander for making their validation methods available (4.h and A.c). Javier Vargas and Jordi Burguet would like to thank the Spanish Ministry of Science and Innovation for financial support through the call 2019 Proyectos de I+D+i – RTI Tipo A (PID2019-108850RA-I00) and Arrate Muñoz Barrutia, PID2019-109820RB-I00, MCIN/AEI/10.13039/501100011033/, cofinanced by European Regional Development Fund (ERDF), “A way of making Europe.”. The authors acknowledge the economic support from MICIN of the Instruct Image Processing Center (I2PC) as part of the Spanish participation in Instruct-ERIC, the European Strategic Infrastructure Project (ESFRI) in the area of Structural Biology, Grant PID2019-104757RB-I00 funded by MCIN/AEI/10.13039/501100011033/ and “ERDF A way of making Europe”, by the European Union and Grant PRE2020 - 093527 funded by MCIN/AEI/ 10.13039/501100011033 and by “ESF Investing in your future”. We also acknowledge support from “Comunidad Autónoma de Madrid” through Grant: S2017/BMD-3817, Instituto de Salud Carlos III (project IMPaCT-Data, exp. IMP/00019), co-funded by the European Union, European Regional Development Fund (ERDF, “A way to make Europe”), and European Union (EU) and Horizon 2020 through grants: EOSC Life (INFRAEOSC-04-2018, Proposal: 824087), High-ResCells (ERC – 2018 – SyG, Proposal: 810057), IMPaCT (WIDESPREAD-03-2018 – Proposal: 857203), EOSC – Synergy (EINFRA-EOSC-5, Proposal: 857647), iNEXT-Discovery (Proposal: 871037), EnLaCES (H2020-MSCA-IF-2020, Proposal: 101024130).

Notes and references

- 1 A. Patwardhan, *Acta Crystallogr., Sect. D: Struct. Biol.*, 2017, **73**, 503–508.
- 2 C. O. S. Sorzano and J. M. Carazo, Cryo-Electron Microscopy: the field of 1000+ methods, *J. Struct. Biol.*, 2022, **214**, 107861.
- 3 M. Woelfle, P. Olliaro and M. H. Todd, *Nat. Chem.*, 2011, **3**, 745–748.
- 4 A. Iudin, P. K. Korir, J. Salavert-Torres, G. J. Kleywegt and A. Patwardhan, *Nat. Methods*, 2016, **13**, 387–388.
- 5 M. Baker, *Nature*, 2016, **533**, 452–454.
- 6 C. L. Lawson, A. Kryshatovych, P. D. Adams, P. V. Afonine, M. L. Baker, B. A. Barad, P. Bond, T. Burnley, R. Cao, J. Cheng, *et al.*, *Nat. Methods*, 2021, **18**, 156–164.
- 7 R. Henderson, A. Sali, M. L. Baker, B. Carragher, B. Devkota, K. H. Downing, E. H. Egelman, Z. Feng, J. Frank, N. Grigorieff, W. Jiang, S. J. Ludtke, O. Medalia, P. A. Penczek, P. B. Rosenthal, M. G. Rossmann, M. F. Schmid, G. F. Schröder, A. C. Steven, D. L. Stokes, J. D. Westbrook, W. Wriggers,



- H. Yang, J. Young, H. M. Berman, W. Chiu, G. J. Kleywegt and C. L. Lawson, *Structure*, 2012, **20**, 205–214.
- 8 C. O. S. Sorzano, A. Jimenez-Moreno, D. Maluenda, M. Martinez, E. Ramirez-Aportela, R. Melero, A. Cuervo, J. Conesa, J. Filipovic, P. Conesa, L. del Caño, Y. C. Fonseca, J. Jiménez-de la Morena, P. Losana, R. Sánchez-García, D. Strelak, E. Fernandez-Gimenez, F. de Isidro, D. Herreros, J. L. Vilas, R. Marabini and J. M. Carazo, On bias, variance, overfitting, gold standard and consensus in Single Particle Analysis by Cryo-electron microscopy, *Acta Crystallogr., Sect. D: Struct. Biol.*, 2022, **D78**, 410–423.
- 9 J. L. Vilas, N. Tabassum, J. Mota, D. Maluenda, A. Jiménez-Moreno, T. Majtner, J. M. Carazo, S. T. Acton and C. O. S. Sorzano, *Curr. Opin. Struct. Biol.*, 2018, **52**, 127–145.
- 10 P. B. Rosenthal and R. Henderson, *J. Mol. Biol.*, 2003, **333**, 721–745.
- 11 E. Ramírez-Aportela, J. Mota, P. Conesa, J. M. Carazo and C. O. S. Sorzano, *IUCrJ*, 2019, **6**, 1054–1063.
- 12 S. Kaur, J. Gomez-Blanco, A. A. Khalifa, S. Adinarayanan, R. Sanchez-Garcia, D. Wrapp, J. S. McLellan, K. H. Bui and J. Vargas, *Nat. Commun.*, 2021, **12**, 1.
- 13 J. García Condado, A. Muñoz-Barrutia and C. O. S. Sorzano, *bioRxiv*, 2022, DOI: [10.1101/2022.03.01.482513](https://doi.org/10.1101/2022.03.01.482513).
- 14 C. O. S. Sorzano, J. Vargas, J. Oton, V. Abrishami, J. M. de la Rosa-Trevin, J. Gomez-Blanco, J. L. Vilas, R. Marabini and J. M. Carazo, *Prog. Biophys. Mol. Biol.*, 2017, **124**, 1–30.
- 15 M. Beckers and C. Sachse, *J. Struct. Biol.*, 2020, **212**, 107579.
- 16 G. Cardone, J. B. Heymann and A. C. Steven, *J. Struct. Biol.*, 2013, **184**, 226–236.
- 17 A. Kucukelbir, F. J. Sigworth and H. D. Tagare, *Nat. Methods*, 2014, **11**, 63–65.
- 18 J. L. Vilas, J. Gómez-Blanco, P. Conesa, R. Melero, J. M. de la Rosa Trevin, J. Otón, J. Cuenca, R. Marabini, J. M. Carazo, J. Vargas and C. O. S. Sorzano, *Structure*, 2018, **26**, 337–344.
- 19 J. L. Vilas, H. D. Tagare, J. Vargas, J. M. Carazo and C. O. S. Sorzano, *Nat. Commun.*, 2020, **11**, 55.
- 20 Y. Z. Tan, P. R. Baldwin, J. H. Davis, J. R. Williamson, C. S. Potter, B. Carragher and D. Lyumkis, *Nat. Methods*, 2017, **14**, 793–796.
- 21 C. O. S. Sorzano, J. Vargas, J. M. de la Rosa-Trevin, A. Zaldívar-Peraza, J. Otón, V. Abrishami, I. Foche, R. Marabini, G. Caffarena and J. M. Carazo, *Proc. Intl. Work-Conference on Bioinformatics and Biomedical Engineering, IWBBIO*, 2014, p. 950.
- 22 J. Méndez, E. Garduño, J. M. Carazo and C. O. S. Sorzano, *J. Struct. Biol.*, 2021, **213**, 107771.
- 23 J. Vargas, J. Otón, R. Marabini, J. M. Carazo and C. O. S. Sorzano, *Sci. Rep.*, 2016, **6**, 21626.
- 24 J. Vargas, R. Melero, J. Gómez-Blanco, J. M. Carazo and C. O. S. Sorzano, *Sci. Rep.*, 2017, **7**, 6307.
- 25 B. Heymann, *AIMS Biophys.*, 2015, **2**, 21–35.
- 26 K. Naydenova and C. J. Russo, *Nat. Commun.*, 2017, **8**, 629.
- 27 P. R. Baldwin and D. Lyumkis, *Prog. Biophys. Mol. Biol.*, 2020, **150**, 160–183.
- 28 R. Sanchez-Garcia, J. Segura, D. Maluenda, C. O. S. Sorzano and J. M. Carazo, *J. Struct. Biol.*, 2020, **210**, 107498.
- 29 G. Pintilie and W. Chiu, *Acta Crystallogr., Sect. D: Struct. Biol.*, 2021, **77**, 1142–1152.



- 30 E. Ramírez-Aportela, D. Maluenda, Y. C. Fonseca, P. Conesa, R. Marabini, J. B. Heymann, J. M. Carazo and C. O. S. Sorzano, *Nat. Commun.*, 2021, **12**, 42.
- 31 M. A. Herzik, J. S. Fraser and G. C. Lander, *Structure*, 2019, **27**, 344–358.
- 32 J. L. Vilas, J. Vargas, M. Martínez, E. Ramírez-Aportela, R. Melero, A. Jiménez-Moreno, E. Garduño, P. Conesa, R. Marabini, D. Maluenda, J. M. Carazo and C. O. S. Sorzano, *J. Struct. Biol.*, 2020, **209**, 107447.
- 33 P. V. Afonine, B. P. Klaholz, N. W. Moriarty, B. K. Poon, O. V. Sobolev, T. C. Terwilliger, P. D. Adams and A. Urzhumtsev, *Acta Crystallogr., Sect. D: Struct. Biol.*, 2018, **74**, 814–840.
- 34 B. A. Barad, N. Echols, R. Y.-R. Wang, Y. Cheng, F. DiMaio, P. D. Adams and J. S. Fraser, *Nat. Methods*, 2015, **12**, 943–946.
- 35 G. Terashi, X. Wang, S. R. M. V. Subramaniya, J. J. G. Tesmer and D. Kihara, 2022, submitted, <https://colab.research.google.com/drive/1Q-Dj42QJV08TCOLXMQBJlv1zInxPkOu?usp=sharing>.
- 36 C. J. Williams, J. J. Headd, N. W. Moriarty, M. G. Prisant, L. L. Videau, L. N. Deis, V. Verma, D. A. Keedy, B. J. Hintze, V. B. Chen, S. Jain, S. M. Lewis, W. B. Arendall, J. Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson and D. C. Richardson, *Protein Sci.*, 2018, **27**, 293–315.
- 37 T. Cragolini, H. Sahota, A. P. Joseph, A. Sweeney, S. Malhotra, D. Vasishtan and M. Topf, *Acta Crystallogr., Sect. D: Struct. Biol.*, 2021, **77**, 41–47.
- 38 J. M. de la Rosa-Trevín, A. Quintana, L. Del Cano, A. Zaldívar, I. Foche, J. Gutiérrez, J. Gómez-Blanco, J. Burguet-Castell, J. Cuenca-Alba, V. Abrishami, J. Vargas, J. Otón, G. Sharov, J. L. Vilas, J. Navas, P. Conesa, M. Kazemi, R. Marabini, C. O. S. Sorzano and J. M. Carazo, *J. Struct. Biol.*, 2016, **195**, 93–99.
- 39 M. Sinnott, S. Malhotra, M. S. Madhusudhan, K. Thalassinou and M. Topf, *Structure*, 2020, **28**, 1061–1070.
- 40 A. Jiménez, S. Jonic, T. Majtner, J. Oón, J. L. Vilas, D. Maluenda, J. Mota, E. Ramírez-Aportela, M. Martínez, Y. Rancel, J. Segura, R. Sánchez-García, R. Melero, L. Del Caño, P. Conesa, L. Skjaerven, R. Marabini, J. M. Carazo and C. O. S. Sorzano, *Bioinformatics*, 2019, **35**, 2427–2433.
- 41 R. Henderson, S. Chen, J. Z. Chen, N. Grigorieff, L. A. Passmore, L. Ciccarelli, J. L. Rubinstein, R. A. Crowther, P. L. Stewart and P. B. Rosenthal, *J. Mol. Biol.*, 2011, **413**, 1028–1046.
- 42 T. Kato, F. Makino, T. Nakane, N. Terahara, T. Kaneko, Y. Shimizu, S. Motoki, I. Ishikawa, K. Yonekura and K. Namba, *Microsc. Microanal.*, 2019, **25**, 998–999.
- 43 R. Melero, C. O. S. Sorzano, B. Foster, J. L. Vilas, M. Martínez, R. Marabini, E. Ramírez-Aportela, R. Sanchez-Garcia, D. Herreros, L. Del Caño, P. Losana, Y. Fonseca-Reyna, P. Conesa, D. Wrapp, P. Chacón, J. S. McLellan, H. D. Tagare and J. M. Carazo, *IUCr*, 2020, **7**, 1059–1069.
- 44 L. Yurkovetskiy, X. Wang, K. E. Pascal, C. Tomkins-Tinch, T. P. Nyalile, Y. Wang, A. Baum, W. E. Diehl, A. Dauphin, C. Carbone, K. Veinotte, S. B. Egri, S. F. Schaffner, J. E. Lemieux, J. B. Munro, A. Rafique, A. Barve, P. C. Sabeti, C. A. Kyratsous, N. V. Dudkina, K. Shen and J. Luban, *Cell*, 2020, **183**, 739–751.
- 45 S. M.-C. Gobeil, K. Janowska, S. McDowell, K. Mansouri, R. Parks, K. Manne, V. Stalls, M. F. Kopp, R. Henderson, R. J. Edwards, B. F. Haynes and P. Acharya, *Cell Rep.*, 2021, **34**, 108630.
- 46 K. M. Yip, N. Fischer, E. Paknia, A. Chari and H. Stark, *Nature*, 2020, **587**, 157–161.

