Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Nat Methods. Author manuscript; available in PMC 2020 August 25.

Published in final edited form as:

Nat Methods. 2020 March ; 17(3): 328–334. doi:10.1038/s41592-020-0731-1.

Measurement of Atom Resolvability in CryoEM Maps with Q-scores

Grigore Pintilie[1,*], **Kaiming Zhang**[1], **Zhaoming Su**[1], **Shanshan Li**[1], **Michael F. Schmid**[2], **Wah Chiu**[1,2,*]

[1.]Department of Bioengineering, James H. Clark Center, Stanford University, Stanford, California, USA

[2.]Division of CryoEM and Bioimaging, SSRL, SLAC National Accelerator Laboratory, Stanford University, Menlo Park, California, USA

Abstract

CryoEM density maps are now at the point where resolvability of individual atoms can be achieved. However, resolvability is not necessarily uniform throughout the map. We introduce a quantitative parameter to characterize the resolvability of individual atoms in cryoEM maps, the map Q-score. Q-scores can be calculated for atoms in proteins, nucleic acids, water, ligands, and other solvent atoms, using models fitted to or derived from cryoEM maps. Q-scores can also be averaged to represent larger features such as entire residues and nucleotides. Averaged over entire models, Q-scores correlate very well with the estimated resolution of cryoEM maps for both protein and RNA. Assuming the models they are calculated from are well-fitted to the map, Q-scores can thus be used as a measure of resolvability in cryoEM maps at various scales, from entire macromolecules down to individual atoms. Q-score analysis of multiple cryoEM maps of

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Corresponding authors: gregp@slac.stanford.edu, wahc@stanford.edu.

Author Contributions

G.P and W.C. conceived Q-scores. G.P. implemented the software and performed testing. K.Z. collected the images and reconstructed the maps (EMDB: 20026, 20027, 20028). Z.S. and S.L. provided additional data (not shown) for testing the Q-score. M.F.S. and W.C. contributed the discussion during the development. G.P. wrote the manuscript with inputs from other authors.

Competing Financial Interests Statement

We declare that none of the authors have competing financial or non-financial interests as defined by Nature Research.

Data Availability Statement

The cryoEM maps of Apoferritin have been deposited in the EMDB with accession codes 20026 (1.75Å), 20027 (2.3Å), and 20028 (3.1Å). The figures show these maps and also other maps and models available in the EMDB and PDB (accession codes specified in Figures text, also Tables 1 and 2). All data including calculations based on these maps that support the findings of this study are available from the corresponding author upon request.

Code Availability Statement

Q-scores are implemented in the MapQ plugin to UCSF Chimera, available at https://github.com/gregdp/mapq. A tutorial is also available at the link under the “tutorials” folder. Pseudo code is provided under the “docs” folder.

Reproduction Instructions

To reproduce Q-scores shown in the various images, the maps and models indicated can first be downloaded from EMDB and PDB. For the Apoferritin examples, the models may first have to be rigidly fitted to the density (e.g. using Chimera/Segger/Fit to Segments), and then refined using Phenix.real_space_refinement. Then, Q-scores can be calculated using the MapQ plugin (https://github.com/gregdp/mapq). For more details, see tutorials/MapQ_Tutorial.pdf at the link.

The statistical analysis (Q-scores vs. map resolution), can be reproduced using the same approach, downloading the maps and models indicated in Tables 1 and 2. Q-scores can also be calculated for maps and model from the command line, making it easier to run a large batch of maps and models (see documentation/QScores_CommandLine.docx at the github link above).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the same proteins derived from different labs confirms the reproducibility of structural features from side chains down to water and ion atoms.

Introduction

CryoEM single-particle methods strive to create accurate, high-resolution 3D maps of macromolecules. Depending on many factors including imaging apparatus, detector, reconstruction method, structure flexibility, sample heterogeneity, and differential radiation damage, resulting maps have varying degrees of resolvability. Accurate quantification of resolvability in cryoEM maps has been a challenge in the field[1]. This task is very important as it can affect the interpretation of such maps.

For every cryoEM map, a resolution is estimated from a Fourier shell correlation (FSC) plot between two independent reconstructions, each reconstruction stemming from a separate half of the data set[2]. It is well recognized that cryoEM maps usually do not have isotropic resolution throughout, and thus local resolution is typically estimated, e.g. with ResMap[3], Bsoft[4], or MonoRes[5]. However such loclal resolutions do not easily translate to particular features of interest such as side chains or individual atoms.

Atomic models can be either fitted or built directly into cryoEM maps[6,7]. Map-model scores are then calculated to assess how well the model fits the map[8]. Real-space refinement[9] or flexible fitting[10,11] can be applied, making sure to not overfit to noise[12,13]. The latter is accomplished through stereochemical restraints, e.g. bond lengths, angles, dihedrals, preferred rotamers and van-der Waals distances, and additional secondary-structure constraints, e.g. in the form of hydrogen bonds[9,11,14,15].

Once an atomic model has been fitted to or derived from a cryoEM map, it can then be used to assess the map itself. This can be done in several ways, including a map-model FSC curve, which requires that the model first be converted to a cryoEM-like map at the same resolution as the original map. Such an FSC plot reflects the entire map volume. Proper masking may be used to assess smaller features such as individual protein chains[12], however it is impractical to assess even smaller features such as side chains or individual atoms using this approach.

Other methods that assess smaller features in a cryoEM map using a fitted model include EMRinger[16] and Z-scores[17]. EMRinger considers map values near carbon-β atoms, while Z-scores can be applied to secondary structure elements (such as α-helices and β-sheets) or side chains. These scores were shown to correlate with map resolution when averaged over entire maps and models. Moreover, they can also identify features in the model (e.g. secondary structure elements or side chains) which are not well-resolved or not fitted properly to the map.

CryoEM maps have reached resolutions nearer to atomic-dimensions, for example apoferritin at 1.54Å (EMD:9865), 1.62Å (EMD:0144)[18], 1.65Å (EMD:9599), and 1.75Å (EMD: 20026). At such resolutions, we may start to assess the resolvability of individual atoms. In crystallography, B-factors or atomic displacement parameters (ADPs) reflect the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

uncertainty in the position of any atom, and are refined from diffraction data[19–21]. ADPs can also be calculated in cryoEM maps[22]. However, since ADPs are typically refined with restraints, they are not dependent only on the map values around the atom. Other ways to measure positional uncertainties include multi-model refinement[23] and molecular dynamics[12,24]; these also assume various restraints on atoms and hence do not reflect map values alone.

In this paper, we introduce Q-scores, which are calculated directly from map values around an atom's position. A similar score is EDIA, which was applied to high-resolution X-ray maps. The EDIA method considers map values within each atom's radius, which is parameterized for different elements and resolutions. In contrast, Q-scores are calculated independently of element type or map resolution. We apply Q-scores to measure resolvability of individual atoms, including solvent atoms, and also of groups of atoms such as side chains in proteins and bases in nucleic acids.

Results

Atomic Map Profiles

The basis of the Q-score is the atomic map profile. Atomic map profiles are calculated by averaging map values at increasing radial distances from an atom's position. The radial distances range from 0Å to 2.0Å, and only points that are closer to the atom in question than to any other atoms in the model are considered. Figure 1A shows example atomic profiles in our two new maps of Apoferritin with resolutions of 1.75Å and 2.32Å, now deposited as EMD:20026, and EMD:20027.

When calculating the profile for an atom, map values at N points are used to calculate the average at a particular distance, r. The N points are distributed evenly across the part of the sphere (centered at the atom, with radius r) that is closer to the atom and not any other atom in the model. At r=0 or the atom center, the map value is duplicated N times, so that N is the same at each radial distance. In all calculations used here, we use N=8. Larger values of N typically create smoother profiles, however have only minor effects on Q-scores described below.

The model in Figure 1 is the X-ray model of Apoferritin, (PDB:3ajo), which was first rigidly fitted to the cryoEM map, and then further refined into each cryoEM map using Phenix real-space refinement[9]. In the examples, atomic profiles have Gaussian-like contours. We consider a Gaussian equation of the form:

$$y = \boldsymbol{A} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} + \boldsymbol{B} \tag{1}$$

Gaussian functions of the form in Eqn.1, where x is the radial distance and y the average map value, fit well to the atomic profiles shown in Figure 1 up to a distance of 2Å, with a mean error of 2.4%. For higher resolution data, e.g. from X-ray crystallography, multiple Gaussians are used to closely represent atomic form factors[25], however we do not consider that here. Past 2Å from the atom, map profiles observed in these and other similar resolution

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

cryoEM maps become noisy and start to increase. This is likely due to effects from other nearby atoms and/or solvent.

When the model is well-fitted to the map, the width of the Gaussian function (Eqn.1) fitted to the profile, $\boldsymbol{\sigma}$, may be considered to be proportional to factors such as the resolution of the map and the overall mobility of the atom. Regardless of the cause, in this paper we assume that the profile seen in the map indicates to what degree the atom is resolved: narrower profiles indicate the atom is better resolved, while wider profiles indicate the atom is less well resolved.

Q-score

The Q-score measures how similar the map profile of an atom is to a Gaussian-like function we would see if the atom is well-resolved. Thus, to calculate it, the atomic map profile is compared to a 'reference Gaussian' as given by Eqn. 1, with the following parameters:

$$\mu = 0 \tag{2}$$

$$\boldsymbol{A} = avg_M + 10\sigma_M \tag{3}$$

$$\boldsymbol{B} = avg_M - 1\sigma_M \tag{4}$$

$$\sigma = 0.6\text{Å} \tag{5}$$

In the above, the mean, μ, is set to 0, as the reference Gaussian is centered at the atom's position. The parameters $\boldsymbol{A}$ and $\boldsymbol{B}$ are obtained using the mean/average across all values in the entire map, avg_M, and the standard deviation of all values around this mean, $\boldsymbol{\sigma}_M$. The width of the reference Gaussian is set as $\boldsymbol{\sigma}$=0.6. These parameters were chosen to make the reference Gaussian roughly match the atomic profile of a well-resolved atom in the 1.54Å cryoEM map as shown in Figure 2B.

The Q-score is then calculated as a correlation between values in the atomic profile obtained from the map, $\boldsymbol{u}$, by trilinear interpolation to nearest 8 grid points, and values obtained from the reference Gaussian, $\boldsymbol{v}$. The following normalized about-the-mean cross-correlation formula is used:

$$Q = \frac{\langle \boldsymbol{u} - \boldsymbol{u}_{mean} \rangle \langle \boldsymbol{v} - \boldsymbol{v}_{mean} \rangle}{|\boldsymbol{u} - \boldsymbol{u}_{mean}||v - \boldsymbol{v}_{mean}|} \tag{6}$$

Several atomic profiles and reference Gaussians are illustrated in Figure 2. At resolutions close to 1.5Å, the atomic profiles are more similar to the reference Gaussian, and hence Q-scores are higher. At lower resolutions, the atomic profiles of the same atom are wider than the reference Gaussian, hence Q-scores are lower. Q-scores would also be low for atomic profiles that are mostly noise (e.g. random values or a sharp peak). In some cases when the atom is not well-placed in the map, the Q-score can be negative if the atomic profile has a shape that increases away from the atom's position.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Q-scores are low when the entire model is placed incorrectly in the map, e.g. during a global search. They can increase if the model-map fit is improved by local refinement (Supplementary Figure 1). Q-scores begin to decrease as resolutions of the map increase beyond 1.30Å, as atomic profiles begin to be much narrower than the reference Gaussian (Supplementary Figure 2). This effect may be useful in cryoEM maps to give very sharp peaks, which are more likely to be noise, lower Q-scores.

Calculating Q-scores is similar to calculating a cross-correlation between the model and a cryoEM map, using a simulated map of the model blurred using a Gaussian function with the parameters in Eqns. 2–5. The main difference is that with Q-scores, the cross-correlation is performed atom-by-atom, separating out parts of the density that are closest to each atom. The cross-correlation about the mean is used so that the Q-scores decrease as resolution also decreases. When not subtracting the mean, this effect would not be ensured, as shown previously[17] and also in Supplementary Figure 3.

We tested the effect of several factors on Q-scores. First, using the cross-correlation about the mean makes the Q-scores insensitive to the height and vertical offset of the reference Gaussian (Supplementary Figure 3). This means that as long as map values are decreasing around an atom, regardless of their relative magnitude in the map, the Q-score for the atom could still be high. Second, small changes in grid step and placement do not affect the Q-score; however if the grid step is too large relative to the resolution of the map, resolvability and also Q-scores can start to decrease (Supplementary Figure 4). Finally, sharpening can increase the visible detail in the map along with Q-scores, but Q-scores start to decrease if excessive sharpening is applied (Supplementary Figure 5).

Q-scores of Atoms in Proteins

Figure 3 shows Q-scores for atoms taken from maps of Apoferritin at various resolutions. One of the maps is an X-ray map at 1.52Å resolution (2fo-fc, PDB:3ajo) as a reference; another is a recent high-resolution map at 1.54Å (EMD:9599). The other three are new maps we reconstructed to 1.75Å (EMD:20026), 2.3Å (EMD:20027), and 3.1Å (EMD:20028) with different numbers of particle images from the same data set. For the cryoEM maps, the X-ray model PDB:3ajo was fitted to the density and refined using Phenix real-space refinement[9].

In Figure 3, Q-scores for each atom correlate well with visual resolvability at the contour level used in each case, i.e. the more resolvable an atom, the higher the Q-score. However, in some cases, the Q-score for an atom can be relatively high even if there is no map contour around it; this is due to the effect mentioned previously that even if the map values around an atom are low, the Q-score can still be high if they are decreasing away from the atom.

Resolvability and Q-scores can decrease for some residues faster than others as a function of resolution. For example, in Figure 3, the Q-score for ASP126 drops more than for ASN25 from 1.52Å to 3.9Å. This effect may be due to several reasons. First, some residue types may be more susceptible to radiation damage (as previously shown using EMRinger[16]). Also, certain residue types may be more conformationally dynamic, or occur in environments that are more dynamic (e.g. solvent accessible), and hence may not resolve as

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

well with fewer number of particles. Finally, the interaction of the electron beam with negatively charged side chains may have a weakening effect on map values around them[22].

Q-scores for Atoms in Nucleic Acids

Q-scores can also be calculated for atoms in nucleic acids. In Figure 4, we used several maps and models containing RNA from the EMDB at resolutions ranging from 2.5Å to 4.0Å. Q-scores were averaged over atoms in bases (labeled with Q_{base}), phosphate-sugar backbones (labeled with Q_{bb}), and entire nucleotides. As with proteins, Q-scores decrease with resolvability and estimated map resolution. Figure 4 also illustrates a general trend that at ~4Å and lower resolutions, stacked bases from adjacent nucleotides are typically not separable in cryoEM maps, whereas at higher than 4Å resolutions, they usually do become separate at some contour levels.

It is also interesting to note that for the examples in Figure 4, at high resolutions (~2.5Å), the difference in Q-score or resolvability of individual bases is higher than that of the backbone (0.84 for base vs. 0.73 for backbone). Going towards lower resolutions in this example, bases become less resolvable (0.45 for bases vs 0.56 for backbone). This may be counter-intuitive as bases can have higher values in the map (i.e. appear first at a high contour level). However, these contours may have overall less detail as adjacent stacked bases are not fully separable at any contour level.

Q-score vs. Resolution

Q-scores can also be averaged across an entire model to represent an average resolvability measure for the entire map. Such average Q-scores were plotted as a function of reported resolution for a number of maps and models obtained from the EMDB. Figure 5 shows these plots for two sets of maps and models, one set using only atoms in proteins, and the other set only atoms in nucleic acids. The full sets are listed in Tables 1 and 2. In both cases, the average Q-score correlates very strongly to reported resolution. This strong correlation indicates that Q-scores closely capture the resolvability of atomic features in cryoEM maps, much as the estimated resolution of a map does. However, Q-scores are useful in quantifying resolvability of small features within each map down to individual atoms.

The linear plots in Figure 5 show that average Q-scores drop toward 0 at ~6–7 Å, however an analysis using simulated maps indicates that they taper off and decrease slowly toward 0 at lower resolutions (Supplementary Figure 6). Negative Q-scores would only be expected if atoms are not placed on peaks, such that map values increase away from their position. Nevertheless, due to the change in rate of decrease, we expect that Q-scores are most useful at resolutions better than 5–6Å.

Q-scores vs. B-factors and ADPs

B-factors and atomic displacement parameters (ADPs) are used in X-ray crystallography to convey the positional uncertainty of atoms[19–21]. They are also dependent to some degree on resolution[27] (Supplementary Figure 7). When refining B-factors and ADPs, various restraints, parameters and initial values can be used, hence the results in each map may vary.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Comparisons of B-factors/ADPs to Q-scores show that they correlate only weakly (Supplementary Figures 8,9). Hence they likely convey somewhat different information.

Q-scores of Solvent Atoms

The X-ray Apoferritin model (PDB:3ajo) contains one protein chain, 229 oxygen (O) atoms (from water) and 12 Mg atoms. A closeup on the 2Fo-Fc map and model with two Mg and three O atoms is shown in Figure 6A. Figure 6 B,C,D shows cryoEM maps at near-atomic resolutions (1.54Å, 1.65Å, and 1.75Å). The model used all cases comes from the X-ray map. It is reassuring to see that some of the solvent atoms in the X-ray structure can also be observed in the cryoEM maps (e.g. Mg183, O280, O236). However, some of the solvent atoms (e.g. Mg184), are not seen equally well in all three maps; for example, in the 1.54Å and 1.65Å maps, Mg184 has low Q-score (0.12 and 0.03 respectively). Such differences may be due to different affinities at some sites and/or different biochemical conditions across the different data sets.

Supplementary Figure 10A shows distributions of Q-scores for solvent atoms in the X-ray map (PDB:3ajo). Most solvent atoms have very high Q-scores of 0.9 and higher. Visual inspection confirmed that all these solvent atoms can be seen in the X-ray map (2fo-fc), e.g. as shown in Figure 6A. Supplementary Figure 10B,C shows Q-score distribution plots for the same model rigidly fitted to, and also refined in, the cryoEM maps at 1.54Å and 1.75Å resolution. The model was refined in the cryoEM maps including solvent atoms, using Phenix real-space refine[9].

For the rigidly fitted model, Q-scores of the solvent atoms are considerably lower than in the X-ray map (Supplementary Figure 10B). For example, in the 1.75Å cryoEM map, only 44 of the 229 O atoms from water have Q-scores of 0.8 and higher. In the 1.54Å map, 68 have Q-scores of 0.8 and higher. Thus some of the solvent atoms in the X-ray structure may not be resolvable in the cryoEM maps or potentially be in different positions.

To explore whether solvent atoms may have different positions in the cryoEM maps, Q-scores of the solvent atoms were also calculated in the X-ray structure after real-space refinement with Phenix[9]. The distributions in the Q-scores for solvent atoms after this procedure are plotted in Supplementary Figure 10B, C for the two cryoEM maps. Q-scores are now higher; 142 water atoms in the 1.54Å map and 145 atoms in the 1.75Å map have Q-scores of 0.8 and higher, compared to 225 water atoms in the X-ray map with Q-scores of 0.8 and higher.

We further consider water atoms with Q-scores of 0.8 and higher after refinement, which can be considered to be resolved in the cryoEM maps. In the 1.54Å map, the 142 water atoms with Q-scores 0.8 and higher moved between 0.1Å and 2.2Å, on average 0.54Å. In the 1.75Å map, the 145 water atoms with Q-scores of 0.8 and higher moved between 0.1Å and 1.6Å, on average 0.67Å. Radial distance plots in Supplementary Figure 11 show sharp peaks at ~2.8Å for water-water and water-protein distances in X-ray maps, but more diffuse peaks around the same distance in cryoEM maps.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Although it is difficult to assess the exact cause of these relatively small distance variations between X-ray and cryoEM structures, it is reasonable to conclude that many of the waters in the X-ray structure are also resolved and near the same positions in cryoEM maps. Water networks have been shown to be important in ligand binding affinities and to vary due to structural differences even in X-ray structures[28]. Further studies with more cryoEM maps at similar resolutions may further elucidate and characterize such variations.

In the above analysis, solvent atom positions were based on those originally observed in the X-ray structure. If one studies a *de novo* map, the identification of solvent atoms would require a protocol used in modeling software[30]. In addition to such a protocol, Q-scores may be useful as an additional parameter to assist in the finding of such solvent atoms.

Q-scores of Solvent Atoms at Different Resolutions

Finally, we looked at the resolvability and Q-scores of solvent atoms in cryoEM maps of Apoferritin at lower resolutions, as shown in Figure 6 E,F. The locations of the solvent atoms are again taken from the X-ray model (PDB:3ajo). Mg183 appears resolved at both 1.75Å and 2.3Å, with separable contours in both maps and high Q-scores (0.93 and 0.80). In the 3.1Å map, the contour is no longer separable from that of the nearby His65 residue, and the Q-score is also considerably lower (0.60). The water atoms are similarly resolved in the 1.75Å and 2.3Å maps and contours around them can be seen, however at 3.1Å they can no longer be seen and Q-scores become very low.

At 3.1Å resolution, both Mg atoms still have relatively high Q-scores, and they are inside the map contour at lower threshold. Thus even at such lower resolutions, it appears ions can significantly influence cryoEM map values. Thus even at these resolutions, solvent atoms perhaps may be considered in the model, particularly if known structures of the same complex at higher resolutions also contain such atoms. Consequently, this may improve the accuracy of side chain positions and rotameric configurations during refinement.

Discussion

Q-scores measure the resolvability of individual atoms in a cryoEM map, using an atomic model fitted to or built into the map. It should be noted that nothing is assumed about the model itself, e.g. whether it has good stereochemistry; this could be deduced with other scores such as the Molprobity score[31,31]. Q-scores averaged over entire models correlate very closely to the reported resolution of the maps in which they are calculated. The score can also be useful to analyze the map and its resolvability in different regions, and also test whether the model may need further refinement in some areas as indicated by low Q-scores. Here, Q-scores were also applied to various maps at different resolutions to show quantifiable trends across different side chains in proteins, bases in nucleic acids, and also to assess the resolvability of solvent atoms and ions. Q-scores should continue to be a useful metric in the analysis of cryoEM maps and models.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Online Methods

CryoEM

Human apoferritin samples were provided by F. Sun and X.J. Huang (Institute of Biophysics, CAS). Images of the sample were collected in Titan Krios electron microscope (Thermo Fisher) at 300 keV, equipped with BioQuantum energy filter and K2 director detector (Gatan). A total of 1,100 images were recorded in movie mode. Motion correction was performed with MotionCor2[1] (v1.1.0). Particles were picked using the EMAN2 neural network particle picker[2] (EMAN2 v2.22). 3D reconstruction was performed using Relion[3] (v3.0). Map resolution was estimated from two independently reconstructed maps. Three maps of apoferritin were reconstructed using different number of particles: 1.75Å using 70,648 particles, 2.3Å using 9,600 particles, and 3.1Å using 495 particles. All three maps were reconstruction with octahedral symmetry.

Models

The X-ray model PDB:3ajo of human apoferritin was rigidly fitted to each new apoferritin cryoEM map using the Segger[4] plugin in UCSF Chimera[5], (v2.3), and refined using Phenix real-space refinement[6] (v1.14 build 3260). Q-score calculations were performed with the MapQ plugin to UCSF Chimera (v1.2).

Statistical Analysis

The Pearson correlation (r) values for Q-scores vs. reported resolution (plotted in Figure 5) were calculated using *python* and the *scipy.stats.linregress* function. The reported *r_value* was squared to obtain r^2 in each case. In these figures, the number of data points is the number of entries in the respective table (Table 1 for Figure 5A an Table 2 for Figure 5B). For all figures, since the methods used are deterministic, the measurements were only performed once to obtain the displayed values.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research has been supported by NIH grants (R01GM079429, P41GM103832, and S10OD021600). Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41GM103311). We thank F. Sun and X.J. Huang (Institute of Biophysics, CAS) for the human Apoferritin samples. We also thank C. Lawson for discussions related to B-factors and ADPs.

References

1. Challenges for cryo-EM. Nat. Methods 15, 985–985 (2018).
2. Henderson R et al. Outcome of the first electron microscopy validation task force meeting. Struct. Lond. Engl 1993 20, 205–214 (2012).
3. Kucukelbir A, Sigworth FJ & Tagare HD Quantifying the local resolution of cryo-EM density maps. Nat. Methods 11, 63–65 (2014). [PubMed: 24213166]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

4. Heymann JB Guidelines for using Bsoft for high resolution reconstruction and validation of biomolecular structures from electron micrographs. Protein Sci. Publ. Protein Soc 27, 159–171 (2018).
5. Local resolution of cryo-EM maps with MonoRes. Nature Methods https://www.nature.com/articles/nmeth.4651 (2018) 10.1038/nmeth.4651.
6. Pintilie G & Chiu W Comparison of Segger and other methods for segmentation and rigid-body docking of molecular components in cryo-EM density maps. Biopolymers 97, 742–760 (2012). [PubMed: 22696409]
7. 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 (2018). [PubMed: 30377346]
8. Joseph AP, Lagerstedt I, Patwardhan A, Topf M & Winn M Improved metrics for comparing structures of macromolecular assemblies determined by 3D electron-microscopy. J. Struct. Biol 199, 12–26 (2017). [PubMed: 28552721]
9. Afonine PV et al. Real-space refinement in PHENIX for cryo-EM and crystallography. Acta Crystallogr. Sect. Struct. Biol 74, 531–544 (2018).
10. Trabuco LG et al. Applications of the molecular dynamics flexible fitting method. J. Struct. Biol 173, 420–427 (2011). [PubMed: 20932910]
11. Joseph AP et al. Refinement of atomic models in high resolution EM reconstructions using Flex-EM and local assessment. Methods San Diego Calif 100, 42–49 (2016).
12. Pintilie G, Chen D-H, Haase-Pettingell CA, King JA & Chiu W Resolution and Probabilistic Models of Components in CryoEM Maps of Mature P22 Bacteriophage. Biophys. J 110, 827–839 (2016). [PubMed: 26743049]
13. DiMaio F, Zhang J, Chiu W & Baker D Cryo-EM model validation using independent map reconstructions. Protein Sci 22, 865–868 (2013). [PubMed: 23592445]
14. Trabuco LG, Villa E, Schreiner E, Harrison CB & Schulten K Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. Methods San Diego Calif 49, 174–180 (2009).
15. DiMaio F, Tyka MD, Baker ML, Chiu W & Baker D Refinement of protein structures into low-resolution density maps using rosetta. J. Mol. Biol 392, 181–190 (2009). [PubMed: 19596339]
16. 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]
17. Pintilie G & Chiu W Assessment of structural features in Cryo-EM density maps using SSE and side chain Z-scores. J. Struct. Biol 204, 564–571 (2018). [PubMed: 30144506]
18. Zivanov J et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. eLife 7, e42166 (2018). [PubMed: 30412051]
19. Dunitz JD, Schomaker V & Trueblood KN Interpretation of atomic displacement parameters from diffraction studies of crystals. J. Phys. Chem 92, 856–867 (1988).
20. Trueblood KN et al. Atomic Dispacement Parameter Nomenclature. Report of a Subcommittee on Atomic Displacement Parameter Nomenclature. Acta Crystallogr. A 52, 770–781 (1996).
21. Carugo O How large B-factors can be in protein crystal structures. BMC Bioinformatics 19, (2018).
22. Hryc CF et al. Accurate model annotation of a near-atomic resolution cryo-EM map. Proc. Natl. Acad. Sci 201621152 (2017) 10.1073/pnas.1621152114.
23. Herzik MA, Fraser JS & Lander GC A Multi-model Approach to Assessing Local and Global Cryo-EM Map Quality. Struct. Lond. Engl 1993 27, 344–358.e3 (2019).
24. Singharoy A et al. Molecular dynamics-based refinement and validation for sub-5 Å cryo-electron microscopy maps. eLife 5, e16105 (2016). [PubMed: 27383269]
25. Wahab MA Essentials of Crystallography (Alpha Science International, 2009).
26. Graef MD Introduction to Conventional Transmission Electron Microscopy by Marc De Graef. Cambridge Core /core/books/introduction-to-conventional-transmission-electron-microscopy/257FBB684B79174B1C6D3ACCC256ECC0 (2003) 10.1017/CBO9780511615092.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

27. Holton JM A beginner's guide to radiation damage. J. Synchrotron Radiat 16, 133–142 (2009). [PubMed: 19240325]
28. Darby JF et al. Water Networks Can Determine the Affinity of Ligand Binding to Proteins. J. Am. Chem. Soc 141, 15818–15826 (2019). [PubMed: 31518131]
29. Adams PD et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr (2010).
30. Emsley P & Cowtan K Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr 60, 2126–2132 (2004). [PubMed: 15572765]
31. Chen VB et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr 66, 12–21 (2010). [PubMed: 20057044]

Methods-only References

1. Zheng SQ et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods 14, 331–332 (2017). [PubMed: 28250466]
2. Bell JM, Chen M, Durmaz T, Fluty AC & Ludtke SJ New software tools in EMAN2 inspired by EMDatabank map challenge. J. Struct. Biol 204, 283–290 (2018). [PubMed: 30189321]
3. Zivanov J et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. eLife 7, e42166 (2018). [PubMed: 30412051]
4. Pintilie G & Chiu W Comparison of Segger and other methods for segmentation and rigid-body docking of molecular components in cryo-EM density maps. Biopolymers 97, 742–760 (2012). [PubMed: 22696409]
5. Pettersen EF et al. UCSF Chimera - A visualization system for exploratory research and analysis. J. Comput. Chem 25, 1605–1612 (2004). [PubMed: 15264254]
6. Afonine PV et al. Real-space refinement in PHENIX for cryo-EM and crystallography. Acta Crystallogr. Sect. Struct. Biol 74, 531–544 (2018).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

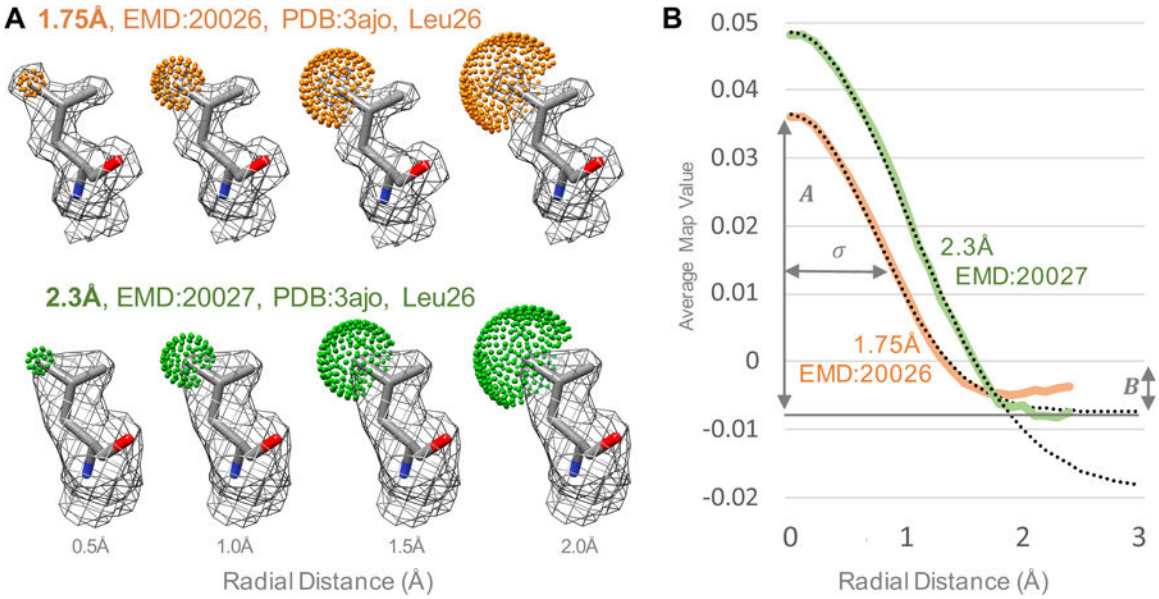

Figure 1.

Atomic map profiles in cryoEM two maps of Apoferritin. (A) The residue Leu26 in the fitted model (PDB:3ajo) is shown, along with contour surface of the cryoEM map around this residue. Spherical shells of points centered on the CD2 atom are shown at increasing radial distances. Only points that are closer to the CD2 atom than to any other atom in the model are used to calculate an average map value at each radial distance. (B) Plots of average map value vs. radial distance; these are the atomic map profiles. The dotted lines represent Gaussian functions which are fitted to each profile.

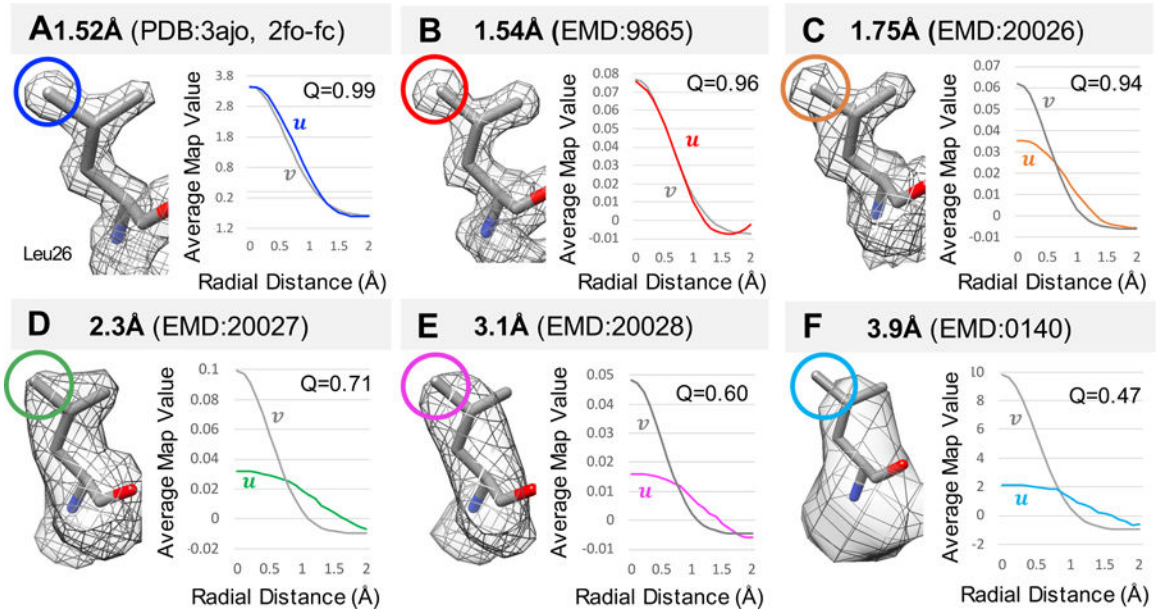

Figure 2.

Calculation of Q-scores for an atom in 6 maps at different resolutions, including an X-ray map (PDB:3ajo). The atom is CD2 from Leu 26 in the X-ray model PDB:3ajo fitted to each map. The atomic profile in each map is marked with the letter ***u***, while the reference Gaussian is marked with ***v***.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

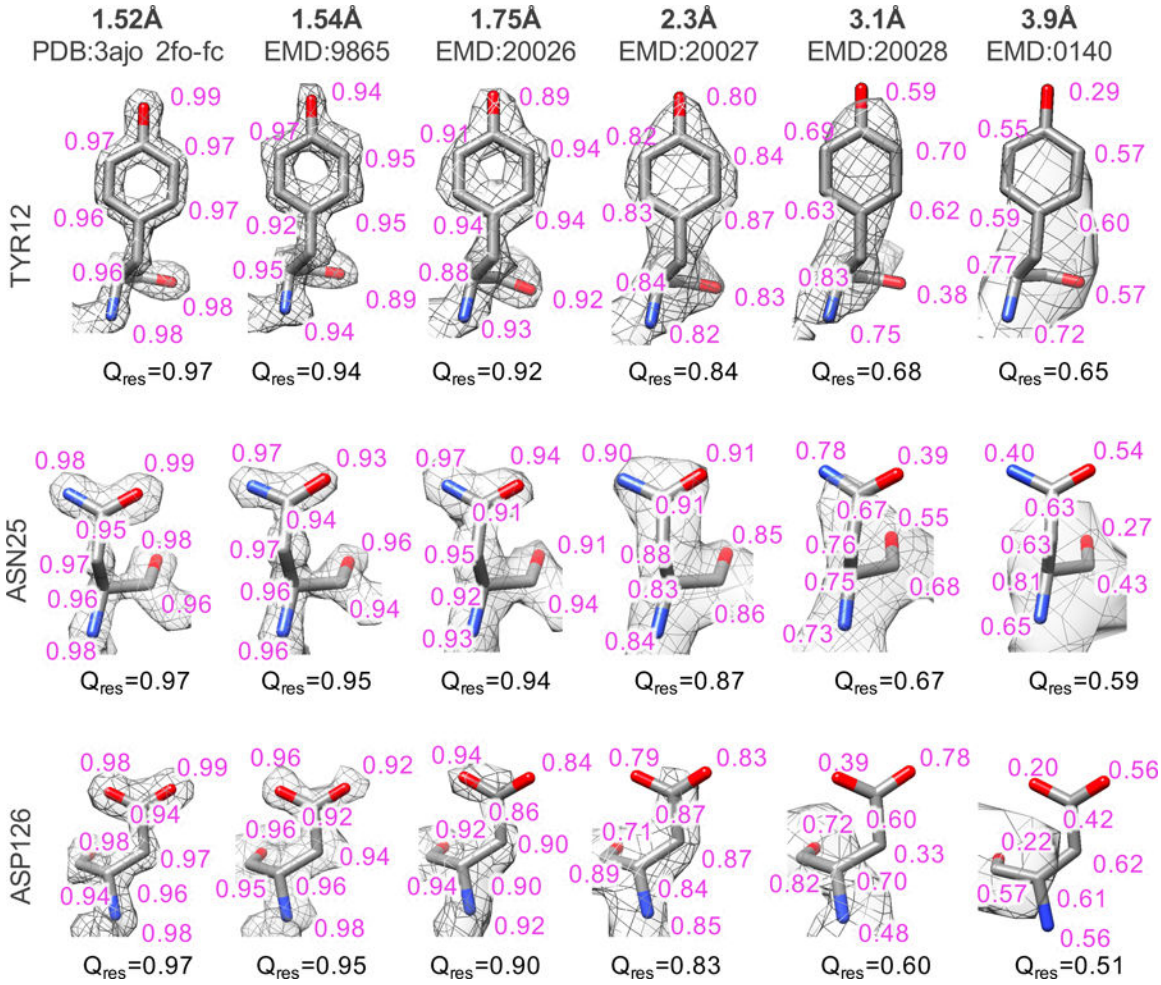

Figure 3.

Atom Q-scores for three residues taken from Apoferritin maps at various resolutions. Atom Q-scores are shown close to each atom, and the average Q-score is shown under each residue.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

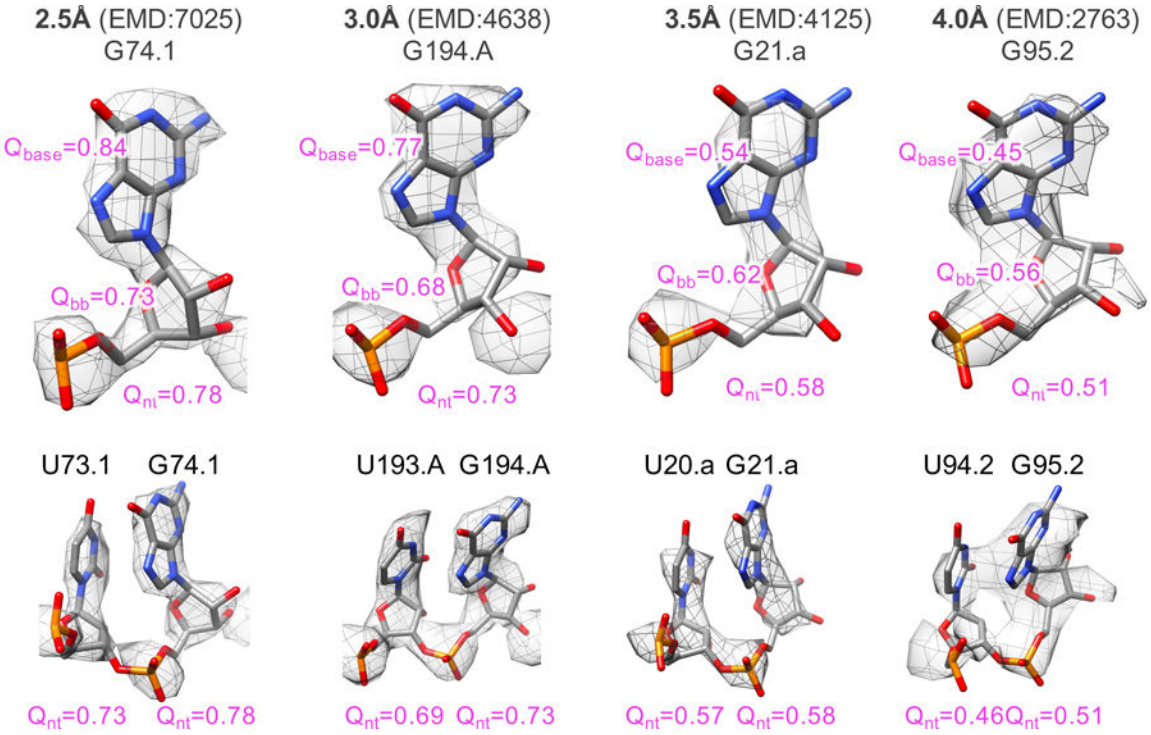

Figure 4.

Q-scores averaged over nucleotides (Q_{nt}) in cryoEM maps and models of ribosomes from the EMDB at four different resolutions. Q-scores are also averaged for base (Q_{base}) and phosphate-sugar backbone (Q_{bb}) atoms.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

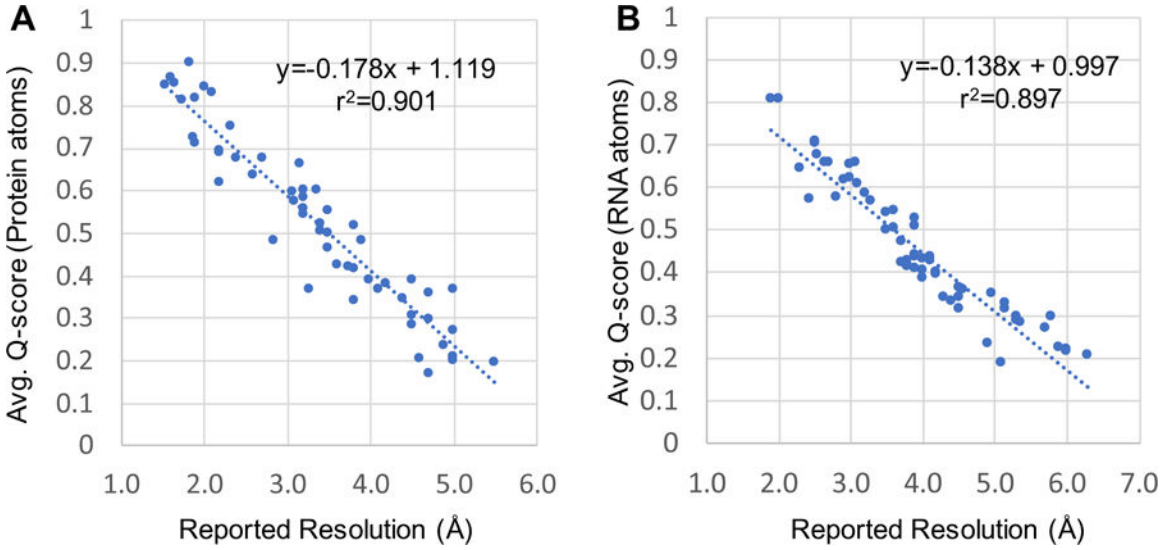

Figure 5.

Average Q-scores vs. reported resolution for maps and models obtained from EMDB. (A) Q-scores averaged over only protein atoms in maps and models listed in Table 1. (B) Q-scores averaged over only nucleic acid atoms in maps and models listed in Table 2. Linear functions fitted to the points are drawn with a dotted line in both plots; equations and r^2 value are inset.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

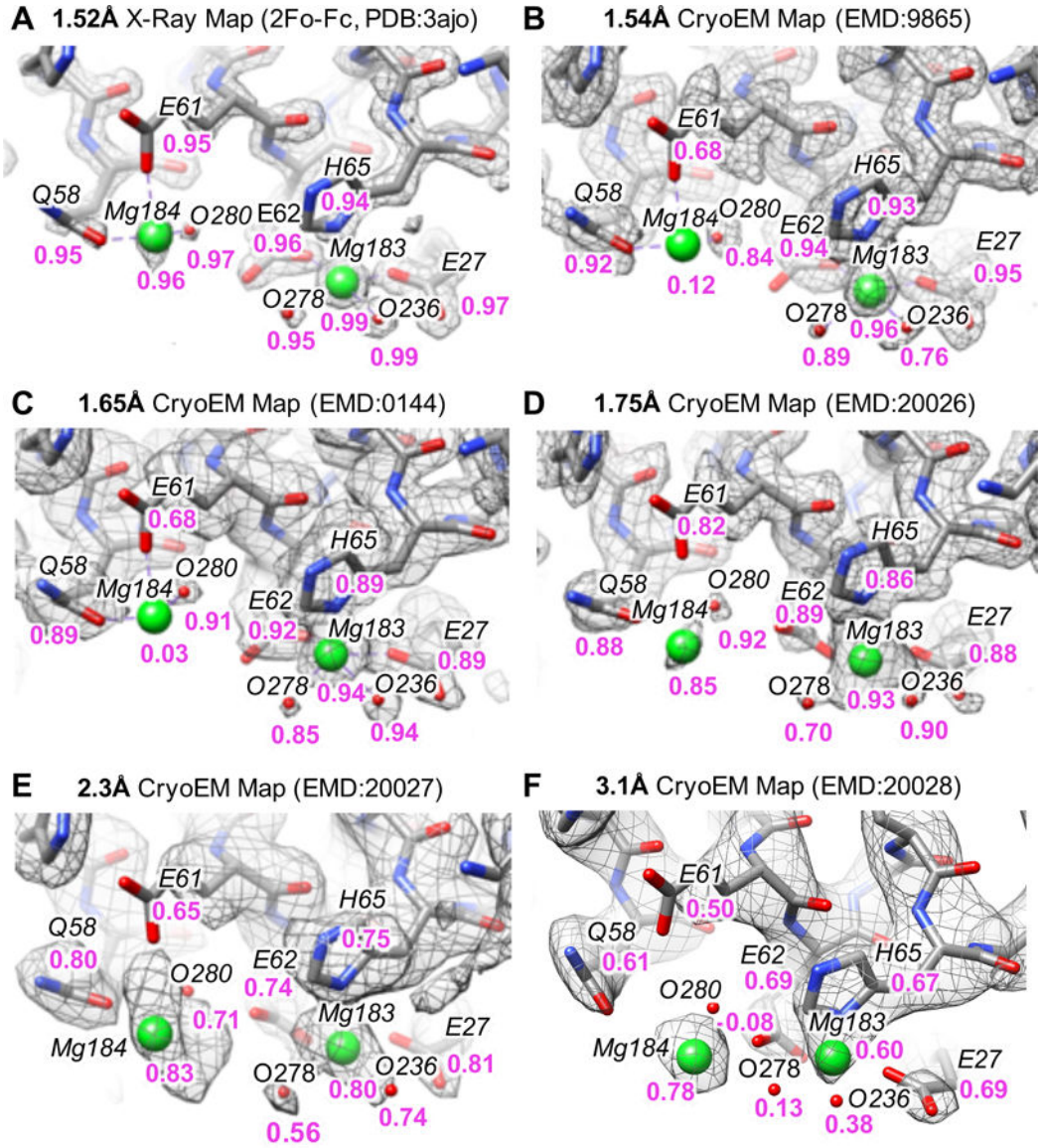

Figure 6.

A close up in Apoferritin models showing solvent atoms (Mg and O from water), along with calculated Q-scores in purple under each atom and nearby residue. The initial model comes from the X-ray map (PDB:3ajo) shown in A. It was further refined into each of the three cryoEM maps, B–F.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1.

Maps from EMDB for which Q-scores of protein components are calculated for the plot in Figure 5A. The entries marked with * were also in the original EMRinger analysis[16]. All others are maps of Apoferritin and β-galactosidase at resolutions up to 1.54Å.

	EMD ID	PDB	Resolution (Å)	Q-score	# Protein Atoms
1	9865	3ajo	1.54	0.85	1,473
2	9599	3wnw	1.62	0.87	1,433
3	144	3ajo	1.65	0.85	1,473
4	20026	3ajo	1.75	0.81	1,473
5	10101	6s61	1.84	0.90	2,799
6	0153	5a1a	1.89	0.72	32,828
7	9890	3ajo	1.9	0.82	1,473
8	7770	5a1a	1.9	0.71	32,828
9	9914	3wnw	2.01	0.84	1,433
10	4905	6rjh	2.1	0.83	1,364
11	4116	5a1a	2.2	0.69	1,364
12	4415	5a1a	2.2	0.69	32,828
13	8908	5a1a	2.2	0.69	32,828
14	2984	5a1a	2.2	0.62	32,828
15	20027	3ajo	2.32	0.75	1,473
16	4414	5a1a	2.4	0.68	32,828
17	6840	5a1a	2.6	0.64	32,828
18	4701	3wnw	2.7	0.67	1,433
19	20227	3ajo	2.85	0.48	1,473
20	20028	3ajo	3.08	0.60	1,473
21	5256*	3izx	3.1	0.57	32,209
22	3854	3ajo	3.15	0.66	1,473
23	5160*	3iyl	3.2	0.56	80,835
24	5623*	3j9i	3.2	0.60	46,228
25	5995*	3j7h	3.2	0.58	32,824
26	5995	5a1a	3.2	0.54	32,828
27	5778*	3j5p	3.27	0.37	18,424
28	2513*	4ci0	3.36	0.60	6,867
29	2762*	3j7y	3.4	0.52	60,863
30	2787*	4v19,4v1a	3.4	0.51	66,810
31	2278*	3j2v	3.5	0.47	4,629
32	5764*	3j4u	3.5	0.55	24,653
33	6035*	3j7w	3.5	0.50	17,829
34	5925*	3j6j	3.6	0.43	6,344
35	2764*	3j80	3.75	0.42	39,871

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

	EMD ID	PDB	Resolution (Å)	Q-score	# Protein Atoms
36	2773*	4uy8	3.8	0.34	26,960
37	5830*	3j63	3.8	0.42	10,590
38	6000*	3j7l	3.8	0.52	3,613
39	0140	3ajo	3.9	0.48	1,473
40	2763*	3j81	4	0.39	43,848
41	5600*	3j3i	4.1	0.37	7,515
42	2824	5a1a	4.2	0.38	32,828
43	2364*	4btg	4.4	0.34	11,840
44	2273*	3zif	4.5	0.30	94,377
45	2677*	4upc	4.5	0.28	3,127
46	5678*	3j40	4.5	0.39	24,066
47	5645*	3j3x	4.6	0.21	61,264
48	2788*	4v1w	4.7	0.36	32,736
49	5646*	3j3x	4.7	0.17	61,264
50	5895*	3j6e	4.7	0.29	60,318
51	5391*	3j1b	4.9	0.24	62,992
52	5886*	3jbd	5	0.37	7,560
53	5896*	3j6f	5	0.27	60,318
54	6187*	3j8x	5	0.21	9,235
55	6188*	3j8y	5	0.20	9,343

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2.

Maps from EMDB containing RNA for which Q-scores vs. resolution are plotted in Figure 5B.

	EMD ID	PDB File	Resolution (Å)	Q-score	# Nucleic Acid Atoms
1	10129	4udv	1.9	0.81	67
2	10130	4udv	2	0.80	67
3	10077	6s0z	2.3	0.64	97,227
4	10076	6s0x	2.43	0.57	64,722
5	7025	6az3-pdb-bundle1	2.5	0.70	34,068
6	7025	6az3-pdb-bundle2	2.5	0.70	39,212
7	8361	5t5h-pdb-bundle1	2.54	0.68	60,092
8	0243	6hma	2.65	0.66	63,217
9	7024	6az1	2.7	0.66	42,699
10	6583	3jcs-pdb-bundle1	2.8	0.57	72,130
11	20173	6ore-pdb-bundle1	2.9	0.62	97,294
12	4638	6qul	3	0.65	62,760
13	0600	6ole-pdb-bundle3	3	0.62	80,776
14	0233	6hiz-pdb-bundle1	3.08	0.66	31,798
15	4560	6qik-pdb-bundle1	3.1	0.61	3,030
16	10068	6rzz-pdb-bundle1	3.2	0.58	67,292
17	0101	6gzq-pdb-bundle1	3.28	0.56	67,292
18	4125	5lze-pdb-bundle1	3.5	0.50	65,324
19	4125	5lze-pdb-bundle2	3.5	0.54	64,391
20	2938	4ug0-pdb-bundle1	3.6	0.54	37,311
21	2938	4ug0-pdb-bundle2	3.6	0.50	38,504
22	6559	3jcj-pdb-bundle1	3.7	0.47	34,577
23	6559	3jcj-pdb-bundle2	3.7	0.42	63,932
24	8620	5uyq-pdb-bundle1	3.8	0.42	33,012
25	8620	5uyq-pdb-bundle2	3.8	0.43	70,155
26	0076	6gwt-pdb-bundle1	3.8	0.42	34,656
27	0076	6gwt-pdb-bundle2	3.8	0.41	36,969
28	0192	6hcf-pdb-bundle1	3.9	0.52	64,900
29	0192	6hcf-pdb-bundle2	3.9	0.51	83,585
30	0192	6hcf-pdb-bundle3	3.9	0.41	2,109
31	8279	5kps-pdb-bundle1	3.9	0.43	33,016
32	8279	5kps-pdb-bundle2	3.9	0.44	68,569
33	8618	5uyn-pdb-bundle1	4	0.38	33,012
34	8618	5uyn-pdb-bundle2	4	0.39	70,133
35	4080	5lmu	4	0.43	34,527
36	2763	3j81	4	0.40	39,828
37	4350	6g51	4.1	0.43	19,905

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

	EMD ID	PDB File	Resolution (Å)	Q-score	# Nucleic Acid Atoms
38	8280	5kpv-pdb-bundle1	4.1	0.44	33,016
39	8280	5kpv-pdb-bundle2	4.1	0.43	70,236
40	0643	6o7k	4.2	0.40	34,777
41	20188	6ost-pdb-bundle1	4.2	0.40	97,110
42	4382	6gc7	4.3	0.34	40,850
43	0083	6gxp-pdb-bundle1	4.4	0.33	64,749
44	4349	6g4w	4.5	0.31	18,753
45	3133	5ady	4.5	0.36	12,104
46	4351	6g53	4.5	0.34	19,905
47	0104	6gzx-pdb-bundle1	4.57	0.36	65,324
48	4083	5lmv	4.9	0.23	34,527
49	3553	5mrf-pdb-bundle1	4.97	0.35	57,598
50	8473	5tzs	5.1	0.18	13,410
51	3661	5no2	5.16	0.33	32,930
52	3662	5no3	5.16	0.31	32,930
53	4122	5lzb-pdb-bundle1	5.3	0.28	37,309
54	4427	6i7o-pdb-bundle1	5.3	0.29	72,803
55	4075	5lmp	5.35	0.28	32,964