Public Data-based Report

Validation Report Service

Cryo-EM Map Validation Report

Report to assess Cryo-EM Volume Map at Level(s) 1, 0

This report has been generated based on data publicly available at EMDB.

Basic Entry Information:

EMDB ID: EMD-57017
Title: focused refined map; map for 40S body from human 48S PIC on Kozak mRNA (Kozak_1)
Authors: See EMDB entry link
Deposited on: 2026-03-03T00:00:00
Reported Resolution: 2.78 Å

Contact Us:

Instruct Image Processing Center (I^2PC)
Biocomputing Unit (BCU)
i2pc@cnb.csic.es
VRS Website

National Center for Biotechnology (CNB)
St/ Darwin, 3 (Autonomous University of Madrid)
28049 Cantoblanco, Madrid (Spain)

Last update: **September 5, 2026, 7:28am**

Context

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 and keeping the quality of the submitted maps is essential to maintain the scientific quality of the field. The ultimate quality measure is the consistency of the map and an atomic model. However, this is only possible for high resolution maps. Over the years there have been many suggestions about validation measures of 3DEM maps. Unfortunately, most of these measures are not currently in use for their spread in multiple software tools and the associated difficulty to access them. To alleviate this problem, we made available a validation grading system that evaluate the information provided to assess the map.
This system grades a map from 0 to 5 depending on the amount of information available. 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). In addition, we can have three optional qualifiers: A (... + atomic model), W (... + image processing workflow), and O (... + other techniques). To know more about this service read this paper

This Validation Report Service uses Scipion (see this link for more detail) as workflow engine and ChimeraX (see this link for more detail) to generate the 3D views. For more information about the different methods and softwares used for this report, see the references here.

SCIPION

Summarized overall quality

The map seems to have some problem in its centering or extra space (see Sec. 2.1). There is no problem with the suggested threshold. There seems to be a problem with the map's background (see Sec. 2.3).

The overall score (passing tests) of this report is 2 out of 4 evaluable items.

0.a Mass analysis	Sec. 2.1	1 warnings
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

Summary of the warnings across sections.

Section 2.1 (0.a Mass analysis)

1. **The center of mass in X may be significantly shifted. This is common when the refinement is applied exclusively to one protein region.**

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=1037.73**)

Contents

1 Input data **6**

2 Level 0 analysis **13**

2.1 Level 0.a Mass analysis 13

2.2 Level 0.b Mask analysis 14

2.3 Level 0.c Background analysis 18

2.4 Level 0.d B-factor analysis 20

3 Half maps **23**

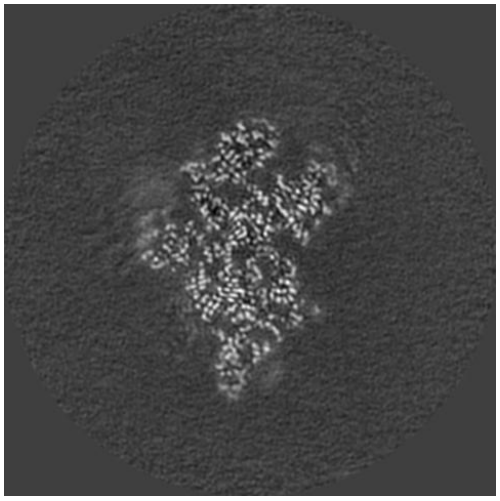
1 Input data

Input map: emd_57017.map
SHA256 hash: c0bb4db806890ad6d977830f86a73629199cec474e6eabc5154266cbd2593b51
Voxel size: 0.720000 (Å)
Visualization threshold: 0.004900
Resolution estimated by user: 2.78

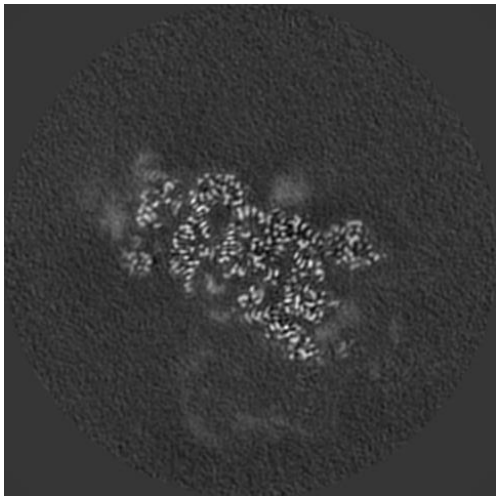
Orthogonal slices of the input map

Explanation:
In the orthogonal slices of the map, the noise outside the protein should not have any structure (stripes going out, small blobs, particularly high or low densities, ...)

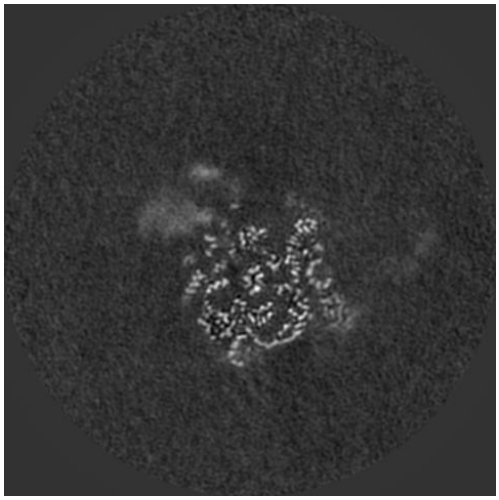
Results:
See Fig. 1.



(a) X Slice 240



(b) Y Slice 240



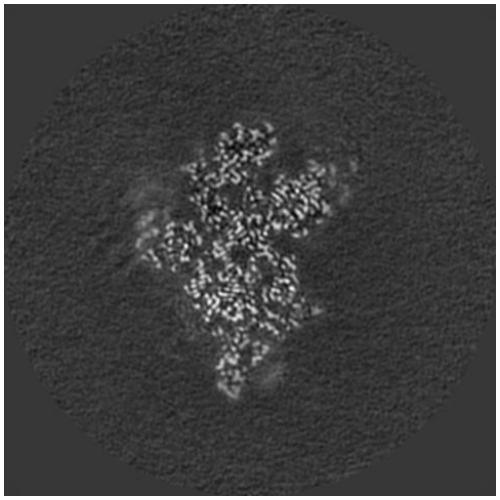
(c) Z Slice 240

Figure 1: Central slices of the input map in the three dimensions

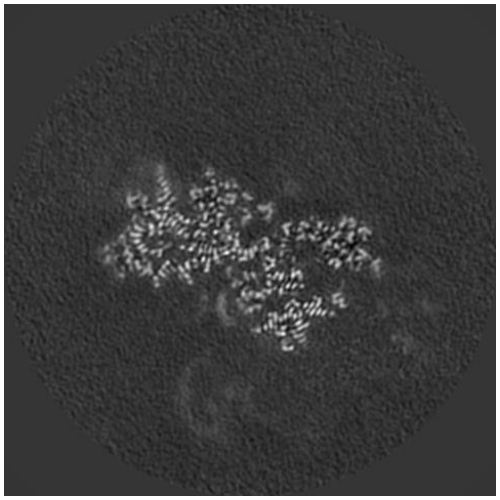
Orthogonal slices of maximum variance of the input map

Results:

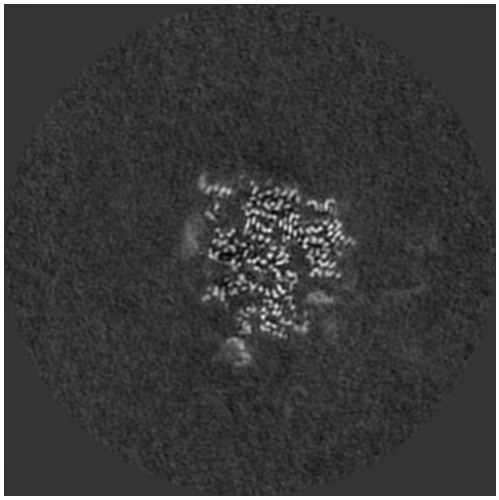
See Fig. 2.



(a) X Slice 238



(b) Y Slice 219



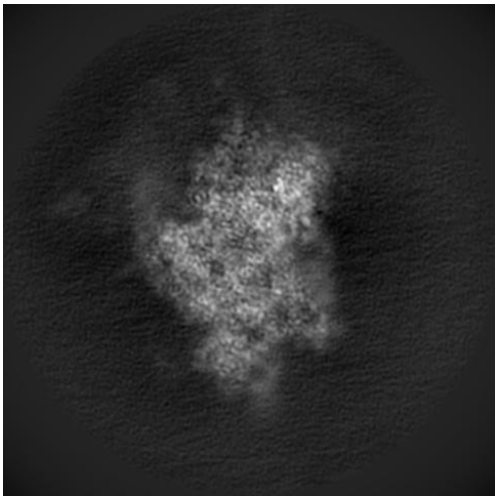
(c) Z Slice 264

Figure 2: Slices of maximum variation in the three dimensions

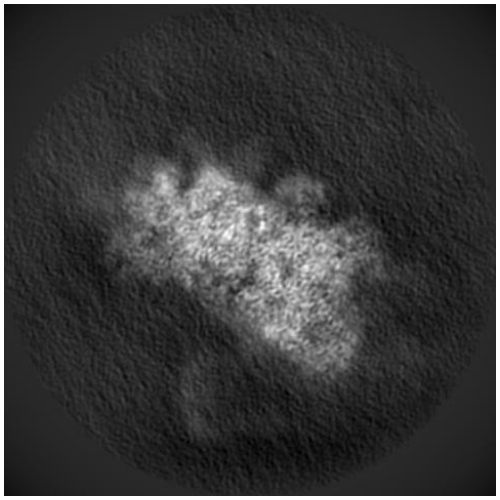
Orthogonal projections of the input map

Explanation:
In the projections there should not be stripes (this is an indication of directional overweighting, or angular attraction), and there should not be a dark halo around or inside the structure (this is an indication of incorrect CTF correction or the reconstruction of a biased map).

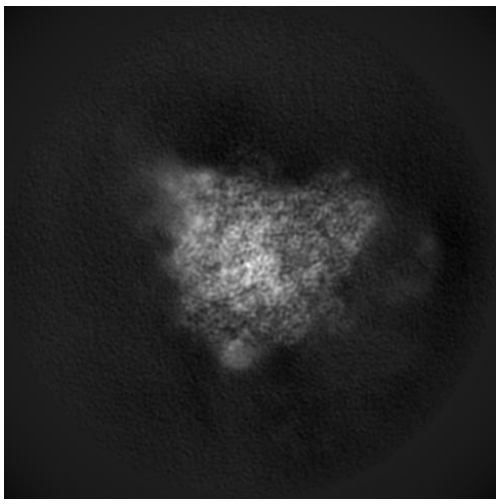
Results:
See Fig. 3.



(a) X Projection



(b) Y Projection



(c) Z Projection

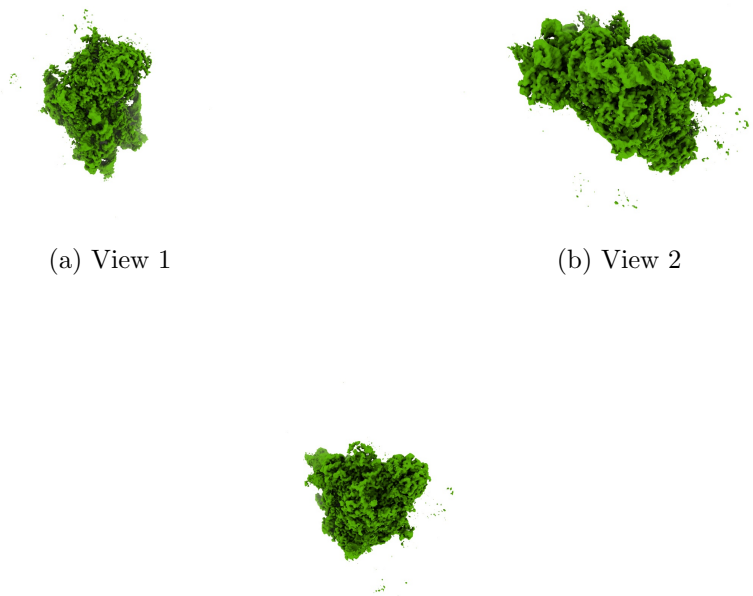
Figure 3: Projections in the three dimensions

Isosurface views of the input map

Explanation:

An isosurface is the surface of all points that have the same gray value. In these views there should not be many artifacts or noise blobs around the map.

Results:
See Fig. 4.



(a) View 1

(b) View 2

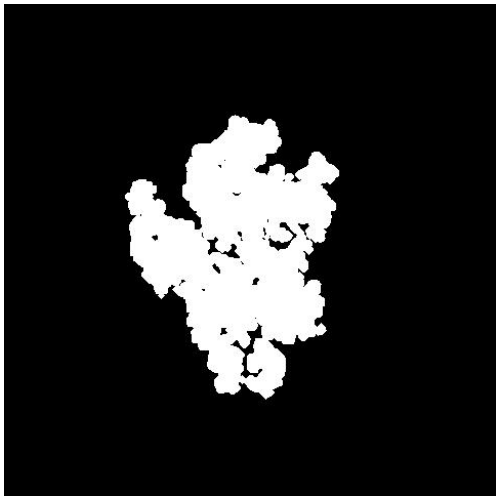
(c) View 3

Figure 4: Isosurface at threshold=0.004900. Views generated by ChimeraX at a the following X, Y, Z angles: View 1 (0, -90, -90), View 2 (-90, 0, -90), View 3 (0, 0, 0).

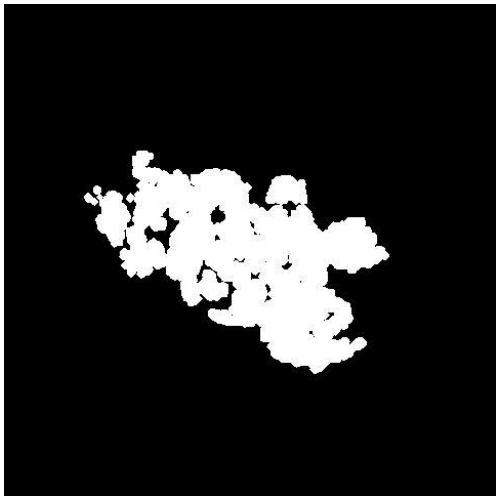
Orthogonal slices of maximum variance of the mask with hard borders

Explanation:
The mask with hard borders has been calculated at the suggested threshold 0.004900, the largest connected component was selected, and then dilated by 2Å.

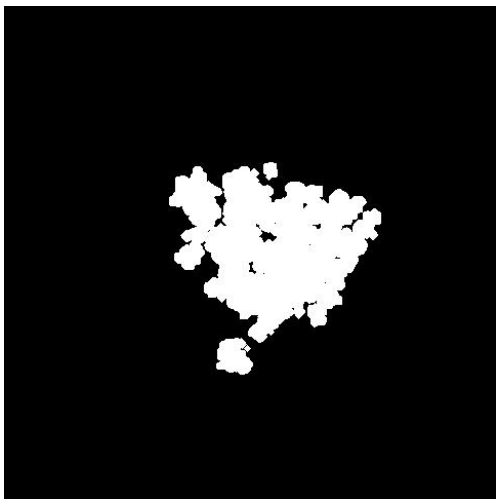
Results:
See Fig. 5.



(a) X Slice 229



(b) Y Slice 243



(c) Z Slice 287

Figure 5: Slices of maximum variation in the three dimensions of the mask with hard borders

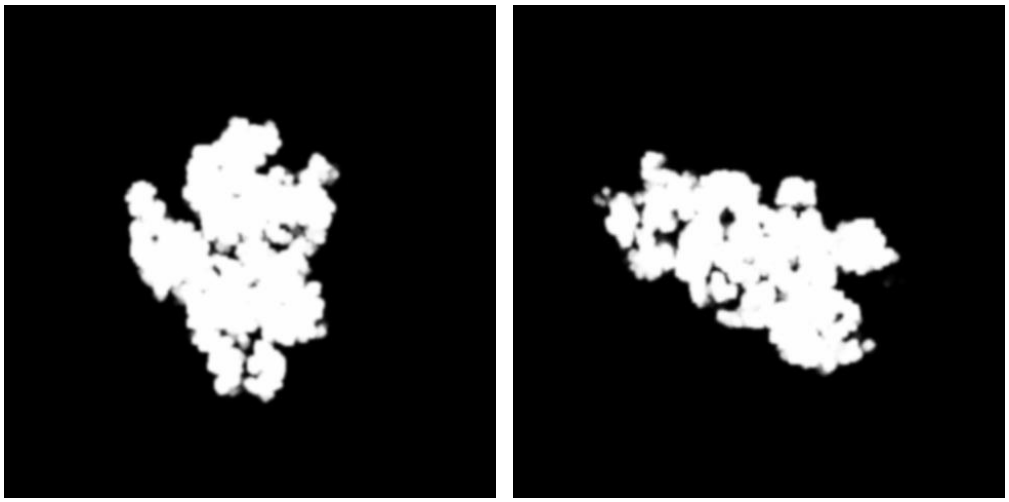
Orthogonal slices of maximum variance of the mask with soft borders

Explanation:

The mask with soft borders has been calculated at the suggested threshold 0.004900, the largest connected component was selected, and then dilated by

2Å.

Results:
See Fig. 6.



(a) X Slice 228 (b) Y Slice 242

(c) Z Slice 287

Figure 6: Slices of maximum variation in the three dimensions of the mask with soft borders

2 Level 0 analysis

2.1 Level 0.a Mass analysis

Explanation:
The reconstructed map must be relatively well centered in the box, and there should be at least 30Å (the exact size depends on the CTF) on each side to make sure that the CTF can be appropriately corrected.

Results:
The space from the left and right in X are 89.28 and 75.60 Å, respectively. There is a decentering ratio (abs(Right-Left)/Size)% of 3.96%

The space from the left and right in Y are 79.20 and 98.64 Å, respectively. There is a decentering ratio (abs(Right-Left)/Size)% of 5.62%

The space from the left and right in Z are 53.28 and 64.80 Å, respectively. There is a decentering ratio (abs(Right-Left)/Size)% of 3.33%

The center of mass is at (x,y,z)=(114.41,300.32,177.42). The decentering of the center of mass (abs(Center)/Size)% is 26.16, 12.57, and 13.04, respectively.

Automatic criteria: The validation is OK if 1) the decentering and center of mass less than 20% of the map dimensions in all directions, and 2) the extra space on each direction is more than 20% of the map dimensions. For local reconstruction, focused refinement, or similar, warnings are expected.

WARNINGS: 1 warnings

1. **The center of mass in X may be significantly shifted. This is common when the refinement is applied exclusively to one protein region.**

2.2 Level 0.b Mask analysis

Explanation:
The map at the suggested threshold should have most of its mass concentrated in a single connected component. It is normal that after thresholding there are a few thousands of very small, disconnected noise blobs. However, there total mass should not exceed 10%. The raw mask (just thresholding) and the mask constructed for the analysis (thresholding + largest connected component + dilation) should significantly overlap. Overlap is defined by the overlapping coefficient (size(Raw AND Constructed)/size(Raw)) that is a number between 0 and 1, the closer to 1, the more they agree.

Results:

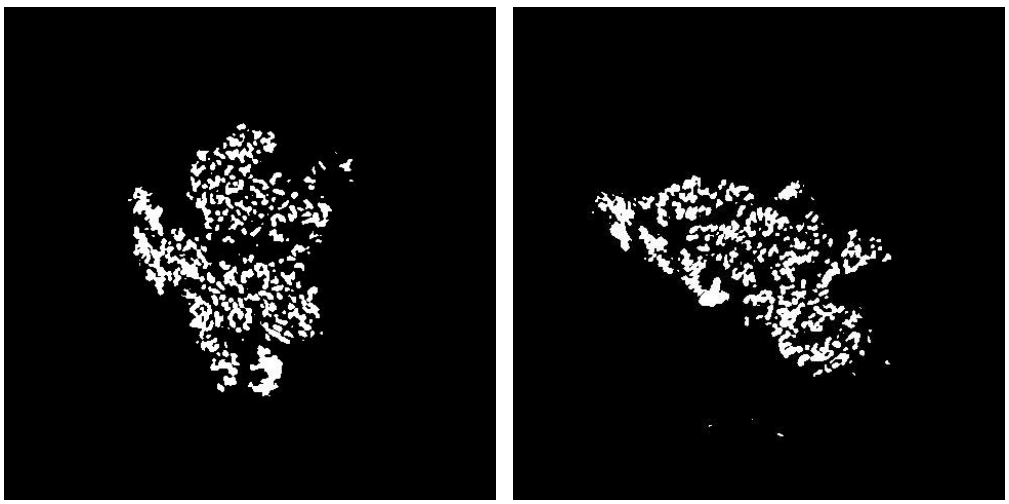
Raw mask: At threshold 0.004900, there are 1158 connected components with a total number of voxels of 1286832 and a volume of 480307.47 $Å^3$ (see Fig. 7). The size and percentage of the total number of voxels for the raw mask are listed below (up to 95% of the mass or the first 100 clusters, whatever happens first), the list contains (No. voxels (volume in $Å^3$), percentage, cumulated percentage):

(1270440 (474189.19), 98.73, 98.73)

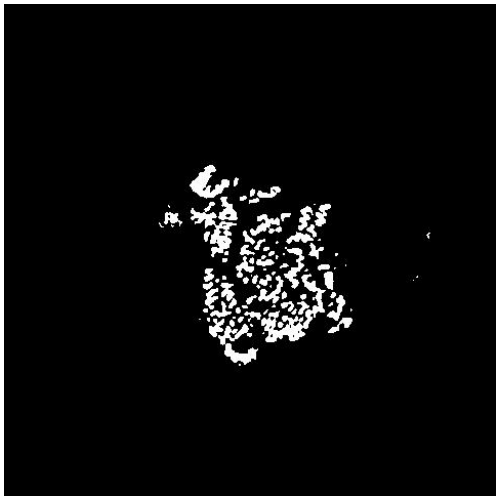
Number of components to reach 95% of the mass: 1

The average size of the remaining 1157 components is 14.17 voxels (0.37 Å^3). Their size go from 1270440 voxels (474189.19 Å^3) to 1 voxels (0.37 Å^3).

The slices of the raw mask can be seen in Fig. 7.



(a) X Slice 227 (b) Y Slice 250



(c) Z Slice 251

Figure 7: Maximum variance slices in the three dimensions of the raw mask

The following table shows the variation of the mass enclosed at different thresholds (see Fig. 8):

Threshold	Voxel mass	Molecular mass(kDa)	# Aminoacids
0.0014	7251675.00	2242.48	20386.18
0.0027	2375038.00	734.45	6676.80
0.0041	1530676.00	473.34	4303.09
0.0055	1155090.00	357.20	3247.23
0.0068	903367.00	279.35	2539.58
0.0082	715744.00	221.33	2012.13
0.0096	569776.00	176.20	1601.78
0.0109	453292.00	140.17	1274.31
0.0123	357496.00	110.55	1005.01
0.0136	277610.00	85.85	780.43
0.0150	210595.00	65.12	592.03
0.0164	155197.00	47.99	436.30
0.0177	110634.00	34.21	311.02
0.0191	76002.00	23.50	213.66
0.0205	50298.00	15.55	141.40
0.0218	32174.00	9.95	90.45
0.0232	19333.00	5.98	54.35
0.0246	10925.00	3.38	30.71
0.0259	5598.00	1.73	15.74
0.0273	2552.00	0.79	7.17
0.0287	998.00	0.31	2.81
0.0300	311.00	0.10	0.87
0.0314	72.00	0.02	0.20
0.0327	8.00	0.00	0.02
0.0341	0.00	0.00	0.00

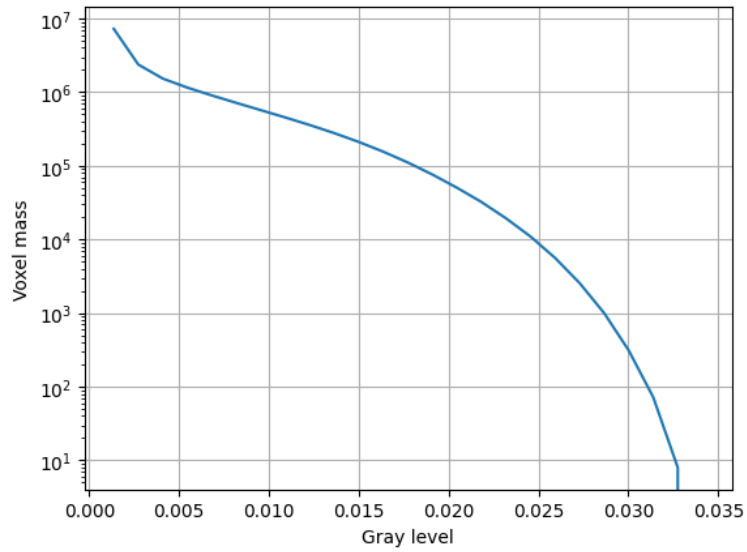


Figure 8: Voxel mass as a function of the gray level.

Constructed mask: After keeping the largest component of the previous mask and dilating it by 2Å, there is a total number of voxels of 3751290 and a volume of 1400161.49 Å^3. The overlap between the raw and constructed mask is 0.99.

Automatic criteria: The validation is OK if 1) to keep 95% of the mass we need to keep at most 5 connected components; and 2) the average volume of the blobs outside the given threshold has a size smaller than 5Å^3; and 3) the overlap between the raw mask and the mask constructed for the analysis is larger than 75%.

STATUS: OK

2.3 Level 0.c Background analysis

Explanation:
Background is defined as the region outside the macromolecule mask. The background mean should be zero, and the number of voxels with a very low or very high value (below 5 standard deviations of the noise) should be very

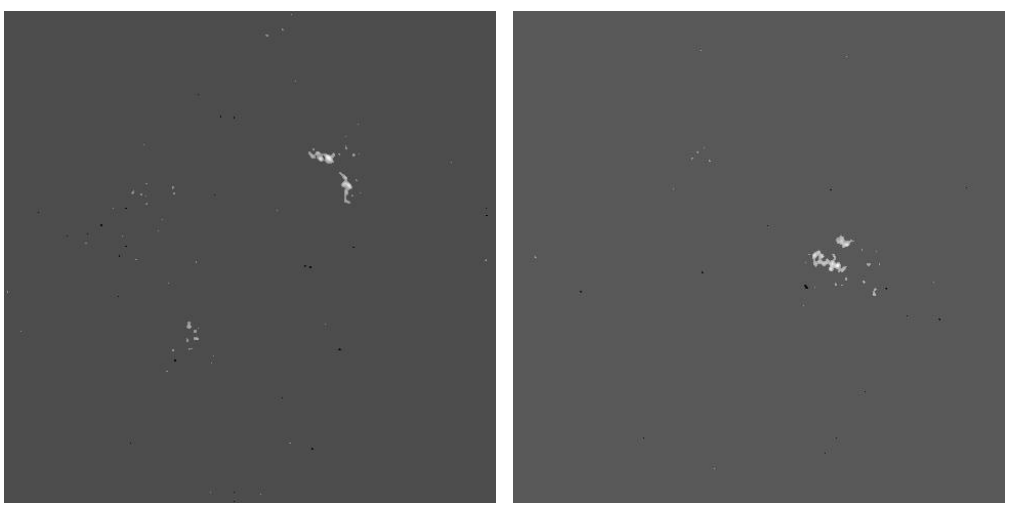
small and they should be randomly distributed without any specific structure. Sometimes, you can see some structure due to the symmetry of the structure.

Results:

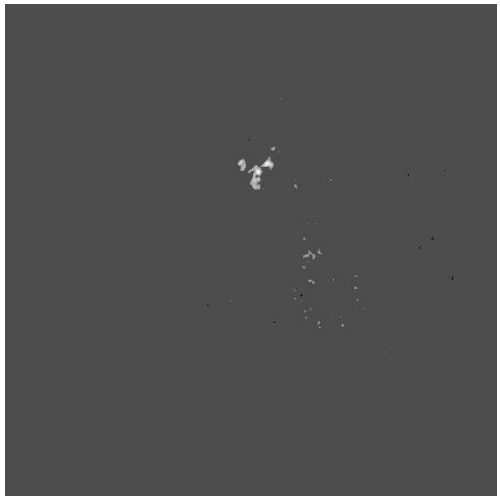
The null hypothesis that the background mean is 0 was tested with a one-sample Student's t-test. The resulting t-statistic and p-value were -1676.69 and 0.000000, respectively.

The mean and standard deviation (sigma) of the background were -0.000128 and 0.000787. The percentage of background voxels whose absolute value is larger than 5 times the standard deviation is 0.06 % (see Fig. 9). The same percentage from a Gaussian would be 0.000057% (ratio between the two percentages: 1037.727433).

Slices of the background beyond 5*sigma can be seen in Fig. 9.



(a) X Slice 246 (b) Y Slice 333



(c) Z Slice 337

Figure 9: Maximum variance slices in the three dimensions of the parts of the background beyond 5*sigma

Automatic criteria: The validation is OK if 1) the p-value of the null hypothesis that the background has 0 mean is larger than 0.001; and 2) the number of voxels above or below 5 sigma is smaller than 20 times the amount expected for a Gaussian with the same standard deviation whose mean is 0.

WARNINGS: 2 warnings

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=1037.73)**

2.4 Level 0.d B-factor analysis

Explanation:
The B-factor line (see this link for more details) fitted between 15Åand the resolution reported should have a slope that is between 0 and 300 $Å^2$.
Results:

Fig. 10 shows the logarithm (in natural units) of the structure factor (the module squared of the Fourier transform) of the experimental map, its fitted line, and the corrected map. The estimated B-factor was -72.3. The fitted line was $\log(|F|^2) = -18.1/R^2 + (-13.8)$.

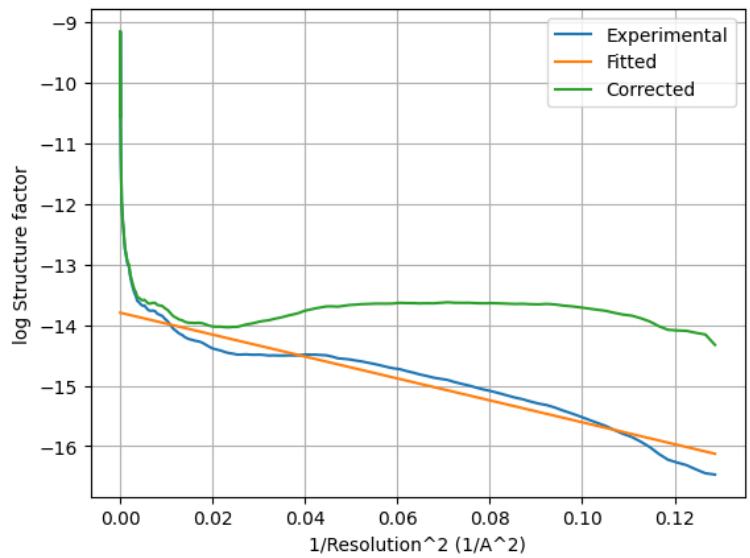
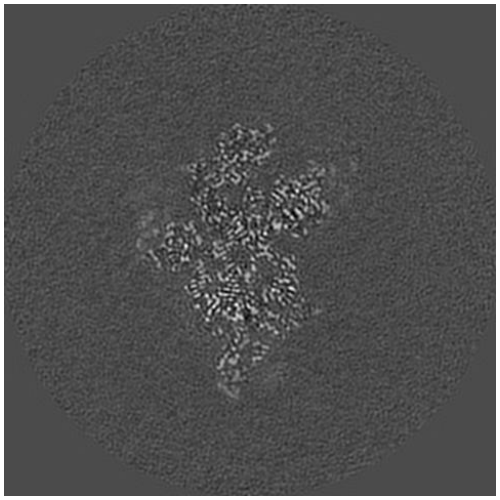
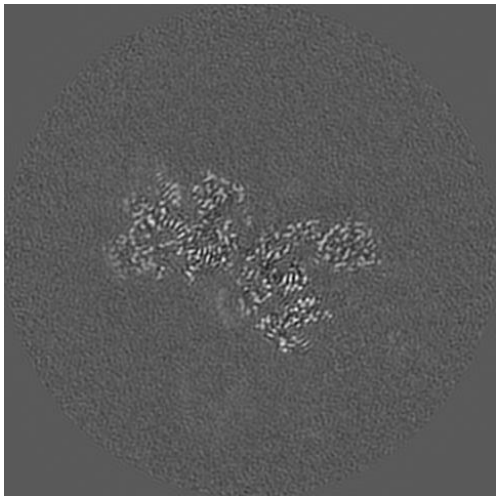


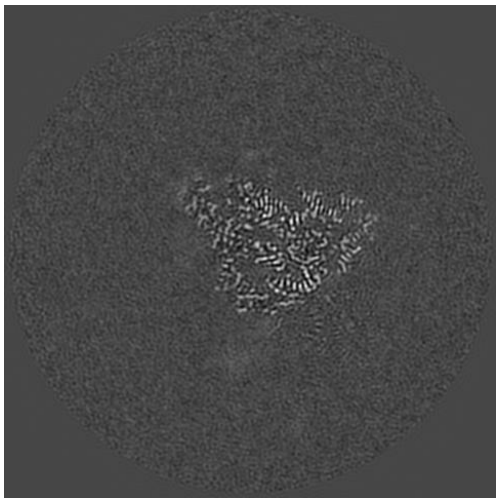
Figure 10: Guinier plot. The X-axis is the square of the inverse of the resolution in Å.



(a) X Slice 238



(b) Y Slice 223



(c) Z Slice 293

Figure 11: Slices of maximum variation in the three dimensions of the B-factor corrected map

Automatic criteria: The validation is OK if the B-factor is in the range [-300,0].

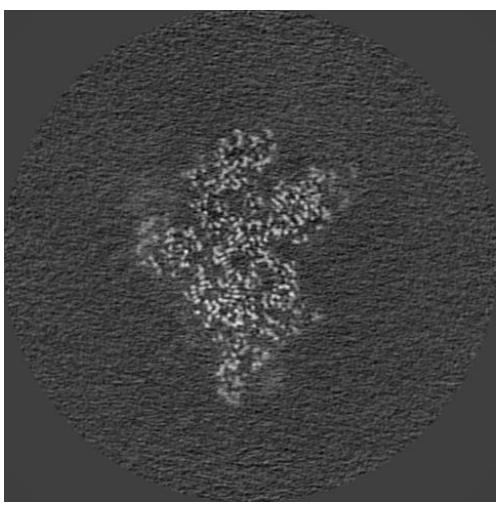
STATUS: OK

3 Half maps

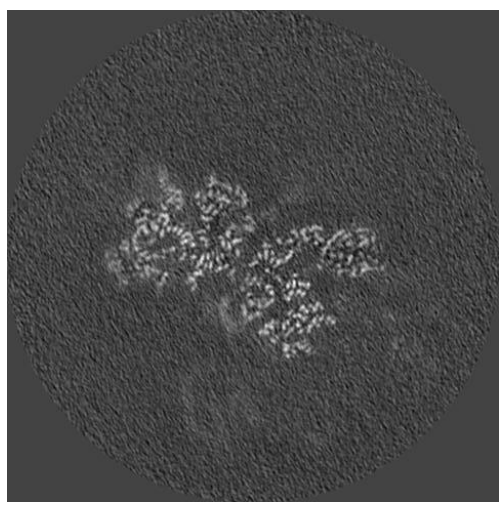
Half map 1: emd_57017_half_map_1.map
SHA256 hash: a1861228aea573f9d0eeb313f7fb5628788b7043a17e3619af6f54420935ed69

Half map 2: emd_57017_half_map_2.map
SHA256 hash: 3a2cd49eba0dbe36d83903c9cbdee66e8d42d8064f3bcaaa0c6647565c76d01e

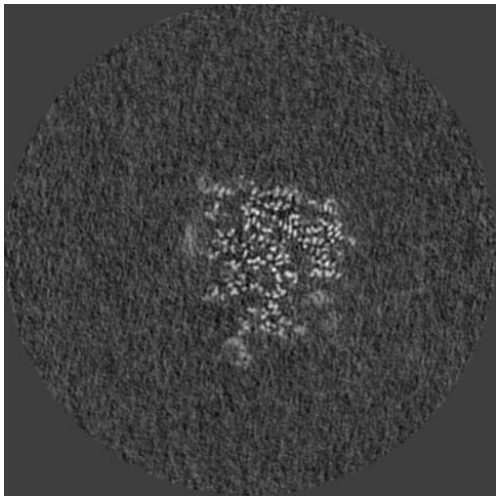
Slices of the first half map can be seen in Fig. 12.
Slices of the second half map can be seen in Fig. 13.
Slices of the difference between both maps can be seen in Fig. 14. There should not be any structure in this difference. Sometimes some patterns are seen if the map is symmetric.



(a) X Slice 238

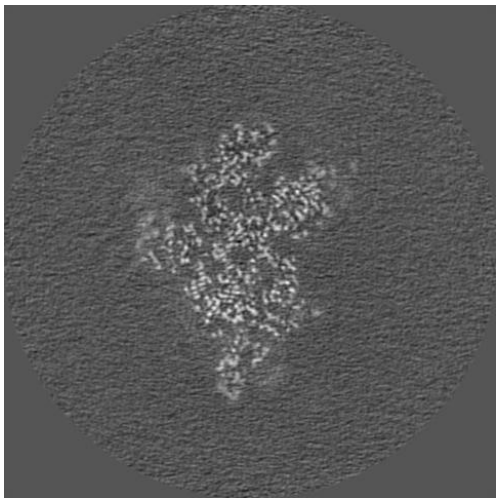


(b) Y Slice 222

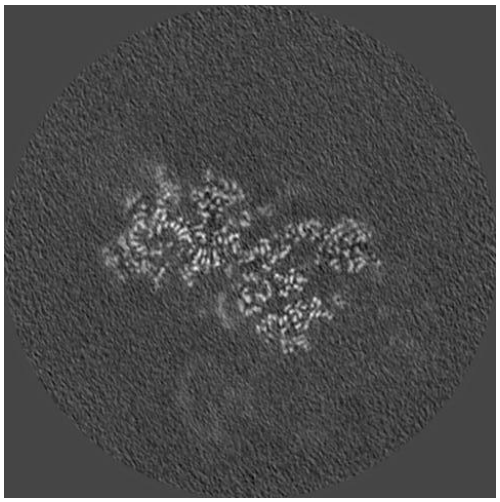


(c) Z Slice 265

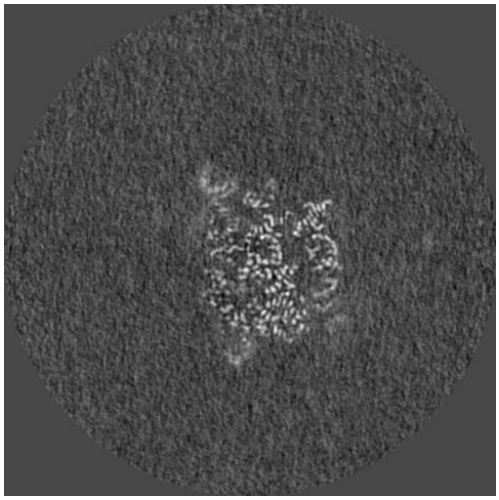
Figure 12: Slices of maximum variation in the three dimensions of Half 1



(a) X Slice 238

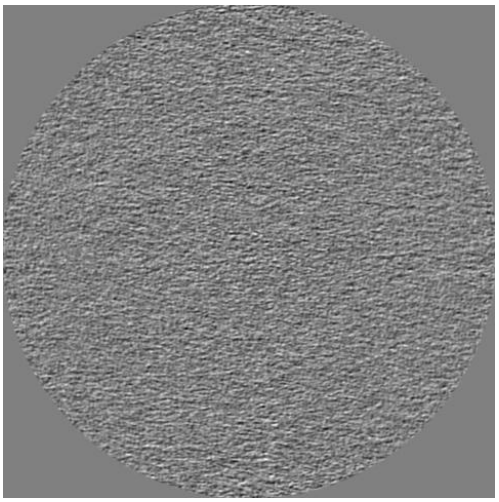


(b) Y Slice 221

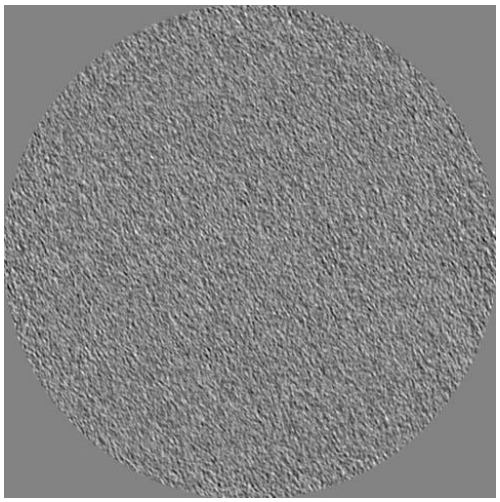


(c) Z Slice 257

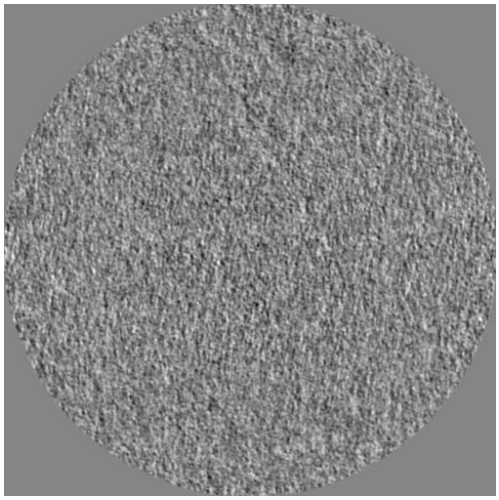
Figure 13: Slices of maximum variation in the three dimensions of Half 2



(a) X Slice 236



(b) Y Slice 234



(c) Z Slice 235

Figure 14: Slices of maximum variation in the three dimensions of the difference Half1-Half2.