Public Data-based Report

Validation Report Service

Cryo-EM Map Validation Report

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

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

Basic Entry Information:

EMDB ID: EMD-1475
Title: Structure of bacteriophage N4
Authors: See EMDB entry link
Deposited on: 2008-02-22T00:00:00
Reported Resolution: 29.0 Å

Contact Us:

Instruct Image Processing Center (I^{2}PC)
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 17, 2024, 12:35 Noon**

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).

This Validation Report Service is explained in more detail in this paper. For more information about the different methods and softwares used for this report, see the references here.

Summarized overall quality

The map seems to have some problem in its centering or extra space (see Sec. 2.1). There seems to be a problem with the suggested threshold (see Sec. 2.2). There seems to be a problem with the map's background (see Sec. 2.3). There seems to be a problem with its local B-factor (see Sec. 2.6).

The overall score (passing tests) of this report is 2 out of 6 evaluable items.

0.a Mass analysis	Sec. 2.1	3 warnings
0.b Mask analysis	Sec. 2.2	1 warnings
0.c Background analysis	Sec. 2.3	2 warnings
0.d B-factor analysis	Sec. 2.4	OK
0.e DeepRes	Sec. 2.5	Does not apply
0.f LocBfactor	Sec. 2.6	1 warnings
0.g LocOccupancy	Sec. 2.7	OK
0.h Deep hand	Sec. 2.8	Does not apply

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.**
2. **The center of mass in Y may be significantly shifted. This is common when the refinement is applied exclusively to one protein region.**
3. **The center of mass in Z may be significantly shifted. This is common when the refinement is applied exclusively to one protein region.**

Section 2.2 (0.b Mask analysis)

1. **There might be a problem with noise and artifacts, because the average noise blob has a volume of 44.361864 Å^3.**

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=13168.82)**

Section 2.6 (0.f LocBfactor)

1. **The median B-factor is out of the interval [-300,0]**

Contents

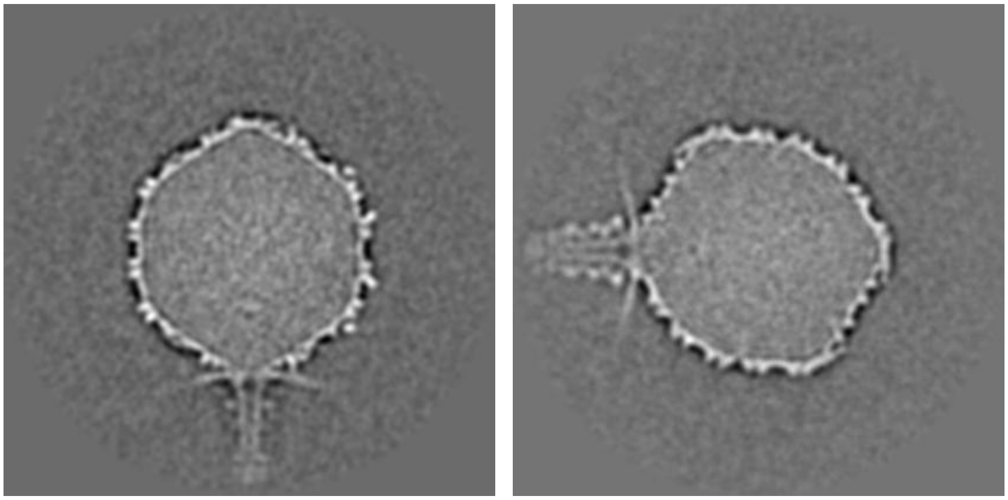
1 Input data

Input map: emd_1475.map
SHA256 hash: a6bece8ed4b5c5913f4298d5619645a605bdea318661d48f81ad67dc0c5edd41
Voxel size: 3.540000 (Å)
Visualization threshold: 2.300000
Resolution estimated by user: 29.0

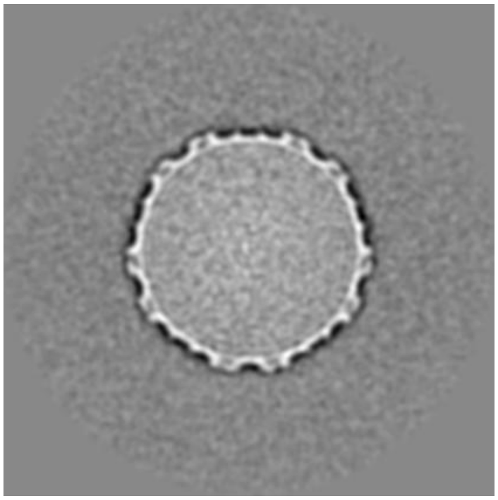
Orthogonal slices of the input map

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 210 (b) Y Slice 210

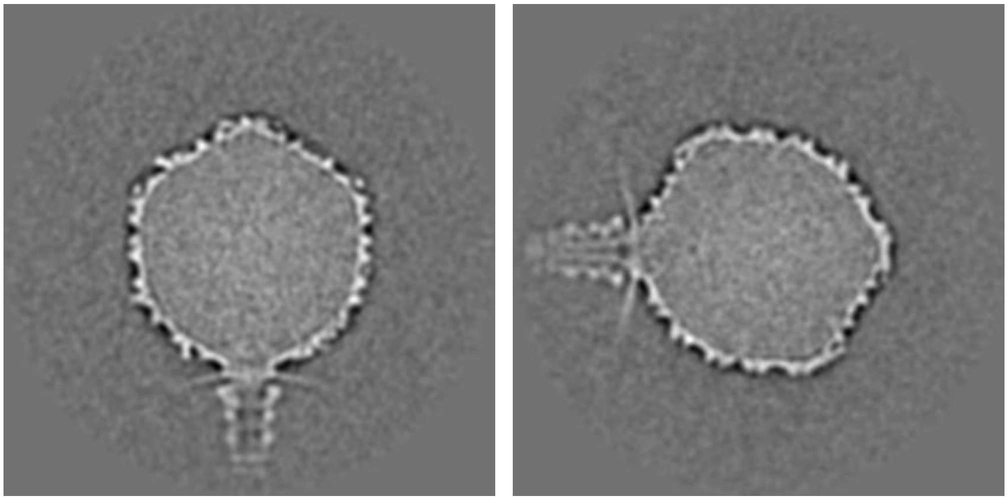


(c) Z Slice 210

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

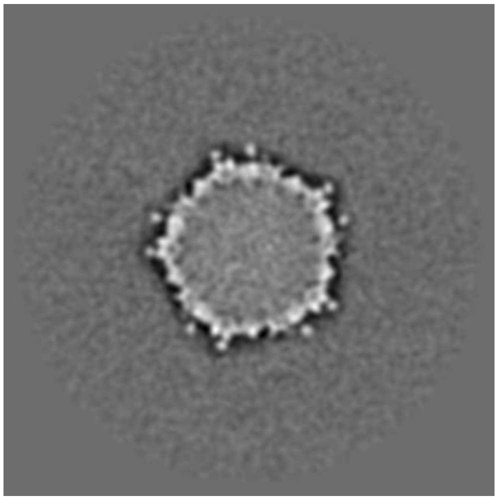
Orthogonal slices of maximum variance of the input map

Results:
See Fig. 2.



(a) X Slice 220

(b) Y Slice 209



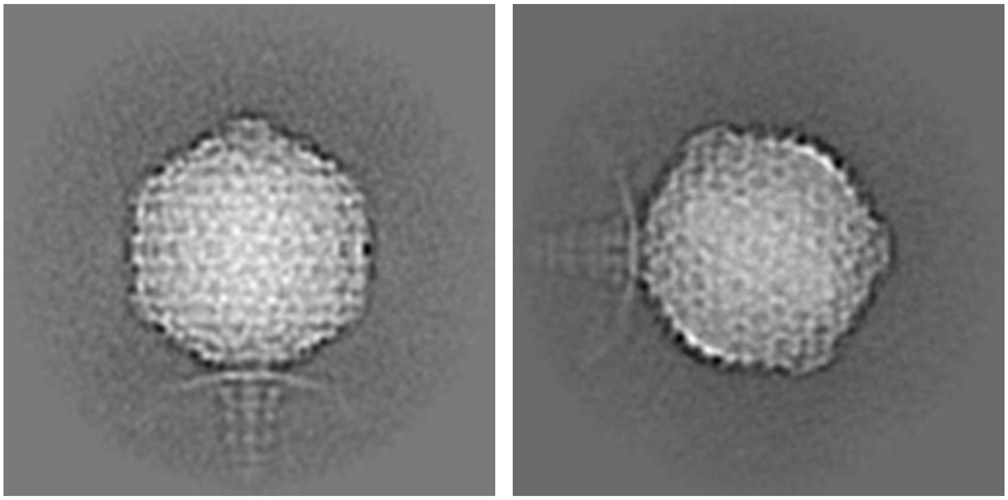
(c) Z Slice 141

Figure 2: Slices of maximum variation in the three dimensions

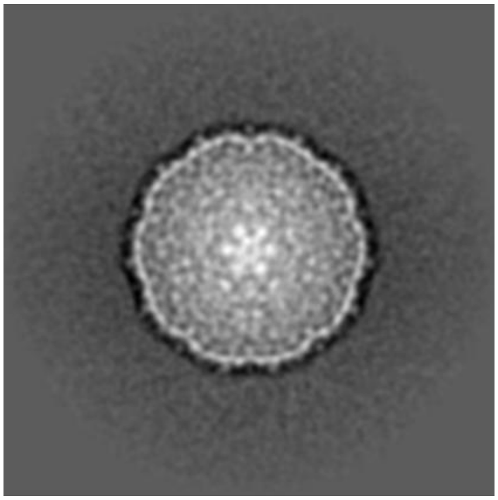
Orthogonal projections of the input map

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

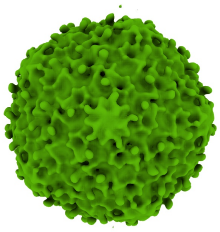
Isosurface views of the input map

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



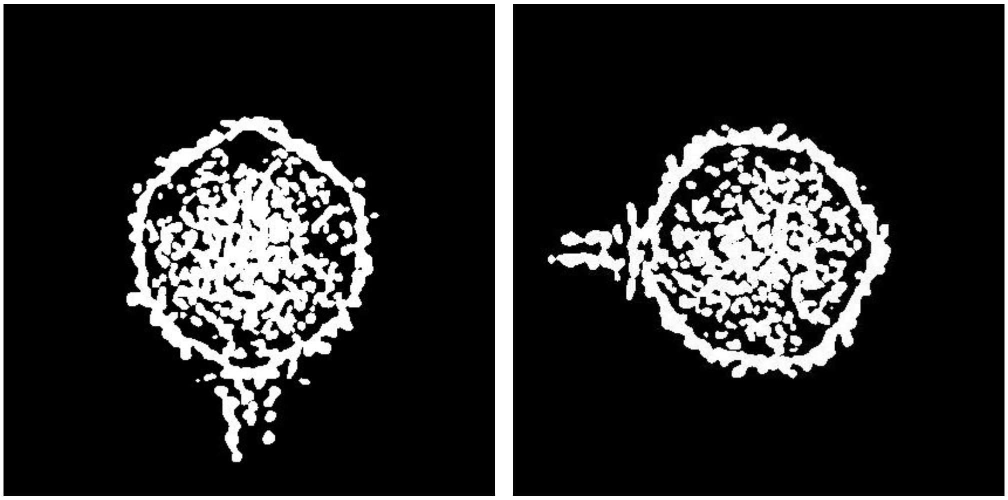
(c) View 3

Figure 4: Isosurface at threshold=2.300000. 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).

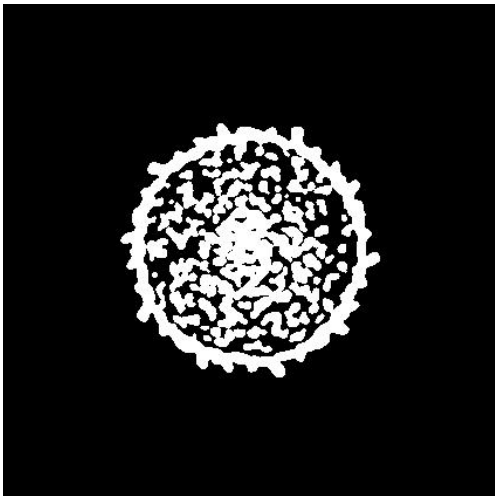
Orthogonal slices of maximum variance of the mask with hard borders

Explanation:
The mask with hard borders has been calculated at the suggested threshold 2.300000, the largest connected component was selected, and then dilated by 2Å.

Results:
See Fig. 5.



(a) X Slice 216 (b) Y Slice 195



(c) Z Slice 222

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

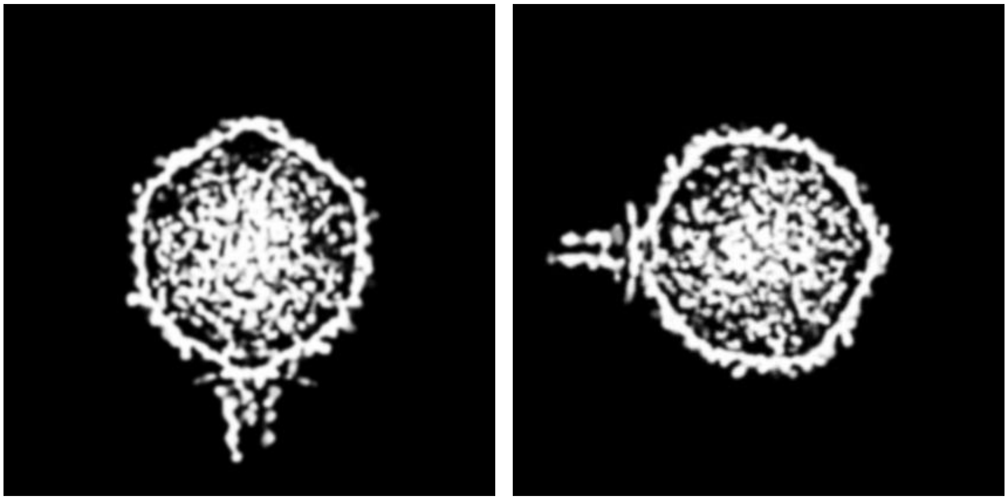
Orthogonal slices of maximum variance of the mask with soft borders

Explanation:

The mask with soft borders has been calculated at the suggested threshold 2.300000, the largest connected component was selected, and then dilated by

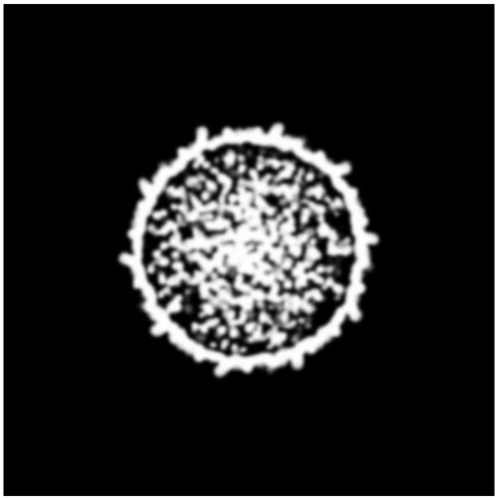
2Å.

Results:
See Fig. 6.



(a) X Slice 216

(b) Y Slice 195



(c) Z Slice 199

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 346.92 and 346.92 Å, respectively. There is a decentering ratio (abs(Right-Left)/Size)% of 0.00%

The space from the left and right in Y are 357.54 and 350.46 Å, respectively. There is a decentering ratio (abs(Right-Left)/Size)% of 0.48%

The space from the left and right in Z are 102.66 and 339.84 Å, respectively. There is a decentering ratio (abs(Right-Left)/Size)% of 15.95%

The center of mass is at (x,y,z)=(37399.81,162388.89,-371938.46). The decentering of the center of mass (abs(Center)/Size)% is 8854.72, 38614.02, and 88606.78, 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 and focused refinement, or similar, warnings are expected.

WARNINGS: 3 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. **The center of mass in Y may be significantly shifted. This is common when the refinement is applied exclusively to one protein region.**
3. **The center of mass in Z 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:

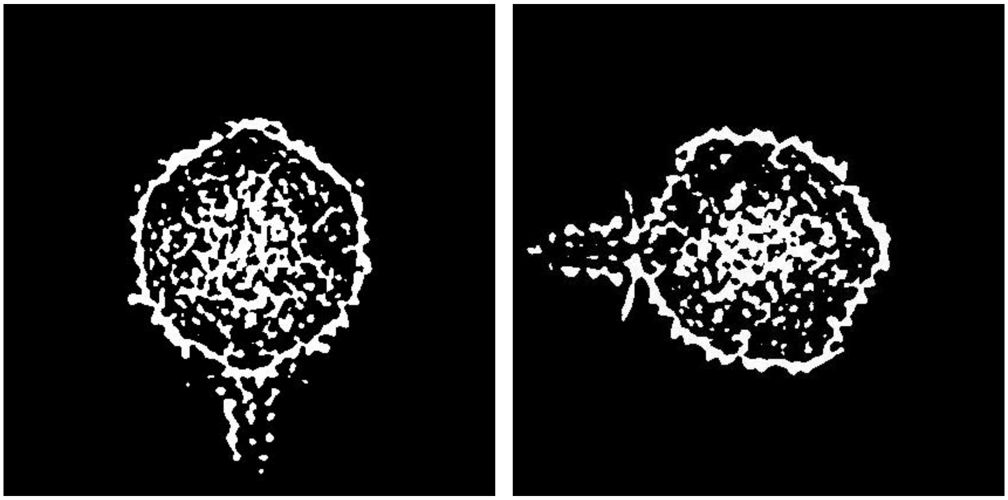
Raw mask: At threshold 2.300000, there are 287 connected components with a total number of voxels of 1838630 and a volume of 81565054.01 Å^3 (see Fig. 7). The size and percentage of the total number of voxels for the raw mask are listed below (up to 95% of the mass or the first 100 clusters, whatever happens first), the list contains (No. voxels (volume in Å^3), percentage, cumulated percentage):

(1818408 (80667968.39), 98.90, 98.90)

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

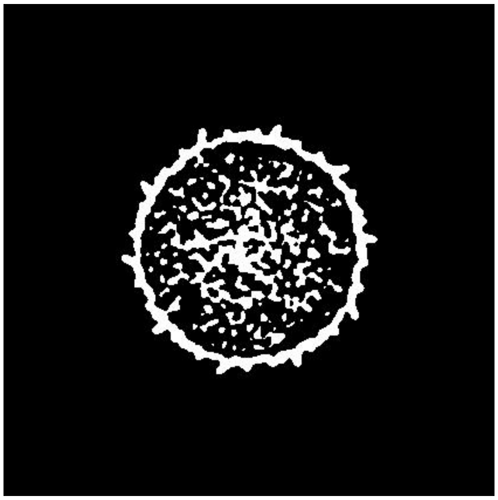
The average size of the remaining 286 components is 70.71 voxels (44.36 Å^3). Their size go from 1818408 voxels (80667968.39 Å^3) to 1 voxels (44.36 Å^3).

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



(a) X Slice 216

(b) Y Slice 209



(c) Z Slice 199

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**
0.5613	6293944.00	231326.50	2102968.20
1.1226	3849989.00	141501.81	1286380.12
1.6839	2809296.00	103252.37	938657.88
2.2452	1913833.00	70340.68	639460.72
2.8065	1285658.00	47252.85	429571.33
3.3678	940245.00	34557.60	314159.98
3.9291	756921.00	27819.74	252906.73
4.4904	634369.00	23315.49	211958.96
5.0517	531837.00	19547.04	177700.39
5.6130	437732.00	16088.32	146257.49
6.1743	351009.00	12900.92	117281.11
6.7356	273272.00	10043.79	91307.19
7.2969	205289.00	7545.16	68592.32
7.8582	148461.00	5456.51	49604.63
8.4195	102955.00	3783.99	34399.91
8.9808	68307.00	2510.54	22823.12
9.5421	42229.00	1552.08	14109.79
10.1034	24116.00	886.36	8057.77
10.6647	12486.00	458.91	4171.89
11.2260	5728.00	210.53	1913.87
11.7873	2217.00	81.48	740.76
12.3486	716.00	26.32	239.23
12.9099	176.00	6.47	58.81
13.4712	37.00	1.36	12.36

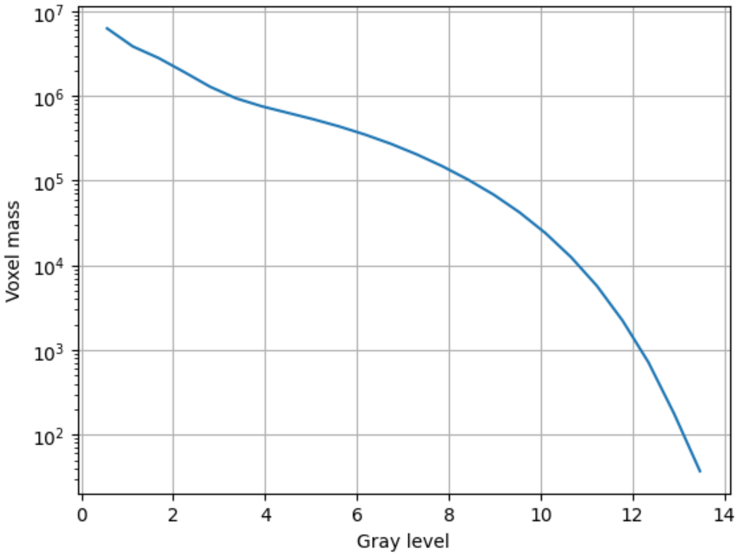


Figure 8: Voxel mass as a function of the gray level.

Constructed mask: After keeping the largest component of the previous mask and dilating it by 2Å, there is a total number of voxels of 3022648 and a volume of 134090299.50 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%.

WARNINGS: 1 warnings

1. **There might be a problem with noise and artifacts, because the average noise blob has a volume of 44.361864 Å^3.**

2.3 Level 0.c Background analysis

Explanation:

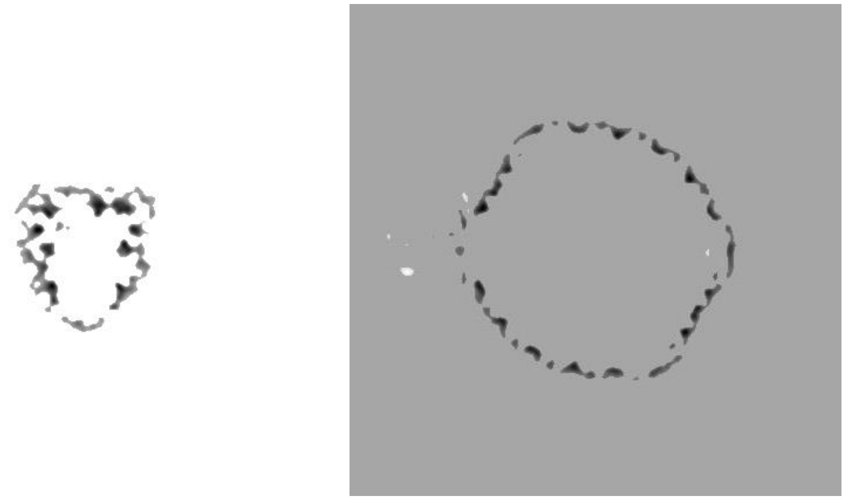
Background is defined as the region outside the macromolecule mask. The background mean should be zero, and the number of voxels with a very low or very high value (below 5 standard deviations of the noise) should be very small and they should be randomly distributed without any specific structure. Sometimes, 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 -1841.06 and 0.000000, respectively.

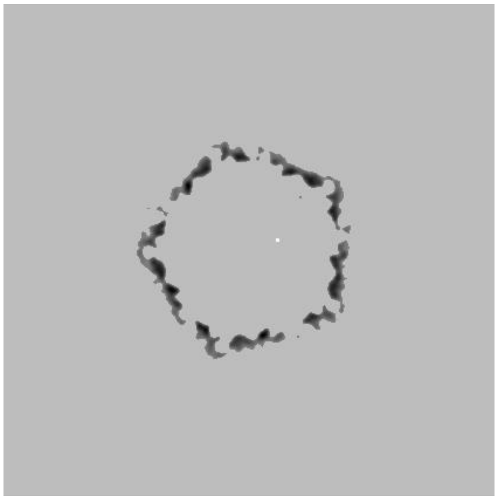
The mean and standard deviation (sigma) of the background were -0.136109 and 0.623229. The percentage of background voxels whose absolute value is larger than 5 times the standard deviation is 0.75 % (see Fig. 9). The same percentage from a Gaussian would be 0.000057% (ratio between the two percentages: 13168.821743).

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



(a) X Slice 116

(b) Y Slice 210



(c) Z Slice 142

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=13168.82)**

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 nan. The fitted line was $\log(|F|^2) = nan/R^2 + (nan)$.

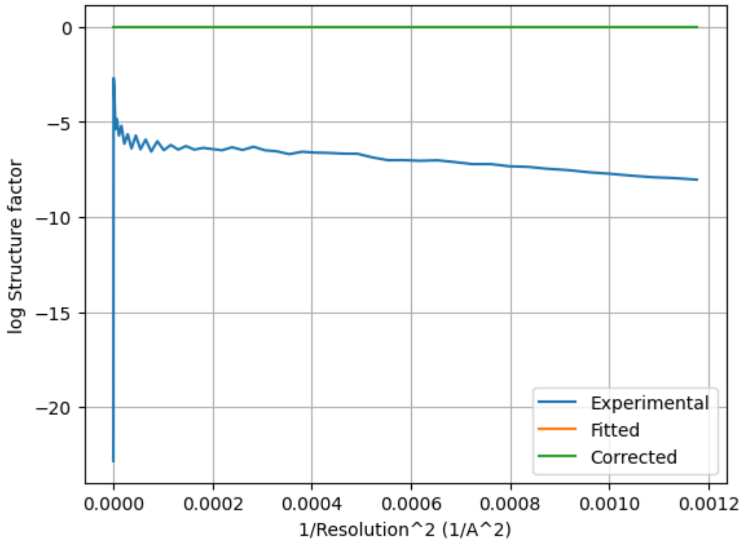
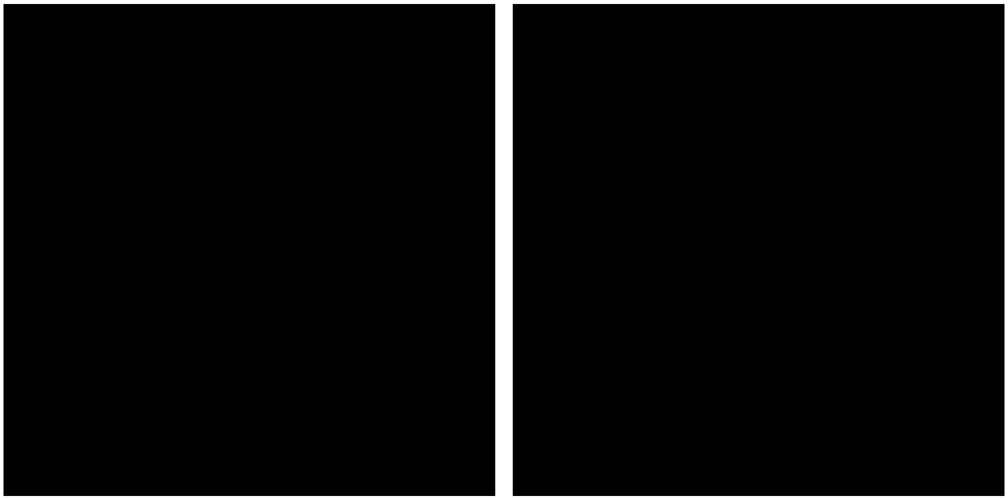
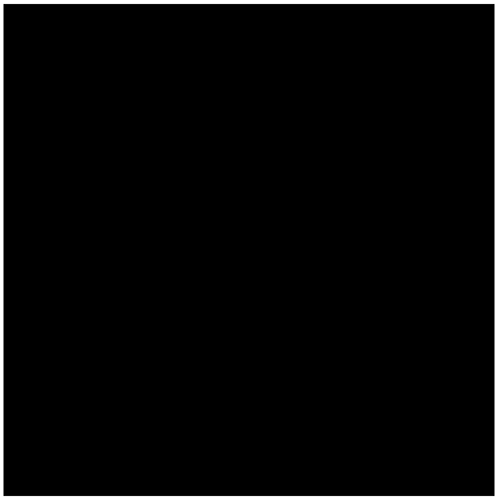


Figure 10: Guinier plot. The X-axis is the square of the inverse of the resolution in Å.



(a) X Slice 0 (b) Y Slice 0



(c) Z Slice 0

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

2.5 Level 0.e Local resolution with DeepRes

Explanation:
DeepRes (see this link for more details) measures the local resolution using a neural network that has been trained on the appearance of atomic structures at different resolutions. Then, by comparing the local appearance of the input map to the appearance of the atomic structures a local resolution label can be assigned.

Results:

This method cannot be applied to maps with a resolution worse than 13Å.

STATUS: Does not apply

2.6 Level 0.f Local B-factor

Explanation:
LocBfactor (see this link for more details) estimates a local resolution B-factor by decomposing the input map into a local magnitude and phase term using the spiral transform.

Results:

Fig. 12 shows the histogram of the local B-factor according to LocBfactor. Some representative percentiles are:

Percentile	**Local B-factor ($Å^{-2}$)**
2.5%	-22666.63
25%	-21086.03
50%	-20286.93
75%	-19506.63
97.5%	-18029.65

Fig. 13 shows some representative views of the local B-factor.

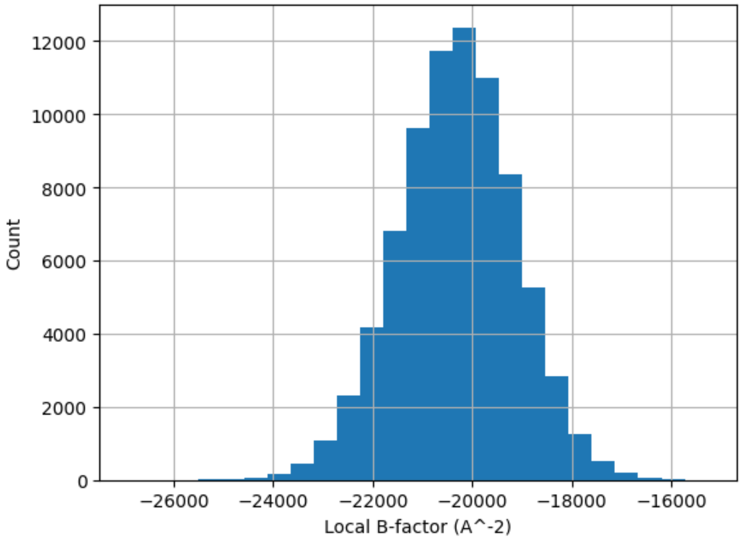
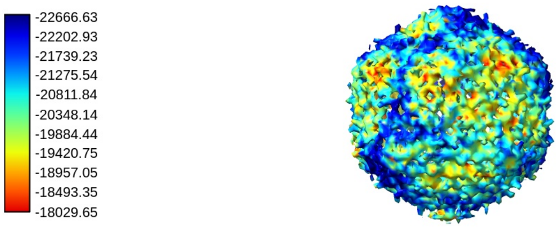
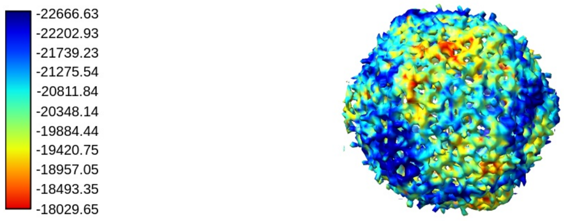


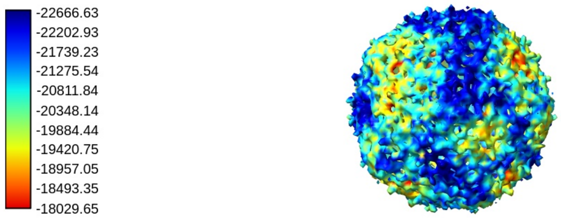
Figure 12: Histogram of the local B-factor according to LocBfactor.



(a) View 1



(b) View 2



(c) View 3

Figure 13: Local B-factor according to LocBfactor. 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).

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

WARNINGS: 1 warnings

1. **The median B-factor is out of the interval [-300,0]**

2.7 Level 0.g Local Occupancy

Explanation:
LocOccupancy (see this link for more details) estimates the occupancy of a voxel by the macromolecule.

Results:

Fig. 14 shows the histogram of the local occupancy according to LocOccupancy. Some representative percentiles are:

Percentile	**Local Occupancy [0-1]**
2.5%	0.00
25%	0.25
50%	1.00
75%	1.00
97.5%	1.00

Fig. 15 shows some representative views of the local occupancy.

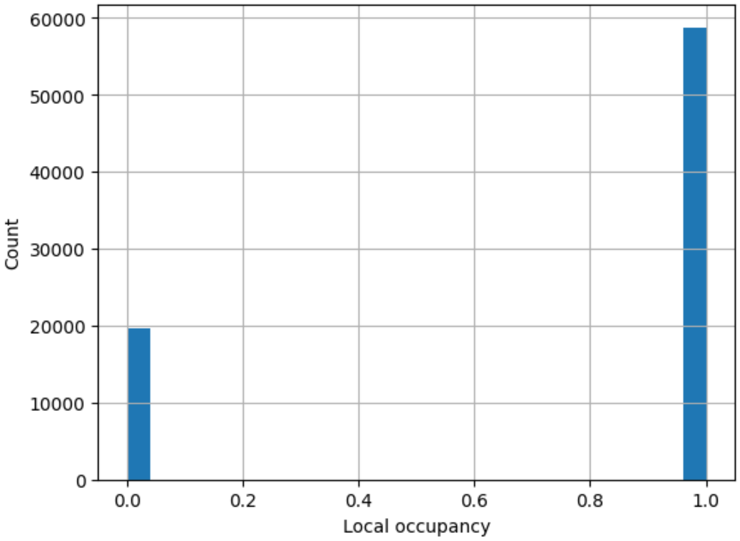
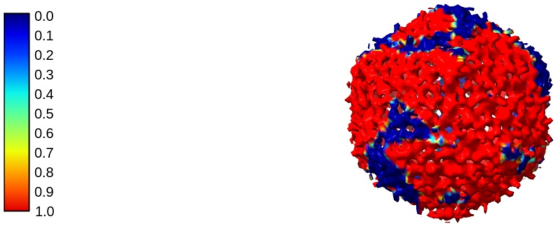
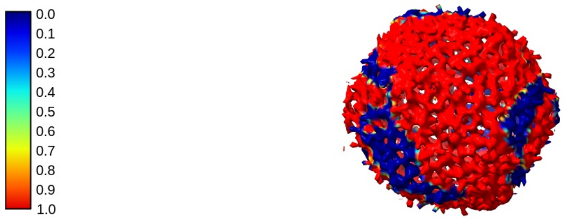


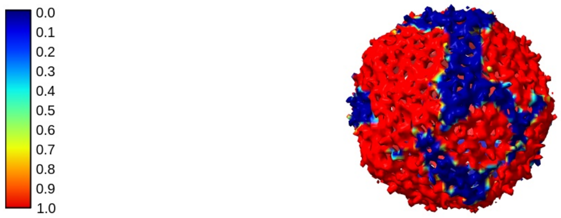
Figure 14: Histogram of the local occupancy according to LocOccupancy.



(a) View 1



(b) View 2



(c) View 3

Figure 15: Local occupancy according to LocOccupancy. 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).

Automatic criteria: The validation is OK if the median occupancy is larger than 50%.

STATUS: OK

2.8 Level 0.h Hand correction

Explanation:
Deep Hand (see this link for more details) determines the correction of the hand for those maps with a resolution smaller than 5Å. The method calculates a value between 0 (correct hand) and 1 (incorrect hand) using a neural network to assign its hand.

Results:

This method cannot be applied to maps with a resolution worse than 5Å.

STATUS: Does not apply