Validation report of Level(s) 0, 1, A

I^{2}PC Validation server

February 25, 2022
8:40pm

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). According to FSC-Q, it seems that there is a mismatch between the map and its model (see Sec. 6.2). According to phenix, it seems that there might be some mismatch between the map and its model (see Sec. 6.4).

The average resolution of the map estimated by various methods goes from 0.3Å to 10.2Å with an average of 4.9Å. The resolution provided by the user was 3.3Å. The resolution reported by the user may be overestimated.

The overall score (passing tests) of this report is 14 out of 20 evaluable items.

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	1 warnings
1.b FSC permutation	Sec. 4.2	OK
1.c Blocres	Sec. 4.3	OK
1.d Resmap	Sec. 4.4	Could not be measured
1.e MonoRes	Sec. 4.5	OK
1.f MonoDir	Sec. 4.6	2 warnings
1.g FSO	Sec. 4.7	OK
1.h FSC3D	Sec. 4.8	Could not be measured
A.a MapQ	Sec. 6.1	OK
A.b FSC-Q	Sec. 6.2	1 warnings
A.d Map-Model Guinier	Sec. 6.3	OK
A.e Phenix validation	Sec. 6.4	1 warnings
A.f EMRinger	Sec. 6.5	OK
A.g DAQ	Sec. 6.6	OK

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=4280.03)**

Section 2.5 (0.e DeepRes)

1. **The reported resolution, 3.30 Å, is particularly with respect to the local resolution distribution. It occupies the 0.03 percentile**

Section 4.1 (1.a Global resolution)

1. **The reported resolution, 3.30 Å, is particularly high with respect to the resolution calculated by the FSC, 7.54 Å**

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.**
2. **The resolution reported by the user, 3.30Å, is at least 80% smaller than the average directional resolution, 7.31 Å.**

Section 6.2 (A.b FSC-Q)

1. **The percentage of voxels that have a FSC-Qr larger than 1.5 in absolute value is 10.2, that is larger than 10%**

Section 6.4 (A.e Phenix validation)

1. **The percentage of residues that have a cross-correlation below 0.5 is 20.8, that is larger than 10%**

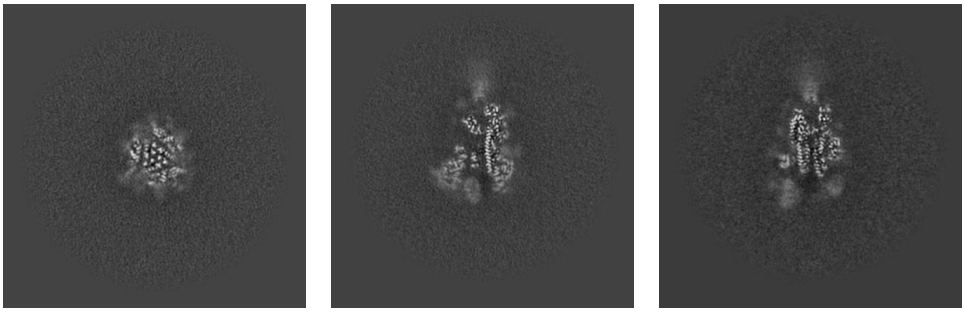
Contents

1 Input data

Input map: /home/coss/data/Dropbox/Aplicaciones/ShareLaTeX/MapValidation/-EMDB11337/emd_11337.map
SHA256 hash: d969dcfa8853ce92e8d9932e578ff2db49e7d00d1cfa3921607bcbbd4ff3cc23
Voxel size: 1.047000 (Å)
Visualization threshold: 0.165000
Resolution estimated by user: 3.300000

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, ...)

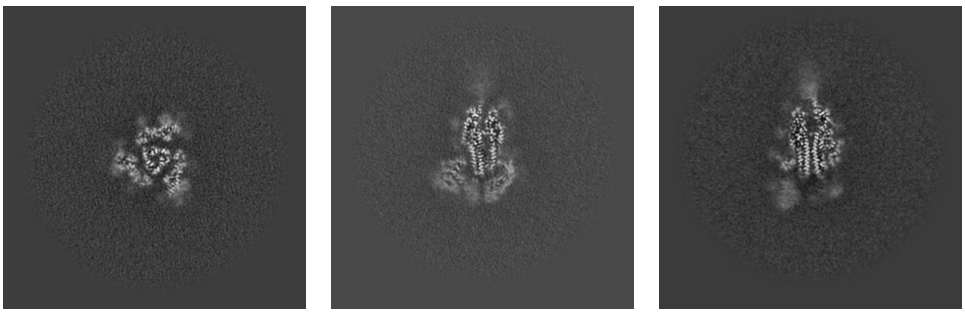
Results:
See Fig. 1.



(a) X Slice 216 (b) Y Slice 216 (c) Z Slice 216

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.



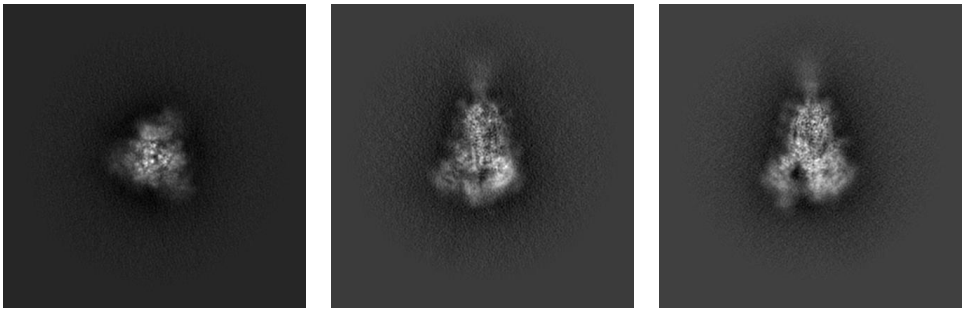
(a) X Slice 231 (b) Y Slice 225 (c) Z Slice 213

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.

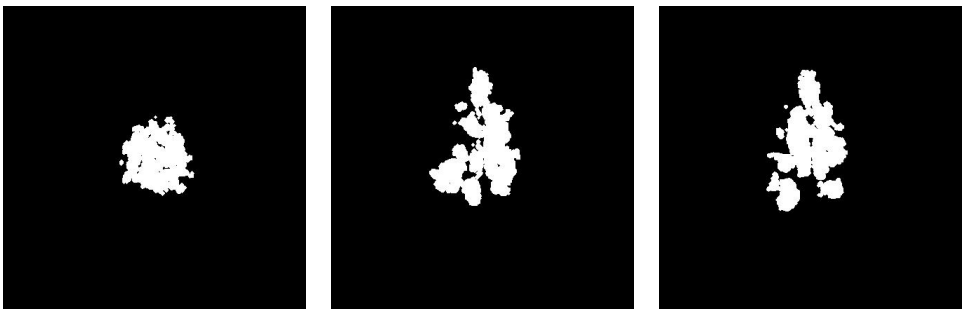
(a) View 1 (b) View 2 (c) View 3

Figure 4: Isosurface at threshold=0.165000. Views generated by ChimeraX at a the following X, Y, Z angles: View 1 (0,0,0), View 2 (90, 0, 0), View 3 (0, 90, 0).

Orthogonal slices of maximum variance of the mask

Explanation:
The mask has been calculated at the suggested threshold 0.165000, the largest connected component was selected, and then dilated by 2Å.

Results:
See Fig. 5.



(a) X Slice 216 (b) Y Slice 216 (c) Z Slice 216

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

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 91.09 and 136.11 Å, respectively. There is a decentering ratio (abs(Right-Left)/Size)% of 9.95%

The space from the left and right in Y are 145.53 and 151.81 Å, respectively. There is a decentering ratio (abs(Right-Left)/Size)% of 1.39%

The space from the left and right in Z are 146.58 and 162.28 Å, respectively. There is a decentering ratio (abs(Right-Left)/Size)% of 3.47%

The center of mass is at (x,y,z)=(212.02,228.50,203.12). The decentering of the center of mass (abs(Center)/Size)% is 0.92, 2.89, and 2.98, 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.

STATUS: OK

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:

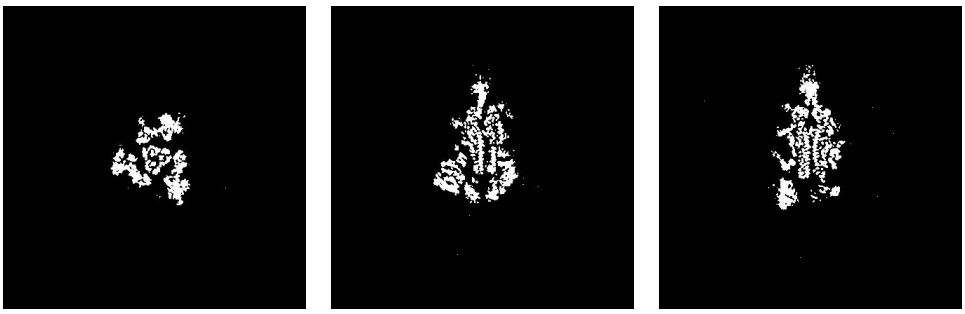
Raw mask: At threshold 0.165000, there are 874 connected components with a total number of voxels of 341566 and a volume of 392025.83 Å^3 (see Fig. 6). The size and percentage of the total number of voxels for the raw mask are listed below (up to 95% of the mass), the list contains (No. voxels (volume in Å^3), percentage, cumulatedPercentage):

(338472 (388474.75), 99.09, 99.09)

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

The average size of the remaining 873 components is 3.54 voxels (1.15 Å^3). Their size go from 248 voxels (284.64 Å^3) to 1 voxels (1.15 Å^3).

The slices of the raw mask can be seen in Fig. 6.



(a) X Slice 232 (b) Y Slice 223 (c) Z Slice 211

Figure 6: 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. 7):

Threshold	**Voxel mass**	**Molecular mass(kDa)**	**# Aminoacids**
0.0622	1664392.00	1582.66	14387.84
0.1245	456945.00	434.51	3950.06
0.1867	300097.00	285.36	2594.19
0.2490	211046.00	200.68	1824.39
0.3112	149710.00	142.36	1294.17
0.3734	107653.00	102.37	930.61
0.4357	79118.00	75.23	683.94
0.4979	58725.00	55.84	507.65
0.5601	43849.00	41.70	379.05
0.6224	32674.00	31.07	282.45
0.6846	24175.00	22.99	208.98
0.7469	17504.00	16.64	151.31
0.8091	12522.00	11.91	108.25
0.8713	8618.00	8.19	74.50
0.9336	5718.00	5.44	49.43
0.9958	3462.00	3.29	29.93
1.0580	1918.00	1.82	16.58
1.1203	931.00	0.89	8.05
1.1825	384.00	0.37	3.32
1.2448	167.00	0.16	1.44
1.3070	58.00	0.06	0.50
1.3692	20.00	0.02	0.17
1.4315	6.00	0.01	0.05
1.4937	3.00	0.00	0.03

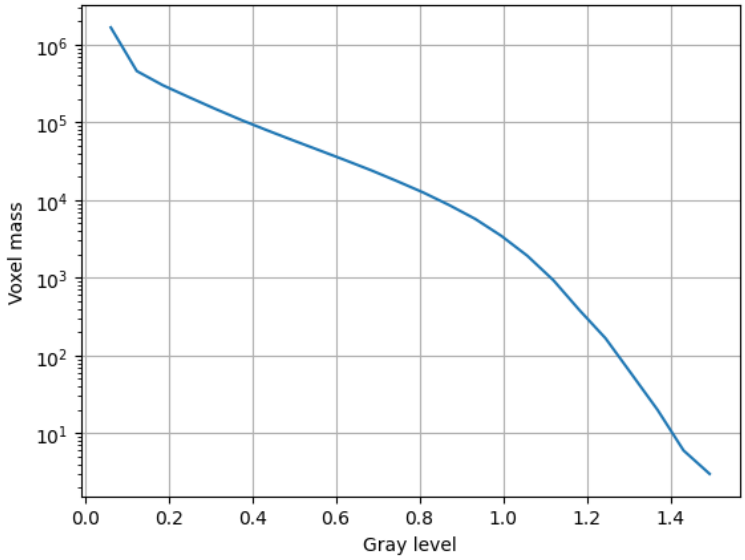


Figure 7: Voxel mass as a function of the gray level.

Constructed mask: After keeping the largest component of the previous mask and dilating it by 2Å, there is a total number of voxels of 799000 and a volume of 917036.93 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

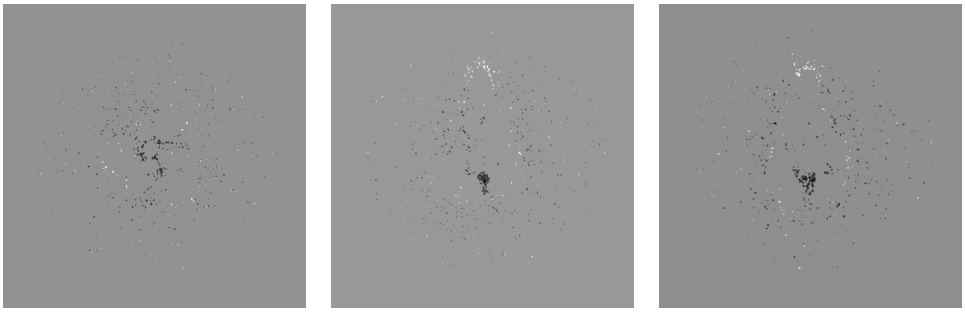
small and they should be randomly distributed without any specific structure. Sometimes, you can see some structure due to the symmetry of the structure.

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 -839.68 and 0.000000, respectively.

The mean and standard deviation (sigma) of the background were -0.002207 and 0.023486. The percentage of background voxels whose absolute value is larger than 5 times the standard deviation is 0.25 % (see Fig. 8). The same percentage from a Gaussian would be 0.000057% (ratio between the two percentages: 4280.030049).

Slices of the background beyond 5*sigma can be seen in Fig. 8.



(a) X Slice 240 (b) Y Slice 218 (c) Z Slice 220

Figure 8: 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=4280.03)**

2.4 Level 0.d B-factor analysis

Explanation:
The B-factor line [Rosenthal and Henderson, 2003] fitted between 15Åand the resolution reported should have a slope that is between 0 and 300 Å^2.

Results:
Fig. 9 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 -89.2. The fitted line was $\log(|F|^2) = -22.3/R^2 + (-10.9)$.

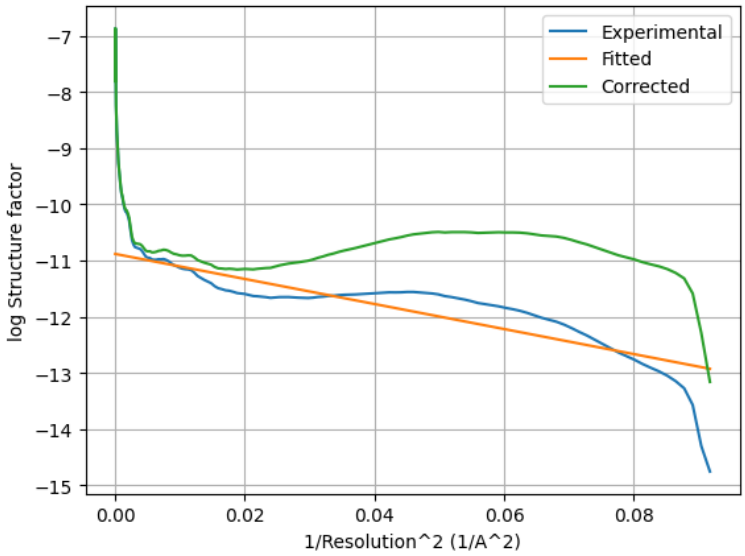
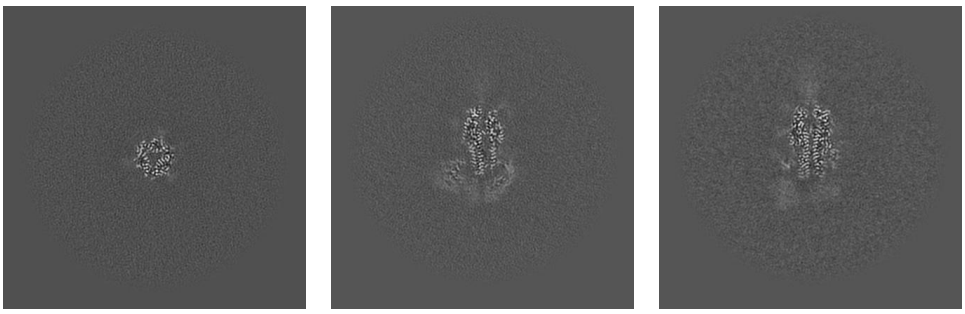


Figure 9: Guinier plot. The X-axis is the square of the inverse of the resolution in Å.



(a) X Slice 165 (b) Y Slice 225 (c) Z Slice 211

Figure 10: 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 [Ramírez-Aportela et al., 2019] 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:

Fig. 11 shows the histogram of the local resolution according to DeepRes. Some representative percentiles are:

Percentile	Resolution(Å)
2.5%	3.88
25%	4.34
50%	4.64
75%	5.02
97.5%	6.03

The reported resolution, 3.30 Å, is at the percentile 0.0. Fig. 12 shows some representative views of the local resolution.

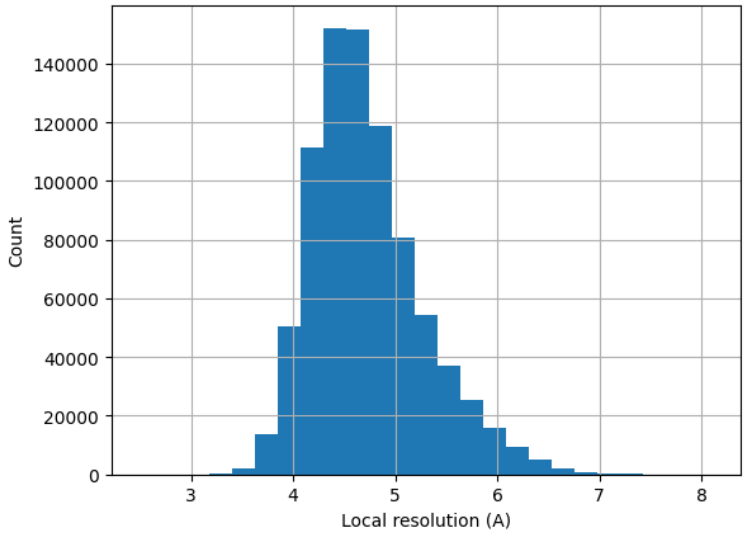
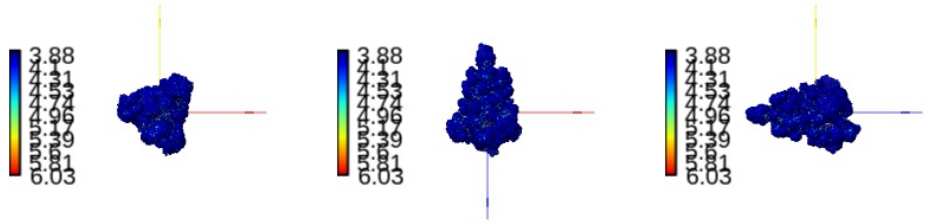


Figure 11: Histogram of the local resolution according to deepres.



(a) View 1 (b) View 2 (c) View 3

Figure 12: Local resolution according to DeepRes. Views generated by ChimeraX at a the following X, Y, Z angles: View 1 (0,0,0), View 2 (90, 0, 0), View 3 (0, 90, 0).

Automatic criteria: The validation is OK if the percentile of the user provided resolution is larger than 0.1% of the percentile of the local resolution as estimated by DeepRes.

WARNINGS: 1 warnings

1. **The reported resolution, 3.30 Å, is particularly with respect to the local resolution distribution. It occupies the 0.03 percentile**

2.6 Level 0.f Local B-factor

Explanation:
LocBfactor [Kaur et al., 2021] estimates a local resolution B-factor by decomposing the input map into a local magnitude and phase term using the spiral transform.

Results:

Fig. 13 shows the histogram of the local B-factor according to LocBfactor. Some representative percentiles are:

Percentile	Local B-factor ($Å^{-2}$)
2.5%	-307.82
25%	-262.94
50%	-239.10
75%	-212.06
97.5%	-149.18

Fig. 14 shows some representative views of the local B-factor.

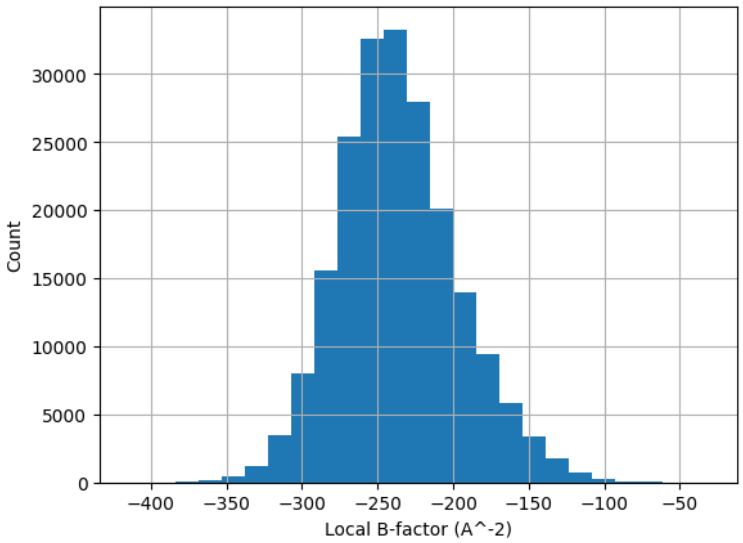
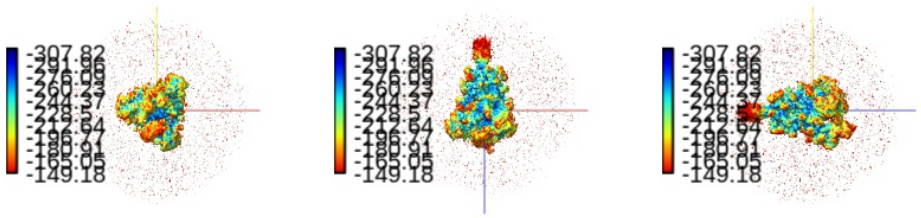


Figure 13: Histogram of the local B-factor according to LocBfactor.



(a) View 1 (b) View 2 (c) View 3

Figure 14: Local B-factor according to LocBfactor. Views generated by ChimeraX at a the following X, Y, Z angles: View 1 (0,0,0), View 2 (90, 0, 0), View 3 (0, 90, 0).

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

STATUS: OK

2.7 Level 0.g Local Occupancy

Explanation:
LocOccupancy [Kaur et al., 2021] estimates the occupancy of a voxel by the macromolecule.

Results:

Fig. 15 shows the histogram of the local occupancy according to LocOccupancy. Some representative percentiles are:

Percentile	Local Occupancy [0-1]
2.5%	0.23
25%	0.55
50%	0.86
75%	0.95
97.5%	1.00

Fig. 16 shows some representative views of the local occupancy.

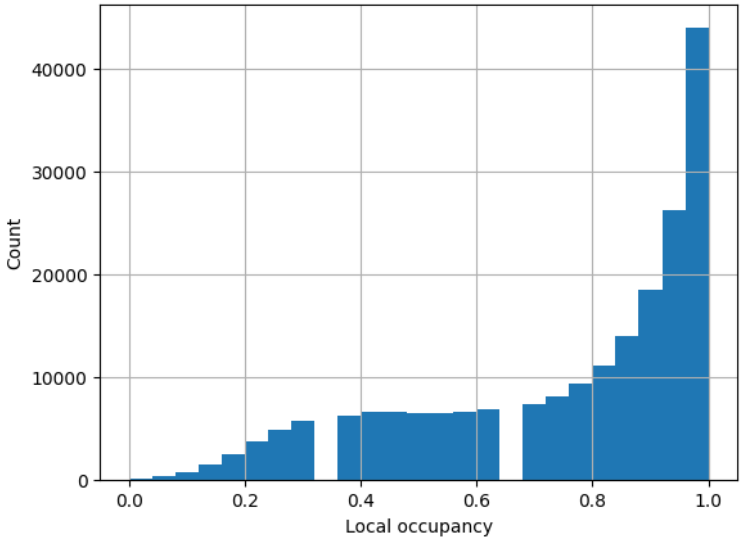
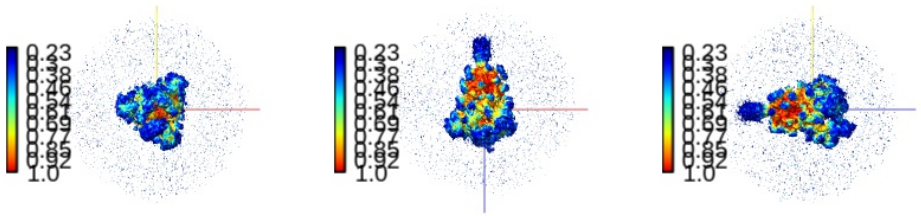


Figure 15: Histogram of the local occupancy according to LocOccupancy.



(a) View 1 (b) View 2 (c) View 3

Figure 16: Local occupancy according to LocOccupancy. Views generated by ChimeraX at a the following X, Y, Z angles: View 1 (0,0,0), View 2 (90, 0, 0), View 3 (0, 90, 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 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:

Deep hand assigns a score of 0.254 to the input volume.
Automatic criteria: The validation is OK if the deep hand score is smaller than 0.5.

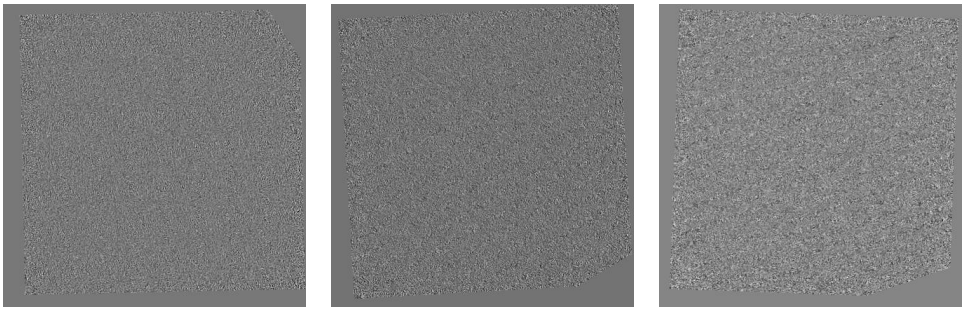
STATUS: OK

3 Half maps

Half map 1: /home/coss/data/Dropbox/Aplicaciones/ShareLaTeX/MapValidation/-EMDB11337/emd_11337_half_map_1.map
SHA256 hash: 17945f1afcb4d373fe3b41d2a904ea0721dc63f78fa3956d012dda9e537390a6

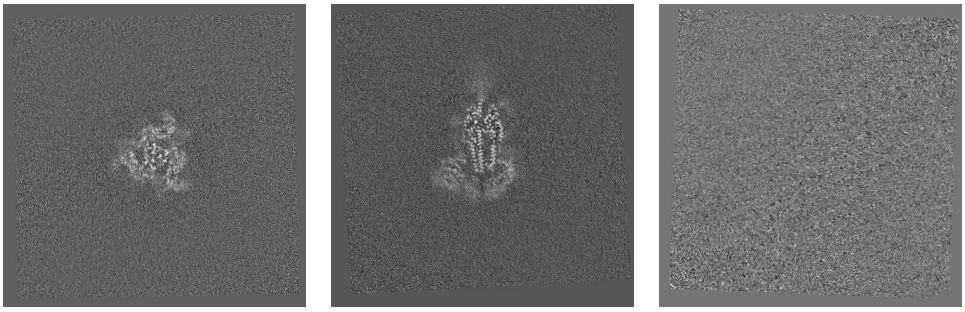
Half map 2: /home/coss/data/Dropbox/Aplicaciones/ShareLaTeX/MapValidation/-EMDB11337/emd_11337_half_map_2.map
SHA256 hash: ded62a6026f96d71c7c487135471c7443aca06d1d32f2afdf909d3152f7f79e2

Slices of the first half map can be seen in Fig. 17.
Slices of the second half map can be seen in Fig. 18.
Slices of the difference between both maps can be seen in Fig. 19. There should not be any structure in this difference. Sometimes some patterns are seen if the map is symmetric.



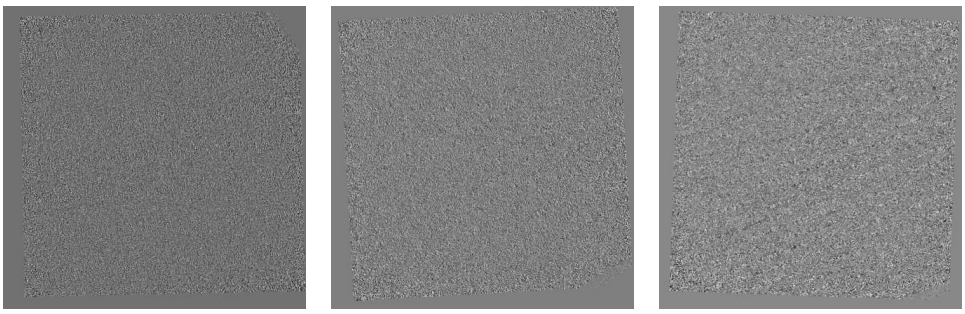
(a) X Slice 386 (b) Y Slice 407 (c) Z Slice 30

Figure 17: Slices of maximum variation in the three dimensions of Half 1



(a) X Slice 222 (b) Y Slice 224 (c) Z Slice 31

Figure 18: Slices of maximum variation in the three dimensions of Half 2



(a) X Slice 386 (b) Y Slice 407 (c) Z Slice 31

Figure 19: Slices of maximum variation in the three dimensions of the difference Half1-Half2.

4 Level 1 analysis

4.1 Level 1.a Global resolution

Explanation: 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 right thresholds for these measures. Probably, the most clear threshold is the one of the SSNR (SSNR=1). For the DPR we have chosen 103.9° and for the FSC, the standard 0.143. For a deep discussion of all these thresholds, see [Sorzano et al., 2017]. Note that these thresholds typically result in resolution values that are at the lower extreme of the local resolution range, meaning that this resolution is normally in the first quarter. It should not be understood as the average resolution of the map.

Except for the noise, the FSC and DPR should be approximately monotonic. They should not have any "coming back" behavior. If they have, this is typically due to the presence of a mask in real space or non-linear processing.

Results:
Fig. 20 shows the FSC and the 0.143 threshold. The resolution according to the FSC is 7.54Å. The map information is well preserved (FSC>0.9) up to

41.58Å.
Fig. 21 shows the DPR and the 103.9° threshold. The resolution according to the DPR is 3.82Å.
Fig. 22 shows the SSNR and the SSNR=1 threshold. The resolution according to the SSNR is 3.79Å.
The mean resolution between the three methods is 5.05Å and its range is within the interval [3.79, 7.54]Å.

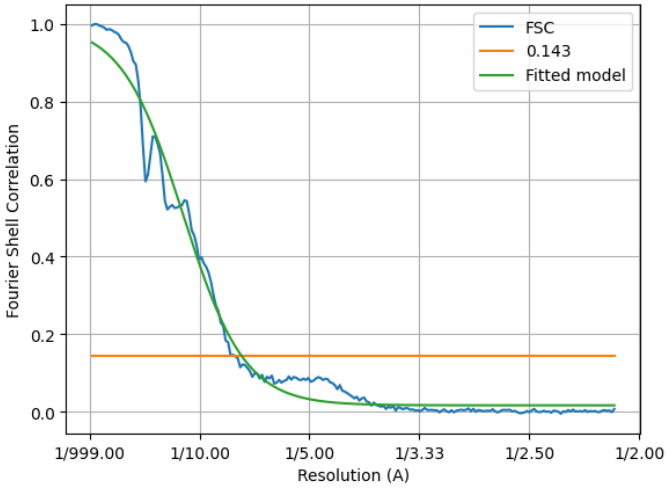


Figure 20: Fourier Shell correlation between the two halves.

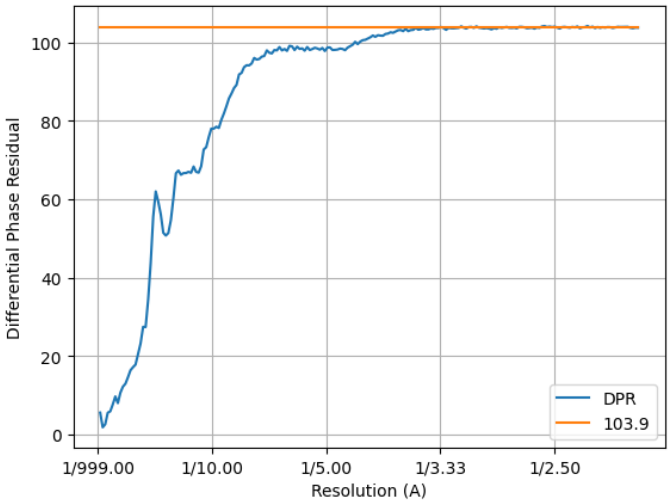


Figure 21: Differential Phase Residual between the two halves.

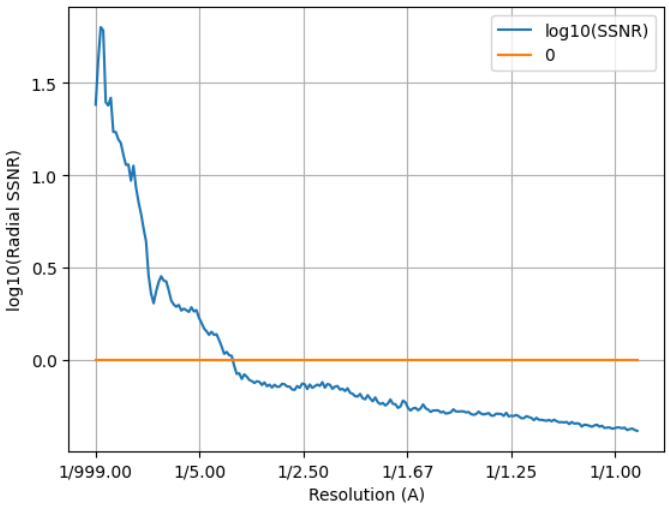


Figure 22: Spectral Signal-to-Noise Ratio estimated from the two halves.

Automatic criteria: The validation is OK if the user provided resolution is larger than 0.8 times the resolution estimated by 1) FSC, 2) DPR, and 3) SSNR.

WARNINGS: 1 warnings

1. **The reported resolution, 3.30 Å, is particularly high with respect to the resolution calculated by the FSC, 7.54 Å**

4.2 Level 1.b FSC permutation

Explanation:
This method [Beckers and Sachse, 2020] calculates a global resolution by formulating a hypothesis test in which the distribution of the FSC of noise is calculated from the two maps.

Results:

The resolution at 1% of FDR was 3.4. The estimated B-factor was -36.4. Fig. 23 shows the estimated FSC and resolution.

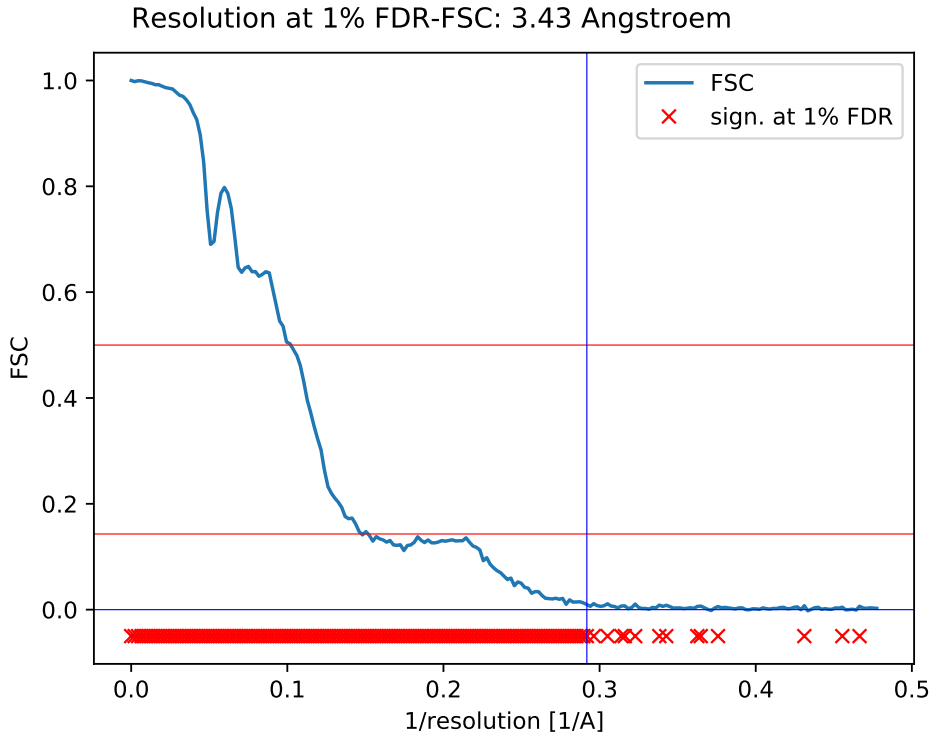


Figure 23: FSC and resolution estimated by a permutation test.

Automatic criteria: The validation is OK if the user provided resolution is larger than 0.8 times the resolution estimated by FSC permutation.

STATUS: OK

4.3 Level 1.c Local resolution with Blocres

Explanation:
This method [Cardone et al., 2013] computes a local Fourier Shell Correlation (FSC) between the two half maps.

Results:

Fig. 24 shows the histogram of the local resolution according to Blocres. Some representative percentiles are:

Percentile	Resolution(Å)
2.5%	3.19
25%	3.79
50%	5.34
75%	7.22
97.5%	9.38

The reported resolution, 3.30 Å, is at the percentile 7.0. Fig. 25 shows some representative views of the local resolution.

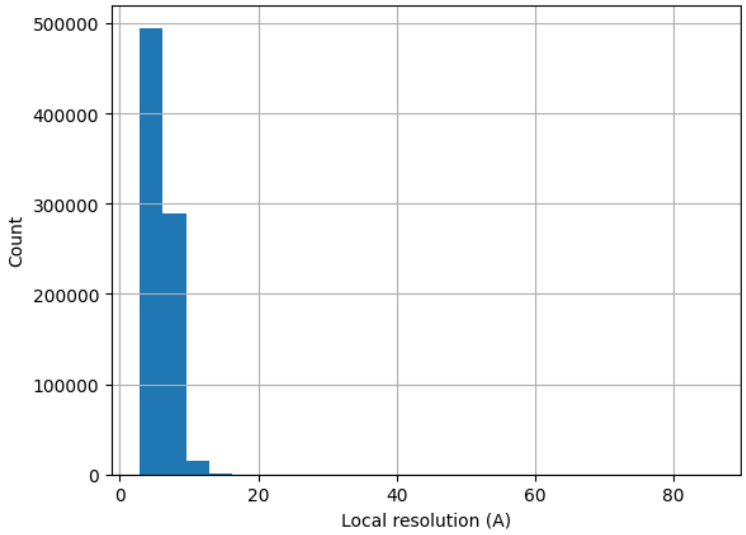
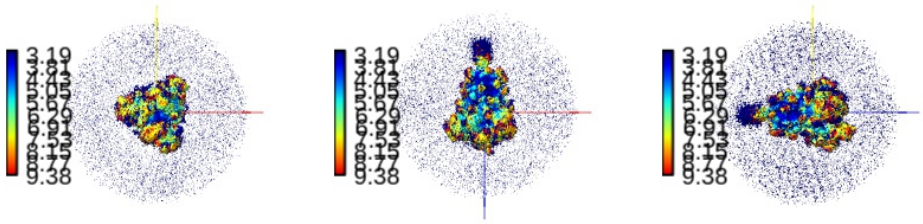


Figure 24: Histogram of the local resolution according to blocres.



(a) View 1 (b) View 2 (c) View 3

Figure 25: Local resolution according to Blocres. Views generated by ChimeraX at a the following X, Y, Z angles: View 1 (0,0,0), View 2 (90, 0, 0), View 3 (0, 90, 0).

Automatic criteria: The validation is OK if the percentile of the user provided resolution is larger than 0.1% of the percentile of the local resolution as estimated by BlocRes.

STATUS: OK

4.4 Level 1.d Local resolution with Resmap

Explanation:
This method [Kucukelbir et al., 2014] is based on a test hypothesis testing of the superiority of signal over noise at different frequencies.

Results:

ERROR: The protocol failed.

4.5 Level 1.e Local resolution with MonoRes

Explanation:
MonoRes [Vilas et al., 2018] evaluates the local energy of a point with respect 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. Then the energy of the amplitude within the map is compared to the amplitude distribution observed in the noise, and a hypothesis test is run for every voxel to check if its energy is signficantly above the level of noise.

Results:

Fig. 26 shows the histogram of the local resolution according to MonoRes. Some representative percentiles are:

Percentile	**Resolution(Å)**
2.5%	3.38
25%	5.09
50%	10.17
75%	14.74
97.5%	16.00

The reported resolution, 3.30 Å, is at the percentile 2.5. Fig. 27 shows some representative views of the local resolution

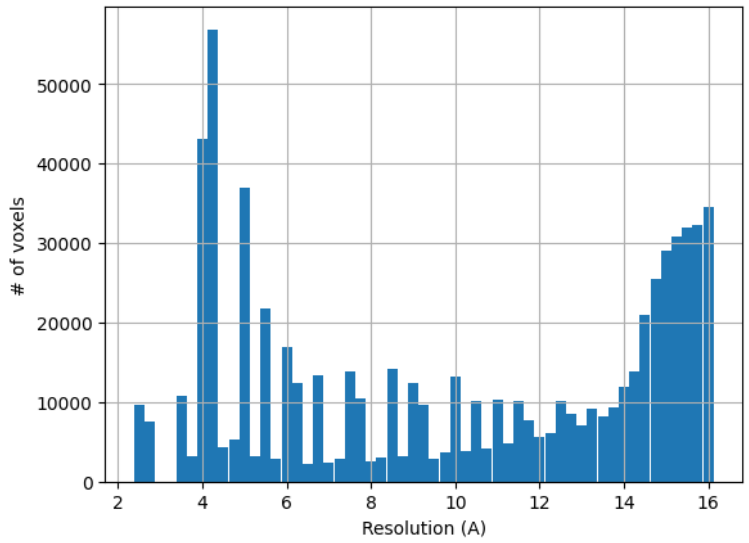
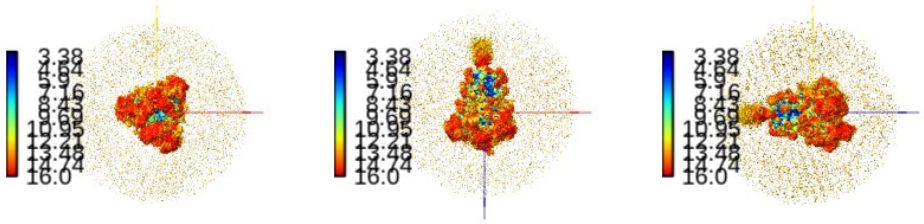


Figure 26: Histogram of the local resolution according to MonoRes.



(a) View 1 (b) View 2 (c) View 3

Figure 27: Local resolution according to Monores. Views generated by ChimeraX at a the following X, Y, Z angles: View 1 (0,0,0), View 2 (90, 0, 0), View 3 (0, 90, 0).

Automatic criteria: The validation is OK if the percentile of the user provided resolution is larger than 0.1% of the percentile of the local resolution as estimated by MonoRes.

STATUS: OK

4.6 Level 1.f Local and directional resolution with MonoDir

Explanation:
MonoDir [Vilas et al., 2020] 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.

The histogram of best resolution voxels per direction (Directional Histogram 1D) shows how many voxels in the volume have their maximum resolution in that direction. Directions are arbitrarily numbered from 1 to N. This histogram should be relatively flat. We perform a Kolmogorov-Smirnov test to check its uniformity. If the null hypothesis is rejected, then the directional resolution is not uniform. It does not mean that it is wrong, and it could be caused by several reasons: 1) the angular distribution is not uniform, 2) there are missing directions, 3) there is some anisotropy in the data (including some preferential directional movement).

Ideally, the radial average of the minimum, maximum, and average res-

olution at each voxel (note that these are spatial radial averages) should be flat and as low as possible. If they show some slope, this is associated with inaccuracies in the angular assignment. These averages make sense when the shells are fully contained within the protein. As the shells approach the outside of the protein, these radial averages make less sense.

Results:

Fig. 28 shows the 1D directional histogram and Fig. 29 the 2D directional histogram. We compared the 1D directional histogram to a uniform distribution using a Kolmogorov-Smirnov test. The D statistic was 0.048545, and the p-value of the null hypothesis 0.000000.

The radial average of the minimum, maximum and average resolution at each voxel is shown in Fig. 30. The overall mean of the directional resolution is 7.31

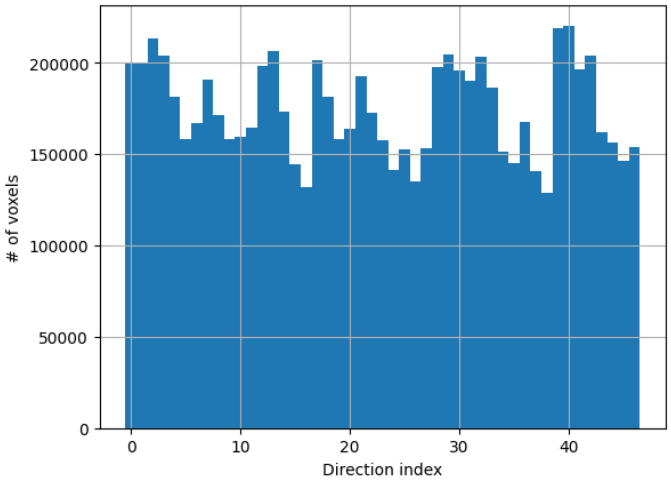


Figure 28: Histogram 1D of the best direction at each voxel.

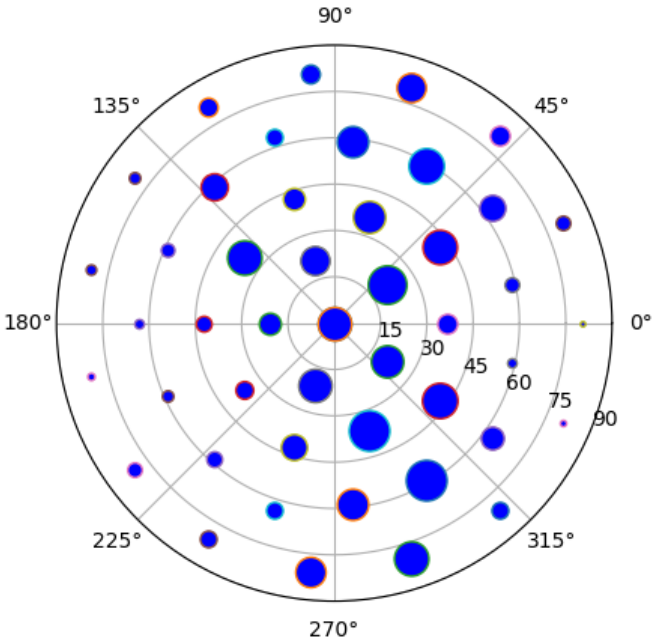


Figure 29: Histogram 2D of the best direction at each voxel. The azimuthal rotation is circular, while the tilt angle is the radius. The size of the point is proportional to the number of voxels whose maximum resolution is in that direction (this count can be seen in Fig. 28.

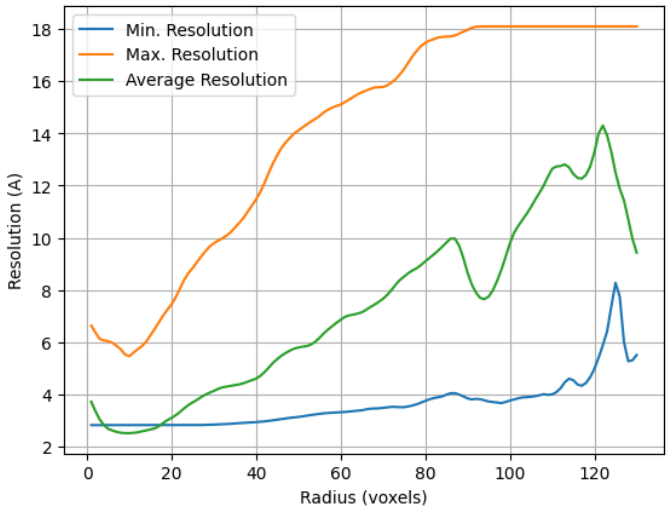


Figure 30: Radial averages (in space) of the minimum, maximum and average resolution at each voxel.

Automatic criteria: The validation is OK if 1) the null hypothesis that the directional resolution is not uniform is not rejected with a threshold of 0.001 for the p-value, and 2) the resolution provided by the user is not smaller than 0.8 times the average directional resolution.

WARNINGS: 2 warnings

1. **The distribution of best resolution is not uniform in all directions. The associated p-value is 0.000000.**
2. **The resolution reported by the user, 3.30Å, is at least 80% smaller than the average directional resolution, 7.31 Å.**

4.7 Level 1.g Fourier Shell Occupancy

Explanation:
This method calculates the anisotropy of the energy distribution in Fourier shells. This is an indirect measure of anisotropy of the angular distribution or the presence of heterogeneity. A natural threshold for this measure is 0.5. However, 0.9 and 0.1 are also interesting values that define the frequency at which the occupancy is 90% and 10%, respectively. This region is shaded in

the plot.
Results:

Fig. 31 shows the Fourier Shell Occupancy and its anisotropy. The directional resolution is shown in Fig. 32. The resolution according to the FSO is 3.39Å. Fourier shells are occupied at between 90 and than 10% in the range [3.58, 3.29]Å.

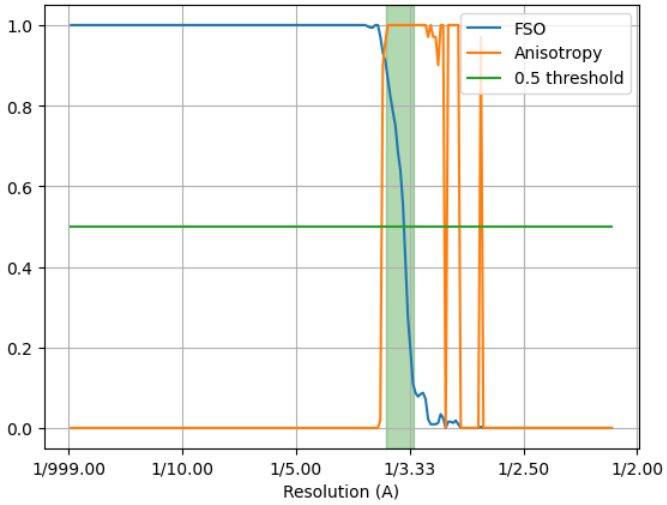


Figure 31: FSO and anisotropy.

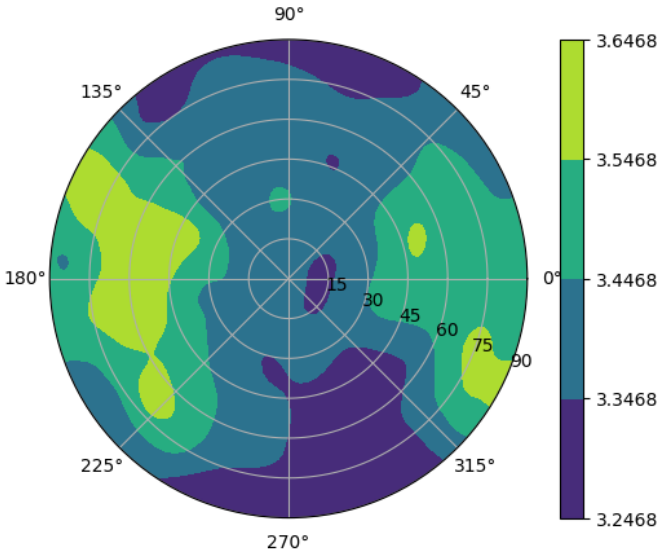


Figure 32: Directional resolution in the projection sphere.

Automatic criteria: The validation is OK if the resolution provided by the user is not smaller than 0.8 times the resolution estimated by the first cross of FSO below 0.5.

STATUS: OK

4.8 Level 1.h Fourier Shell Correlation 3D

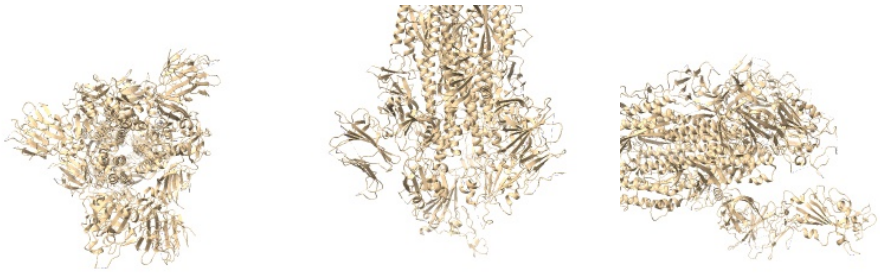
Explanation:
This method analyzes the FSC in different directions and evaluates its homogeneity.
Results:

ERROR: The protocol failed.

5 Atomic model

Atomic model: /home/coss/data/Dropbox/Aplicaciones/ShareLaTeX/MapValidation/-EMDB11337/6zp7_updated_centered.pdb

See Fig. 33.



(a) View 1 (b) View 2 (c) View 3

Figure 33: Input atomic model Views generated by ChimeraX at a the following X, Y, Z angles: View 1 (0,0,0), View 2 (90, 0, 0), View 3 (0, 90, 0).

6 Level A analysis

6.1 Level A.a MapQ

Explanation:
MapQ [**?**] computes the local correlation between the map and each one of its atoms assumed to have a Gaussian shape.

Results:

Fig. 34 shows the histogram of the Q-score according calculated by MapQ. Some representative percentiles are:

Percentile	MapQ score [0-1]
2.5%	-0.54
25%	-0.14
50%	0.10
75%	0.33
97.5%	0.67

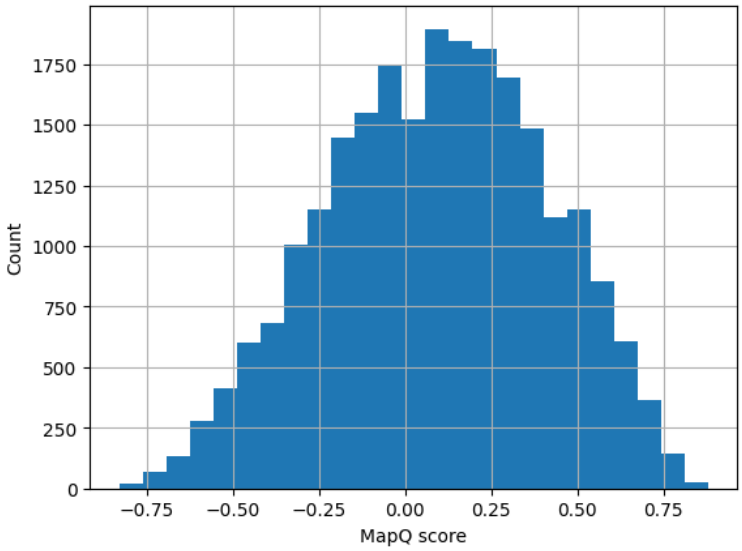


Figure 34: Histogram of the Q-score.

The following table shows the average Q score and estimated resolution for each chain.

Chain	Average Q score [0-1]	Estimated Resol. (Å)
A	0.15	5.4
A	0.16	0.0
A	0.12	0.0
A	0.09	0.0
B	0.04	6.1
B	0.01	0.0
B	0.19	0.0
C	0.07	5.9
C	0.17	0.0
D	0.03	0.0
E	0.29	0.0
E	0.37	0.0
E	0.32	0.0
F	0.06	0.0
F	0.24	0.0
F	-0.03	0.0
G	0.21	0.0
H	0.20	0.0
I	0.17	0.0
J	0.04	0.0
K	0.36	0.0
L	0.09	0.0
M	0.16	0.0
N	0.32	0.0
N	0.29	0.0
O	0.07	0.0
O	0.24	0.0
P	0.26	0.0
Q	0.37	0.0
Q	0.28	0.0
R	0.33	0.0
R	0.15	0.0
R	0.10	0.0
S	0.10	0.0
T	0.30	0.0
U	0.14	0.0
U	0.05	0.0
U	0.05	0.0
V	0.15	0.0
W	0.18	0.0
W	-0.05	0.0
X	0.04	0.0
Y	0.17	0.0
Y	0.14	0.0
Z	-0.09	0.0

Automatic criteria: The validation is OK if the median Q-score is larger than 0.1.

STATUS: OK

6.2 Level A.b FSC-Q

Explanation:
FSC-Q [Ramírez-Aportela et al., 2021] compares the local FSC between the map and the atomic model to the local FSC of the two half maps. FSC-Qr is the normalized version of FSC-Q to facilitate comparisons. Typically, FSC-Qr should take values between -1.5 and 1.5, being 0 an indicator of good matching between map and model.

Results:

Fig. 35 shows the histogram of FSC-Qr and Fig. 36 the colored isosurface of the atomic model converted to map. The average FSC-Qr is 0.82, its 95% confidence interval is [-1.00, 2.91]. The percentage of values whose FSC-Qr absolute value is beyond 1.5 is 10.2 %.

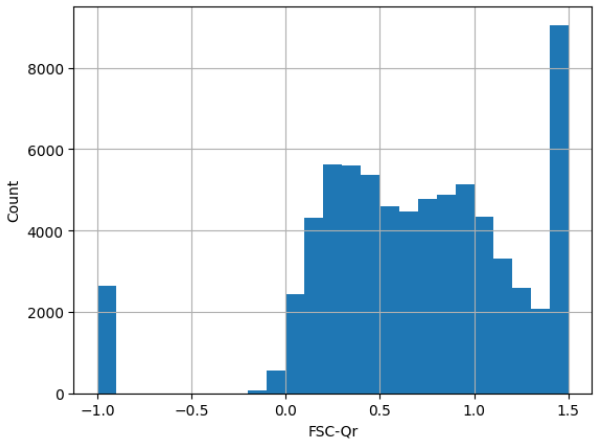
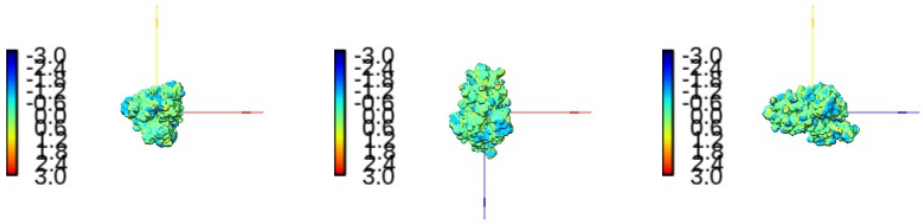


Figure 35: Histogram of the FSC-Qr limited to -1.5 and 1.5.



(a) View 1 (b) View 2 (c) View 3

Figure 36: Isosurface of the atomic model colored by FSC-Qr between -1.5 and 1.5 Views generated by ChimeraX at a the following X, Y, Z angles: View 1 (0,0,0), View 2 (90, 0, 0), View 3 (0, 90, 0).

Automatic criteria: The validation is OK if the percentage of residues whose FSC-Q is larger than 1.5 in absolute value is smaller than 10%.

WARNINGS: 1 warnings

1. **The percentage of voxels that have a FSC-Qr larger than 1.5 in absolute value is 10.2, that is larger than 10%**

6.3 Level A.d Map-Model Guinier analysis

Explanation:
We compared the Guinier plot [Rosenthal and Henderson, 2003] of the atomic model and the experimental map. We made the mean of both profiles to be equal (and equal to the mean of the atomic model) to make sure that they had comparable scales.

Results:
Fig. 37 shows the logarithm (in natural units) of the structure factor (the module squared of the Fourier transform) of the atom model and the experimental map. The correlation between the two profiles was 0.970.

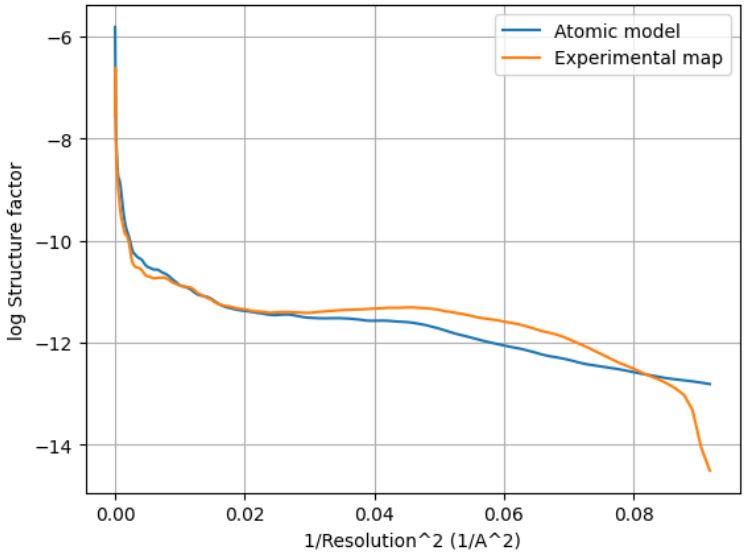


Figure 37: Guinier plot of the atom model and experimental map. The X-axis is the square of the inverse of the resolution in Å.

Automatic criteria: The validation is OK if the correlation between the two Guinier profiles is larger than 0.5.

STATUS: OK

6.4 Level A.e Phenix validation

Explanation:
Phenix provides a number of tools to assess the agreement between the experimental map and its atomic model [Afonine et al., 2018]. There are several cross-correlations to assess the quality of the fitting:

- CC (mask): Model map vs. experimental map correlation coefficient calculated considering map values inside a mask calculated around the macromolecule.
- CC (box): Model map vs. experimental map correlation coefficient calculated considering all grid points of the box.

- CC (volume) and CC (peaks) compare only map regions with the highest density values and regions below a certain contouring threshold level are ignored. CC (volume): The map region considered is defined by the N highest points inside the molecular mask. CC (peaks): In this case, calculations consider the union of regions defined by the N highest peaks in the model-calculated map and the N highest peaks in the experimental map.
- Local real-space correlation coefficients CC (main chain) and CC (side chain) involve the main skeleton chain and side chains, respectively.

There are also multiple ways of measuring the resolution:

- d99: Resolution cutoff beyond which Fourier map coefficients are negligibly small. Calculated from the full map.
- d_model: Resolution cutoff at which the model map is the most similar to the target (experimental) map. For d_model to be meaningful, the model is expected to fit the map as well as possible. d_model (B factors = 0) tries to avoid the blurring of the map.
- d_FSC_model; Resolution cutoff up to which the model and map Fourier coefficients are similar at FSC values of 0, 0.143, 0.5.

In addition to these resolution measurements the overall isotropic B factor is another indirect measure of the quality of the map.

Results:

To avoid ringing in Fourier space a smooth mask with a radius of 6.6 Å has been applied.

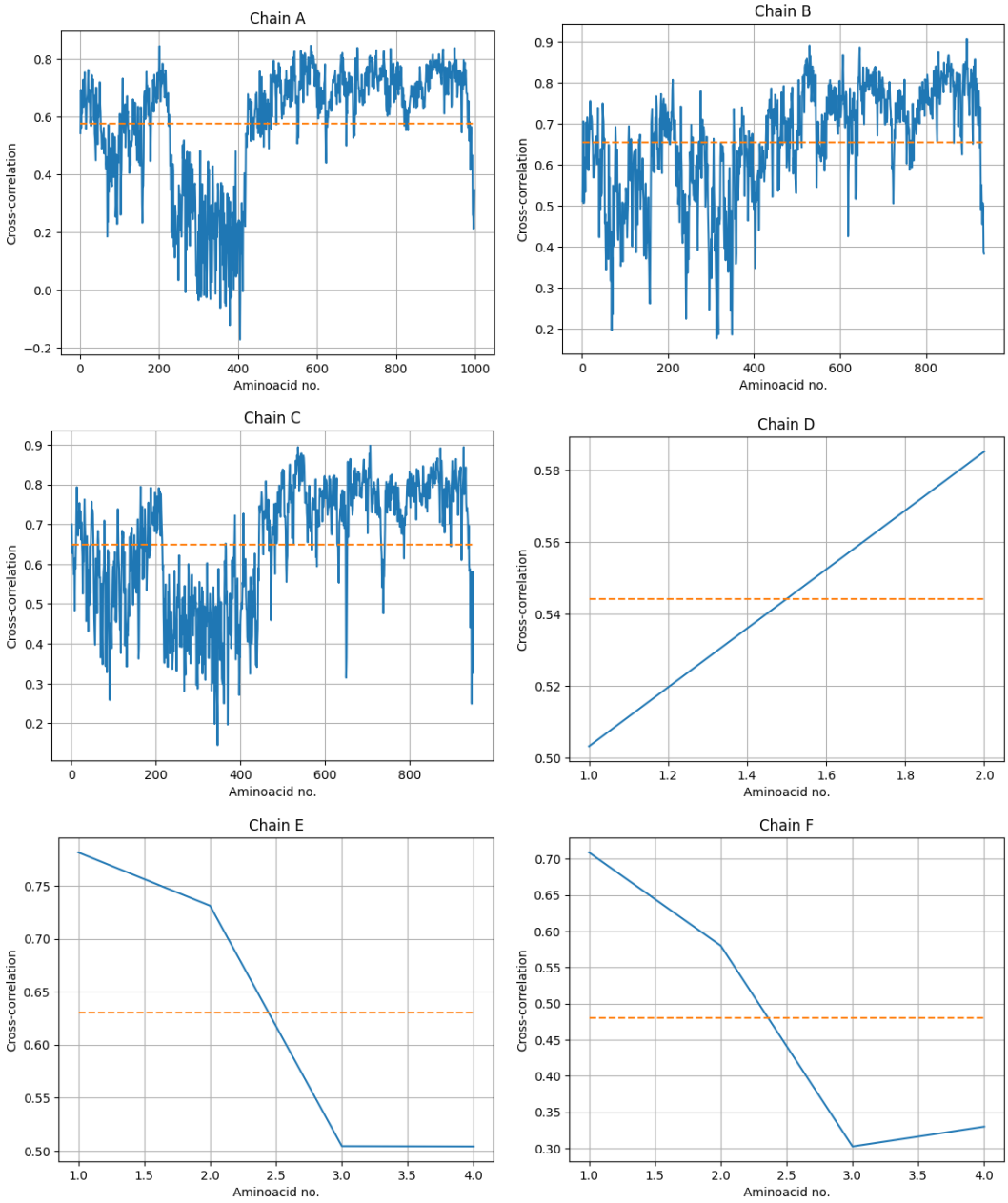
Overall correlation coefficients:

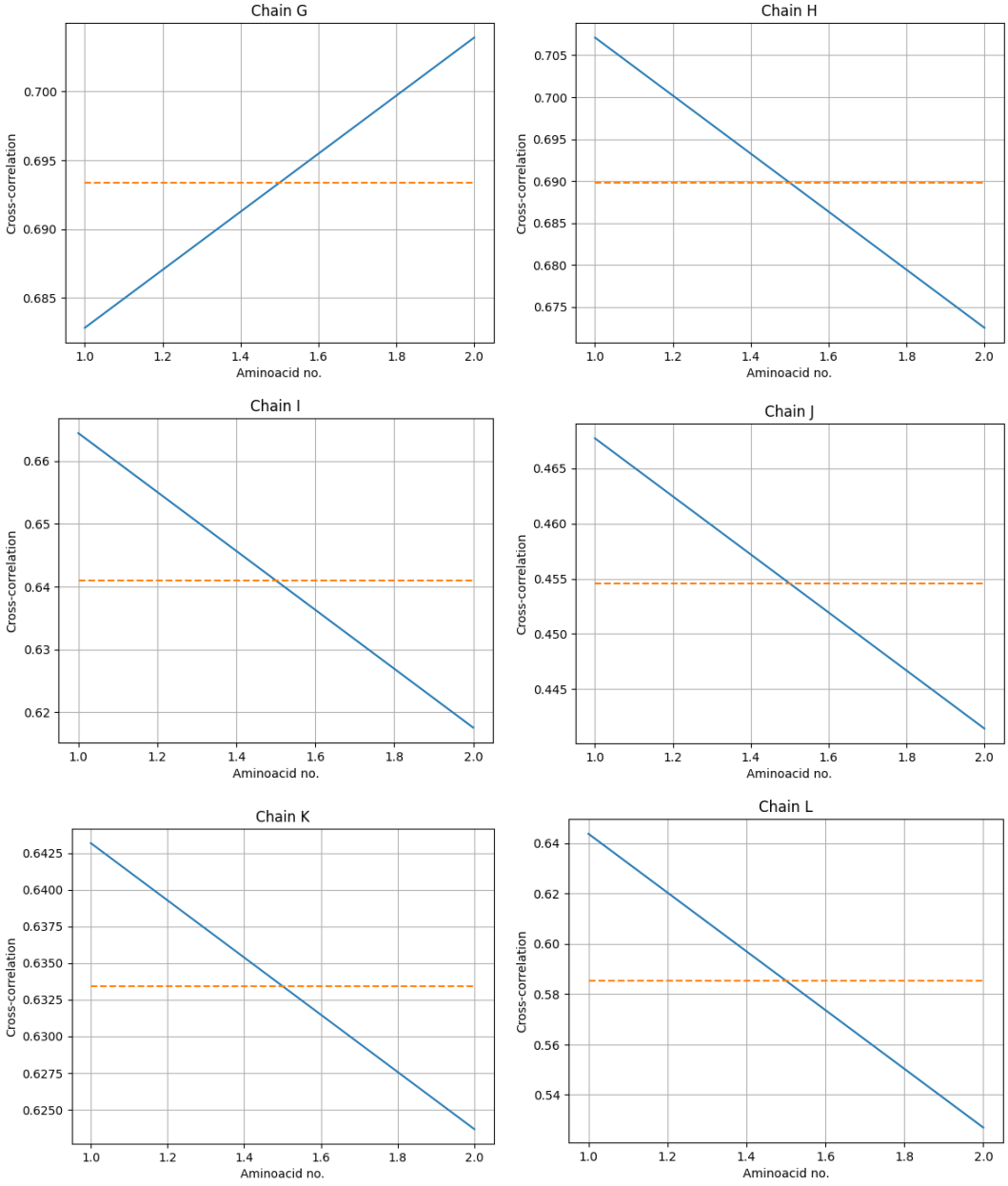
CC (mask) =	0.592
CC (box) =	0.643
CC (volume) =	0.671
CC (peaks) =	0.559
CC (main chain) =	0.611
CC (side chain) =	0.598

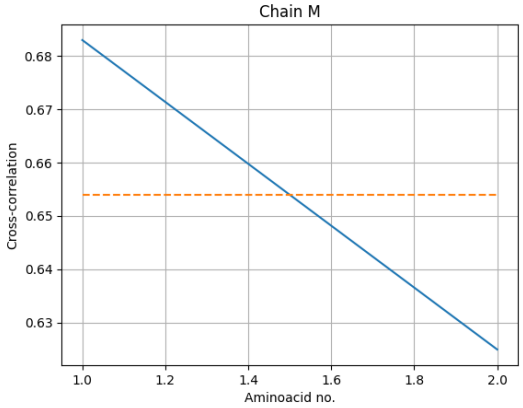
Correlation coefficients per chain:

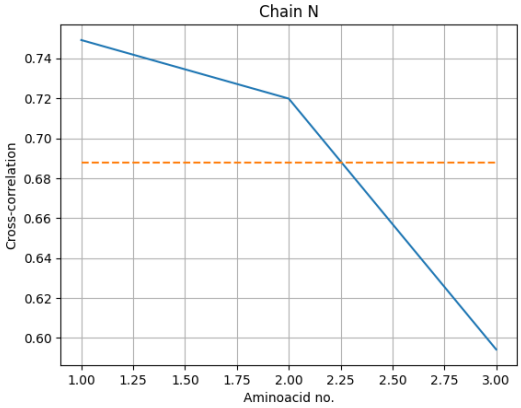
Chain	Cross-correlation
A	0.585339
C	0.627294
B	0.633729
D	0.492978
E	0.629175
F	0.536286
G	0.653338
H	0.679420
I	0.621025
J	0.467733
K	0.622740
L	0.614352
M	0.652587
N	0.658026
O	0.506773
P	0.670467
Q	0.610106
R	0.653794
S	0.446171
T	0.390683
U	0.552064
V	0.576380
W	0.502179
X	0.599782
Y	0.607647
Z	0.604608
a	0.507833
b	0.374672
c	0.414481
d	0.294232

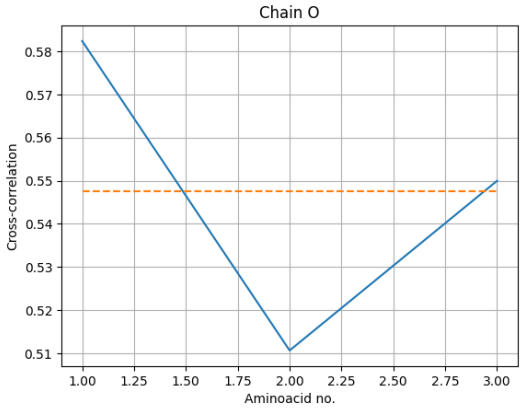
We now show the correlation profiles of the different chain per residue.

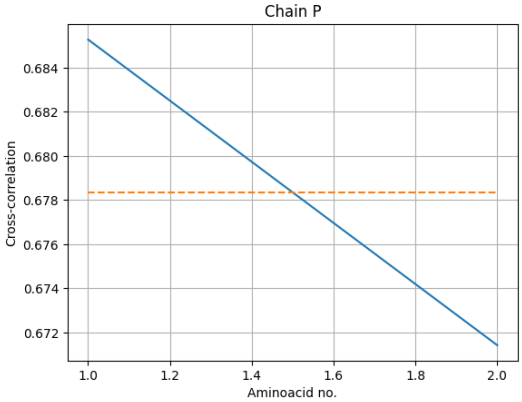

Chain A
Cross-correlation
Aminoacid no.
Chain B
Cross-correlation
Aminoacid no.
Chain C
Cross-correlation
Aminoacid no.
Chain D
Cross-correlation
Aminoacid no.
Chain E
Cross-correlation
Aminoacid no.
Chain F
Cross-correlation
Aminoacid no.

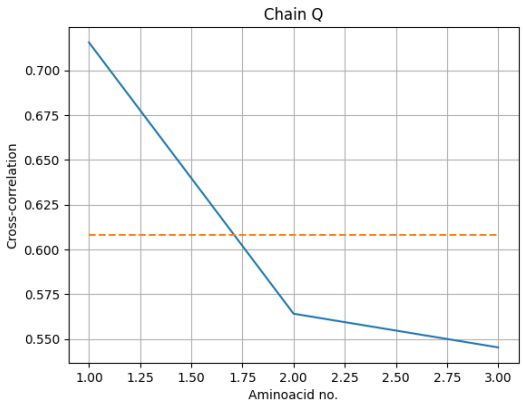

Chain G
Cross-correlation
Aminoacid no.
Chain H
Cross-correlation
Aminoacid no.
Chain I
Cross-correlation
Aminoacid no.
Chain J
Cross-correlation
Aminoacid no.
Chain K
Cross-correlation
Aminoacid no.
Chain L
Cross-correlation
Aminoacid no.

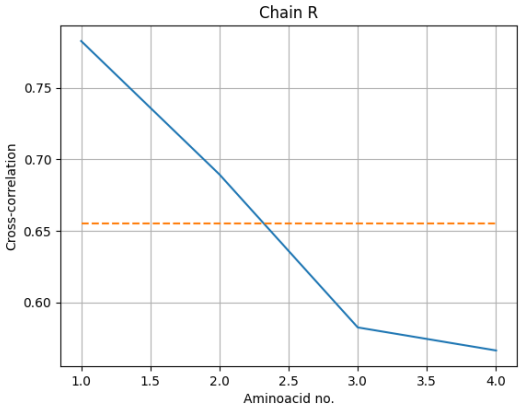

Chain M
Cross-correlation
Aminoacid no.

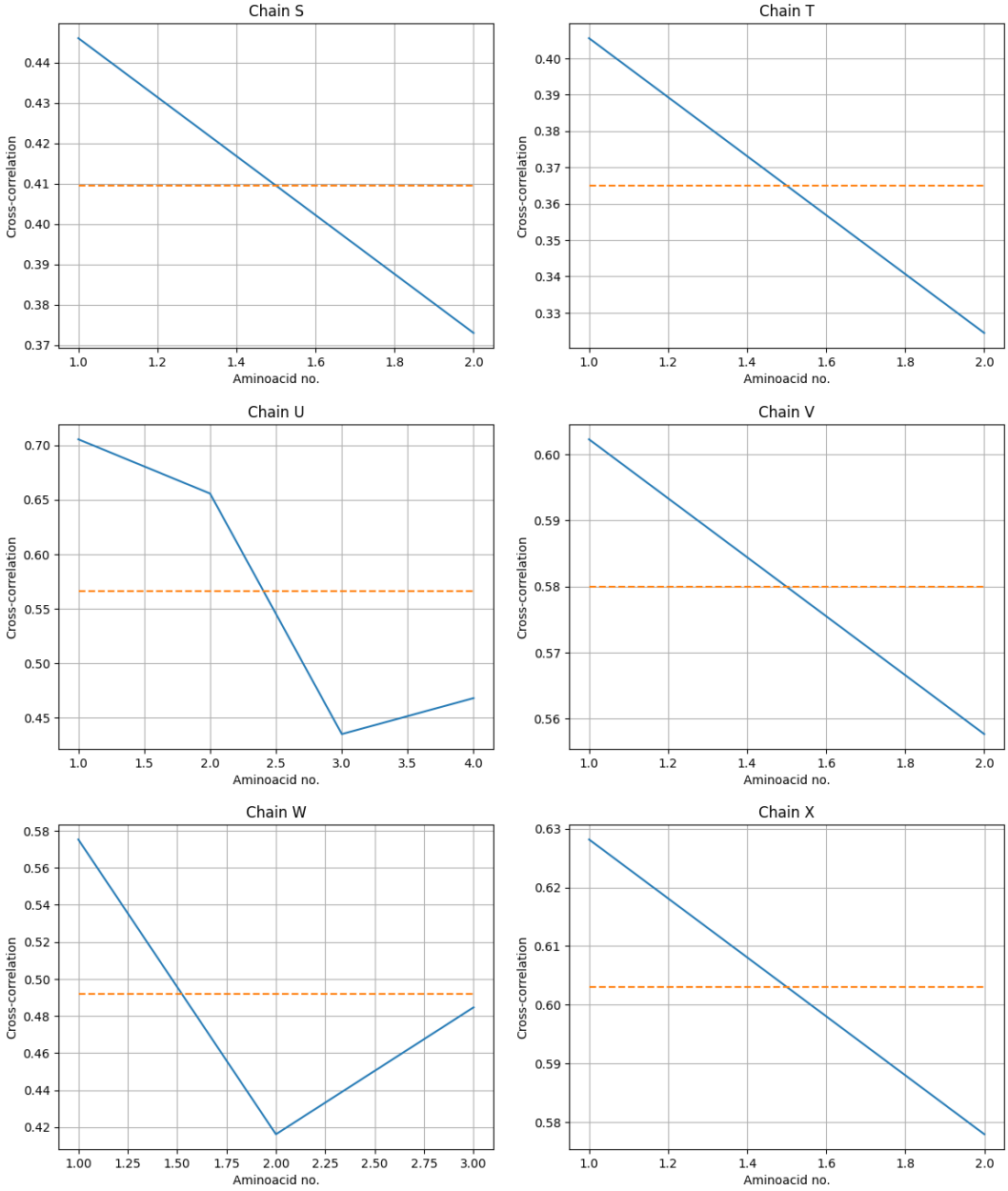

Chain N
Cross-correlation
Aminoacid no.


Chain O
Cross-correlation
Aminoacid no.


Chain P
Cross-correlation
Aminoacid no.


Chain Q
Cross-correlation
Aminoacid no.


Chain R
Cross-correlation
Aminoacid no.


Chain S
Cross-correlation
Aminoacid no.
Chain T
Cross-correlation
Aminoacid no.
Chain U
Cross-correlation
Aminoacid no.
Chain V
Cross-correlation
Aminoacid no.
Chain W
Cross-correlation
Aminoacid no.
Chain X
Cross-correlation
Aminoacid no.

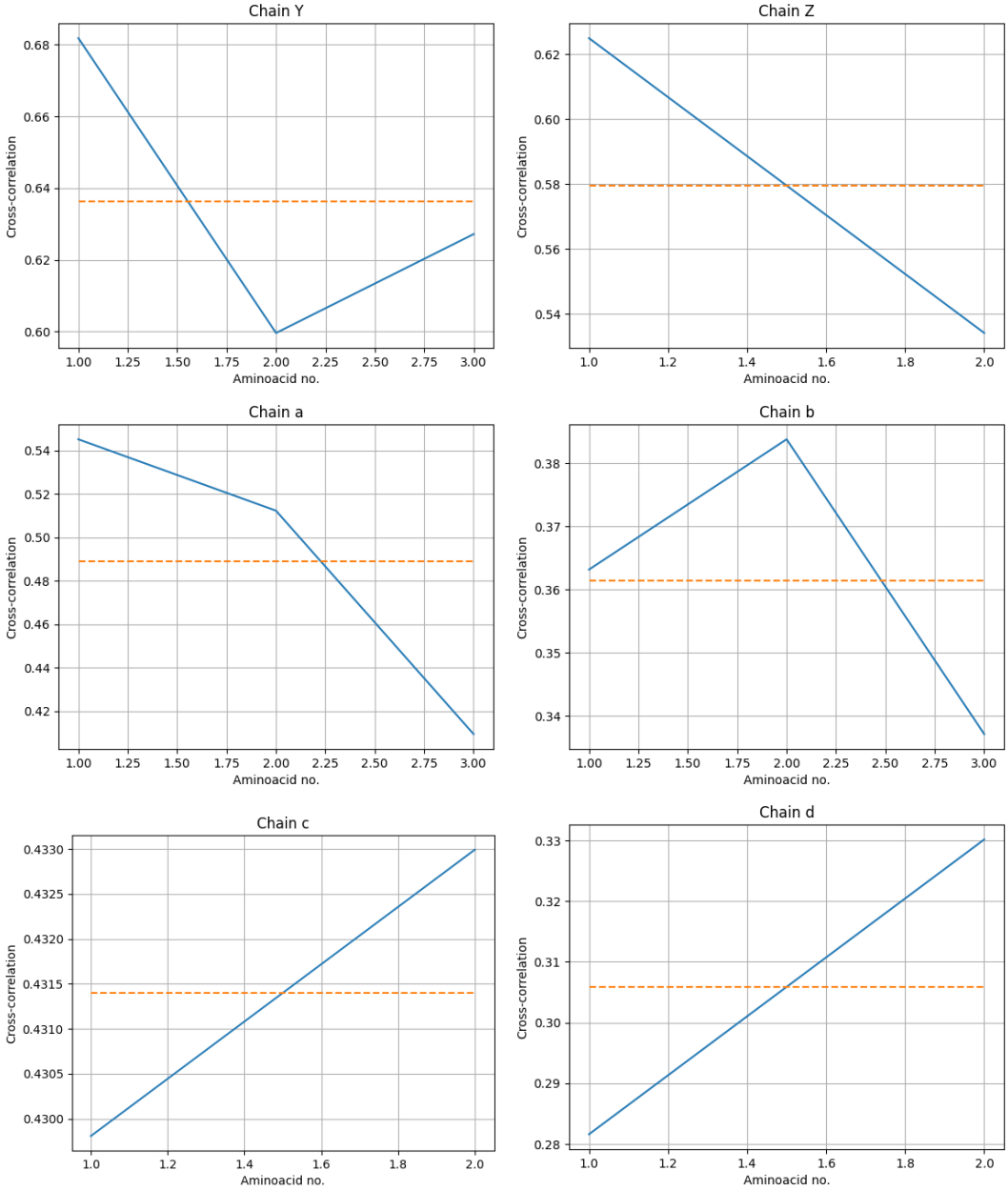


Fig. 38 shows the histogram of all cross-correlations evaluated at the residues. The percentage of residues whose correlation is below 0.5 is 20.8 %.

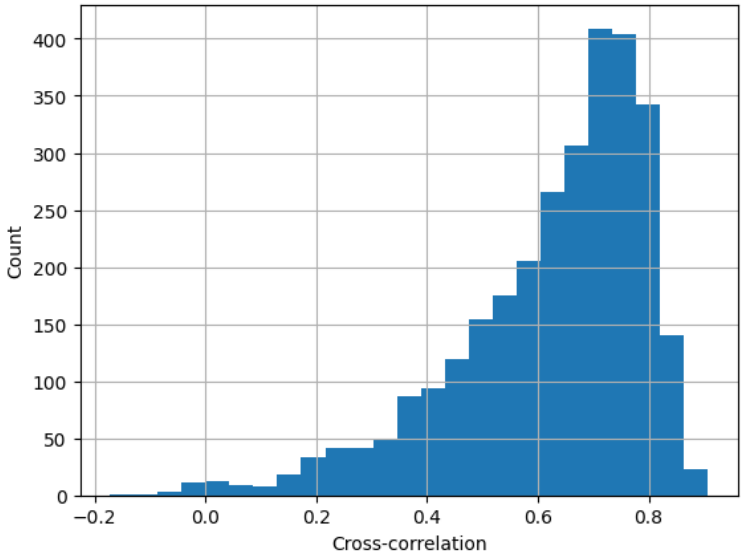


Figure 38: Histogram of the cross-correlation between the map and model evaluated for all residues.

Resolutions estimated from the model:

Resolution (Å)	**Masked**	**Unmasked**
d99	4.0	3.9
d_model	3.7	3.7
d_model (B-factor=0)	6.9	7.1
FSC_model=0	3.3	3.3
FSC_model=0.143	3.4	3.4
FSC_model=0.5	4.0	4.3

Overall isotropic B factor:

B factor	**Masked**	**Unmasked**
Overall B-iso	110.0	115.0

Fig. 39 shows the FSC between the input map and the model.

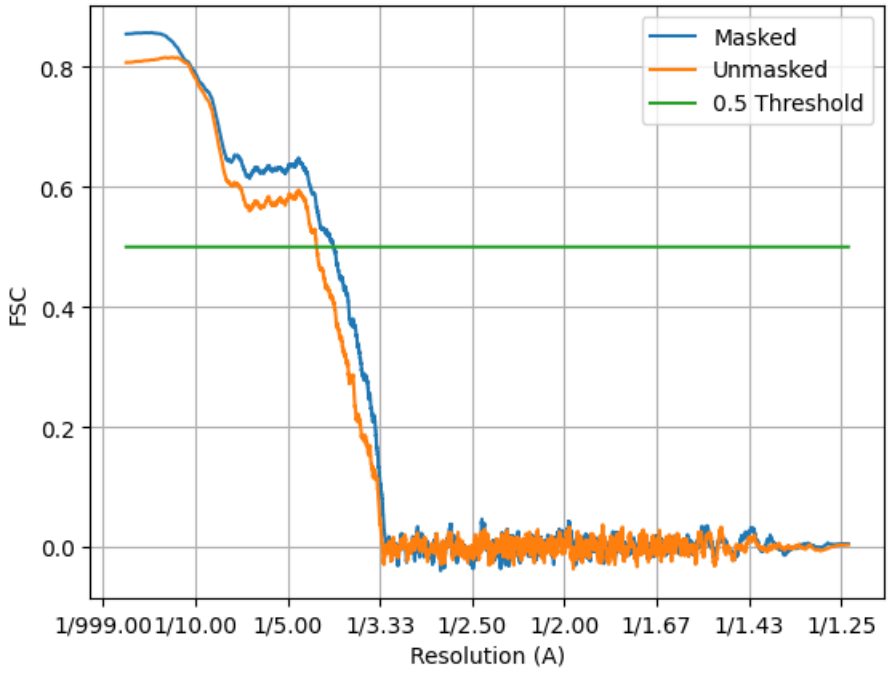


Figure 39: FSC between the input map and model with and without a mask constructed from the model. The X-axis is the square of the inverse of the resolution in Å.

Automatic criteria: The validation is OK if 1) the percentage of residues whose correlation is smaller than 0.5 is smaller than 10%, and 2) the resolution reported by the user is larger than 0.8 times the resolution estimated between the map and model at FSC=0.5.

WARNINGS: 1 warnings

1. **The percentage of residues that have a cross-correlation below 0.5 is 20.8, that is larger than 10%**

6.5 Level A.f EMRinger validation

Explanation:
EMringer [Barad et al., 2015] compares the side chains of the atomic model to the CryoEM map. The following features are reported:

- Optimal Threshold: Electron potential map cutoff value at which the maximum EMRinger score was obtained.
- Rotamer Ratio: Fraction of rotameric residues at the Optimal threshold value.
- Max Zscore: Z-score computed to determine the significance of the distribution at the Optimal threshold value.
- Model Length: Total of non-gamma-branched, non-proline aminoacids with a non-H gamma atom used in global EMRinger score computation.
- EMRinger Score: Maximum EMRinger score calculated at the Optimal Threshold.

A rotameric residue is one in which EMRinger peaks that fall within defined rotamers based on chi1, this often suggests a problem with the modelling of the backbone. In general, the user should look at the profiles and identify regions that may need improvement.

Results:

General results:

Optimal threshold	0.601310
Rotamer ratio	0.713
Max. Zscore	6.12
Model length	1723
EMRinger Score	1.474

Fig. 40 shows the EMRinger score and fraction of rotameric residues as a function of the map threshold. The optimal threshold was selected looking for the maximum EMRinger score in this plot.

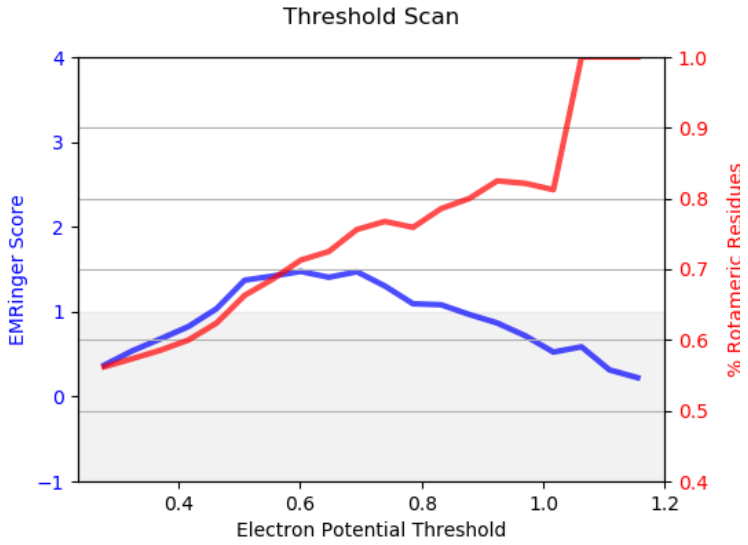


Figure 40: EMRinger score and fraction of rotameric residues as a function of the map threshold.

Fig. 41 shows the histogram for rotameric (blue) and non-rotameric (red) residues at the optimal threshold.

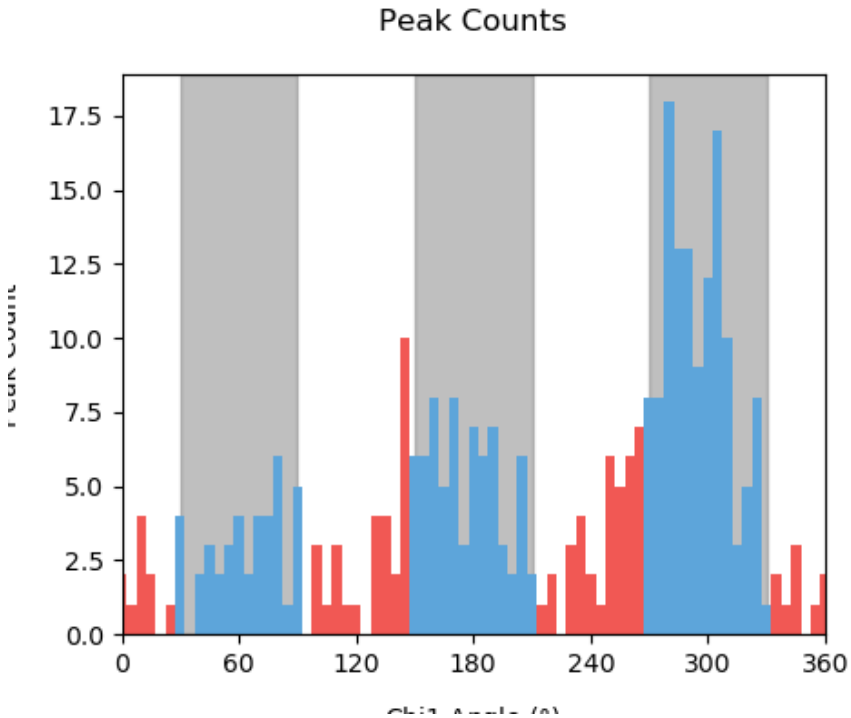
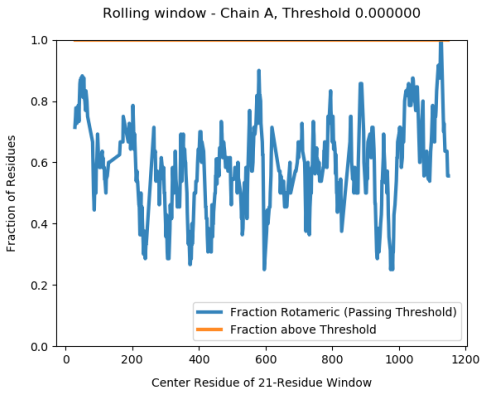
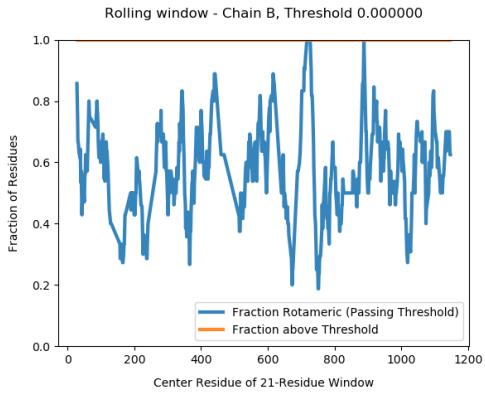
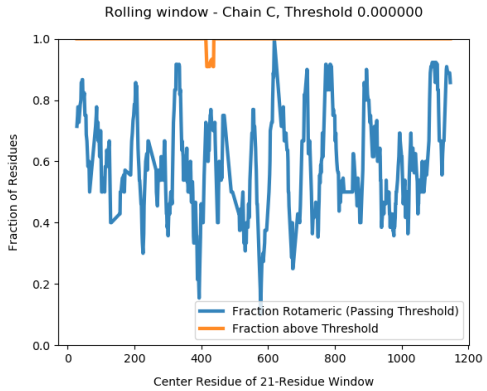


Figure 41: Histogram for rotameric (blue) and non-rotameric (red) residues at the optimal threshold as a function of the angle Chi1.

The following plots show the rolling window EMRinger analysis of the different chains to distinguish regions of improved model quality. This analysis was performed on rolling sliding 21-residue windows along the primary sequence of the protein chains.







Automatic criteria: The validation is OK if the EMRinger score and Max. Zscore are larger than 1.

STATUS: OK

6.6 Level A.g DAQ validation

Explanation:
DAQ [Terashi et al., 2022] is a computational tool using deep learning that can estimate the residue-wise local quality for protein models from cryo-Electron Microscopy maps. The method calculates the likelihood that a given density feature corresponds to an aminoacid, atom, and secondary structure. These likelihoods are combined into a score that ranges from -1 (bad quality) to 1 (good quality).

Results:
Fig. 42 shows the histogram of the DAQ values. The mean and standard deviation were 0.5 and 0.4, respectively.

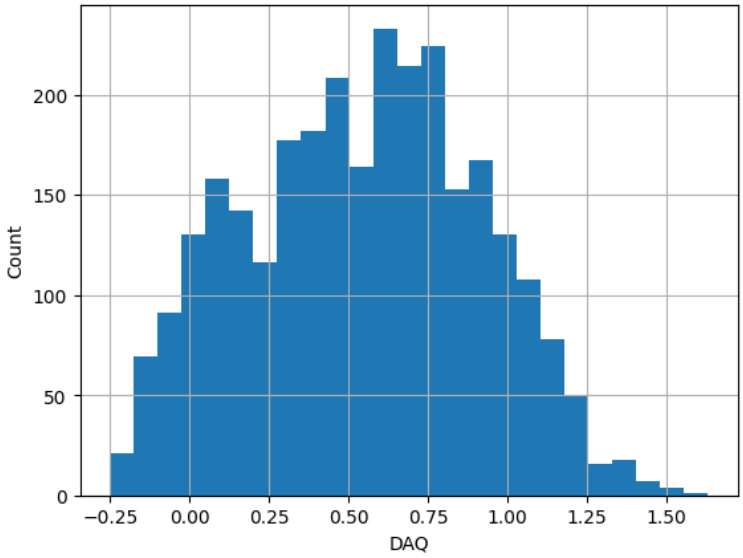
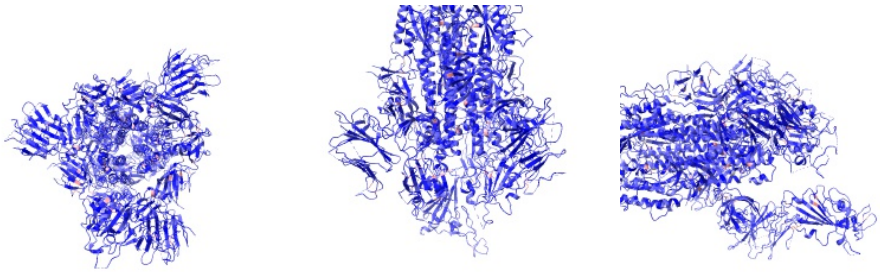


Figure 42: Histogram of the DAQ values.

The atomic model colored by DAQ can be seen in Fig. 43.



(a) View 1 (b) View 2 (c) View 3

Figure 43: Atomic model colored by DAQ Views generated by ChimeraX at a the following X, Y, Z angles: View 1 (0,0,0), View 2 (90, 0, 0), View 3 (0, 90, 0).

Automatic criteria: The validation is OK if the average DAQ score is larger than 0.5.

STATUS: OK

References

[Rosenthal and Henderson, 2003] Rosenthal, P. B. and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single particle electron-cryomicroscopy. *J. Molecular Biology*, 333:721–745.

[Ramírez-Aportela et al., 2019] Ramírez-Aportela, E., Mota, J., Conesa, P., Carazo, J. M., and Sorzano, C. O. S. (2019). DeepRes: a new deep-learning- and aspect-based local resolution method for electron-microscopy maps. *IUCRj*, 6:1054–1063.

[Kaur et al., 2021] Kaur, S., Gomez-Blanco, J., Khalifa, A. A., Adinarayanan, S., Sanchez-Garcia, R., Wrapp, D., McLellan, J. S., Bui, K. H., and Vargas, J. (2021). Local computational methods to improve the interpretability and analysis of cryo-EM maps. *Nature Communications*, 12(1):1–12.

[Sorzano et al., 2017] Sorzano, C. O. S., Vargas, J., Oton, J., Abrishami, V., de la Rosa-Trevin, J. M., Gomez-Blanco, J., Vilas, J. L., Marabini, R., and Carazo, J. M. (2017). A review of resolution measures and related aspects in 3D electron microscopy. *Progress in biophysics and molecular biology*, 124:1–30.

[Beckers and Sachse, 2020] Beckers, M. and Sachse, C. (2020). Permutation testing of fourier shell correlation for resolution estimation of cryo-em maps. *J. Structural Biology*, 212(1):107579.

[Cardone et al., 2013] Cardone, G., Heymann, J. B., and Steven, A. C. (2013). One number does not fit all: Mapping local variations in resolution in cryo-em reconstructions. *J. Structural Biology*, 184:226–236.

[Kucukelbir et al., 2014] Kucukelbir, A., Sigworth, F. J., and Tagare, H. D. (2014). Quantifying the local resolution of cryo-EM density maps. *Nature Methods*, 11:63–65.

[Vilas et al., 2018] Vilas, J. L., Gómez-Blanco, J., Conesa, P., Melero, R., de la Rosa Trevín, J. M., Otón, J., Cuenca, J., Marabini, R., Carazo, J. M., Vargas, J., and Sorzano, C. O. S. (2018). MonoRes: automatic and unbiased estimation of local resolution for electron microscopy maps. *Structure*, 26:337–344.

[Vilas et al., 2020] Vilas, J. L., Tagare, H. D., Vargas, J., Carazo, J. M., and Sorzano, C. O. S. (2020). Measuring local-directional resolution and local anisotropy in cryo-EM maps. *Nature communications*, 11:55.

[Pintilie et al., 2020] Pintilie, G., Zhang, K., Su, Z., Li, S., Schmid, M. F., and Chiu, W. (2020). Measurement of atom resolvability in cryo-em maps with q-scores. *Nature methods*, 17(3):328–334.

[Ramírez-Aportela et al., 2021] Ramírez-Aportela, E., Maluenda, D., Fonseca, Y. C., Conesa, P., Marabini, R., Heymann, J. B., Carazo, J. M., and Sorzano, C. O. S. (2021). Fsc-q: A cryoem map-to-atomic model quality validation based on the local fourier shell correlation. *Nature Communications*, 12(1):1–7.

[Afonine et al., 2018] Afonine, P. V., Klaholz, B. P., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., Adams, P. D., and Urzhumtsev, A. (2018). New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallographica D, Struct. Biol.*, 74:814–840.

[Barad et al., 2015] Barad, B. A., Echols, N., Wang, R. Y.-R., Cheng, Y., DiMaio, F., Adams, P. D., and Fraser, J. S. (2015). EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nature Methods*, 12(10):943–946.

[Terashi et al., 2022] Terashi, G., Wang, X., Subramaniya, S.R.M.V., Tesmer, J.J.G. and Kihara, D. (2022). Residue-Wise Local Quality Estimation for Protein Models from Cryo-EM Maps. (submitted).