Electronic Supplementary Material (ESI) for Faraday Discussions.


# Validation report of Level(s) 0, A

I$^2$PC Validation server

March 2, 2022
6:09am

**Abstract**

The map seems to be well centered. 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 B-factor (see Sec. 2.4). There seems to be a problem with its MapQ scores (see Sec. 4.1). It seems that the Guinier plot of the map and its model do not match (see Sec. 4.2). According to phenix, it seems that there might be some mismatch between the map and its model (see Sec. 4.3). DAQ detects some mismatch between the map and its model (see Sec. 4.5).

The average resolution of the map estimated by various methods goes from 1.2Å to 6.6Å with an average of 3.9Å. The resolution provided by the user was 1.1Å. The resolution reported by the user may be overestimated.

**The overall score (passing tests) of this report is 3 out of 10 evaluable items.**

| | | |
|---|---|---|
| 0.a Mass analysis | Sec. 2.1 | OK |
| 0.b Mask analysis | Sec. 2.2 | 2 warnings |
| 0.c Background analysis | Sec. 2.3 | 2 warnings |
| 0.d B-factor analysis | Sec. 2.4 | 1 warnings |
| 0.e DeepRes | Sec. 2.5 | Does not apply |
| 0.f LocBfactor | Sec. 2.6 | Could not be measured |
| 0.g LocOccupancy | Sec. 2.7 | Could not be measured |
| 0.h DeepHand | Sec. 2.8 | OK |
| A.a MapQ | Sec. 4.1 | 1 warnings |
| A.d Map-Model Guinier | Sec. 4.2 | 1 warnings |
| A.e Phenix validation | Sec. 4.3 | 1 warnings |
| A.f EMRinger | Sec. 4.4 | OK |
| A.g DAQ | Sec. 4.5 | 1 warnings |

**Summary of the warnings across sections.**

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

Section 2.2 (0.b Mask analysis)

1. **There might be a problem of connectivity at this threshold because more than 5 connected components are needed to reach 95% of the total mask.**
2. **There might be a problem in the construction of the mask, because the overlap is smaller than 0.75. A common reason is that the suggested threshold causes too many disconnected components.**

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=1081.35**)

Section 2.4 (0.d B-factor analysis)

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

Section 4.1 (A.a MapQ)

1. **The median Q-score is less than 0.1.**

Section 4.2 (A.d Map-Model Guinier)

1. **The correlation is smaller than 0.5, it is 0.497.**

Section 4.3 (A.e Phenix validation)

1. **The percentage of residues that have a cross-correlation below 0.5 is 100.0, that is larger than 10%**

Section 4.5 (A.g DAQ)

1. **The average DAQ is smaller than 0.5.**

# Contents

**1 Input data** **5**

**2 Level 0 analysis** **8**
2.1 Level 0.a Mass analysis . . . . . 8
2.2 Level 0.b Mask analysis . . . . . 9
2.3 Level 0.c Background analysis . . . . . 13
2.4 Level 0.d B-factor analysis . . . . . 14
2.5 Level 0.e Local resolution with DeepRes . . . . . 16
2.6 Level 0.f Local B-factor . . . . . 16
2.7 Level 0.g Local Occupancy . . . . . 16
2.8 Level 0.h Hand correction . . . . . 17

**3 Atomic model** **17**

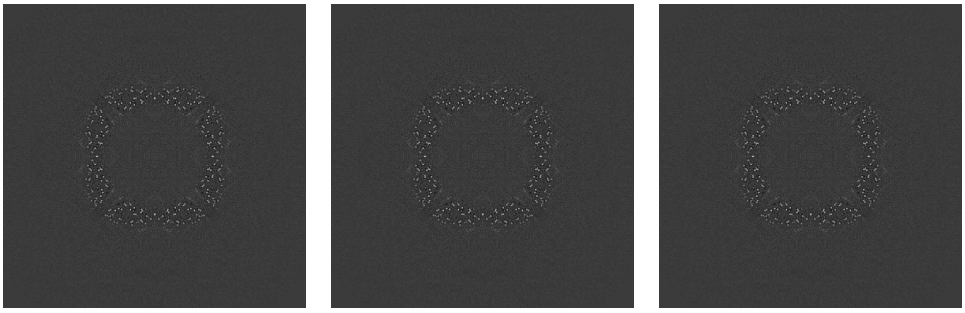
**4 Level A analysis** **18**
4.1 Level A.a MapQ . . . . . 18
4.2 Level A.d Map-Model Guinier analysis . . . . . 21
4.3 Level A.e Phenix validation . . . . . 22
4.4 Level A.f EMRinger validation . . . . . 30
4.5 Level A.g DAQ validation . . . . . 37

# 1 Input data

Input map: /home/coss/data/Dropbox/Aplicaciones/ShareLaTeX/MapValidation/-EMDB11668/emd_11668.map
SHA256 hash: 69a72c5b39bb0573f60a4289b4e17063ebd26cee331f5f018d153aa06f184813
Voxel size: 0.492000 (Å)
Visualization threshold: 0.050000
Resolution estimated by user: 1.150000

**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, ...)

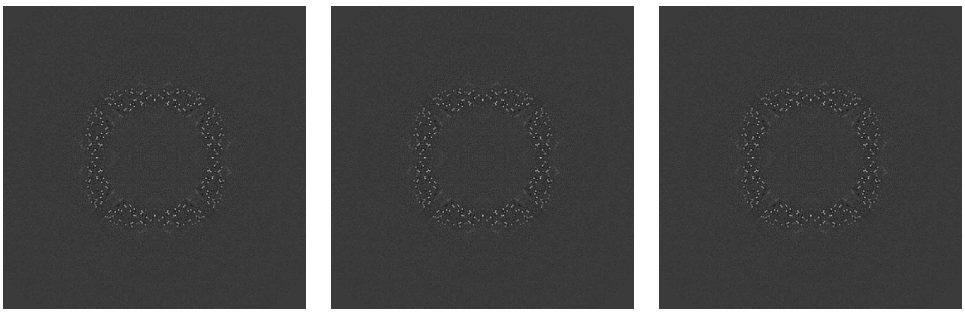
**Results**:
See Fig. 1.



(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.



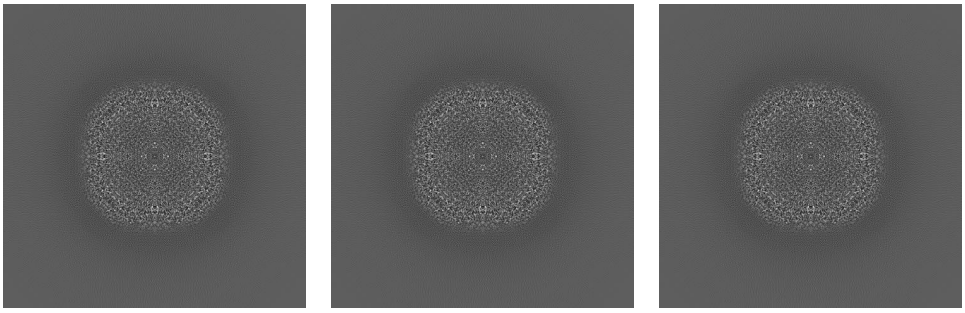
(a) X Slice 240 (b) Y Slice 240 (c) Z Slice 240

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

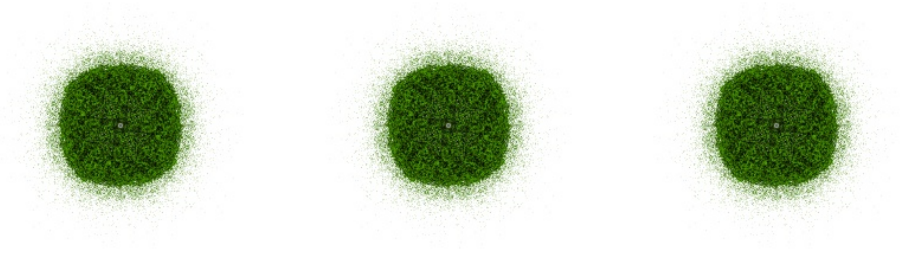
Figure 3: Projections in the three dimensions

**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.



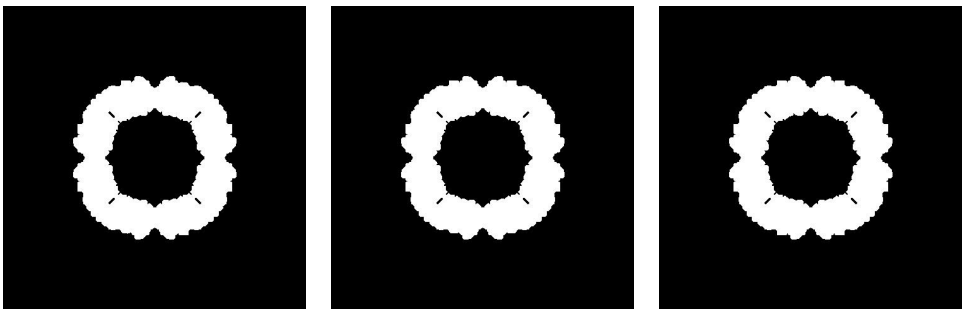
(a) View 1 (b) View 2 (c) View 3

Figure 4: Isosurface at threshold=0.050000. 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 0.050000, the largest connected component was selected, and then dilated by 2Å.

**Results**:
See Fig. 5.



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

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 54.61 and 54.61 Å, respectively. There is a decentering ratio (abs(Right-Left)/Size)% of 0.00%

The space from the left and right in Y are 54.61 and 54.61 Å, respectively. There is a decentering ratio (abs(Right-Left)/Size)% of 0.00%

The space from the left and right in Z are 54.61 and 54.61 Å, respectively. There is a decentering ratio (abs(Right-Left)/Size)% of 0.00%

The center of mass is at (x,y,z)=(240.51,240.51,240.51). The decentering of the center of mass (abs(Center)/Size)% is 0.11, 0.11, and 0.11, 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 0.050000, there are 478289 connected components with a total number of voxels of 1531522 and a volume of 182397.36 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 or the first 100 clusters, whatever happens first), the list contains (No. voxels (volume in Å$^3$), percentage, cumulatedPercentage):

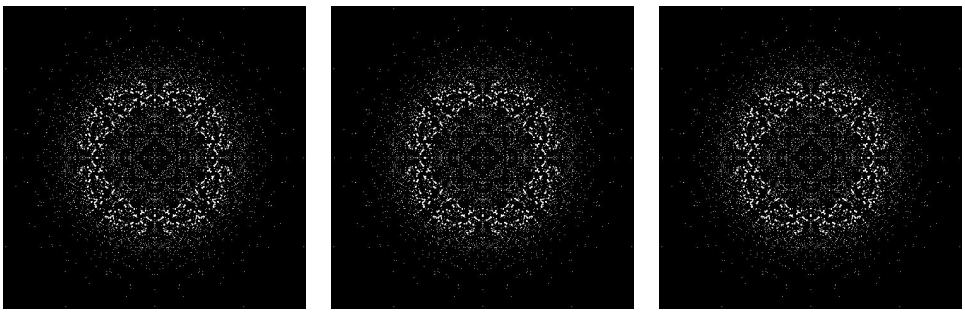
, (650384 (77457.80), 42.47, 42.47), (106 (12.62), 0.01, 42.47), (106 (12.62), 0.01, 42.48), (106 (12.62), 0.01, 42.49), (106 (12.62), 0.01, 42.49), (106 (12.62), 0.01, 42.50), (106 (12.62), 0.01, 42.51), (106 (12.62), 0.01, 42.51), (106 (12.62), 0.01, 42.52), (106 (12.62), 0.01, 42.53), (106 (12.62), 0.01, 42.54), (106 (12.62), 0.01, 42.54), (106 (12.62), 0.01, 42.55), (85 (10.12), 0.01, 42.56), (85 (10.12), 0.01, 42.56), (85 (10.12), 0.01, 42.57), (85 (10.12), 0.01, 42.57), (85 (10.12), 0.01, 42.58), (85 (10.12), 0.01, 42.58), (85 (10.12), 0.01, 42.59), (85 (10.12), 0.01, 42.59), (68 ( 8.10), 0.00, 42.60), (68 ( 8.10), 0.00, 42.60), (68 ( 8.10), 0.00, 42.61), (68 ( 8.10), 0.00, 42.61), (68 ( 8.10), 0.00, 42.62), (68 ( 8.10), 0.00, 42.62), (68 ( 8.10), 0.00, 42.63), (68 ( 8.10), 0.00, 42.63), (68 ( 8.10), 0.00, 42.63), (68 ( 8.10), 0.00, 42.64), (68 ( 8.10),

0.00, 42.64), (68 ( 8.10), 0.00, 42.65), (68 ( 8.10), 0.00, 42.65), (68 ( 8.10), 0.00, 42.66), (68 ( 8.10), 0.00, 42.66), (68 ( 8.10), 0.00, 42.67), (68 ( 8.10), 0.00, 42.67), (68 ( 8.10), 0.00, 42.67), (68 ( 8.10), 0.00, 42.68), (68 ( 8.10), 0.00, 42.68), (68 ( 8.10), 0.00, 42.69), (68 ( 8.10), 0.00, 42.69), (68 ( 8.10), 0.00, 42.70), (68 ( 8.10), 0.00, 42.70), (65 ( 7.74), 0.00, 42.70), (65 ( 7.74), 0.00, 42.71), (65 ( 7.74), 0.00, 42.71), (65 ( 7.74), 0.00, 42.72), (65 ( 7.74), 0.00, 42.72), (65 ( 7.74), 0.00, 42.73), (65 ( 7.74), 0.00, 42.73), (65 ( 7.74), 0.00, 42.73), (65 ( 7.74), 0.00, 42.74), (65 ( 7.74), 0.00, 42.74), (65 ( 7.74), 0.00, 42.75), (65 ( 7.74), 0.00, 42.75), (65 ( 7.74), 0.00, 42.76), (65 ( 7.74), 0.00, 42.76), (65 ( 7.74), 0.00, 42.76), (65 ( 7.74), 0.00, 42.77), (65 ( 7.74), 0.00, 42.77), (65 ( 7.74), 0.00, 42.78), (65 ( 7.74), 0.00, 42.78), (65 ( 7.74), 0.00, 42.79), (65 ( 7.74), 0.00, 42.79), (65 ( 7.74), 0.00, 42.79), (65 ( 7.74), 0.00, 42.80), (65 ( 7.74), 0.00, 42.80), (54 ( 6.43), 0.00, 42.81), (54 ( 6.43), 0.00, 42.81), (54 ( 6.43), 0.00, 42.81), (54 ( 6.43), 0.00, 42.82), (54 ( 6.43), 0.00, 42.82), (54 ( 6.43), 0.00, 42.82), (54 ( 6.43), 0.00, 42.83), (54 ( 6.43), 0.00, 42.83), (54 ( 6.43), 0.00, 42.83), (54 ( 6.43), 0.00, 42.84), (54 ( 6.43), 0.00, 42.84), (54 ( 6.43), 0.00, 42.84), (54 ( 6.43), 0.00, 42.85), (54 ( 6.43), 0.00, 42.85), (54 ( 6.43), 0.00, 42.86), (54 ( 6.43), 0.00, 42.86), (54 ( 6.43), 0.00, 42.86), (54 ( 6.43), 0.00, 42.87), (54 ( 6.43), 0.00, 42.87), (54 ( 6.43), 0.00, 42.87), (54 ( 6.43), 0.00, 42.88), (54 ( 6.43), 0.00, 42.88), (54 ( 6.43), 0.00, 42.88), (54 ( 6.43), 0.00, 42.89), (53 ( 6.31), 0.00, 42.89), (53 ( 6.31), 0.00, 42.89), (53 ( 6.31), 0.00, 42.90), (53 ( 6.31), 0.00, 42.90), (53 ( 6.31), 0.00, 42.90), (53 ( 6.31), 0.00, 42.91)

Number of components to reach 95% of the mass: 401714

The average size of the remaining 76575 components is 1.00 voxels ( 0.12 Å$^3$). Their size go from 1 voxels ( 0.12 Å$^3$) to 1 voxels ( 0.12 Å$^3$).

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



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

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** |
|---|---|---|---|
| 0.0422 | 2333120.00 | 230.21 | 2092.82 |
| 0.0845 | 536788.00 | 52.97 | 481.50 |
| 0.1267 | 352884.00 | 34.82 | 316.54 |
| 0.1690 | 266092.00 | 26.26 | 238.69 |
| 0.2112 | 202304.00 | 19.96 | 181.47 |
| 0.2535 | 155937.00 | 15.39 | 139.88 |
| 0.2957 | 117288.00 | 11.57 | 105.21 |
| 0.3380 | 86874.00 | 8.57 | 77.93 |
| 0.3802 | 63930.00 | 6.31 | 57.35 |
| 0.4224 | 45810.00 | 4.52 | 41.09 |
| 0.4647 | 32418.00 | 3.20 | 29.08 |
| 0.5069 | 21234.00 | 2.10 | 19.05 |
| 0.5492 | 13692.00 | 1.35 | 12.28 |
| 0.5914 | 8316.00 | 0.82 | 7.46 |
| 0.6337 | 5253.00 | 0.52 | 4.71 |
| 0.6759 | 2748.00 | 0.27 | 2.46 |
| 0.7181 | 1308.00 | 0.13 | 1.17 |
| 0.7604 | 630.00 | 0.06 | 0.57 |
| 0.8026 | 246.00 | 0.02 | 0.22 |
| 0.8449 | 102.00 | 0.01 | 0.09 |
| 0.8871 | 54.00 | 0.01 | 0.05 |
| 0.9294 | 54.00 | 0.01 | 0.05 |
| 0.9716 | 24.00 | 0.00 | 0.02 |
| 1.0139 | 24.00 | 0.00 | 0.02 |

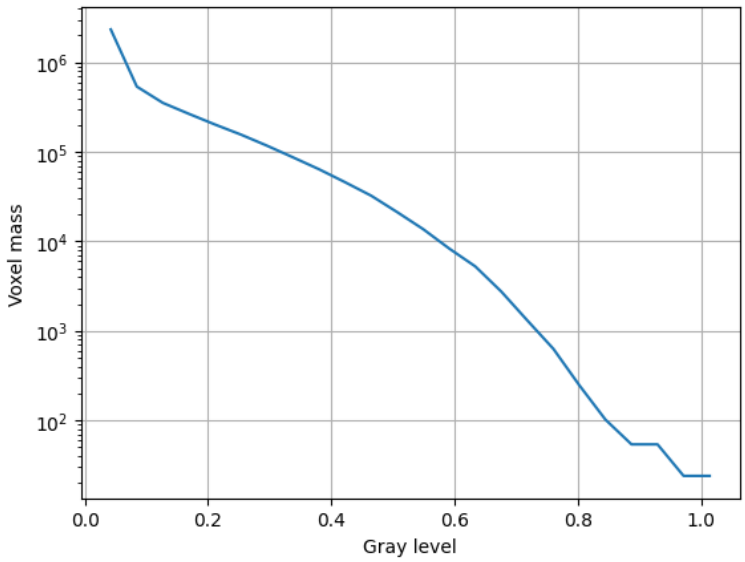



Figure 7: Voxel mass as a function of the gray level.

Constructed mask: After keeping the largest component of the previous mask and dilating it by 2Å, there is a total number of voxels of 7330938 and a volume of 873081.64 Å$^3$. The overlap between the raw and constructed mask is 0.60.

**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**: 2 warnings

1. **There might be a problem of connectivity at this threshold because more than 5 connected components are needed to reach 95% of the total mask.**
2. **There might be a problem in the construction of the mask, because the overlap is smaller than 0.75. A common reason is that the suggested threshold causes too many disconnected components.**

## 2.3 Level 0.c Background analysis

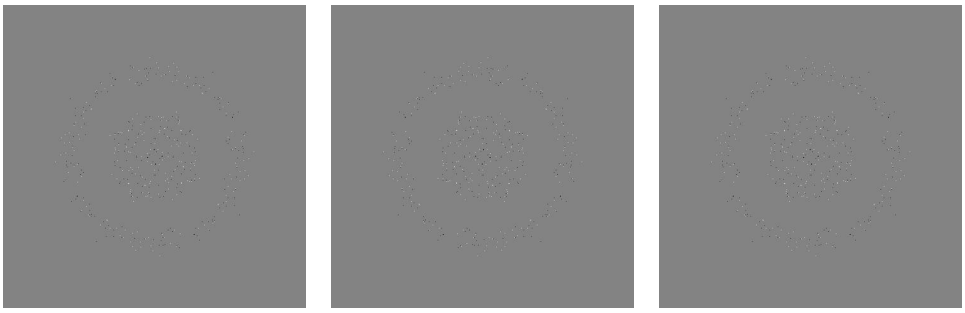
**Explanation:**
Background is defined as the region outside the macromolecule mask. The background mean should be zero, and the number of voxels with a very low or very high value (below 5 standard deviations of the noise) should be very small and they should be randomly distributed without any specific 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 -203.22 and 0.000000, respectively.

The mean and standard deviation (sigma) of the background were -0.000338 and 0.016924. The percentage of background voxels whose absolute value is larger than 5 times the standard deviation is 0.06 % (see Fig. 8). The same percentage from a Gaussian would be 0.000057% (ratio between the two percentages: 1081.353334).

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



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

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=1081.35)**

## 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 0.1. The fitted line was $\log(|F|^2) = 0.0/R^2 + (-12.6)$.

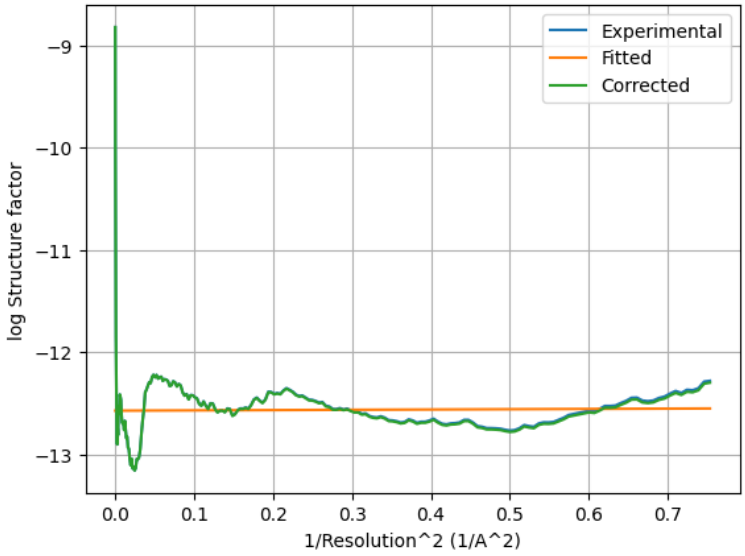
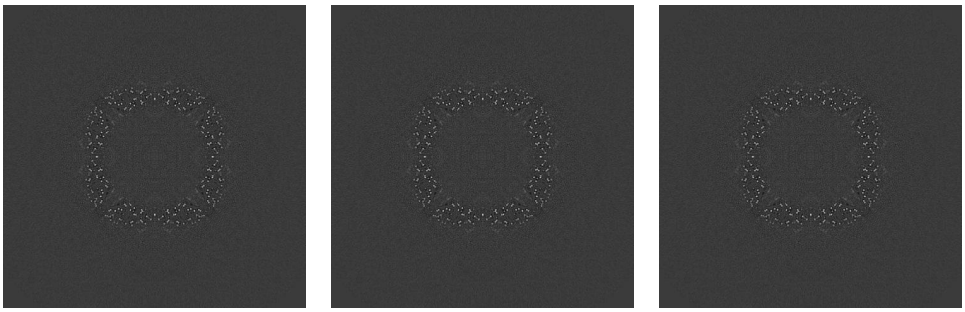



Figure 9: Guinier plot. The X-axis is the square of the inverse of the resolution in Å.



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

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].

**WARNINGS**: 1 warnings

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

## 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:**

This method cannot be applied to maps with a resolution better than 2Å.

## 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:**

**ERROR: The protocol failed.**

## 2.7 Level 0.g Local Occupancy

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

**Results:**

**ERROR: The protocol failed.**

### 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:**

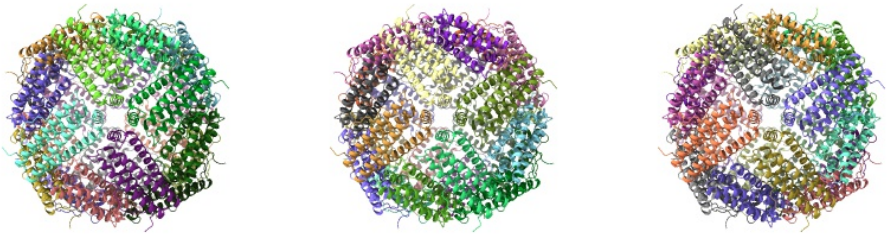
Deep hand assigns a score of 0.327 to the input volume.
**Automatic criteria**: The validation is OK if the deep hand score is smaller than 0.5.

**STATUS**: OK

## 3 Atomic model

Atomic model: /home/coss/data/Dropbox/Aplicaciones/ShareLaTeX/MapValidation/-EMDB11668/7a6a_updated.cif

See Fig. 11.



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

Figure 11: 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).

# 4 Level A analysis

## 4.1 Level A.a MapQ

**Explanation**:
MapQ [Pintilie et al., 2020] computes the local correlation between the map and each one of its atoms assumed to have a Gaussian shape.

**Results:**

Fig. 12 shows the histogram of the Q-score according calculated by MapQ. Some representative percentiles are:

| **Percentile** | **MapQ score [0-1]** |
|---|---|
| 2.5% | -0.23 |
| 25% | 0.00 |
| 50% | 0.00 |
| 75% | 0.00 |
| 97.5% | 0.94 |

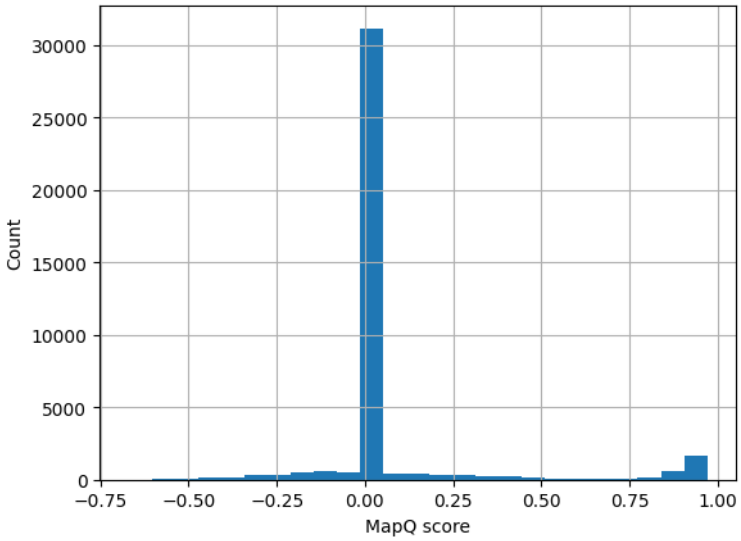



Figure 12: Histogram of the Q-score.

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

| Chain | Average Q score [0-1] | Estimated Resol. (Å) |
|---|---|---|
| 1 | 0.00 | 11.2 |
| 1 | 0.00 | 9.8 |
| 1 | 0.13 | 5.6 |
| 1 | 0.00 | 0.0 |
| 2 | 0.01 | 11.1 |
| 2 | 0.05 | 9.4 |
| 2 | 0.12 | 5.6 |
| 2 | -0.09 | 0.0 |
| 4 | 0.00 | 11.1 |
| 4 | 0.00 | 9.8 |
| 4 | 0.13 | 5.6 |
| 4 | 0.05 | 0.0 |
| 6 | 0.00 | 11.2 |
| 6 | 0.00 | 9.8 |
| 6 | 0.13 | 5.6 |
| 6 | 0.00 | 0.0 |
| A | 0.00 | 11.2 |
| A | 0.00 | 9.8 |
| A | 0.13 | 5.6 |
| A | 0.00 | 0.0 |
| B | 0.00 | 11.2 |
| B | 0.00 | 9.8 |
| B | 0.13 | 5.6 |
| B | 0.00 | 0.0 |
| E | 0.00 | 11.2 |
| E | 0.00 | 9.8 |
| E | 0.13 | 5.6 |
| E | 0.00 | 0.0 |
| F | 0.00 | 11.2 |
| F | 0.00 | 9.8 |
| F | 0.13 | 5.6 |
| F | 0.00 | 0.0 |
| G | -0.00 | 11.2 |
| G | 0.00 | 9.8 |
| G | 0.13 | 5.6 |
| G | 0.01 | 0.0 |
| H | 0.00 | 11.2 |
| H | 0.00 | 9.8 |
| H | 0.13 | 5.6 |
| H | 0.00 | 0.0 |
| I | -0.00 | 11.2 |
| I | -0.05 | 10.3 |
| I | 0.13 | 5.6 |
| I | -0.11 | 0.0 |
| K | -0.00 | 11.2 |

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

**WARNINGS**: 1 warnings

1. **The median Q-score is less than 0.1.**

## 4.2 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. 13 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 0.497.

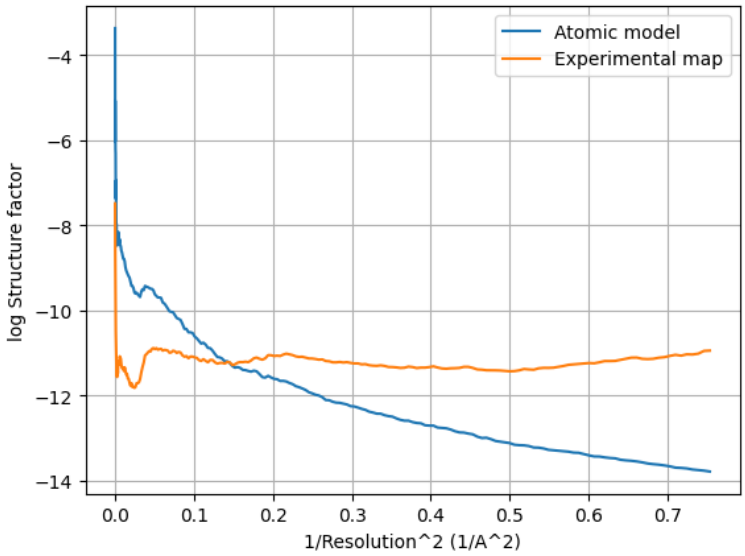



Figure 13: Guinier plot of the atom model and experimental map. The X-axis is the square of the inverse of the resolution in Å.

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

**WARNINGS**: 1 warnings

1. **The correlation is smaller than 0.5, it is 0.497.**

## 4.3 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:**

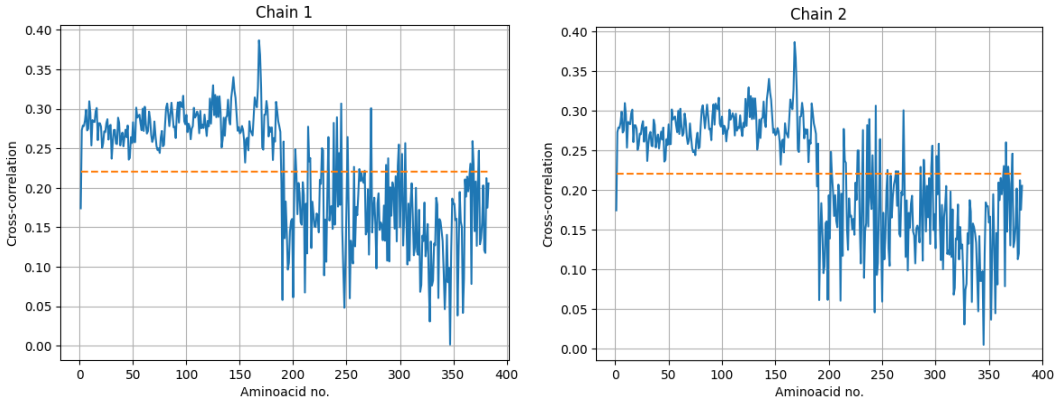
To avoid ringing in Fourier space a smooth mask with a radius of 6.0 Å has been applied.
Overall correlation coefficients:

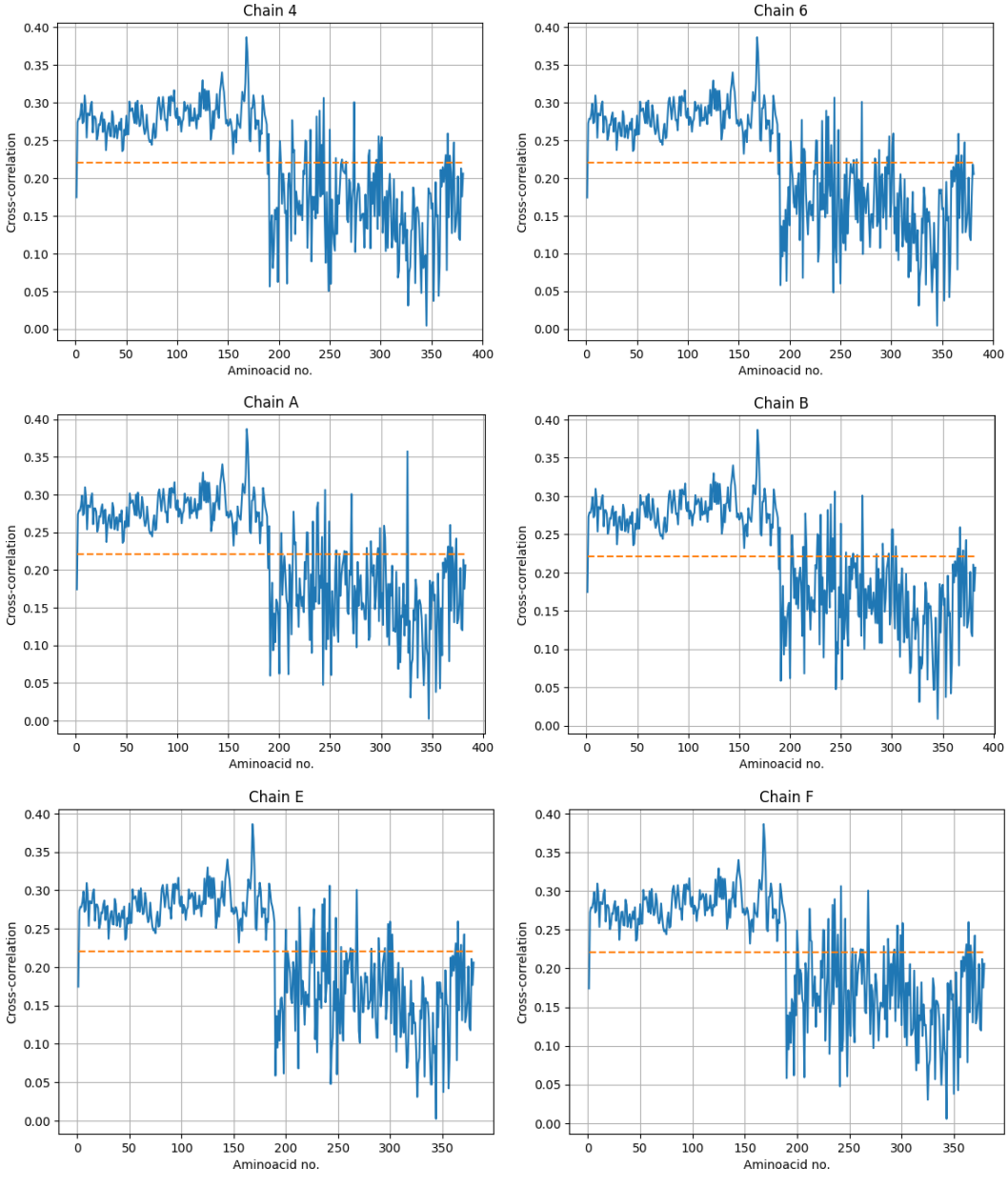
| | |
|---|---|
| CC (mask) = | 0.303 |
| CC (box) = | 0.236 |
| CC (volume) = | 0.292 |
| CC (peaks) = | 0.124 |
| CC (main chain) = | 0.301 |
| CC (side chain) = | 0.291 |

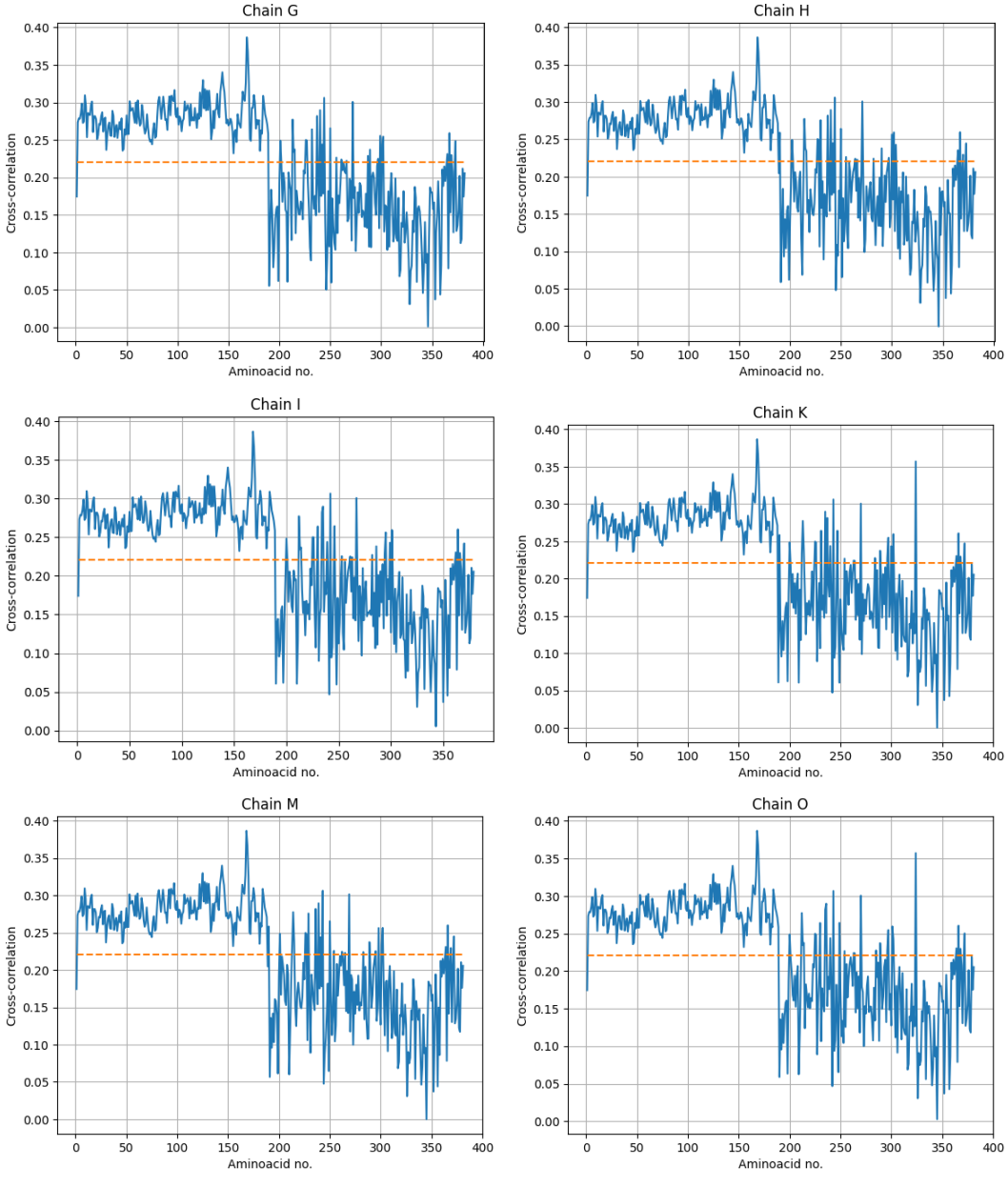
Correlation coefficients per chain:

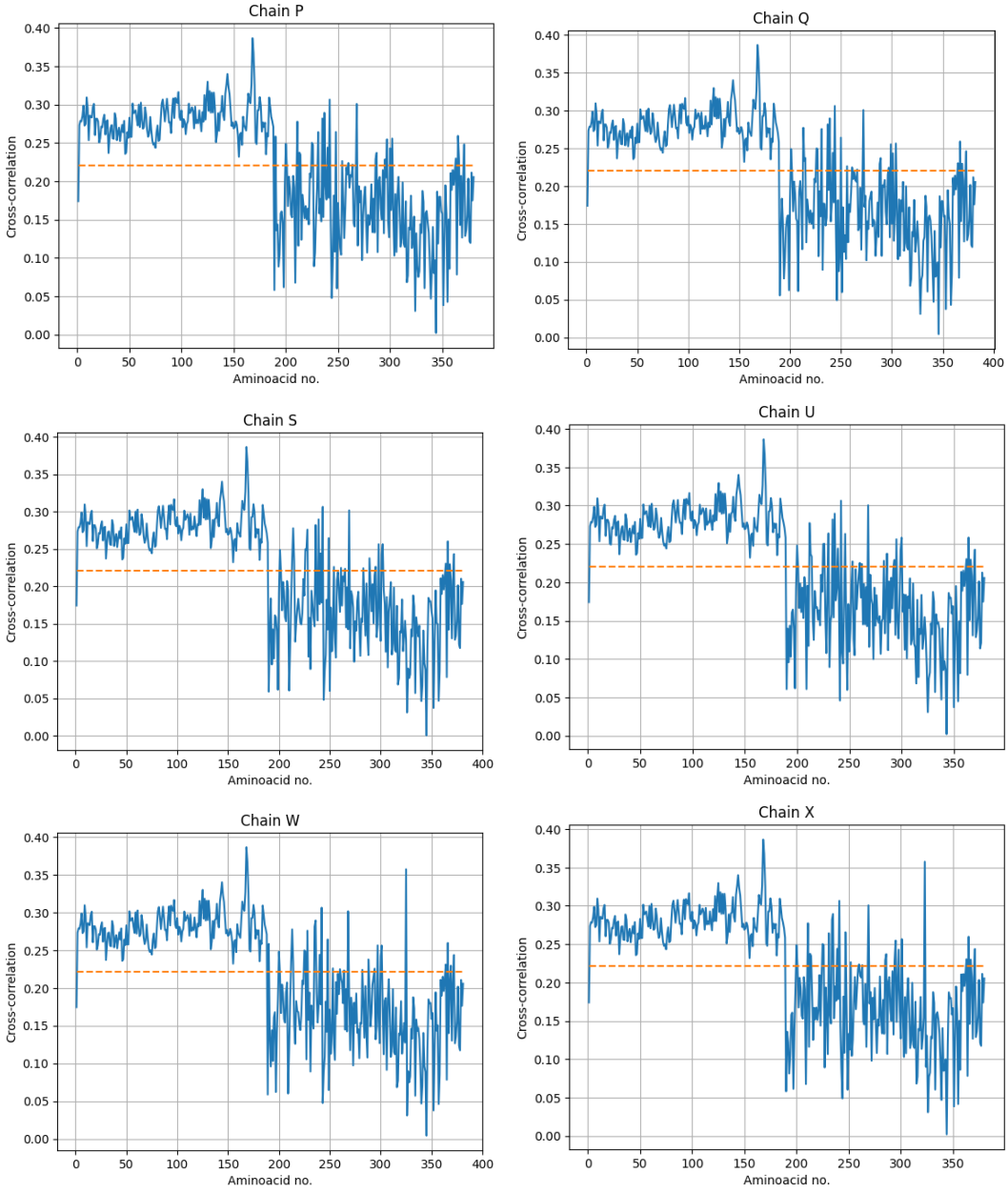
| Chain | Cross-correlation |
|---|---|
| A | 0.289322 |
| 1 | 0.289309 |
| K | 0.289362 |
| a | 0.289360 |
| B | 0.289327 |
| E | 0.289332 |
| e | 0.289314 |
| r | 0.289360 |
| G | 0.289330 |
| I | 0.289307 |
| M | 0.289322 |
| O | 0.289368 |
| Q | 0.289324 |
| S | 0.289330 |
| U | 0.289175 |
| W | 0.289317 |
| Y | 0.289379 |
| 2 | 0.289310 |
| 4 | 0.289200 |
| F | 0.289296 |
| H | 0.289320 |
| P | 0.289311 |
| X | 0.289314 |
| 6 | 0.289361 |

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




Chain 4
Chain 6
Chain A
Chain B
Chain E
Chain F
Cross-correlation
Aminoacid no.


Chain G
Chain H
Chain I
Chain K
Chain M
Chain O
Cross-correlation
Aminoacid no.


Chain P
Chain Q
Chain S
Chain U
Chain W
Chain X
Cross-correlation
Aminoacid no.

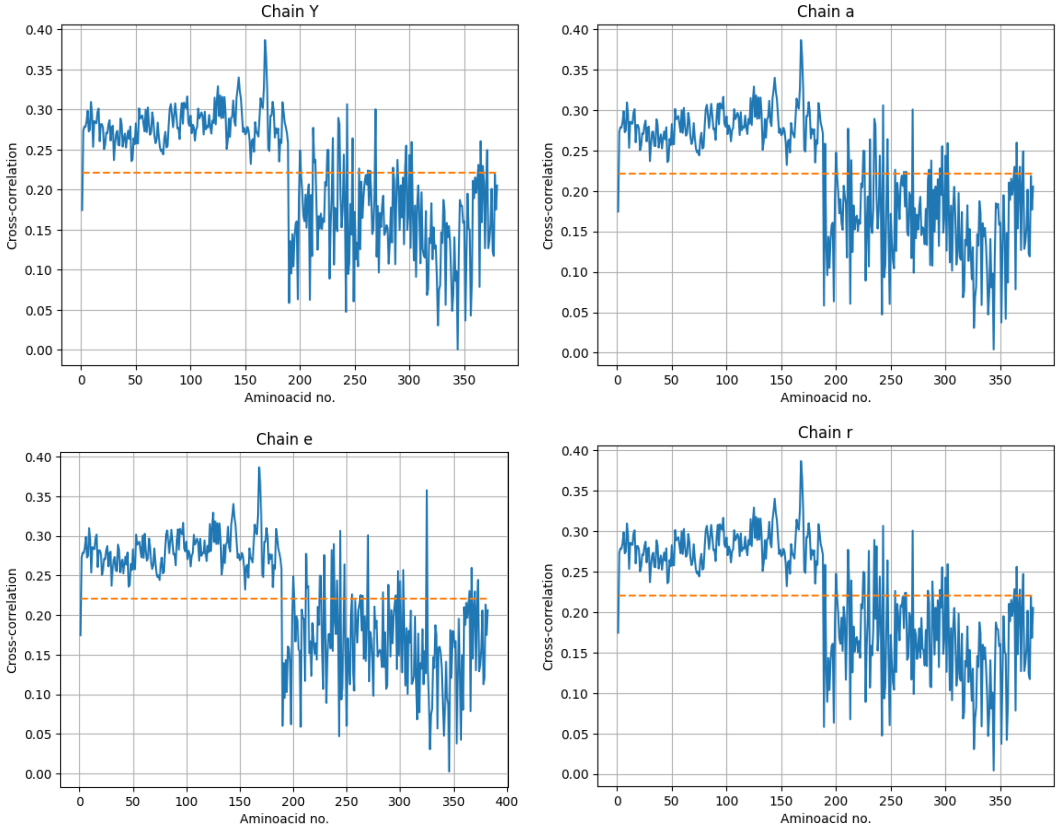



Fig. 14 shows the histogram of all cross-correlations evaluated at the residues. The percentage of residues whose correlation is below 0.5 is 100.0 %.

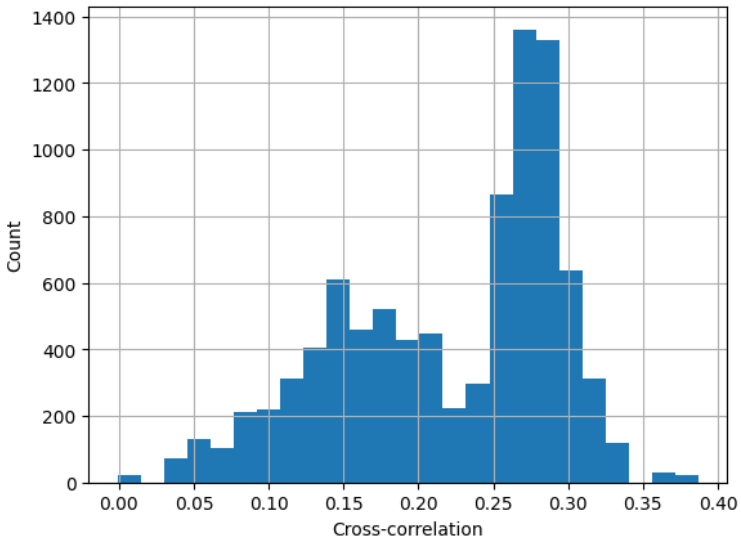



Figure 14: Histogram of the cross-correlation between the map and model evaluated for all residues.

Resolutions estimated from the model:

| **Resolution** (Å) | **Masked** | **Unmasked** |
|---:|:---:|:---:|
| d99 | 1.1 | 1.1 |
| d_model | 1.2 | 1.2 |
| d_model (B-factor=0) | 1.2 | 1.2 |
| FSC_model=0 | 1.1 | 1.1 |
| FSC_model=0.143 | 1.1 | 1.1 |
| FSC_model=0.5 | 1.2 | 1.2 |

Overall isotropic B factor:

| **B factor** | **Masked** | **Unmasked** |
|---:|:---:|:---:|
| Overall B-iso | 0.0 | 0.0 |

Fig. 15 shows the FSC between the input map and the model.

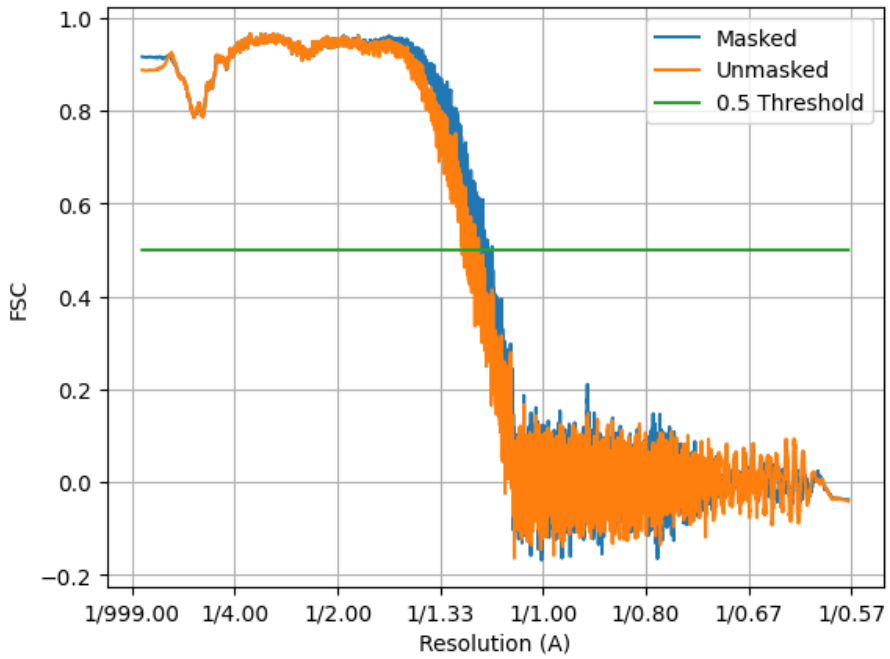



Figure 15: FSC between the input map and model with and without a mask constructed from the model. The X-axis is the square of the inverse of the resolution in Å.

**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 100.0, that is larger than 10%**

## 4.4 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 | 0.094601 |
| Rotamer ratio | 1.000 |
| Max. Zscore | 50.58 |
| Model length | 3144 |
| EMRinger Score | 9.021 |

Fig. 16 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.

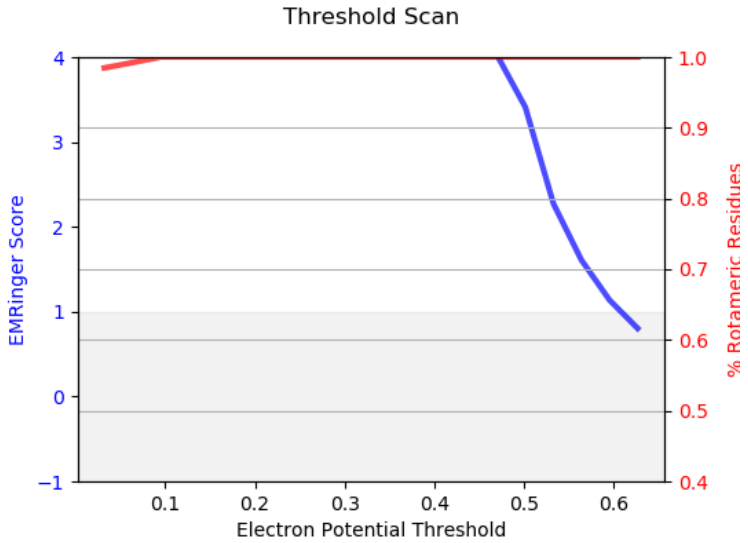



Figure 16: EMRinger score and fraction of rotameric residues as a function of the map threshold.

Fig. 17 shows the histogram for rotameric (blue) and non-rotameric (red) residues at the optimal threshold.

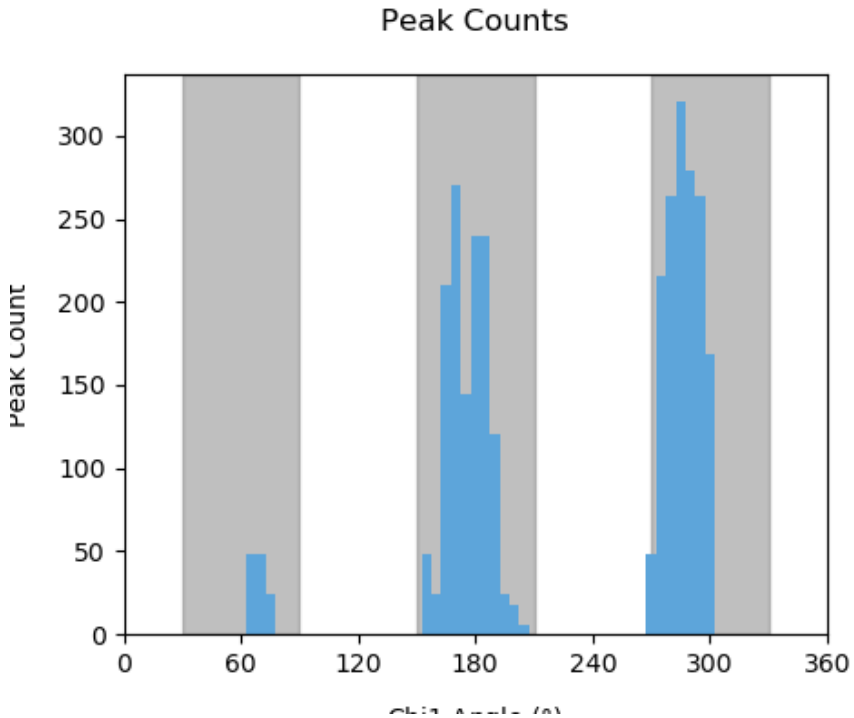
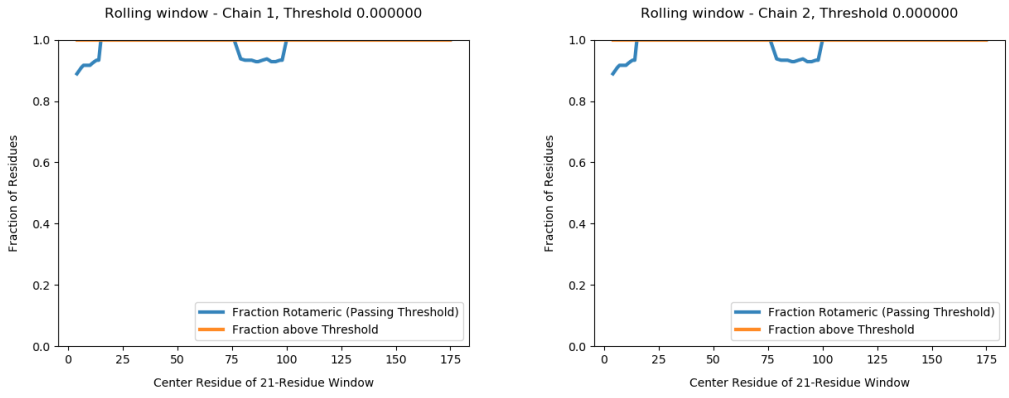
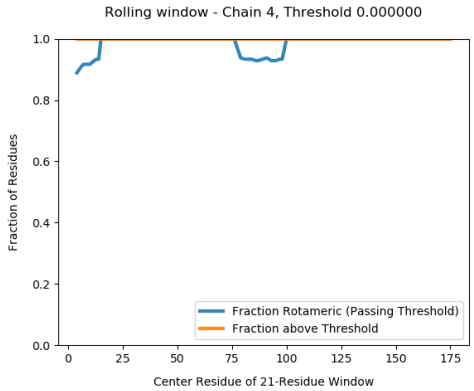
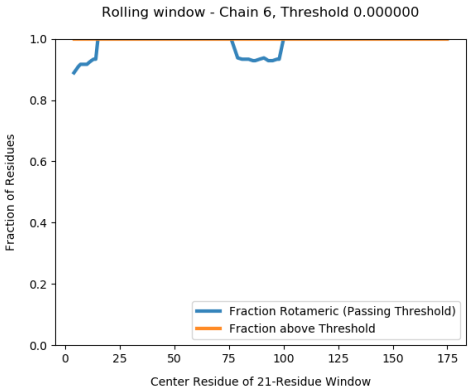


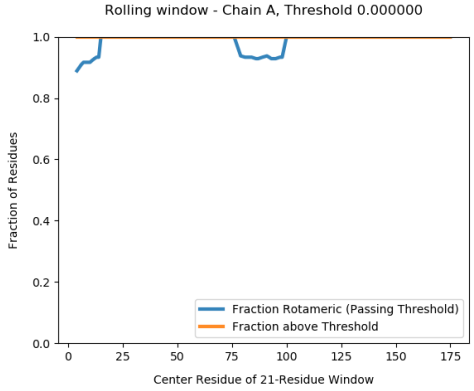

Figure 17: Histogram for rotameric (blue) and non-rotameric (red) residues at the optimal threshold as a function of the angle Chi1.

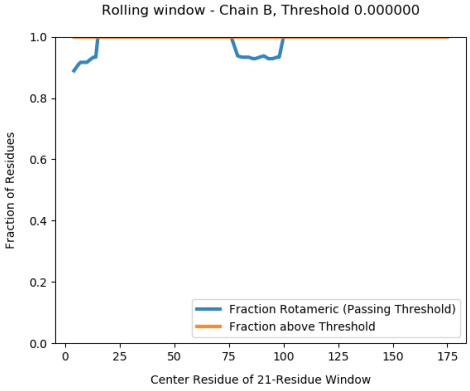
The following plots show the rolling window EMRinger analysis of the different chains to distinguish regions of improved model quality. This analysis was performed on rolling sliding 21-residue windows along the primary sequence of the protein chains.

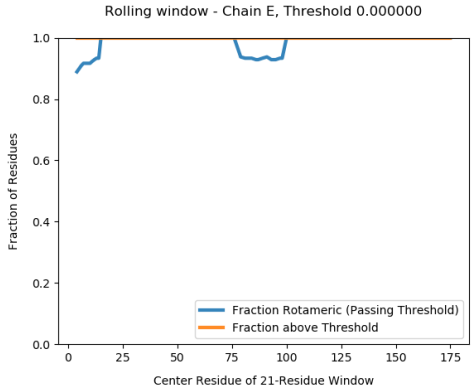


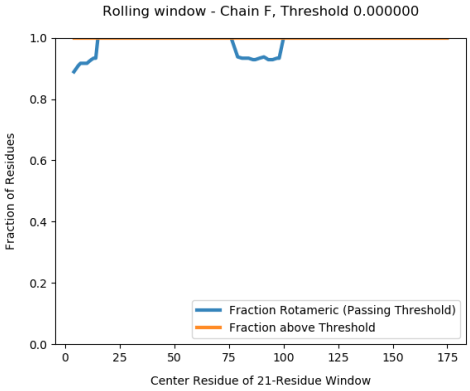

Rolling window - Chain 4, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold

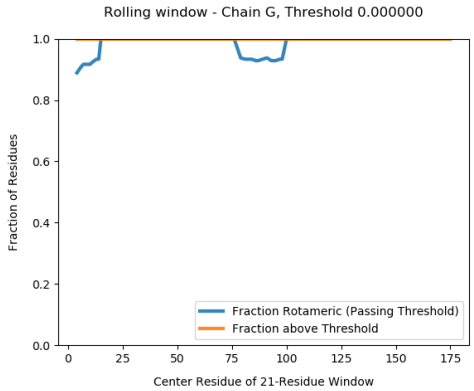

Rolling window - Chain 6, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold

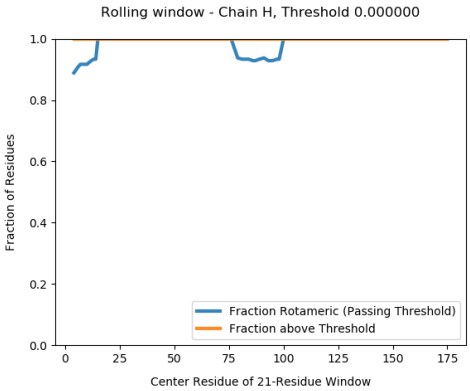

Rolling window - Chain A, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold

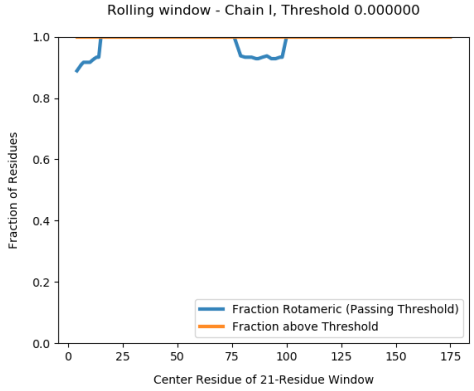

Rolling window - Chain B, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold

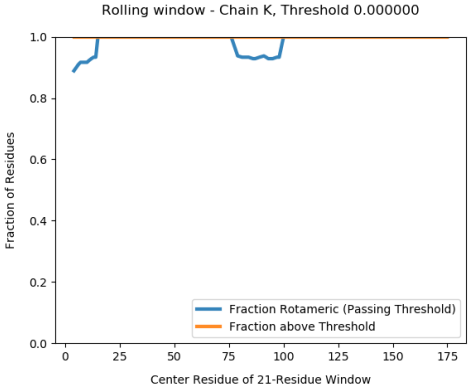

Rolling window - Chain E, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold

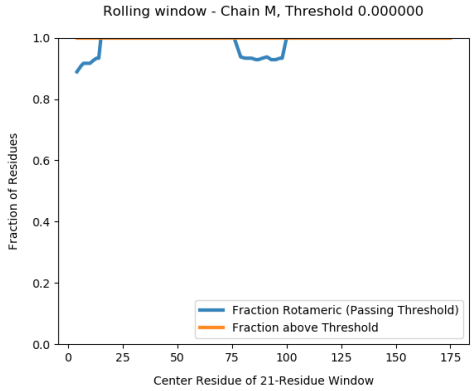

Rolling window - Chain F, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold

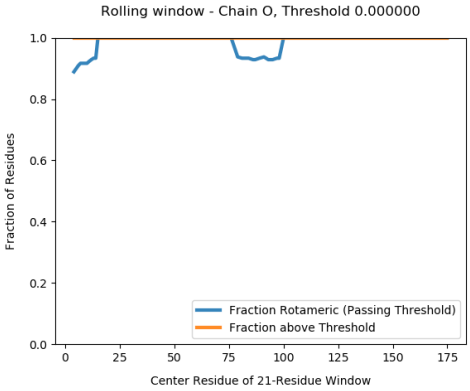

Rolling window - Chain G, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold

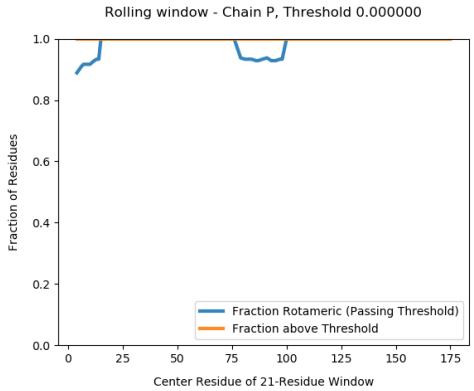

Rolling window - Chain H, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold

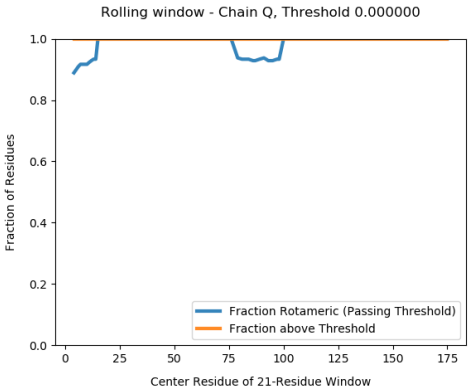

Rolling window - Chain I, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold

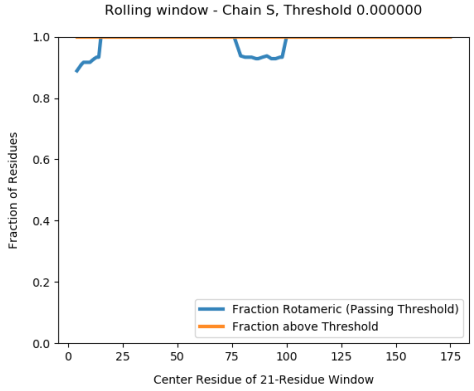

Rolling window - Chain K, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold

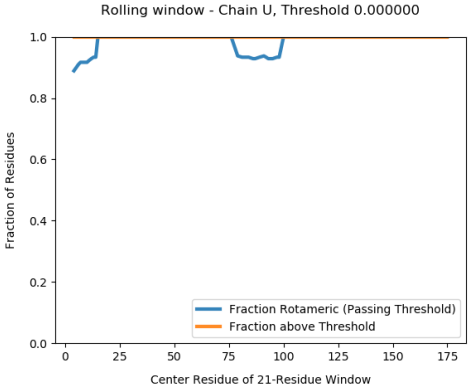

Rolling window - Chain M, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold

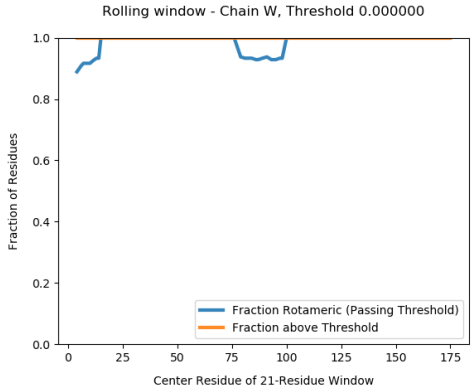

Rolling window - Chain O, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold

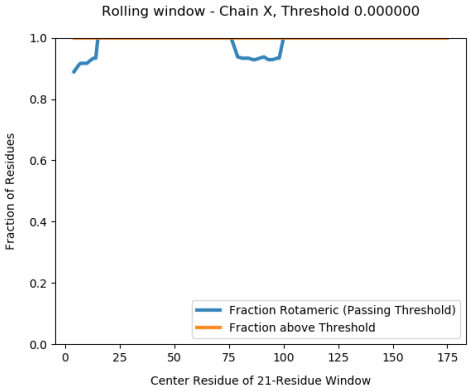

Rolling window - Chain P, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold

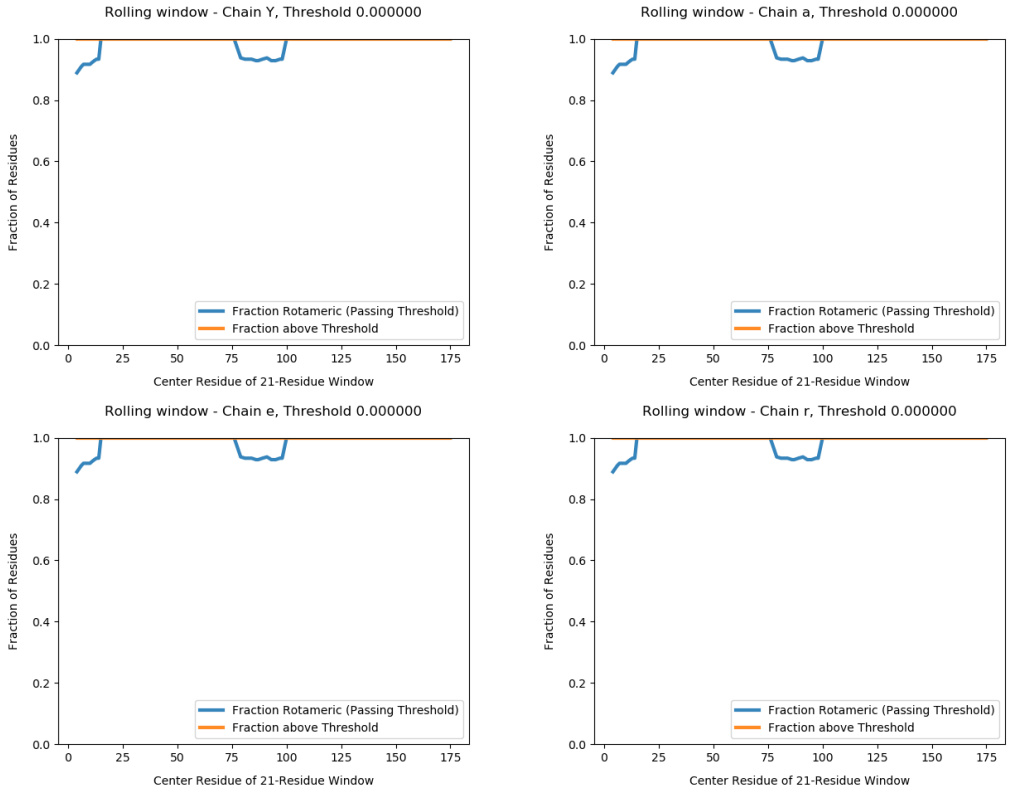

Rolling window - Chain Q, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold


Rolling window - Chain S, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold


Rolling window - Chain U, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold


Rolling window - Chain W, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold


Rolling window - Chain X, Threshold 0.000000
Fraction of Residues
Center Residue of 21-Residue Window
Fraction Rotameric (Passing Threshold)
Fraction above Threshold




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

**STATUS**: OK

## 4.5 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. 18 shows the histogram of the DAQ values. The mean and standard deviation were -0.1 and 0.2, respectively.

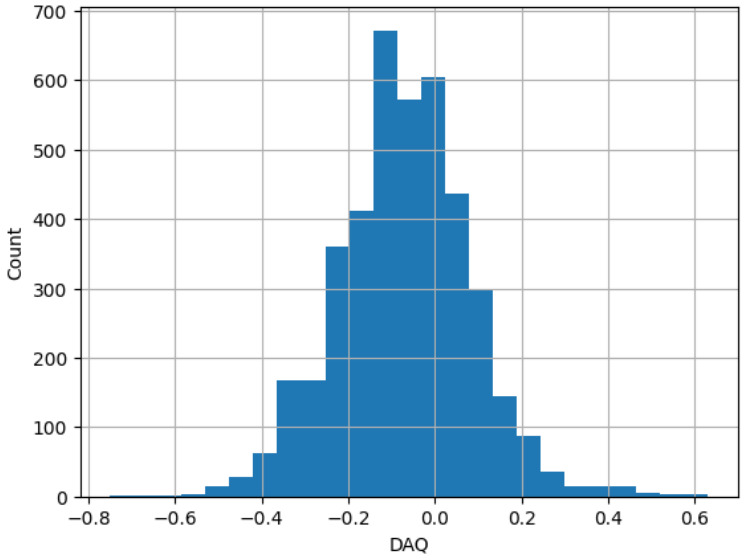
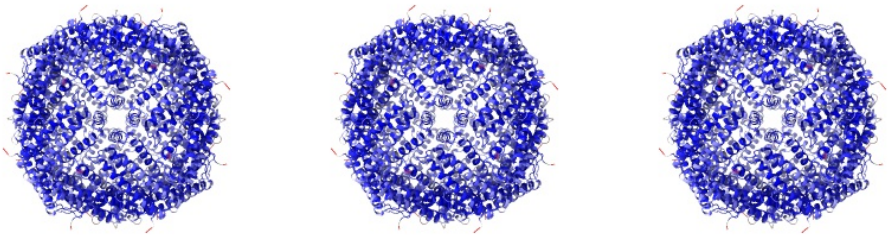



Figure 18: Histogram of the DAQ values.

The atomic model colored by DAQ can be seen in Fig. 19.



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

Figure 19: 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.

**WARNINGS**: 1 warnings

1. **The average DAQ is smaller than 0.5.**

# 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.

[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.

[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).