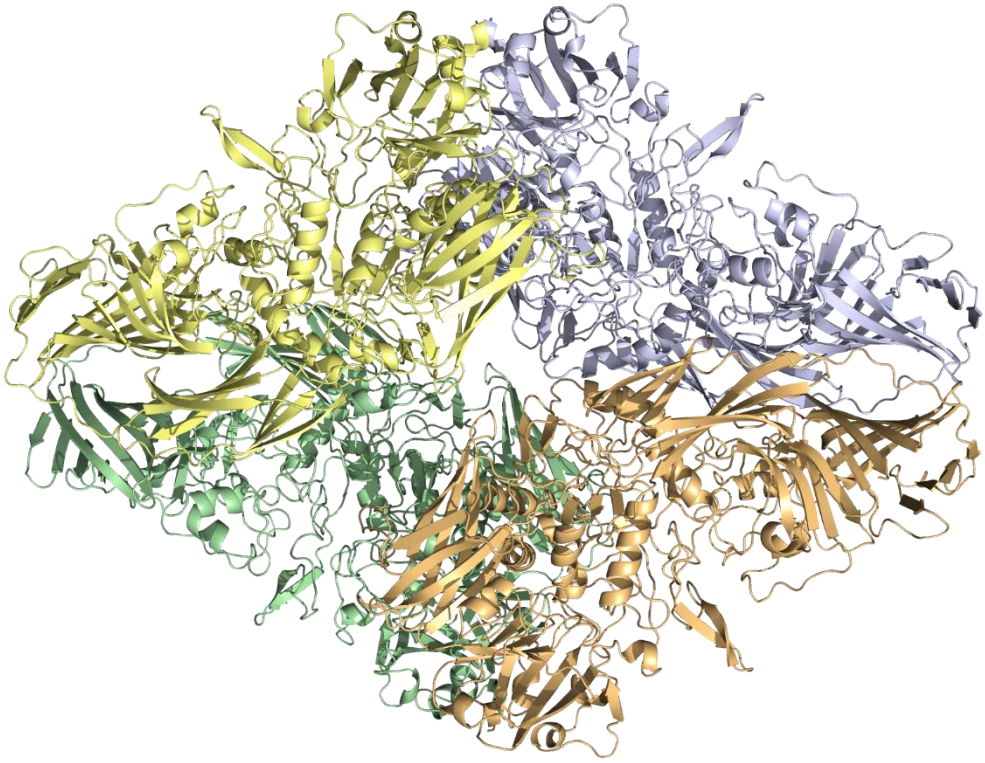
Structural Analysis of β-Galactosidase with Single Particle Cryo-EM



Block course “Macromolecular Structure Determination using Modern Methods”

Supervisor: Rossitza Irobalieva, Locher Lab

Christopher Weingarten, BSc Biology 6. Semester

cweingar@student.ethz.ch

26.05.2021

Abstract

In this block course, β-galactosidase was characterized using cryo-EM, reaching a resolution of 3.03 Å. This report will give a short introduction to β-galactosidase and the history of cryo-EM. It will also show the general pipeline of sample preparation, data processing, and model building and validation, with an analysis of the final model at the end.

Introduction

B-Galactosidase from *Escherichia coli* is a well-known enzyme catalyzing several reactions, including the hydrolysis of lactose into galactose and glucose [1]. It is often used as a reporter gene in cellular biology, cleaving Xgal into galactose and an insoluble, dark blue dye [2], or as a calibrating structure for new cryo-EMs, due to the high rigidness of the structure. The enzyme is encoded in the lacZ gene, which can be found in the lac operon [1].
B-Galactosidase is homotetramer of about 464 kDa with a D2 symmetry. Each subunit consists of 1'023 amino acids with 5 domains. Domain 1 is a jelly-roll type β-barrel, domain 2 and 4 fibronectin type III-like barrels, domain 3 a distorted TIM-type barrel, and domain 5 a β-sandwich. Each subunit interacts with two of the other subunits, leading to a stable and rigid diamond-like structure. [3]

The method of single particle cryo-EM is defined by two key factors: The vitrification of a sample by plunge freezing, and the reconstruction of a 3D structure by combining the 2D projections of many particles. This method started showing usage in the 1980s, driven by the promise of a structural determination method without the need for crystallization. First structures were however very low resolution, serving more as a reference or a general skeleton to fit in protein subunits determined by X-ray crystallography.
In the 2000s, a "resolution revolution" was driven by the introduction of direct electron detectors in cryo-EM. Previously, charged coupled device (CCD) cameras were used, which converted electrons into photons using a scintillator. This conversion blurs a single electron into several photons, reducing high-resolution signal. Before the introduction of direct electron detectors, single particle cryo-EM was even dubbed "blobology", due to its low resolution. However, with the breakthrough of direct electron detectors, single particle cryo-EM started reaching atomic resolution on a routine basis.
Higher resolution was also driven by the implementation of new image processing algorithms, such as per particle CTF (contrast transfer function) and motion correction. The implementation of direct electron detectors was however the main driving force. [4]

Sample preparation

For image acquisition under the cryo-EM, the sample first had to be purified and vitrified on a grid. The purification step was done before the block course by the supervisors. The purity of the sample was however confirmed by size exclusion chromatography (SEC) (Figure 1) and a SDS PAGE gel (Figure 2).

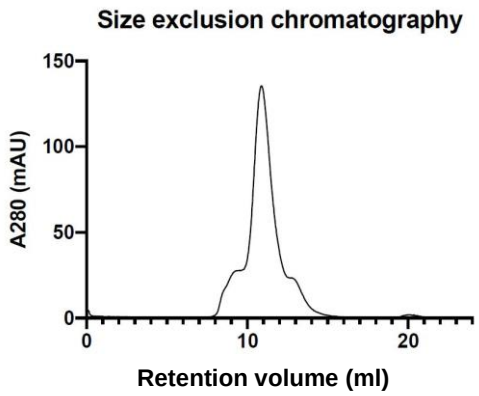


Figure 1: A chromatogram of the SEC performed on the purified β-galactosidase sample, with the absorption units on the y-axis and the retention volume on the x-axis. While there are small void and trailing peaks, the sample is pure enough for cryo-EM if the fractions are taken from the middle of the peak.

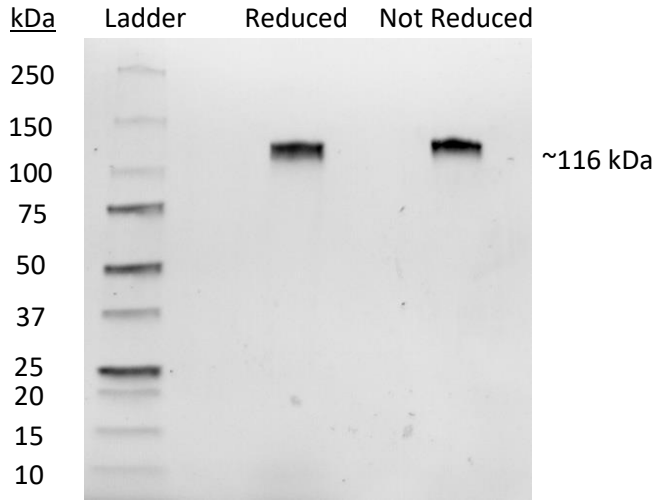


Figure 2: SDS PAGE gel with a ladder (BioRad Precision Plus Protein standards) and a reduced and non-reduced sample of β-galactosidase. Both samples are at the same height, indicating that there are no sulfur bonds, and where the monomer would be expected (116 kDa).

The sample had to then be plunge frozen on EM grids. The grids consist of a copper mesh skeleton with a thin sheet of carbon containing small holes (~2 microns diameter) on top. The grids are first glow discharged, making the carbon sheet hydrophilic. The grid is then inserted into a Vitrobot (Mark IV FEI), which is used for plunge freezing. The grid is kept at room temperature at 100% humidity, where 2 microns of the sample is then applied using a pipette. The grid is then blotted and plunged into a mixture of ethane and isopropane at the temperature of liquid nitrogen, instantly freezing the sample and avoiding water crystallization. The plunging was performed on 8 different grids, with different blot times and sample concentrations (see Table 1). The samples were then ready for image acquisition. This was however not in the scope of this block course, which is why this step was skipped.

Table 1: Different blot times and sample concentrations of grids. With the varying blot times and sample concentrations, the ideal ice thickness and particle density can be "screened" for.

	Box 1				Box 2			
	Grid 1	Grid 2	Grid 3	Grid 4	Grid 1	Grid 2	Grid 3	Grid 4
Blottime [s]	2	2.5	2	2.5	3	3	3	3.5
Concentration [mg/ml]	1	1	2	2	1	1	2	2

Data Processing

For data processing, the RELION-3.1 tutorial was followed [5]. The dataset used is a subset from a previous publication [6], with only 24 movies. The data was collected on a 200 kV JOEL cryo-ARM microscope.
The movies were imported into RELION. As a first step, motion correction was performed, correcting for beam-induced sample motion. This could be done since each movie consists of 40 frames, which allows the program to track the movement of different sectors over a certain timespan.
Before particle picking, CTF estimation and correction was performed, which corrects for certain aberrations such as defocus.
The particles in the micrographs were then picked using an automatic, reference free picking algorithm called Laplacian of Gaussian (LoG)-picking. The picked particles were then extracted in box sizes of 256x256 pixels, which were then binned by a factor 16, leading to 64x64 pixel boxes. While some high-resolution information does get lost in the process, this information is not required for the early steps of data processing, and it significantly increases computational speed.

After the extraction, 2D classification could be performed. Here the program tries to superimpose different particles and assigns them into classes. The number of classes can be determined by the user, and in this case was set to 50 classes. The ten classes with highest populations (see Figure 3).
With these picked particles, an initial 3D model was created using a stochastic gradient descent algorithm. Using this initial model as a reference, 3D classification is performed. The program looks for a 3D structure which could create the 2D projections, which you can see in the 2D classes. This is again an iterative process. The number of classes was set to four. The resulting classes (Figure 4) all have some characteristics of β-Galactosidase, with class 3 having the highest population but also the most detail and features. Consequently, further steps were done using the particles from class 3.

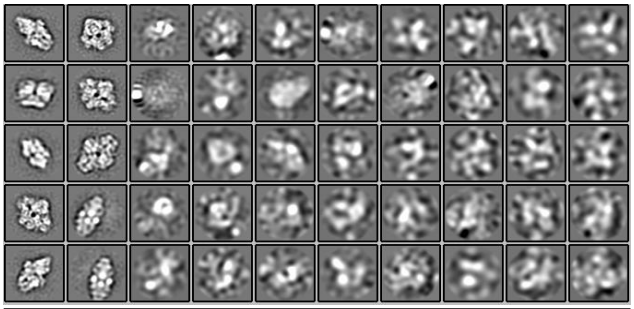


Figure 3: The 50 classes determined by 2D classification, ordered from the highest (top left) to the lowest (bottom right) population.

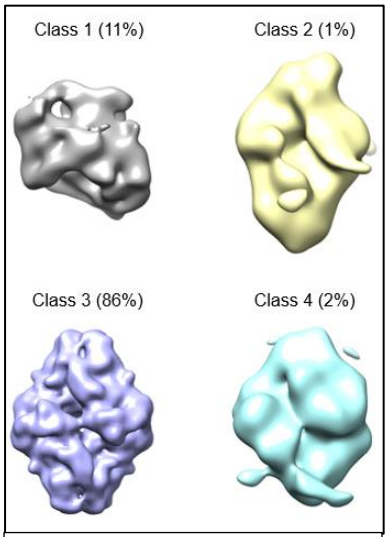


Figure 4: 3D classes with population percentages in brackets.

The particles were re-extracted, unbinning them to access more detailed information, and a 3D refinement was performed. While his process is similar to 3D classification, the program doesn't split the particles into different classes. There are different ways to perform a refinement, but in this case gold standard refinement was used. The program splits the particles in two, and each half refines the particles according to the template. When both halves have converged, the two maps are compared by Fourier shell correlation (FSC), which compares the shell in Fourier space over the spatial frequency. The estimated resolution is concluded as the value where the correlation between the two shells drops below 0.143.

After 3D refinement, a mask was created around the protein, defining an area outside of which all signal is ignored. This helps to reduce the noise and increase computational speed. The mask is then implemented in a post-processing job, which also makes the map noticeably sharper. To further increase resolution and resolvability, per particle CTF refinement and motion correction was performed. While this is done for the whole micrograph or a certain area respectively at the beginning of data processing, repeating these steps for every particle can be beneficial. This resulted in a final map with a resolution of 3.03 Å, which was used further for model building.
A summary of data processing and some additional information (such as number of particles) can be found in figure 6.

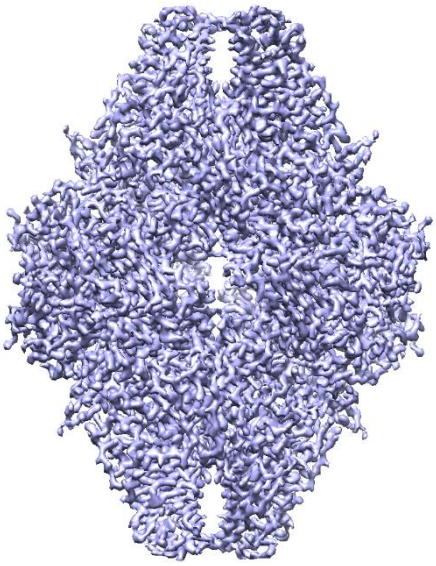


Figure 5: Final map output from RELION, presented in UCSF Chimera.

Model Building & Validation

To save time, a truncated subunit from a previous publication [6] was fit into the map in UCSF Chimera. The missing residues (481-572) were rebuilt in COOT. The completed subunit was then copied and fit into the rest of the subunits of the map.

Using PHENIX, a real-space refinement was performed on the constructed model. The resulting model was then analyzed using different parameters, such as the Ramachandran plot Z-score. Residues with a bad score were looked at manually in COOT and corrected as best as possible. After manual corrections, another real space refinement was performed. This was repeated several times, until satisfactory validation score where achieved. A summary of the most important scores can be found in table 2.

Table 2: Selected validation scores after the first and last real space refinement. **Bond Angles**: general angles of bonds; **Clash score**: score for non-bonded atom proximities; **Ramachandran plot**: validation of peptide backbone angles per residue; **Cis proline/general**: percentage of prolines and other residues in a cis conformation; **CaBLAM outliers**: conformational outliers in secondary structures.

	First Refinement	**Last Refinement**
Bond Angles (°)	0.610	0.521
Clash score	9.24	7.90
Ramachandran plot (%):		
Outliers	0.47	0.00
Allowed	3.14	2.77
Favored	96.40	97.23
Cis proline/general (%)	8.1/0.7	8.1/0.2
CaBLAM outliers (%)	3.32	2.31

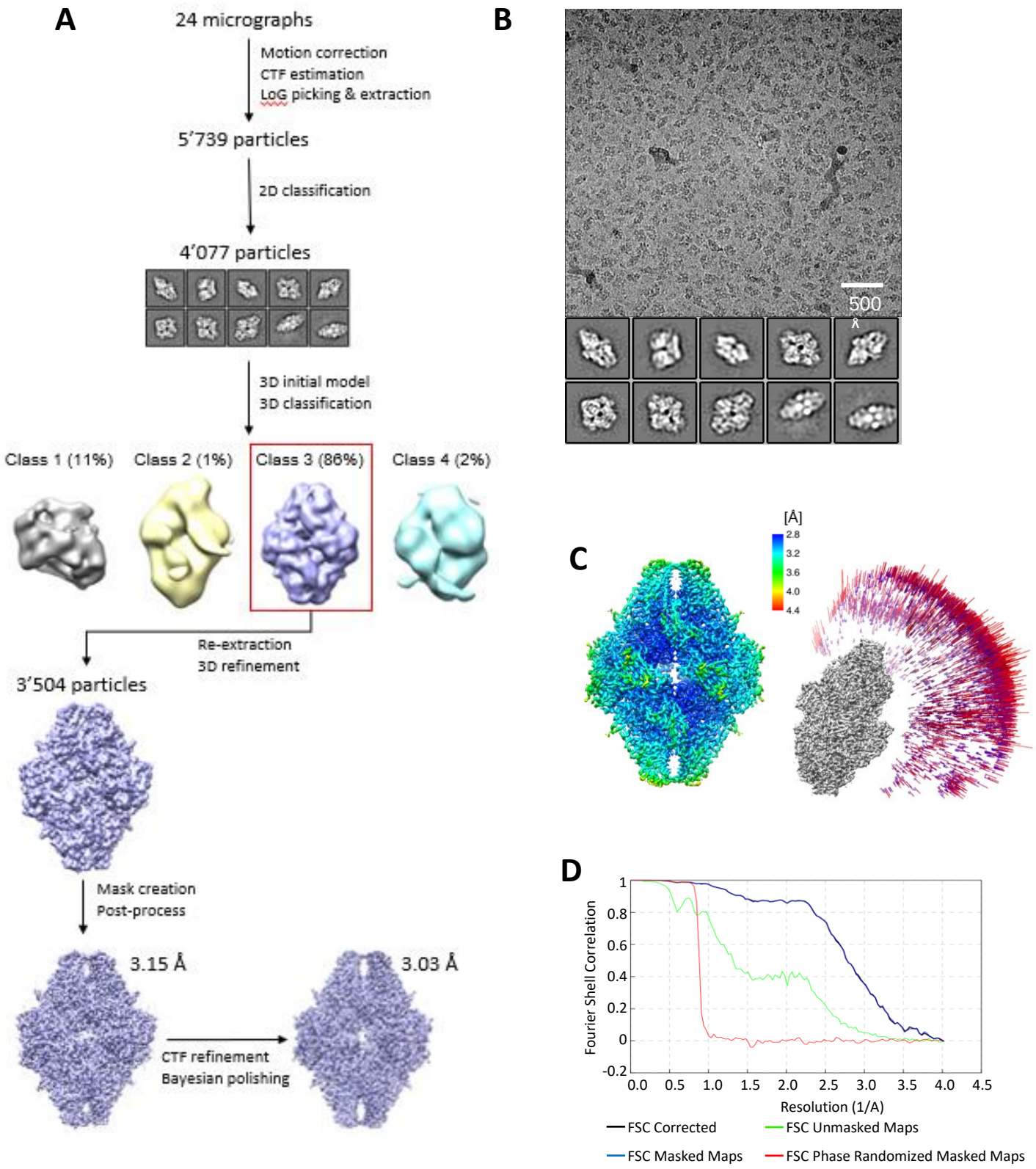


Figure 6: A) Pipeline of data processing in RELION B) A representative micrograph and a larger view of the chosen 2D classes. C) Local resolution estimation with color bar resolution scale (left) and angular distribution (right). D) Fourier shell correlation (FSC) with marked resolution estimation.

Model Analysis

As a final step, the model was analyzed through visualization in PyMOL. The complete model can be seen in figure 6, colored by different subunits. The general structure seems to coincide with previous models of β-Galactosidase. Taking a closer look at a subunit, the different domains can be identified (Figure 7).

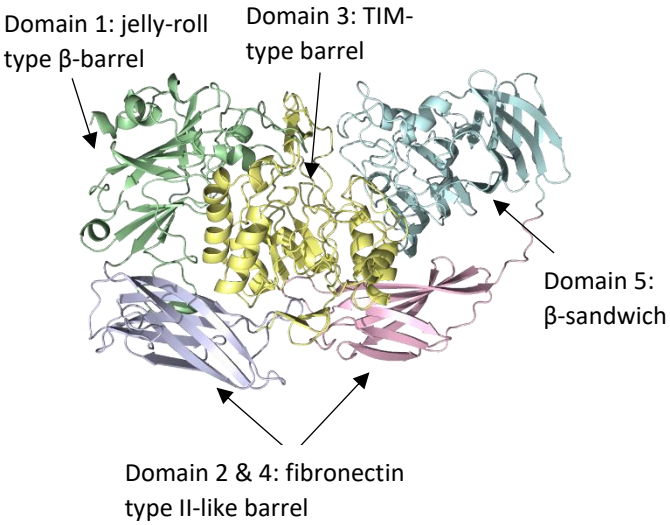


Figure 8: Labeled and colored protein domains of a single β-Galactosidase subunit in cartoon

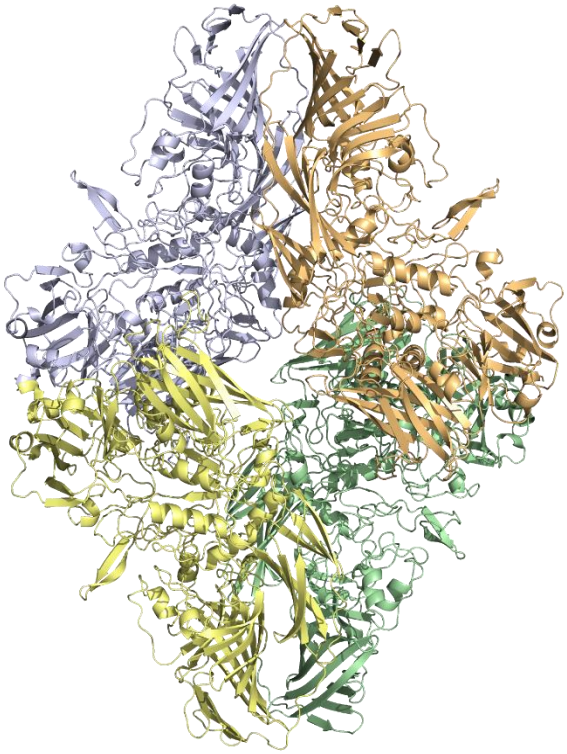


Figure 7: Entire model of β-galactosidase, colored by chain in cartoon representation.

The fragment which was rebuilt in COOT is in the third domain of the protein, and is colored green in figure 8. Looking at the closeup in figure 9 and comparing it to the reference structure [6], it can be seen that most of the model overlaps. However, there are two β-strands in the reference model which are not recognized as such in the built model.

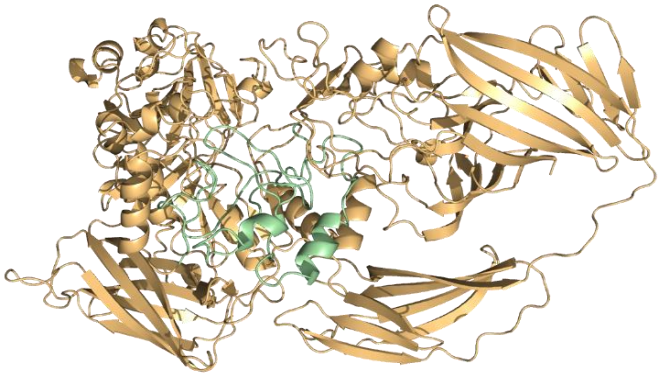


Figure 9: One subunit of β-Galactosidase in cartoon representation, with the rebuilt fragment in green.

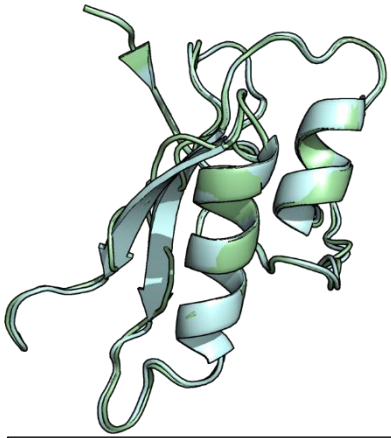


Figure 10: Rebuilt fragment (green) superimposed with a reference fragment (cyan) in cartoon representation.

To identify why PyMOL does not recognize the β-strands, a closer look at the stick representation of the built fragment was taken. As can be seen in figure 10, while the general conformation of the two models is very similar, there are several discrepancies (circled in figure 10) in the backbone. These are easily noted by the different direction of the carboxyl group. Since a β-strand is defined by alternating directions of the carboxyl group, this is most likely the reason why these two β-strands are not recognized in PyMOL. The actual conformation could be determined by a higher resolution map, though considering the energetic advantages of secondary structure, the reference is most likely correct in this case.
While the differences in the backbone are the most significant differences, there are also some deviations in the sidechains, as can be seen in figure 11 with W523 as an example. Here, the tryptophan of the built fragment does not fit into the EM density map, while the reference fits nicely. The conformation of the reference is therefore clearly more likely, and the rebuilt fragment should be corrected manually in COOT.

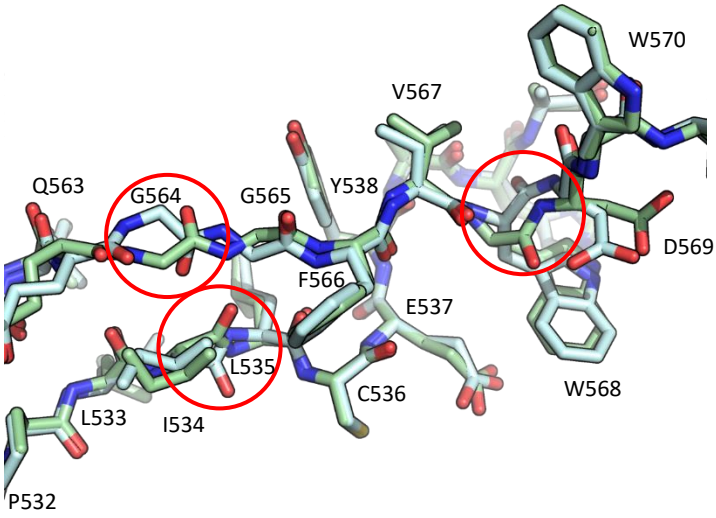


Figure 11: The rebuilt fragment (green) superimposed with the reference (cyan) in cartoon representation. The areas where the fragment deviates from the reference are highlighted with a red circle.

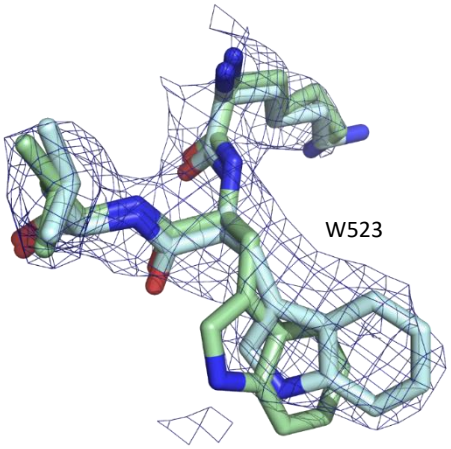


Figure 12: W523 of the rebuilt fragment (green) superimposed with the reference (cyan) in cartoon representation, with the EM density map displayed as a mesh.

Conclusion

In conclusion, β-galactosidase was successfully purified, and samples were prepared for cryo-EM. An EM density map was created in RELION using a provided sub-dataset, and a fragment of the protein successfully rebuilt in COOT. While there was some deviation from the reference model, the two models overlapped very well.

References

[1]: Juers DH, Hakda S, Matthews BW, Huber RE; Biochemistry 2003; Structural Basis for the Altered Activity of Gly794 Variants of Escherichia coli β-Galactosidase; https://doi.org/10.1021/bi035506j

[2]: Schmidt A, Tief K, Foletti A, Hunziker A, Penna D, Hummler E, Beermann F; Brain Research Protocols 1998; lacZ transgenic mice to monitor gene expression in embryo and adult; DOI: 10.1016/s1385-299x(98)00021-x

[3]: Jacobson RH, Zhang XJ, DuBose RF, Matthews BW; Nature 1994; Three-dimensional structure of beta-galactosidase from E.coli; DOI: 10.1038/369761a0

[4]: Cheng Y; Science 2018; Single particle cryo-EM – how did it get here and where will it go; DOI: 10.1126/science.aat4346

[5]: Scheres SHW; High Performance Computing at the NIH 2019, Single-particle processing in RELION-3.1; https://hpc.nih.gov/apps/RELION/relion31_tutorial.pdf

[6]: Kato T, Terahara N, Namba K; Electron Microscopy Public Image Archive 2018; The first reconstruction of beta-galactosidase solved by cryoARM200; EMPIAR-10204; PDB 5A1A