Electron cryo-microscopy studies of protein-carbohydrate complexes and glycoproteins

Manal Mohammed Alzahrani

Doctor of Philosophy

University of York

Chemistry

September 2023

Abstract

The recent resolution revolution experienced by electron cryo-microscopy (Cryo-EM) has meant the technique has begun to displace X-ray crystallography (XRD) as the method of choice for elucidating the structure of large complexes and, more recently, single proteins. While most of the advances that have pushed the technique's resolution and size limits have been made using highest-end 300 keV instrumentation, it should be possible to complete complex structural biology projects with more accessible hardware. One particular area of interest and of inherent complexity is structural glycobiology: glycoproteins and protein-carbohydrate complexes have traditionally been affected by shortfalls in methodology and the intrinsic difficulties associated with the mobility of oligosaccharide chains. Recent studies comparing the quality of carbohydrate structure derived by XRD and Cryo-EM have demonstrated differences, which might be attributable to the nature of the experiments. In this work, we aim to further test these differences using an enzyme-ligand complex (beta-galactosidase with 4-epi cyclophellitol, chapter 3) and a small glycoprotein (human transferrin, chapter 4), using a middle-range 200 keV electron cryo-microscope (TFS Glacios).

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Å	Angstrom
aa	Amino acid
ABPs	Activity-based probes
β-Gal	β-Galactosidase
CAP	Catabolite activator protein
CAZy	Carbohydrate-Active Enzymes database
CBB	Coomassie Brilliant Blue
CBE	Conduritol B epoxide
CCP4	Collaborative Computational Project No. 4
CCP-EM	Collaborative Computational Project for electron cryo-microscopy
cold-FEG	Cold Field Emission Electron Gun
CPU	Central processing unit
CRP	cAMP receptor protein
Cryo-EM	Electron cryo-microscopy
CTF	Contrast transfer function
CV	Column volume
DED	Direct Electron Detectors
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
DQE	detective quantum efficiency
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EER	Electron Event Representation
FSC	Fourier Shell Correlation
FEG	Field Emission Guns
GBA1	Glucocerebrosidase

GPUs	Graphical processing units
HEPES	4-(2-hydroxyelthyl)-1-piperazineethanesulfonic acid
hTF	Human serum transferrin
HPLC	High Performance Liquid Chromatography method
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KX	(K) camera constant, (X) magnification factor
kDa	Kilo-Dalton
keV	kilo-electron volt
mRNA	messenger ribonucleic acid
MWCO	Molecular weight cut off
NaCl	Sodium chloride
OD	Optical density
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PETG	phenylethyl β-D-thiogalactopyranoside
RCSB PDB	Research Collaboratory for Structural Bioinformatics Protein Data Bank
RNA	Ribonucleic acid
scFv13R4	Fv antibody domains
SDS-PAGE	Sodium dodecyl sulphate–polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
TEM	Transmission electron microscope
TFS	Thermo Fisher Scientific
UV	Ultraviolet
XRD	X-ray crystallography
6Az2FGalF	6-azido-2,6-dideoxy-2-fluoro-β-D-galactosyl fluoride

Acknowledgements

Declaration

This thesis is an original work of my effort for which I am the only author, except as stated in the acknowledgement and indicated in the text. This work has never been nominated for an award at University of York or any other university.

The beta-galactosidase complex structure described in Chapter 3 has been deposited as an EM map and atomic model under the title:

***E. coli* beta-galactosidase with covalently attached 4-epi cyclophellitol**
Corresponding authors: Jon Agirre

The deposition session D_1292133439 resulted in the following accession code(s): (PDB code: 8QLZ), EMD-18488.

All structural data including validation reports and commands to restraint and refine the structure have been deposited in Zenodo: 10.5281/zenodo.8384171

Chapter 1   Introduction

1.1 Biomolecular structure analysis

1.1.1 X-ray crystallography

In the last five decades, this technique has maintained its position as a stepping stone to determining the three-dimensional structure of complex and biological molecules. X-rays collide with the electrons present in biological macromolecules in their crystalline forms; as a result, the X-rays diffract (Jaeger 2005), causing a distinctive pattern that encodes the structure of the target. This diffraction pattern must be converted back into a 3D structure through phasing (as only intensities and no phases are recovered during the diffraction experiment), either by using a structure of a related target or experimentally by the addition of heavy atoms that cause anomalous diffraction. The difficulty in growing protein crystals can range from trivial to nearly impossible and often requires extensive trial and error. All the first structures of β-galactosidase (the topic of a later chapter) were resolved by X-ray crystallography (XRC). (Table 1.1) presents the known *E. coli* β-galactosidase structures obtained through XRC.

Table 1.1 Selected relatively high-resolution XRC structures of liganded *E. coli* β-galactosidase from the RCSB PDB.

PDB	Å	Complex	References
1DP0	1.7	NA	(Douglas H. Juers 2000)
1JYV [A]	1.75	Complex with ONPG	(Douglas H. Juers et al. 2001)
1JZ7 [A]	1.50	Complex with galactose	(Douglas H. Juers et al. 2001)
3T0D [A, B,C,D]	1.93	Complex with galactonolactone	(Jancewicz et al. 2012)
4DUX [A,B,C,D]	2.3	Complex with L-ribosyl molecules	(Wheatley et al. 2013)

1.1.1.1 Resolution estimation in X-ray crystallography

X-ray photons irradiate the studied crystals in XRC, producing diffraction spots (or reflections), which are measured. The distance between reflections and the centre of the detector is inversely proportional to the resolution of the data; therefore, the further away the reflections are, the higher the corresponding resolution is. However, the signal decreases commensurately with disorder and atomic movement, limiting the resolution of the studied dataset alongside quality parameters of the data, such as signal-to-noise ratio, completeness, and R-factors (Dubach and Guskov 2020). Resolution cut-off remains a dilemma in XRC. A signal-to-noise ratio ≥ 2 is traditionally widely recommended, keeping the data where the signal doubles the intensity of the estimated noise. Rmerge, a measure of multiple measurements for each reflection, is another cut-off value (*e.g.* 40% Rmerge cut-off value used over 20 years ago). However, Evans and Murshudov state, correlation coefficients (*e.g.* $CC_{1/2}$) are less dependent on data distribution and more on the linear dependence between data (Evans and Murshudov 2013).

1.1.2 Cryo-EM

This structural technique is based on vitrifying the macromolecules in a layer of 'vitreous ice' by cooling them so quickly that there is insufficient time for the water molecules to crystallise(Cabra and Samsó 2015), preventing air-water interface that can denature protein and prevent macromolecules adopting a single preferred orientation (Lyumkis 2019). Liquid ethane or propane is used to vitrify the sample because they have better heat capacity than liquid nitrogen. The supercooled target is then imaged using an electron microscope, and its structure is computed by averaging images of thousands of particles together.

The single-particle analysis (SPA) approach of structure determination has developed into a powerful technique rivalling X-ray crystallography (XRC), initially becoming crucial for specific targets that have eluded crystallisation and, more recently, undertaking routine structural biology work. SPA is not based on the reconstruction of a structure from a single biological molecule but on taking many pictures of the same type of molecule in different orientations and computationally calculating a 3D electron-potential map. The 3D reconstruction software uses a picture projection algorithm after determining the orientation of the images. The

projection data can be collected from the algorithm and reconstructed in 3D so that the whole volume of the biomolecule can be examined (Murata and Wolf 2018).

Cryo-EM has become a mainstay in structural biology, constantly pushing the limits of resolution and expanding the types of samples that can be examined; its development led to the award of the 2017 Nobel Prize in Chemistry to Jacques Dubochet, Joachim Frank and Richard Henderson (Cheng, Glaeser, and Nogales 2017; Nogales 2018; Shen 2018).

1.1.2.1 Electron microscopy: The basics

In the case of light microscopes (LMs), the resolution is restricted to ~200 nm by the relatively long wavelength of incident light. Due to scientific necessity, the first electron microscope (EM) was developed to overcome this technical problem in examining the smallest samples. Scientists clearly hypothesised that an electron microscope would give a far better and higher image resolution in comparison to the visible light microscope because of the large difference between the wavelengths of the two: the wavelength of electrons is about 300,000 times shorter than that of light photons(Inkson 2016). However, the record ~1.5 Å resolutions reached for some biological molecules by an electron TEM are much lower than the wavelength of 0.02 Å at 300 keV. Thus, the wavelength of an electron beam does not limit the resolution of images of biological molecules, which are limited by other factors discussed later (Milne et al. 2013).

The first EM was a transmission electron microscope (TEM) developed in 1931 by Knoll and Ruska, operating under the same principles as an LM. TEM uses a highly energetic electron beams instead of photons to 'traverse' the sample, reach the detector, and obtain an image (Knoll and Ruska 1932). Later, in 1942, the first scanning electron microscope (SEM) was invented, which employed a different approach; the SEM uses the electron beam to 'scan' the sample (Zworykin 1942).

The electron gun creates a stream of electrons under a high vacuum that is accelerated towards the sample. The apertures and magnetic lenses inside the microscope allow the electrons to be focused into a thin beam that irradiates and interacts with the sample. From

these electron–specimen interactions, an image can be created. Once an electron reaches the specimen, scattering can occur in two forms, *i.e.*, elastic and inelastic. Elastic, where the trajectory is modified, keeping velocity and kinetic energy constant, and inelastic, where electrons displace other electrons from their orbits in the sample and lose some of their energy. Elastic scattering is utilised in TEM to examine thin samples, while SEM benefits the most from inelastic interactions to study thick specimens (Bauer 1994).

As a result, SEM is well suited to study the surface topology of a sample. SEM can also help in the determination of surface chemical composition. This is because increasing atomic number results in an increased brightness of the backscatters. The challenging part of SEM technology is that the resolution is rather limited.

TEM is a challenging and expensive technique that requires a high vacuum andhigh voltage applied to fragile samples; however, this allows the visualisation of structures down to atomic scales through elastic scattering. The sample can be tilted to image different projections of the same molecules to allow reconstruction of a 3D structure (Echlin 2011).

1.1.2.2 The Principles of Cryo-EM

The determination of macromolecular structures is crucial for the mechanistic comprehension of biochemical and cellular systems. This field has been dominated by XRC for decades, especially in the case of high-resolution structures. However, XRC presents many technical limitations. It can be challenging to produce adequate quantities or concentrations of individual macromolecular complexes for crystallisation. Some macromolecules are laborious to crystallise, and others are hard to purify in specific functional states (Nogales and Scheres 2015). The SPA method of 3D electron microscopy (3D-EM) provides an alternative approach to determine the structure; it is based on, transmission electron microscopy (TEM) -and the 'projection-slice theorem'. This theorem states that a 3D image can be reconstructed by collecting a group of 2D projections at various angles. So, if the -orientations of the 2D projections are known, the corresponding slices of the 2D Fourier can be positioned in the 3D Fourier transform, and the original 3D structure can be determined computation ally by applying the inverse Fourier transform (Nogales and Scheres 2015).

The use of high vacuum in electron microscopes means that biological samples have to be frozen and kept at a low temperature during imaging, to prevent them from drying and losing their original structure. Secondly, biological samples have to be protected from radiation damage; this can be accomplished using a very low electron flux. Once a sample is prepared for TEM, it can be studied through negative staining – requiring a further preparation step in which the grid is covered by a solution of a heavy metal (*e.g.* uranyl acetate) or Cryo-EM. The latter requires blotting the sample onto a thin layer and quickly freezing it to achieve vitrification, preserving the biological material in its real state and protecting biological samples from radiation damage. However, this process only reduces the wear, and it is still necessary to use as few electrons as possible to further minimise radiation exposure, usually within the range of 20–40 e$^-$/Å², according to Nogales and Scheres (2015). This translates into noisy images with an extremely low signal to noise ratio: the averaging of thousands of images is required to heighten the signal from the noise floor. Another limitation of 3D-EM is the unknown relative orientation of the individual particles in the images; these must be determined computationally, and additionally, there may be a problem if that sample is not captured in different orientations (Nogales and Scheres 2015).

Finally, some macromolecular complexes may present multiple conformational or compositional states in one sample. This is a problem because the analysis of the 2D images requires homogeneous subsets of an identical 3D structure so that they may be averaged together. To overcome this, the Valle group generated many reference structures from pre-existing knowledge of ribosomal ratcheting (Valle et al. 2002). Another method that does not require prior knowledge of the sample's structural variability has been proposed (Scheres et al. 2007). This problem has been transformed into new possibilities through methods that use the maximum-likelihood (ML) approach; this allows a better understanding of molecular function through the concurrent analysis of many structural states(Nogales and Scheres 2015).

1.1.2.3 Resolution estimation of Cryo-EM

Fourier shell correlation (FSC) is the main criterion to determine resolution in Cryo-EM; however, there is no consensus on a threshold. It consists of independently calculating maps from half the datasets, comparing them across resolution and observing where they stop

correlating significantly. A value of 1.0 in the FSC curve represents perfectly correlating maps; as spatial frequency increases (and thus resolution), the correlation decreases, dependent on the strength of the signal in the Fourier domain: in poor resolution maps, FSC will decay rapidly with spatial frequency; in high resolution maps, the FSC will decay much more slowly (Dubach and Guskov 2020). In the gold standard approach, where each half dataset is refined independently, the value of 0.143 has been proposed to be used for estimating the resolution (Figure 1.1) (Rosenthal and Henderson 2003; Scheres and Chen 2012). Despite this, it is being debated whether this threshold is ideal and whether FSC is an adequate measure for resolution in cryo-EM (Dubach and Guskov 2020).
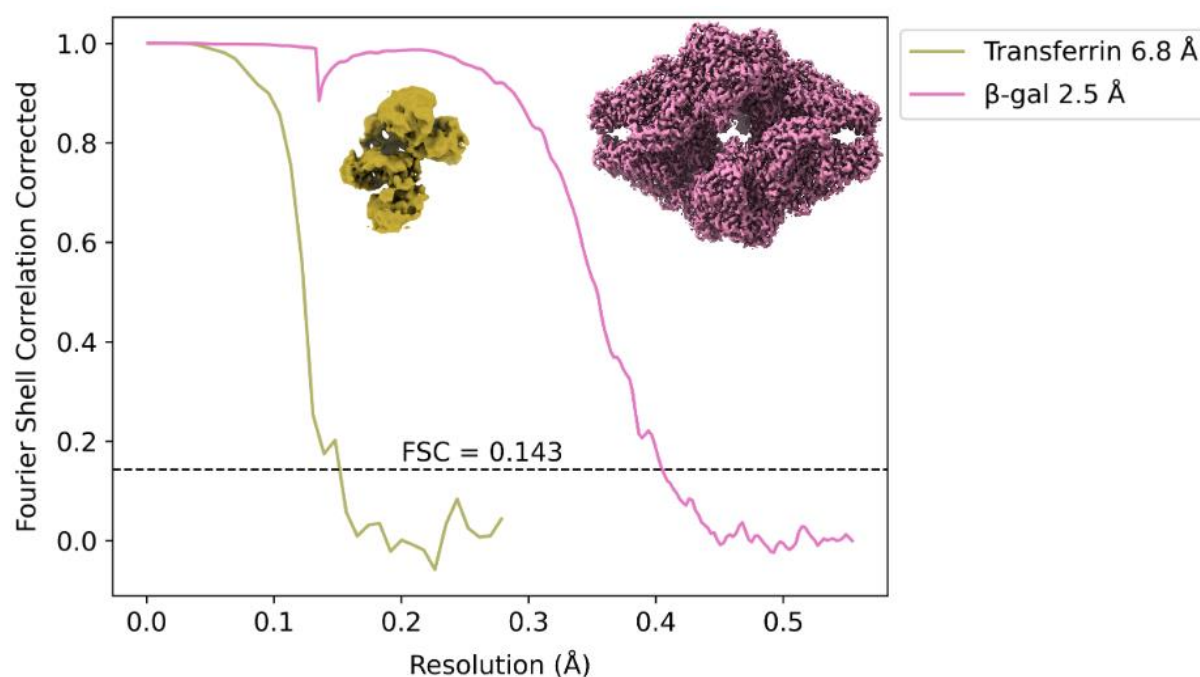
Figure 1.1 Comparison between two independent masked Fourier shell correlation (FSC) curves of low and high-resolution cryo-EM maps.

FSC of β-galactosidase for a map at 2.5 Å and FSC Human serum transferrin (hTF) for map 6.8 Å. Graphs data obtained with *RELION* (Zivanov et al. 2018), maps drawn with Chimera X (Pettersen et al. 2021).

1.1.2.4 FEGs, DEDs, Energy Filters and Improved Grid Technology

Recently, advances have been made in Cryo-EM in terms of resolution, contrast and data acquisition through the development of Field Emission Guns (FEG), Direct Electron Detectors (DED), energy filters and by improving grid technology, sample preparation, and by the development of software such as *RELION*. These advances have permitted the use of Cryo-EM to solve biomolecule structures at near-atomic resolution, reaching levels close to 1.2 Å (Nakane et al. 2020)**.**

1.1.4.5 Field Emission Guns (FEGs)

A field emission gun (FEG) is the modern electron microscope's source of electrons. It is a device capable of emitting a beam of finely tuned electrons with a very narrow energy spread, which goes through the sample, being scattered or not along the way, to be later detected with the use of a Direct Electron Detector (Krivanek et al. 1995; Tanaka et al. 1999). The beam can be scattered inelastically, in which case there is a net transfer of variable amounts of energy from the electrons to the atoms in the sample, which in turn causes noise in the image. It can also be scattered elastically, or it can go through without being scattered at all (Egerton 2011). To minimise noise and enhance image quality, it is therefore desirable to minimise the number of inelastic scattering events. This depends on the electrons' free path, which is the mean distance travelled by an electron between two subsequent scattering events and which depends on the electron's energy. The average free path for 200–300 keV electrons is in the range of 200–300 nm, so, limiting the sample thickness to 200 nm or less makes it likely that most electrons will go through unaffected, while those that are scattered will only be scattered once at most (Grimm et al. 1996; Feja and Aebi 1999).

1.1.4.6 Direct Electron Detectors (DEDs): image acquisition and preprocessing

Cryo-EM laboratories are currently employing direct electron-counting cameras. The cutting-edge models include the Gatan K3, Thermo Fisher Scientific Falcon IV, and the Direct Electron DE64. These detectors provide fast readout, and their detective quantum efficiency (DQE) allows for high sensitivity when recording low-dose electron images (McMullan et al. 2009; Faruqi, Henderson, and McMullan 2014). Motion correction in cryo-electron microscopy (Cryo-EM) refers to compensating for specimen and microscope stage motion during data acquisition. During data collection in Cryo-EM, a series of 2D projection images are captured from different specimen orientations using an electron microscope. However, various sources of motion, including specimen drift, stage vibrations, and thermal effects, can cause blurring and misalignment of the images, degrading the final reconstruction (Murata and Wolf 2018). The capability to record movies of many seconds in duration permits the alignment of frames (images) when the sample has moved because of mechanical drift induced by the exposure to the electron beam – termed motion or drift correction (Brilot et al. 2012; Campbell et al. 2012), which is a first step in preprocessing Cryo-EM data. *RELION* has an implementation of

this approach in the *MotionCor2* software (Zheng et al. 2017). Frames can also be aligned automatically with WARP or CryoSparc after recording (Mills 2021).

The second step in Cryo-EM preprocessing is Contrast Transfer Function (CTF) estimation and correction. CTF is a concept used in electron microscopy to represent the modulation transfer effects of an imaging method, relating the transfer of contrast from the object being imaged onto the recorded image (Fernández, Sanjurjo, and Carazo 1997; Murata et al. 2010), influenced by factors such as defocus, astigmatism, spherical aberration, and the wavelength of the electrons used (Zuo et al. 2017). The CTF describes how rapidly the contrast of the image varies as a function of spatial frequency due to such perturbations. It provides information about the contrast and resolution capabilities of the microscope system. It analyses the power spectrum of the acquired images that fit the mathematical model of the CTF parameters. It can be visualised as a plot or curve of the amplitudes of different spatial frequencies present in the image (Figure 1.2). The CTF is a fundamental property of the microscope that affects the quality of the acquired images and must be accurately estimated and corrected to obtain high-resolution 3D reconstructions of biological specimens. CTFFIND is an open source software program used in Cryo-EM to determine a microscope's contrast transfer function (CTF) parameters for each micrograph (Rohou and Grigorieff 2015). Alternatively, GctF is a close source software program that can be used (Kai Zhang 2016) .
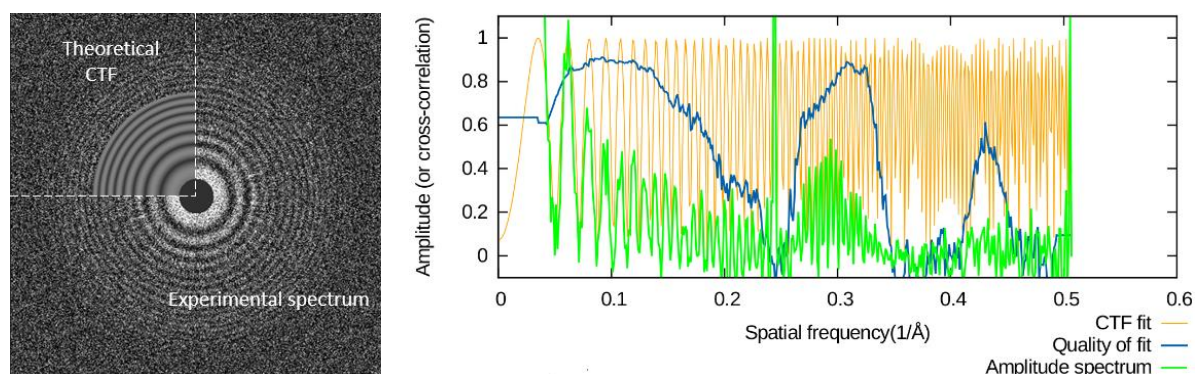
Figure 1.2 Contrast transfer function (CTF).
A thon ring shown at the left upper quadrant is the theoretical CTF. The rest of the plot represents the experimental spectrum. The CTFFIND output example is shown for an image from a β-galactosidase data set that led 3D maps derived at 2.5 Å resolution.CTF parameters were estimated using Alexis Rohou's and Niko Grigorieff's CTFFIND4.1. The left plot is an example showing the radial average of the amplitude spectrum.

1.1.4.7 Energy filters

While working with thin samples helps reduce noise caused by inelastic scattering, as explained above, it is not always possible to reduce sample thickness to the desired levels when working with large structures. A very effective way to reduce this type of noise is by using an energy filter. These filters separate beams of electrons based on differences in energy, much in the same way that an optical prism separates light based on wavelength. The filter has a small slit that only electrons with the specific energy of the original source (the FEG) will go through, while the rest of the electrons are deflected above or below the slit and therefore do not pass.

Phase-contrast comes from the interaction of the elastically scattered and non-scattered beams. Defocusing a phase-contrast microscope increases the contrast but only for high-spatial frequency information. A phase plate changes the phase of the non-scattered beam at the back-focal plane to improve contrast at low-spatial frequencies. Amplitude contrast comes from inelastically scattered electrons being lost somewhere, either by hitting the objective aperture or being filtered by an energy-filter. Importantly, it should be reiterated that the vast majority of contrast in electron microscopy comes from phase-contrast (Malac et al. 2021).

When an electron is scattered inelastically by the sample, its energy changes slightly. An Energy Filter is able to detect such changes and filter the inelastically scattered electrons out. Therefore, only electrons that are scattered elastically or that are not scattered at all (whose energy remains unaltered) make it through the filter. This reduces image noise and therefore increases contrast by about 16% (Fukuda et al. 2015).

1.1.2.8 Grid technology and sample preparation

The grid is an essential part of a Cryo-EM workflow. Its design and the type of substrate used have a big impact on the vitrification process as well as on the distribution of particles within the amorphous ice matrix. It has been proposed that grid stiffness and flatness help reduce beam-induced movement during data acquisition when samples are tilted, which is why some grids are fabricated with doped silicon carbide which has enhanced low-temperature conductivity and can be made very flat (Yoshioka, Carragher, and Potter 2010).

Other grids are made of gold to ensure the highest possible thermal conductivity to ensure quick fixation of the samples in a matrix of vitreous water before it has time to crystallise, allowing researchers to immobilise proteins in an almost native aqueous environment (Passmore and Russo 2016). Additionally, the use of graphene and graphene oxide as substrate materials can improve sample distribution, and antibody-based affinity grids allow the use of low concentration solutions and even native samples when preparing the specimen for imaging (Yu, Li, and Jiang 2016).

The grid itself is not the only part of the sample preparation stage that has evolved. The way the sample is dispensed has also advanced dramatically, allowing the use of sample quantities orders of magnitude lower than the traditional blotting techniques. A recently developed technology uses a pin printer to deposit sub-nanoliter size samples directly on the grid, followed by jet vitrification with liquid ethane to produce a more uniform layer of sample in a predefined pattern (Ravelli et al. 2020).

1.1.2.8.1 Common problems with sample preparation and vitrification

To prepare a biological sample for Cryo-EM, the macromolecules must be embedded in a support film of holey carbon. The excess water is blotted away. After this, the film is quickly plunged into liquid ethane to be vitrified. Ideally, the product in this state will be identical to the original sample before vitrification, but this is not the case most of the time. Four problems may be encountered when preparing these samples: the preferential orientation of macromolecules within the thin film might not be compatible with the measurement (Tan et al. 2017; Carragher et al. 2019; D'Imprima et al. 2019); fewer particles than expected might be present within the holes of the support film (Meyerson et al. 2014; Snijder et al. 2017) ; the particles disintegrate within the aqueous film; and the sample might aggregate.

In 2019 Carragher and colleagues developed a multiple-choice questionnaire about experiences regarding sample preparation (Carragher et al. 2019). The results concluded: optimisation for the process is required, and no current method to overcome the problems of sample preparation can be identified as useful in advance. Successful preparation on the first attempt seems to be more achievable with icosahedral particles (viruses)and helical proteins.

Nonetheless, adequate sample preparation seems rare most of the time, especially with semi-soluble and partly hydrophobic macromolecules (Carragher et al. 2019).

1.1.2.9 Counting vs integrating detectors

When electrons reach the Complementary Metal-Oxide-Semiconductor surface (CMOS), the signal can be employed in two modes to produce an image: integrating and counting modes. The first, the integrating mode, interprets the signals as a value of intensity that is required summed up as multiple electrons traverse a given pixel. In the integration mode, images can be recorded faster, and the throughput is superior. In the counting mode, the signal is interpreted as individual electron events. These are counted and summed up as many electrons per pixel; it requires a lower electron influx to detect specific hits. The counting mode eliminates Landau noise and has a higher DQE (Mendez and Stagg 2018).

Song and his colleagues in their study (2019) have tested the capabilities of a Falcon 3 detector for determining the structure of three distinct particles: for the tobacco mosaic virus, an integrating mode was utilised; for the hepatitis B virus capsid-like particle and β-galactosidase, both integrating and counting modes were utilised, and the results were compared.

In this study, while the counting mode resulted in better outcomes than the integrating mode for the same number of molecular weight particles below 0.5 MDa, such as β-galactosidase (465 kDa), this difference can be compensated with a three-times larger throughput in the integrating mode. Macromolecular complexes weighing several megadaltons gain no resolution advantage by employing the counting mode over the integrating mode (Song et al. 2019).

1.1.2.10 Cryo-EM data analysis

The analysis of images and their reconstruction into a 3D model represent a challenge for many reasons. The captured images represent two-dimensional density projections that emerge as the result of electron–particle interactions; naturally, constructing a 3D model from 2D images is difficult and is mostly resolved through the Fourier slice theorem. This

means that once the 2D projections of an individual 3D particle represent Fourier-analysed and placed at the right angle, they form a 3D model in Fourier space (Nogales and Scheres 2015).

This creates another problem: each projection has to be placed to match the location of the particle. This can be overcome with the maximum-likelihood (ML) approach, iteratively determining the Bayesian likelihood of an image to represent the particle using a reference model for orientation. The possibility of utilising a wrong model is called model bias. Noise can also promote incorrect results when it aligns in a way that resembles the model (Scheres and Chen 2012).

Finally, when the size and appearance of the particles are heterogeneous, which points towards particle aggregation or other types of instabilities (Scheres et al. 2007), the 2D images correspond to different 3D particles. The data can be classified (multireference refinement), so the averaging of projections can yield the corresponding 3D models.

1.1.2.10.1 *RELION* and cryoSPARC

Regularised Likelihood Optimisation (*RELION*) is open source software that employs an empirical Bayesian approach and has become ubiquitous for its accessibility to newcomers to Cryo-EM structure determination, as well as setting a standard for other computer programs in structural biology. In *RELION*-2 (2016), a pipelined approach was used for single-particle workflow, and it employed graphical processing units (GPUs) as hardware acceleration. The latest releases allow *RELION*-3 to automate jobs through Python, making the software more versatile. Central processing unit (CPU) acceleration increases the variety of available hardware to run the program efficiently (Zivanov et al. 2018).

Scheres's group has implemented a detection methodology for the final release, *RELION*-3.1. They have devised corrections of the symmetrical and unsymmetrical optical aberrations and magnification anisotropy aberrations. This was done for four published data sets, establishing its capability for correcting any aberrations for any given Cryo-EM data set (Zivanov, Nakane, and Scheres 2020).

Cryo-EM single-particle *ab initio* reconstruction and classification (cryoSPARC) is a proprietary computer program that minimises user input for refining 3D structures from single-particle images without requiring computer clusters. It employs algorithms and a simple graphical user interface to enable users without expertise in Cryo-EM to process data, even from a consumer-grade desktop computer. The minimum requirement is to employ at least one NVIDIA GPU (consumer or professional grade). The GUI allows multiple users to interact with the program and data, including streaming results being computed in real-time (Punjani et al. 2017).

1.1.3 Model building

1.1.3.1 Software suites for model building

As the intermediate result for both macromolecular X-ray crystallography and single particle Cryo-EM is a map containing information (albeit of different nature, *i.e.*, electron density *vs* electron potential) on the position of electrons, several existing crystallographic software packages have been repurposed to work with single particle Cryo-EM data. The Computational Crystallography Project n4, or CCP4 (Agirre et al. 2023), is a UK-based consortium that compiles and distributes crystallographic software, including Servalcat (Yamashita et al. 2021), *Coot* (Casañal, Lohkamp, and Emsley 2020) and Buccaneer (Hoh, Burnley, and Cowtan 2020). Several of these tools have been released as part of sister project CCP-EM, operating on similar premises. Both suites offer graphical user interfaces and programmatic access to the different tools, and maintain the formats pertinent to the techniques, *e.g.* CCP4 deals with the MTZ format for reciprocal space data representation (map coefficients, reflection data), and CCP-EM deals with the MRC format for real space map data. Outside of Europe, the PHENIX consortium integrates both crystallographic and Cryo-EM software tightly with the Python language, offering an easy-to-use graphical user interface and constant updates, which are released in the form of nightly builds. Coincidentally, the CCP-EM suite also offers nightly builds. All these suites have one model building program in common: The Coot software for model building, refinement and validation.

1.1.3.2 Interactive model building: *Coot*

Interpreting structural data requires model building and refinement; model building is achieved through programs such as *Coot*. This molecular graphics software, originally

designed for crystallography data, presents an interactive environment for model building/refinement alongside validation, which is one of the most important aspects of *Coot*'s functionality.

Coot allows the user to input and refine complete domains or chains. It allows fitting a homologous structure manually near the centre of the domain to be fitted and a local translation/rotation search for model placement in Cryo-EM. The Cryo-EM module – one of many pluggable modules in *Coot* – contains a sharpen/blur tool that may help in map interpretation, particularly the main chain (sharpening) and side chains (blurring); this accomplished by applying a positive or negative B factor correction to the data in reciprocal space. Jiggle-fit can be performed through the 'Morph' tool, found in the 'Modelling' module. Rotation and translation hypotheses are evaluated, selected, and the solutions are subjected to rigid body refinement taking advantage of multicore processing power (Casañal, Lohkamp, and Emsley 2020).

The A and B conformations of Deoxyribonucleic acid (DNA) and Ribonucleic acid (RNA) can be generated with *Coot*. It also incorporates a fitting tool for *N*-linked carbohydrates, but more work is required for building high-quality models of glycoproteins with Cryo-EM data, which has been a difficult task for structural biologists (Crispin et al. 2007).

The validation tools have been essential to model building in *Coot*, helping users detect and resolve problems. Validation provides tools such as Geometry analysis, Density-fit analysis, Rotamer analysis, Difference Map Peaks, and Ramachandran and Kleywegt plots (Casañal, Lohkamp, and Emsley 2020).

1.1.3.3 Automated model building (Buccaneer and PHENIX)

Aside from representing a challenging and time-consuming task, interactive interpretation of Cryo-EM maps is prone to errors, especially with large particles or molecules with unknown sequences or components (Liebschner et al. 2019). This makes automation desirable, so automated model-building pipelines for Buccaneer have been updated for Cryo-EM (Hoh, Burnley, and Cowtan 2020), as well as new programs for PHENIX focused on automated Cryo-EM map/model analysis (Liebschner et al. 2019).

Buccaneer's automated model building requires the user to input a map file (MRC format), a sequence file (FASTA format), an estimated map resolution and a global sharpening parameter (default value is 0, meaning no sharpening will be performed if not modified). Buccaneer iterates model building, and Servalcat iterates refinement tasks; this defaults to five cycles and consists of the model building followed by refinement. The resulting structure is employed as the initial model for the next cycle. The Cowtan group observed improved results when employing a Cryo-EM reference map and model compared with XRC references; however, they reported a decrease in performance with lower overall map resolution (Hoh, Burnley, and Cowtan 2020).

Liebschner et al. (2019) describe the steps for automated structure determination in PHENIX. The starting map needs to be assessed for quality, as a Cryo-EM map can be affected by heterogeneity, radiation damage, and beam-induced sample movement; the estimation can be done with Mtriage (phenix.mtriage). The map is optimised with Autosharpen function (phenix.auto_sharpen), applying map sharpening. The 'Dock in map' (phenix.dock_in_map) tool allows for the docking of models into the map; this is done if the model (or part of it) is known. If the structure (or part of it) is unknown, the model will be built *ab initio*; the tool 'Map to model' (phenix.map_to_model) automatically builds the atomic model by interpreting the Cryo-EM map, and the resulting models are real-space-refined (Liebschner et al. 2019).

1.1.3.4 Chemical restraints for model refinement

Refining low-resolution macromolecules requires providing restraints to minimise overfitting and stabilise refinement; for example, by providing prior chemical knowledge on bond lengths, angles, torsion angles and planar groups. These are called local restraints. Longer-range restraints are frequently generated through the CCP4 ProSMART software (Nicholls et al. 2014) when working with *Coot* or CCP4mg. The program's authors reported using 4.2 Å for local distance restraints, articulated through the use of the Geman McClure robust estimator; however, support for multiple CPUs in *Coot* allows users to work with longer distances (6–7 Å) and many more restraints (Nicholls et al. 2018). *Coot* provides EM-specific tools that aid in the process. From the 'Cryo-EM' module, *Nudge Residues* can be used to

produce alternative hypotheses for the residues register. This is useful when a residue along the chain is in the position that another residue should occupy. *Backrub Rotamers* are another useful tool in *Coot* that is employed for fitting side chains (Casañal, Lohkamp, and Emsley 2020) without causing distortions in the main chain. *Coot* relies on the standard CCP4/CCP-EM monomer library for reading chemical geometry restraints (bond lengths, angles, torsions, planes, chiral volumes…), which are specified in the CIF format. The CCP4/CCP-EM software of choice for the generation of new geometric restraints is AceDRG (Long et al. 2017), which mines chemical information from the Crystallography Open Database (COD) very much in the way the *Grade* software (Global Phasing Ltd) does with the Cambridge Structural Database (CSD). The CCP4 monomer library was originally generated with the LIBCHECK software (Schüttelkopf and van Aalten 2004), although newer versions have been produced with the AceDRG software, including new dictionaries for carbohydrates in the pyranose form (Atanasova et al. 2022).

When it comes to refining structures of carbohydrates, either as ligands or covalently attached to protein (glycosylation), the use of torsion restraints might become necessary if refining a model against low resolution data, otherwise conformational anomalies may occur (Agirre, Davies, et al. 2015). It has been reported that several CCP4 dictionaries for pyranose sugars were erroneous in the past, leading to modelling and refinement errors (Agirre 2017). Recently, a study of pyranose ring conformation covering both XRD and Cryo-EM showed that Cryo-EM seems to be producing far fewer conformational anomalies than XRD despite the median resolution of the structures being lower (higher number) in Cryo-EM compared to XRD (Atanasova, Bagdonas, and Agirre 2020). This, along with the ongoing debate on map resolution, led to speculation that Cryo-EM maps were perhaps more informative than XRD ones at the same calculated/estimated resolution. Considering how easily conformational anomalies are introduced when working with low resolution XRD data in the absence of torsion restraints, a good test for pyranose sugars would perhaps be whether pyranose rings could be refined to a chair conformation against a low (2.5 - 3.0 Å) resolution map without activating torsion restraints.

1.1.3.5 Reciprocal Space Refinement

In PHENIX (Liebschner et al. 2019), as stated above, the automated atomic model building process is followed by real-space refinement with secondary-structure restraints (Liebschner et al. 2019). This is a popular refinement tool for Cryo-EM models, and it is performed with the Phenix.real_space_refine program. Most software was developed for crystallographic data and reciprocal space refinement (or Fourier-space refinement, FSR). However, an atomic model from Cryo-EM data can be refined by employing a reciprocal space target by converting the map into Fourier coefficients. These values are used through standard refinement protocols for crystallographic particle refinement. Afonine and collaborators stated that it is important to fully convert the map, instead of subsets, into Fourier coefficients because FSR is conditioned by all model parameters. In PHENIX, reciprocal space refinement can be implemented with phenix.refine (Afonine et al. 2018).

1.1.4 X-ray Crystallography (XRC) vs single particle Cryo-EM

As mentioned in the section later (3.1.2.1), our protein of interest, β-Galactosidase, as a member of the retaining β-glycosidases enzyme group, has a central role in various biologically essential processes. These enzymes have previously been complexed with inactivators, tagged ligands, untagged ligands and ABPs, allowing for active site resolution by either X-ray crystallography or single particle Cryo-EM. Both are valuable in different situations (Doerr 2015).

Until the beginning of this project, according to our current knowledge, no Cryo-EM 3D structures of complexes for the epoxide- and aziridine-based probes have been targeted for bacterial retaining enzyme β-galactosidase LacZ. Solving an atomic model of a probed glycosidase enzyme with either covalent or noncovalent inhibitors by Cryo-EM requires high enough resolution to decipher the 3D structure and the complex interaction of the active centre channels of the glucosyl moiety. This is a challenging task; however, Cryo-EM has a much lower sample requirement than XRC, making its qualifications suitable for hard-to-synthesise probes and enzymes (*e.g.* mammalian enzymes and membrane proteins; light-sensitive probes). Additionally, there is no need to obtain crystals for systems operations, which can be challenging.

Cryo-EM may reveal why the β-galactosidase can tolerate some substituents on substrate analogues and putative inhibitors and why certain similar enzymes cannot. This can help record the specificity and may ultimately lead to the evolution of activity-based probes and next-generation inhibitors (Renaud et al. 2018).

1.1.4.1 How comparable are XRC and single particle Cryo-EM?

The assessment of Cryo-EM maps is done by the Fourier Shell Correlation (FSC) method in which the two 3D reconstructions are formed after particles are divided into two half-sets, and then the 3D structure is reconstructed independently based on each set (Sorzano et al. 2017). It was proposed that a correlation cutoff value of 0.5 between the two maps derived for each set can be used, as the threshold value of 0.5 is associated with a phase error of about 60 degrees in crystallography (Lunin, Yu. Lunin, and Woolfson 1993). However, at a later stage, to make the resolution estimation for Cryo-EM map comparable to that of X-ray crystallography, a more realistic cutoff value of 0.143 was proposed (Rosenthal and Henderson 2003).

Resolution in Cryo-EM and X-ray crystallography cannot be understood as in light microscopy because XRC and Cryo-EM interpret data through Fourier space (Dubach and Guskov 2020). Lord Rayleigh defined resolution in light microscopy for the first time to be the shortest distance through which two point-sources could be differentiated (Rayleigh 1879). Nevertheless, this description is unsuitable for Cryo-EM and X-ray crystallography because both methods employ Fourier space to estimate data resolution. The characteristics of the experimental data and Fourier space make the resolution determination inherently less precise (Dubach and Guskov 2020).

In terms of resolution, crystallography has achieved atomic resolution, reaching 0.48 Å (Schmidt et al. 2011); single-particle analysis has reached the near-atomic resolution of 1.54 Å (Kato et al. 2019). In a recent study performed by Nakane and coworkers, the utilisation of new equipment allowed high-resolution reconstructions for single-particle Cryo-EM. Employing a new Cold Field Emission Electron Gun (cold-FEG), a new energy filter and a Falcon IV camera, the team achieved a 1.7 Å resolution for β3 GABAa, and a resolution of 1.22 Å for mouse apoferritin (Nakane et al. 2020).
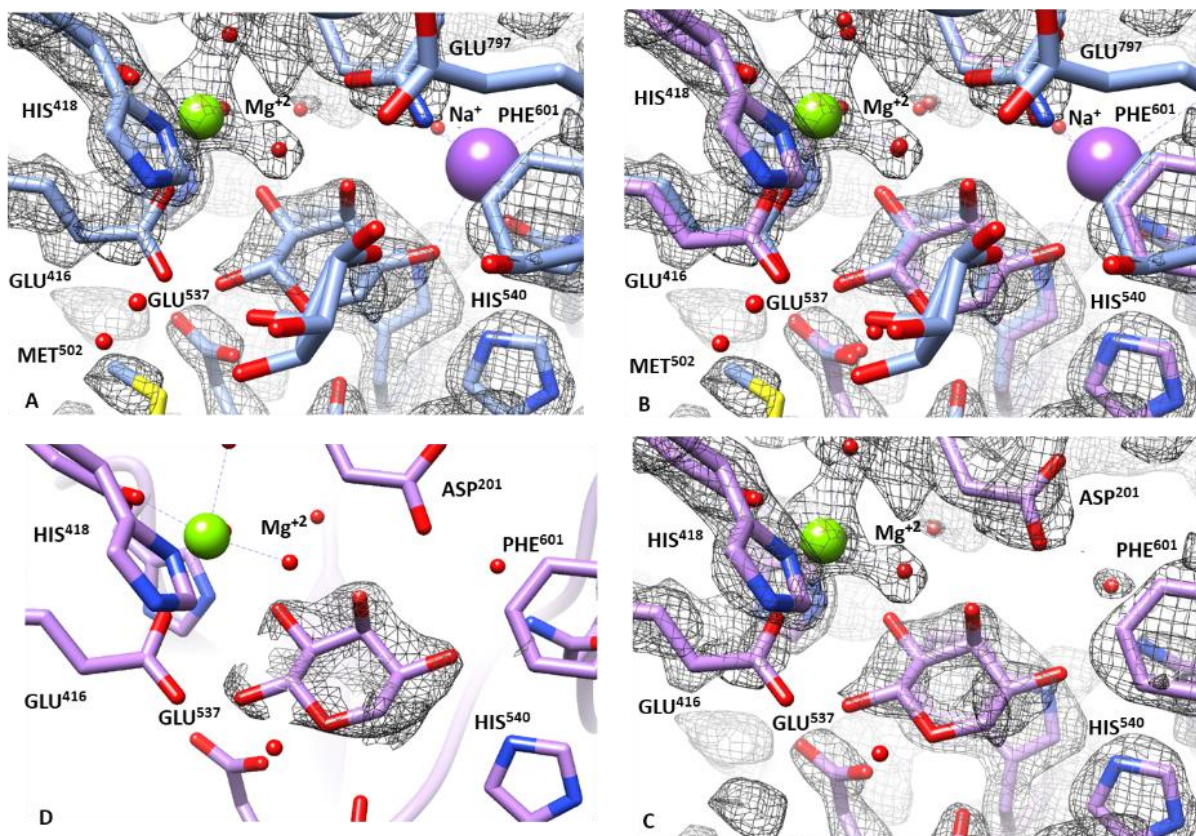
Figure 1.3 Comparison of β-galactosidase Cryo-EM and XRC structures bound to L-ribose.
(A) The XRC blue model (PDB code: 4DUX) (Wheatley et al. 2013) fitted in Cryo-EM map. (B) Both the Cryo-EM purple model (PDB code: 6TSK) (Saur et al. 2020) and the blue XRC model were superimposed on the EM map. (C) The EM model is fitted into the EM map. (D) EM map corresponding to the L-ribose ligand. The figure was generated with Chimera (Pettersen et al. 2004).

Reuben Huber's team has differentiated between a β-galactosidase complex with L-ribose, solved to 2.3 Å by both X-ray crystallography (PDB code: 4DUX; Figure 1.3, blue model) (Wheatley et al. 2013) and Cryo-EM map (PDB code: 6TSK; Figure 1.3, purple model) (Saur et al. 2020). The L-ribose molecules superimpose well, with minor variations in the positions of the sidechains. The 4DUX model indicates a second L-ribose molecule binding, while the Cryo-EM structure does not.

Chapter 2   Materials and Methods

2.1 Over Expression and sample preparations of β-galactosidase

It was one objective of this study to identify experimental protocols that could yield samples with the purity levels required for Cryo-EM. With this aim, the study investigated bacterial strains and growth conditions, Polymerase chain reaction (PCR) cloning, restriction enzyme digestion, agarose gel electrophoresis, expression, purification β-galactosidase and analysis under sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli 1970).

2.1.1 Cloning elements

2.1.1.1 The *lacZ* gene of *E. coli*

β-Galactosidase gene sequence was taken from NCBI GenBank (http://www.ncbi.nlm.nih.gov/) after searching the Protein Data Bank (PDB), which produced PDB ID code 4V40, which pointed to FASTA sequence file (V00296), which can be verified at the EBI site: https://www.ebi.ac.uk/ena/browser/view/V00296. The DNA sequence is 3075 bp, including a stop codon. The amino acid sequence is thus 1024 aa. Comparing the Uniprot sequence to the PDB counterpart showed that the first methionine residue was missing in the PDB file, presumably due to N-terminal methionine being flexible or disordered.

2.1.1.2 Designing primers

The oligonucleotide primers were designed using the York structural biology ligation-independent cloning (YSBLIC) protocol (Bonsor et al. 2006). The gene was cloned into the LIC site of the human rhinovirus (HRV) 3C protease cleavage site, which allowed the His-tag (to be added to the protein's N-terminus) to be cleaved off during purification. Primers with the following sequences were purchased from Eurofins:

Table 2.1 List of primers used in cloning and verifying the construct sequence of β-galactosidase. The vector pET-YSBLIC 3C compatible part of the sequence is bold, and the insert compatible part is in plain text.

Primers	Primer sequence	Purpose
E. coli lacZ gene forward primer	5'**CCAGGGACCAGCA**ATGACCATGATTACGGATTC3'	Amplification of *E. coli lacZ* gene (insert)
E. coli lacZ gene reverse primer	3'**GAGGAGAAGGCGCGTTA**TTTTTGACACCAGACCAAC5'	
T7 forward promoter	5'AAATTAATACGACTCACTATAGGG3'	Verifying N and C-terminus of plasmid sequence
T7 reverse terminator	3'TTATGCTAGTTATTGCTCAGCGGT5'	
Forward primer	5'GTTTGCCGTCTGAATTTGAC3'	Verifying the middle plasmid sequence
Reverse primer	3'ATCAGCAAGTGTATCTGCCG5'	

2.1.1.3 Genomic DNA (template)

The genomic DNA of *E. coli* BL21 (DE3)-Gold cells were extracted and employed as a template. A 100-μL aliquot of competent cell stocks was added to 400 μL *Luria-Bertani* (*LB*) media. Cells were grown for 1 hour with vigorous shaking (250 rpm) at 37°C. The incubated cells were pelleted; then, the genomic DNA was purified using the PureLink Genomic DNA Kit (Invitrogen). The DNA was eluted, and the final concentration was measured at 127.6 ng/μL. Genomic DNA purification was carried out using spin column-based centrifugation with collection tubes used for binding DNA; purified DNA was then stored at -20°C.

2.1.1.4 pET-YSBLIC 3C Vector

The *lacZ* gene was cloned into a pET-YSBLIC 3C vector (Fogg and Wilkinson 2008), Figure 2.1, which is based on the pET-28a (Novagen) derivative vector at size (5369bp), using the ligation-independent cloning (LIC) protocol (Aslanidis and de Jong 1990) method developed by the Structural Genomics Consortium (SGC). The LIC site is for cloning the inserts which are

introduced between the NcoI and NdeI restriction endonuclease cleavage sites which include HRV-3C protease cleavage site and N-terminal hexa histidine tag-coding sequence.

This bacterial expression vector carries a T7 promoter, and the kanamycin resistance genes. The LIC site also contains recognition sequences for a pair of restriction endonucleases, BseRI (5-GAGGAG-3) and AscI (5-GGCGCGCC-3).
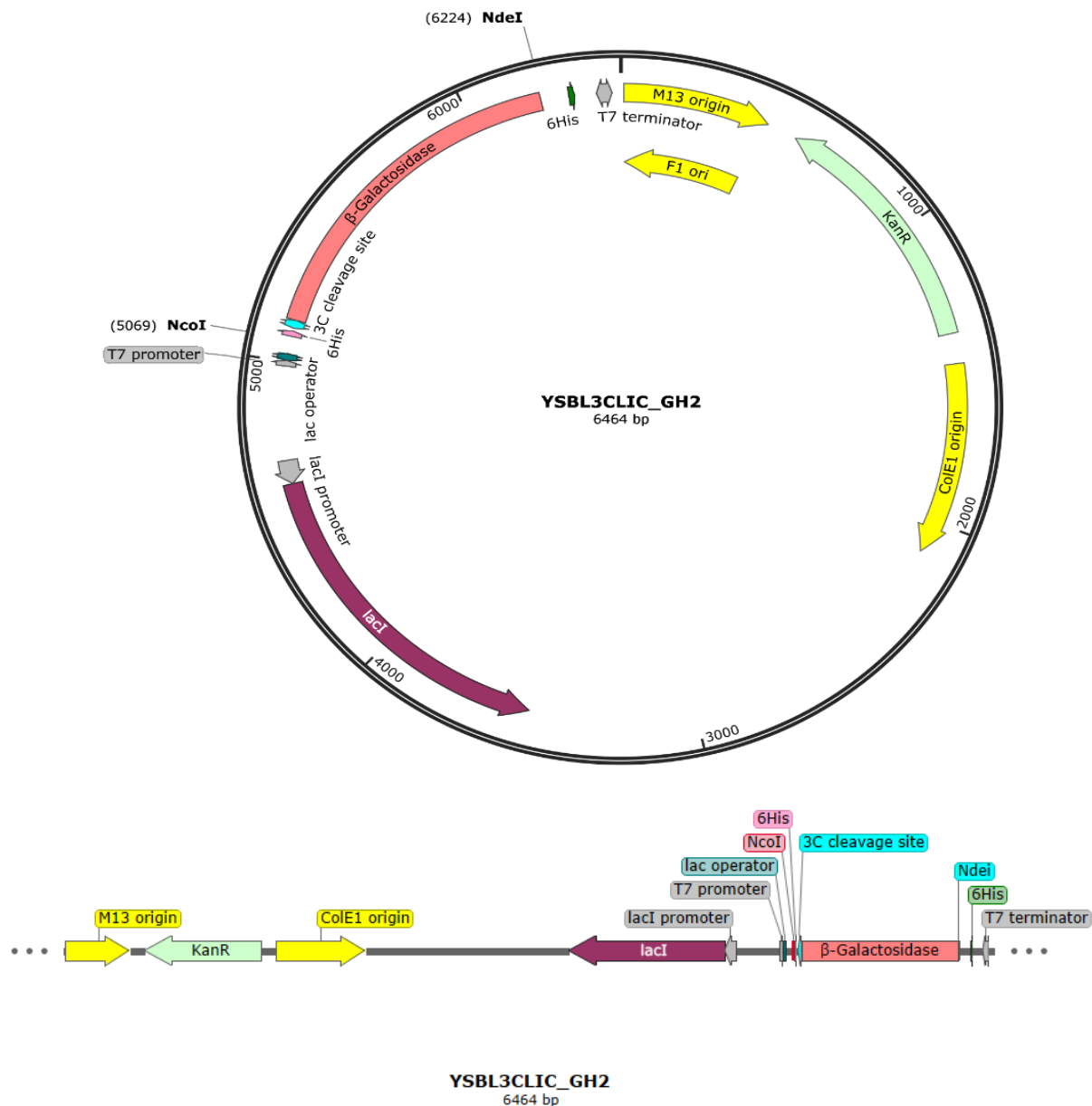
Figure 2.1 Backbone map of a circular and linearised pET-YSBLIC 3C vector.
The *lacZ* gene is present in the pET-YSBLIC 3C vector. It is based on pET-28a(+)vector with a modified cloning position with open reading frame (ORF) and a hexa His-tag (His) cleavable at the HRV 3C protease site (3C). It has a Lac operator (LacO) and repressor gene (LacI) for IPTG induction, a T7 promoter (T7P) and terminator (T7T), a binding site of the ribosome (BSR), a kanamycin resistance gene (KanR) and two replication sites (ori, f1 ori). The figure was obtained from SnapGene.

2.1.2 Transformation of the plasmid into the *E. coli* Tuner (DE3) strain (Novagene) cells

After the sequence of *the lacZ* gene of β-galactosidase from *E. coli* was cloned into the pET-YSBLIC 3C vector (Fogg and Wilkinson, 2008), with a thrombin-cleavable N-terminal hexahistidine tag. The plasmid pET-YSBLIC 3C encoding the β-galactosidase *lacZ* gene was transformed into the *E. coli* Tuner (DE3) strain (Novagene) cells for expression of β-galactosidase.

2.1.3 PCR amplification of the *lacZ* gene

PCR was conducted to amplify the gene using forward and reverse oligonucleotide primers and the genomic DNA template with Q5 high-fidelity DNA polymerase (M0491, New England Biolabs) without using a 5X Q5 high GC enhancer.

2.1.4 Polymerase Reaction of LIC Inserts

The bacteriophage T4 DNA polymerase at volume 0.4 μL from (Novagen) was used to generate the sticky ends of LIC inserts for digestion. Linearised inserts were incubated with mixture of 2 μL T4 polymerase buffer (10 mM), in the presence of 2 μL a single deoxyribonucleotide (dTTP) (Stock 25mM), 1 μL DTT (Stock 100 mM), 0.5 μL H_2O at 22 °C for 30 minutes, followed by 20 minutes of incubation at 75 °C to stop the reaction. The inserts were purified using Qiagen QIAquick Purification Kit and eluted in autoclaved MilliQ water.

2.1.5 Cloning using pET-YSBLIC 3C vector

The T4 polymerase reaction of pET-YSBLIC 3C vector sample was kindly given by a former PhD student Ms Chan in Professor Gideon Davies's group in the (YSBL) at the University of York, UK, before the commencement of this project.

BseRI is a type of IIS restriction endonuclease with the recognition sequence (5-GAGGAG-3) to bind a unique restriction endonuclease site. BseRI was treated pET- YSBLIC vector to produce a linearized molecule with three overhangs caused by the staggered cleavage 8 and 10 base pair (bp) upstream of the recognition sequence on each strand.

2.1.6 Annealing of pET-YSBLIC 3C vector and insert

T4 polymerase was prepared by taking ~50 ng/µL of pET-YSBLIC 3C vector and incubating for 10 min at room temperature (22 °C). After that, 2 µL of insert T4 pol LIC reaction was added to 1 µL of this T4 pol. 1 µL of EDTA was used to make up the volume to 4 µL. After mixing, the reaction was allowed to remain at room temperature for 10 minutes

2.1.7 Transforming pET-YSBLIC 3C plasmid to *E. coli* (Tuner (DE3))

The pET-YSBLIC vector T4 pol reaction was used in annealing and transformation. The plasmids containing the gene of interest, obtained from LIC annealing, were transformed for protein expression into *E. coli* Tuner (DE3) Competent Cells, Novagen Sigma Aldrich (70623) .

2.1.7.1 Transformation and overexpression of lacZ gene

An aliquot of Tuner (DE3) of 100 µL was thawed at room temperature, and 1 µL of plasmid prep was added to the tube, immediately. Then, the tube was incubated in ice for 30 minutes before being shock heated at 42 °C for 30 seconds. It was then incubated on ice for another 2 minutes.

A volume of 950 µL Super Optimal broth with Catabolite (SOC) mediaum (no antibiotics) was incorporated in the tube and incubated with shaking for 1–1.5 hours at 37 °C. Cultures of 50 µL and 100 µL were plated on *LB* agar plates at a concentration of 35 µg/mL kanamycin antibiotics and incubated at 37 °C overnight. Plates were kept in the fridge until doing the expression.

2.1.8 Expression of Bacterial β-Galactosidase

The plasmids pET-YSBLIC-3C were employed for overexpressing the target gene in Tuner (DE3) *E. coli* cells on a large scale. An overnight culture was grown in two 50 mL Falcon tubes containing 10 mL of *LB*, with the 10 µL kanamycin antibiotics at a concentration of 35 mg/mL and at a final concentration of kanamycin in the culture is 0.035 mg/mL.

A sterile loop was used to transfer a single colony of *lacZ* gene transformation, from a freshly streaked agar plate to the tubes to incubate at 37 °C and 180 rpm. Similarly, two glycerol

stocks were stored at -80 °C, which can be used in the future, if need be, using 1.00 mL starter cultures and 0.8 mL of 30% (v/v) sterilised glycerol.

About 10 mL of starter cultures were inoculated in 1 litre of LB, and 1 mL of kanamycin antibiotic (concentration 35 mg/mL) added in two-litre flasks. The tubes were incubated for 2.5 hours at 37 °C with shaking at 200 rpm. The bacterial growth was monitored by measuring the optical density (OD) at 600 nm wavelength until the OD was attained a filter. Just before adding the inducer isopropyl-β-d-thiogalactoside (IPTG), a 1 mL sample of the culture was transferred to a 1.5 mL tube and spun.

The supernatant was thrown away and stored in the cell pellets at -20 °C, to act as a pre-induction or pre-expression control sample for an SDS-PAGE. The vector system of the gene expression was induced with 0.75 mL of 1M IPTG and allowed to express the proteins overnight for 21 hours, at 16 °C and 1800 rpm.

On the third day, the culture was ready to harvest. A 1 mL sample of the culture was taken in a 1.5 mL tube and spun. Then, the supernatant was discarded, and the cell pellets were stored at -20 °C, to act as a post-induction or expression sample for SDS-PAGE. The remaining culture from a one-litre flask was then poured into a one-litre centrifuge tube and centrifuged for 20 min at 5000 g. The supernatant was poured off. Finally, the pellets were scraped and frozen at -20 °C to use at a further date.

Each of the *E. coli* Tuner-DE3 cell pellets of pre-induction and post-induction fractions was resuspended in 50 μL and 100 μL of Bug Buster Protein Extraction reagent Novagen cat 7089-3, USA, respectively. All fractions were vortexed, left for 5-minute and vortexed again; the total should not exceed 10 min. Insoluble and soluble protein fractions were boiled for 5 min at 95 °C then analysed at 12% SDS-PAGE, Figure (3.12).

2.1.9 Purification of LacZ β-galactosidase

A two-step metal affinity chromatography purification was used to remove LacZ β-galactosidase from the soluble fraction followed by size exclusion chromatography.

Prepacked HiTrap Chelating HP column and HiLoad 16/600 Superdex 200 preparation grade column were purchased from GE Healthcare.

Buffers A-C, degassed milli Q water and 20% (v/v) ethanol were filtered through a 0.22 µM Whatman filter before use. Thawing of cell pellets from 2 L of *E. coli* Tuner-DE3-β-galactosidase culture was done at room temperature and then these were again suspended in buffer A (10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5, 0.5 M NaCl and 10 mM imidazole pH 7.5), which contained a Protease Inhibitor Cocktail tablet. Complete Mini-EDTA-free, purchased from Roche Diagnostics, (Mannheim, Germany) was dissolved in buffer A to protect target proteins from protease degradation that would otherwise occur during cell lysis.

About 100 µL of DNase I at a concentration of 2 mg/mL in 50% glycerol and 5 mM of $MgCl_2$ were added to the buffer A to give the final volume about 20 mL. The buffer was added gradually by a filter syringe and the pellets were shaken vigorously to maintain homogeneity in the solution. Cell pellets were finally resuspended in 35 to 40 mL of volume. Sonication of the cells was done on ice for 2 rounds, 10 cycles each with an amplitude of 15 micrometres – each cycle ran for 5 seconds. The debris of the cells was spun down at 150,000 rpm and 4 °C for 30 minutes using a Sorvall RC5B centrifuge with an SS-34 rotor.

2.1.9.1 Affinity chromatography

The supernatant was decanted and filtered through a 0.2 µm ACRODISC filter. A 5 mL prepackaged HiTrap Chelating HP (pre-poured) column was run with 2 mL of milli Q water. Then, equilibrated with 25 mL buffer A at speed 3 with a flow rate of 10 mL/min. The filtered supernatant was subsequently injected onto the HiTrap column using a bench-mounted peristaltic pump (Pharmacia LKB Pump). The resin was then washed with 25 mL of buffer A at rate 3 mL/min.

Unbound material was removed by washing with buffer A, and the peristaltic pump was rinsed with 10 mL of water. The column of prepackaged HiTrap Chelating HP was then equilibrated with 5 mL of buffer A. After that, the HiTrap column shifted to an AKTA start, and a gradient from buffer A to B (50 mM HEPES 4-(2-hydroxyethyl)-1-piperazineethanesulfonic

acid (HEPES) pH 7.5, 0.5 M NaCl, 500 mM imidazole pH 7.5) was run at 100% buffer A to 100% B over 100 mL, with a rate of flow of 1 mL/min. The protein that was bound was eluted with a gradient of buffer B in the Superdex 200 column, and was operated using an AKTA Start.

Fractions of about 14 mL were taken to analyse by SDS-PAGE. The sample fractions of the desired protein were indicated by its approximate monomer molecular weight (~116 kDa). Finally, the Superdex 200 column was washed with filtered 20% (v/v) ethanol.

2.1.9.2 Size exclusion chromatography

Sample fractions containing the desired protein were combined, followed by the determination of its concentration by Eppendorf Bio photometer, with an absorbance at 280 nm of set equal to ~3.6 mg/mL and 50 mg/mL for the total 14 mL fractions. The fractions were concentrated to 2 mL using MWCO 30K Vivaspin 20 tubes, spinning at 45,000 rpm and 4 °C for 3 h using the Sigma refrigerated centrifuge. Before that Vivaspin 20 tube was washed with water to remove glycerol residue.

The protein lysate was then transferred to the Eppendorf tube and spun using Philip Harris's refrigerator centrifuge for 5 min to remove the cell pellet's cell residue and membranes. The soluble protein fractions having β-galactosidase were combined and then concentrated. The buffer was exchanged into buffer C (50 mM HEPES pH 7.5, 0.5 M NaCl and 1 mL of 1 mM DTT), the final concentration of buffer C after adding 1 mL of 1 mM DTT would be approximately 16.67 mM HEPES, 166.67 mM NaCl, and 333.33 mM DTT. Buffer C reduced the concentration of imidazole. After that, the protein was loaded on the Superdex S200 Size exclusion chromatography (SEC) column, which was equilibrated in buffer C. The target protein was eluted at 2 mL/min. Analysis of the protein samples was done with 12% polyacrylamide SDS-PAGE. The fractions containing β-galactosidase were stored at 4 °C.

2.1.9.3 Purification of Commercial β-Galactosidase

A batch of commercial *E. coli* β-galactosidase was purchased (Sigma–Aldrich, catalogue No. G5635), was purchased to act as a control. The lyophilised powder of ≥500 units/mg protein was dissolved in 1 mL of a buffer comprising 25 mM Tris–HCl (pH = 8), 50 mM NaCl, 0.5 mM Tris(2-carboxyethyl)phosphine (TCEP) and 2 mM $MgCl_2$ (Bartesaghi et al. 2014).

Size exclusion chromatography (SEC) was conducted on *E. coli* β-galactosidase by HiLoad 16/600 Superdex 200 preparation grade column equilibrated with the same buffer at 1 mL/min flow rate by an AKTA pure system (GE Healthcare). The highest β-galactosidase protein fraction of the resulting peak was determined to have 2.2 mg/mL protein concentration.

The protein sample was separated into 13 fractions, examined in 12% SDS-PAGE and visualised by the staining dye of Coomassie Brilliant Blue (CBB), depicted in the (Figure 3.15) After purification, the protein peak was collected, centrifuged for 30 min at 4 °C and 4,500 g in a Sigma 3-16KL benchtop centrifuge, and concentrated to 79.23 mg/mL in 20 Vivaspin centrifugal concentrator tubes using 30 kDa molecular weight cut-off (MWCO) (Sartorius Göttingen, Germany). The concentrated protein was stored at -80 °C following snap-freezing in liquid nitrogen.

2.1.10 Preparing complex-covalent inhibitors and activity-based probes

Quantitative evaluations were conducted using β-galactosidase activity and inhibition prior to qualitative analysis with Cryo-EM. The recombinant β-galactosidase was employed in two different substrate assays: one with 4-methylumbelliferyl β-D-galactopyranoside (4-MU-β-D-galactopyranoside) (Figure 2.2) and another with 4-Nitrophenyl β-D-galactopyranoside (4-*p*NP-β-D-galactopyranoside) (Figure 2.3), following an adapted protocol from (Martino et al. 2009).

2.1.10.1 Determining enzyme activity and the efficiency of irreversible inhibition kinetics assays

2.1.10.1.1 4-MU-β-D-galactopyranoside

A fluorometric assay was performed to estimate the hydrolytic activity of β-galactosidase and determine the dilution factor of complexes with available ABP or inhibitor that were reported targeting the β-galactosidase family. Serial dilutions of the (4-MU) with the buffer were measured to provide a calibration curve of fluorescent β-galactose assay standard of the

product (4-MU) concentrations against the average relative fluorescence units (Figure 3.16, A).

4-Methylumbelliferyl-β-D-galactopyranoside **4-Methylumbelliferone**
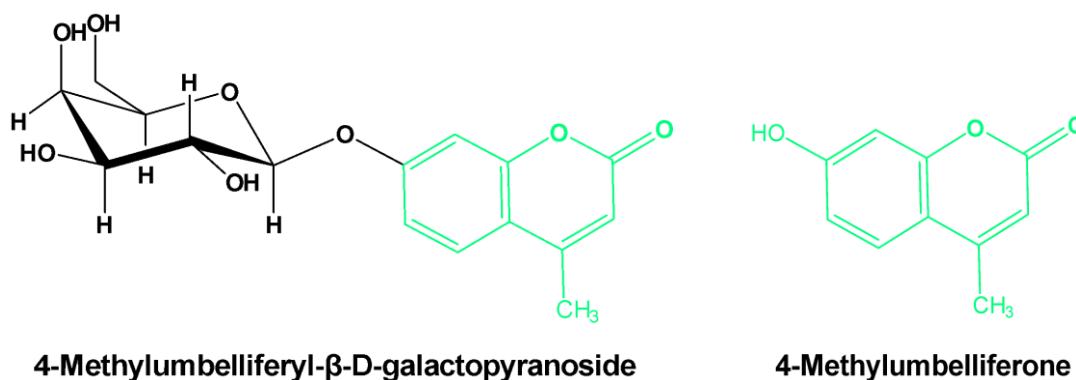
Figure 2.2 4-MU-β-D-galactopyranoside and 4-methylumbelliferone.
On the left, the fluorescent substrate 4-methylumbelliferyl β-D-galactopyranoside (4-MU-β-D-galactopyranoside) is cleaved by β-galactosidase, resulting in the fluorescent product 4-methylumbelliferyl (4-MU), on the right, was used to obtain a calibration curve. Figure created with ChemDraw.

Initially, the substrate was dissolved in Dimethyl sulfoxide (DMSO) to give a 200 mM (200×10^{-3} µM) stock solution, which was diluted in HEPES to provide a final concentration of 1 mM (1×10^{-3} µM).

The second assay involved a three-fold serial dilution of 10 µl of the enzyme with an unknown concentration (1X) (first well), added to 20 µl of the standard buffer (50 mM HEPES and 1 mM NaCl, pH 7.5) (each well) and 20 µl of 2 mM substrate 4-MU-β-D-galactopyranoside (each well). This resulted in a fluorescent intensity over time that corresponded to the product (4-MU). The fluorescence of the latter product was measured at a wavelength of 360 nm (Figure 3.16, B).

The enzyme inhibition and kinetics assays were measured using the substrate 4-methylumbelliferyl β-D-galactopyranoside (4-MU-β-D-galactopyranoside). Two putative inhibitors were dissolved in buffers (TB652) in DMSO and (LWA480) in HEPES, then diluted with HEPES to give a final stock concentration of 200 µM (TB652) and 5 mM (LWA480).

The inhibitor assay was conducted in Eppendorf by taking 1.2 µl of the inhibitor and adding to 4µl of 20 µM enzyme and 194.8 µl HEPES buffer and vortex and incubated for 15 min and read about 20 µL of reaction by adding 20 µL of the substrate then read after 55minutes.

All the final volumes in the wells were 50 µL for the substrate (4-MU-β-D-galactopyranoside); the fluorescence intensity was measured by a microplate reader (CLARIOstar, BMG LABTECH, Offenburg, Germany) at 360 nm.

2.1.10.1.2 4*p*NP-β-D-galactopyranoside

The colourimetric assay of 4-Nitrophenyl β-D-galactopyranoside (4*p*NP-β-D-galactopyranoside), (N1252-250MG, Merck), (Figure 2.3, right panel) was conducted to ensure optimal enzyme activity at pH 7 and 37 °C to, thus maximising the rate of substrate conversion and the sensitivity of the assay. β-Galactosidase catalysed the alteration of the substrate 4-Nitrophenyl β-D-galactopyranoside (4*p*NP-β-D-galactopyranoside) to β-D-galactopyranoside and 4-Nitrophenol, as depicted below (Figure 2.3).

β-Galactosidase can liberate 1 μM (or 1000 nM) of (4*p*NP) from (4*p*NP-β-D-galactopyranoside) per 5 seconds. Therefore, it has been considered the β-galactosidase dilution factor and the incubation time (30 minutes). First, a standard calibration curve was obtained by titration of the product 4-Nitrophenol (4*p*NP) with the buffer to 50 μM to 500 μM (or 50,000 nM to 500,000 nM) against the average relative fluorescence units (Figure 3.17, A). The colour of the yield product was determined with a spectrophotometer at 405 nm (visible).

4-Nitrophenyl-β-D-galactopyranoside   **4-Nitrophenol**
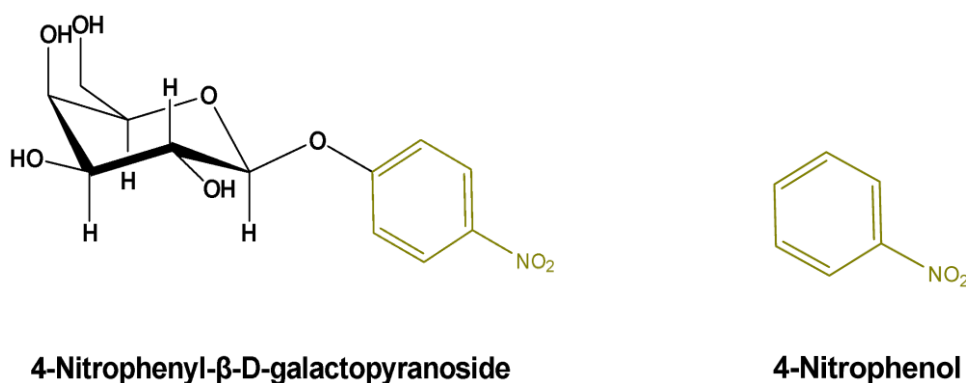
Figure 2.3 4*p*NP-β-D-galactopyranoside and 4-Nitrophenol.
β-Galactosidase was catalysed through the hydrolysis of the colourimetric substrate 4-Nitrophenyl β-D-galactopyranoside (4*p*NP-β-D-galactopyranoside), on the left, to the colourimetric product 4-Nitrophenol (4*p*NP) on the right Figure created with ChemDraw.

Initially, the substrate was dissolved in Dimethyl sulfoxide (DMSO) to give a 200 mM (200×10^{-3} μM) stock solution, which was diluted in HEPES to provide a final concentration of 1 mM (1×10^{-3} μM).

Then, serial dilutions of the enzyme 0.01 μM to 10 μM (or 10 nM to 10,000 nM) were tested with 1 mM (1000 μM) substrate (Figure 3.17, B).

The plot of pNP concentration versus time depicts the serial dilution of the enzyme with substrate. The stock enzyme solution at 10 μg/mL (86 μM) was diluted 10-fold to 1 μg/mL (8.60 μM), and the final target concentration was 0.5 μg/mL (4.30 μM).

Also, the serial dilution of substrates from 10 to 10,000 nM (0.01 µM to 10 µM) was evaluated at the initial rate of enzyme activity.

Kinetic activities were conducted according to the Michaelis-Menten model (Johnson, 2011). The Michaelis-Menten equation is $V = (V_{max} \times [S]) / (K_m + [S])$, where V is reaction velocity and S represents substrate concentration. The experimental results were used to determine the optimal incubation period and the substrate concentration, yielding maximum reaction velocity (Figure 3.16, C).

Michaelis-Menten constant (K_m), which is the substrate concentration at which the initial rate is one-half of the maximum velocity (Eisenthal, Danson, and Hough 2007) and maximal velocity (V_{max}) is the maximum rate of reaction at saturating substrate concentrations (Brooks et al. 2012), were measured at the intermediate and minor activity of the enzyme.

The reaction conducted by using 40 µL of 1mM (1000 µM) 4pNP-β-D-galactopyranosideto then diluted to 10-fold and 100-fold and adding 40 µL of 10 µg/mL (or 86 nM or 0.860 µM) to give final concentration of 0.5 µg/mL (or 4.30 µM) and blank were 40 µL substrate and 40 µL buffer, (Figure 3.16, B) incubated at pH 7 and 37 °C. The test was run in triplicate.

The enzyme inhibition and kinetics assays were measured using the substrates 4-Nitrophenyl β-D-galactopyranoside (4pNP-β-D-galactopyranoside).

The enzyme was diluted in the appropriate amount of the buffer, 50 mM (50×10^{-3} µM) HEPES and 1 mM (1×10^{-3} µM) NaCl pH 7.5.

Serial dilutions of β-galactose epoxide configured Galacto cyclophellitol (LWA480) at final concentrations of 200 µM-1.5 µM were incubated with enzymes at 0.5 µM final concentration for 30 minutes. Then, the substrate at 62 µM final concentration was added to the serial dilution of the inhibitor LWA480 that was incubated with 17 µM of the enzyme, and the reaction was monitored over time.

All the final volumes in wells were 50 µL, and absorbance was determined using an Epoch colourimetry machine at 405 nm for the substrate (4pNP-β-D-galactopyranoside).

IC$_{50}$ was estimated by varying the inhibitor concentration from 0.0122 to 25 μM and fitting the curve to the Hill equation. Once the Hill equation is fitted to experimental data, the resulting parameters can predict enzyme kinetics under different conditions or simulate the enzyme's behaviour and substrate binding. All tests were performed in a 384-well plate and at 37 °C.

2.2 Cryo-EM study of β-Galactosidase

2.2.1 Grid preparation of apo β-Galactosidase

After the size exclusion chromatography (SEC), fresh purified protein samples were used to prepare grids. A fraction of the highest protein concentration produced in the laboratory at 3.00 mg/mL was used for the first data collection, and another sample was diluted with the HEPES buffer to 2.00 mg/mL to collect different data sets. A volume of 2.5 μL was applied on Quantifoil R1.2/1.3 400 mesh copper grids (Quantifoil Micro Tools GmbH, Jena, Germany). The grids were glow discharged, and the blotting was performed with forces of -10 and -5.

Subsequently, after purification, the fresh commercial β-galactosidase specimen was diluted to 1.8 mg/mL final concentration for grid preparation. The purified protein at a volume of 2.5 μL was deposited onto the fresh Quantifoil R1.2/1.3 400 mesh copper grids (Quantifoil Micro Tools GmbH, Jena, Germany).

2.2.2 Labelling and inhibiting Cryo-EM sample

For obtaining complexes of β-galactosidase a protocol from (Saur et al. 2020) was adopted and followed. Snapped frozen recombinant protein at -80 °C was warmed at room temperature, then liganded following this recipe: 27 μL of the protein at a concentration of 2 mg/mL (equal 17 μM) was added to 3 μL of 5 mM of the stock inhibitor Galactocyclophellitol (LWA480) (Figure 2.4, below panel) to give final concentration 0.33 mM of the inhibitor, that was in the buffer of 50 mM HEPES and 1 mM NaCl, pH 7.5. Another complex was obtained with the same protein concentration with a volume of 50 μL and 5 μL of 200 μM of the stock of the probe N-octylnorbornene-Cy5-galactocyclophellitol aziridine (TB652) (Figure 2.4, upper panel) to give a final concentration 20 μM, that was in the buffer of 50 mM of HEPES and 1 mM NaCl, pH 7.5, 10% DMSO.
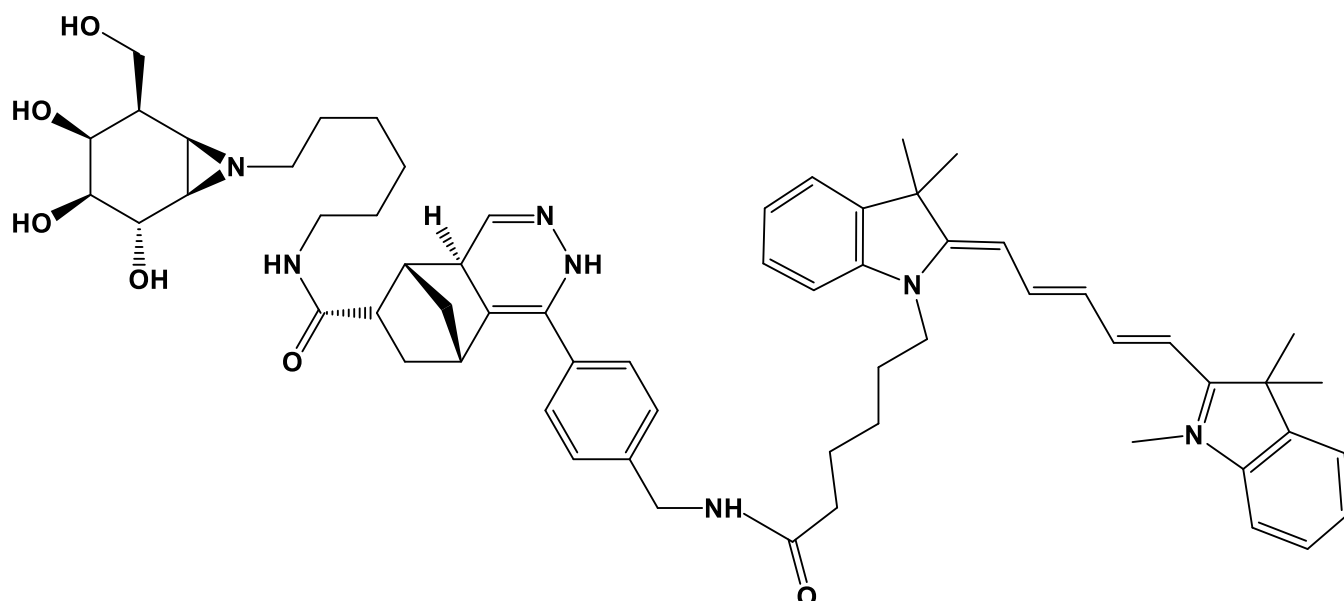
N-octylnorbornene-Cy5-galactocyclophellitol aziridine (TB652)
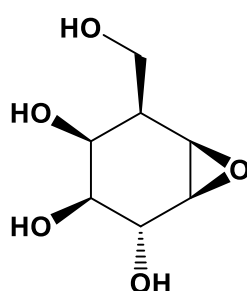
Galactocyclophellitol (LWA480)
or 4-epi cyclophellitol

Figure 2.4 Cyclophellitol aziridine and epoxide ligands.
The cyclophellitol-aziridine fluorescent probe TB652 (above) and cyclophellitol-epoxide inhibitor LWA480 (underneath), specific to galactosidase, were used in this study. Aziridine has a "warhead" that is a chemically active targeting moiety. Aziridine and epoxide ligands, structures were generated using ChemDraw.

Samples were incubated for 30 minutes, then at a volume of 2.5µL was deposited onto the fresh Quantifoil R2/2 300 mesh copper grid (Quantifoil Micro Tools GmbH, Jena, Germany). All grids before use were discharged using Pelcoeasi Glow Discharge Cleaning System (Ted Pella Inc., USA) for hydrophobic to hydrophilic conversion of the carbon-coated grids – this treatment is standard and allows for the easier spreading of water-based mixtures including protein solutions (Thomas et al. 2018; Kontziampasis et al. 2019). The sample was blotted onto the grids using a force of -5 and at approximately 100% humidity; plunge-frozen in liquid ethane, and then cooled by liquid nitrogen on Thermo Fisher (FEI) Vitrobot device Mark IV

(Thermo Fisher Scientific, Waltham, MA). Then, the grids were stored in a liquid nitrogen Dewar.

2.3 β-Galactosidase data acquisition

Four datasets were collected on a Thermo Scientific Glacios 200 keV microscope. Images were collected on a Falcon IV direct electron detector (Thermo Scientific, Eindhoven, Netherlands). Thermo Fisher Scientific EPU software was used to screen all images using a 4×4-hole pattern, and active beam-tilt adjustment was applied. Stage movement was used to conduct target hole centring and autofocusing. After setting the stage position, drift was determined on an adjacent carbon area, and data collection imaging was delayed until the evaluated drift was less than 2 Å s^{-1}.

All images were captured in Electron Event Representation (EER) mode at a nominal magnification of 240,000x, corresponding to a pixel size of ~0.574Å on the specimen, over a target defocus range of −1 to −2.0 μm, frame exposure of 3.88 s and at a dose rate of ~11 e$^-$ physical pixel^{-1} s^{-1} was used, to accrue a total dose of ~50 e$^-$/Å^2. Movie data were collected in the Electron-Event Representation (EER) format, which is a new addition to the Falcon IV line of detectors (FEI). EER fractionation parameters of motion correction were calculated by dividing the total dose or total fluence of ~50 e$^-$/Å^2 by ~1 e$^-$/Å^2, and the total number of EER frames in the movie was obtained by multiplying the fractions number of EPU frames, *e.g.* 133 by 7 (133 × 7=931). Then, the EER frames needed to be grouped, so each frame needed to have a total dose of 1e$^-$/Å^2, *e.g.* for 931/50=18.62 frames; calculations are very dependent on the datasets. All parameters used for data acquisition are provided in (Table 2.4, 2.5 and 2.6 in section 2.8).

2.4 Pipeline of Cryo-EM data-processing (β-Galactosidase)

Four datasets were collected over 24 to 72 h: the first dataset was acquired with a commercial sigma β-galactosidase sample, and three other datasets were from the recombinant β-galactosidase; one of them was in a complex with 4-epi cyclophellitol.

2.4.1 Apo protein (commercial and expressed in-house)

The First dataset of the commercial protein consisted of 1,832 movies. The second dataset of recombinant apo protein contained 2,094 micrographs at a concentration of 3.0 mg/mL, and the third dataset 1,416 micrographs, obtained with protein at a concentration of 2.0 mg/mL. All concentrations were estimated using a nanodrop machine, collecting absorbance at 280 nm.

The open-source program *RELION* (v. 3.1.2) was performed to process all images. Using the wrapper in *RELION*, the beam-induced motion of all the micrograph frames were aligned and corrected using 3 x 3 tiles (*MotionCor2)* (Zheng et al. 2017). CTFFIND 4.1.10 software (Rohou and Grigorieff 2015) was estimated Contrast Transfer Function (CTF) parameters for each corrected micrograph.

404 particles were selected from the initial set of the commercial protein by generating a free auto-picking template, with performance based on a Laplacian-of-Gaussian (LoG) filter. The picked particles were extracted with a box size of 360 pixels (~0.58 Å pixel size) and subjected to two-dimensional (2D) classification. Around 1,000 particles were selected manually to be used for reference-free two-dimensional (2D) classification, which was then used as templates for automated particle picking of the entire data set.

2D class averages representing projections in different orientations were subjected to 2D classification templates to remove junk particles, and those selected 2D class averages were used to auto-pick 138,711 particles. Particles belonging to low abundance classes were withdrawn by one round of 2D classification starting with 17 selected images. The remaining particles were regrouped into five groups and yielded 32,030 selected particles. Particles picked in *RELION* were extracted with a box size of 360 pixels, in agreement with the original pixel size of 0.58 Å.

These selected particles were then used to produce an initial model calculated using the stochastic-gradient-descent (SGD) algorithm of *RELION* and used for a first round of three-dimensional (3D) reconstruction with D2 symmetry (consistent with the reported symmetry of beta-galactosidase), requesting four classes and a maximum resolution of 15 Å, as shown

in Figure 3.24, C. Among all the four classes, three classes including 25,262 selected particles stood out as demonstrating apparent overall features of a tetramer. An iterative procedure was employed for subsequent 3D classification and auto-refinement rounds using the best output 3D maps – this selection is done based both on subjective aspects (map appearance) and objective measures (estimated resolution). The refinement of these particles produced a map at 4.5 Å resolution (FSC 0.143) with D2 symmetry.

After particle refinement against each micrograph, CTF refinement per particle and Bayesian polishing were performed. The final map was improved by running a B-factor estimation job within *RELION*. The FSC of the final masked 3D map was calculated by applying the postprocessing procedure implemented in the *RELION* software. Based on the FSC curve, the resolution was estimated at 0.5 and 0.143 (Rosenthal and Henderson, 2003) (gold standard) FSC cut off criteria from 3.9 Å to 3.4 Å, in Figure 3.24, D. Local resolution calculations were performed with *RELION*'s own post processing program in Figure 3.25, right panels. The structures were visualised using Chimera X (Pettersen et al. 2021).

The first apo recombinant protein dataset, a reference-free 2D class averaging, was used to throw away bad particles from 2,094 micrographs and about 177,309 particles were extracted. Micrographs were subset by 600 micrographs into 4 fractions; the first 600 subset micrographs had extracted 71,547 particles and were used to optimise the auto picking as a template for the first round of 2D class averages.

The first auto picking round used the 2D class template, and a 3D reference was obtained from the final post processed map of the commercial β-galactosidase that we solved at 3.4 Å. The first 3D refinement was the last step for processing this dataset due to encountering a preferred orientation problem. Two rounds of auto picking were performed on 1,416 micrographs of the second recombinant apo protein dataset. First, around 71,634 particles were auto-picked using a map generated from the commercial β-galactosidase solved in this study at 3.4 Å as a reference model.

The 71,634 particles were extracted and generated as free 2D classes; these 2D classes were not used to refine the 2D classes initial model. They were subset to run the second autopicking

round to pick more particles that have a variety of features, and numbers that increased to 23,4781 particles. These particles were extracted to generate 2D class averages, which were obtained from the initial model that was used as a reference model for a new round of 3D classification against the 2D extracted particles.

One best class among the four 3D classes was low-pass filtered to 50 Å and reselected about 77,744 unbinned particles at (0.902222 Å/pixel, 450-pixel box) from that best class without extracting them because the box size was not rescaled. Subsequently, 3D high resolution refinement was performed, resulting in a 4.06 Å resolution structure. In all jobs, D2 symmetry was imposed to improve the Cryo-EM map quality further. The map was post-processed and polished. Additionally, further refinement was run, and that enhanced the final resolution to 2.5 Å, Figure 3.25, left panels.

2.4.2 Enzyme-ligand complex

For the fourth dataset, 5931 movies of recombinant β-galactosidase in complex with the epoxide inhibitor were collected over 72 hours, three shots per marked spot hole. Micrographs were processed in *RELION*; motion corrected in MotionCor2 and CTF parameters were estimated using CTFFIND4 as in section 1.2.5.1 above. Particles were auto picked using the map of recombinant β-galactosidase at 2.5 Å as a 3D reference model, imposing D2 symmetry to improve the picking performance with the following parameters:

Picking threshold:	0.8
Minimum inter-particle distance(A):	120
Maximum stddev noise:	1.2
Minimum avg noise:	-1

Over 736,882 picked particles were subjected to fast subsets run that generated the 2D class-free reference. About ten best different views of 2D classes were selected which brought the total to 643,948 particles. Stochastic gradient descent (SGD) was executed using the extracted 2D classes at 0.889778 Å /pixel, to produce the *de novo 3D model* (not to be confused with *atomic model*).

The initial model was used to yield four 3D classifications. The best 3D classes were selected according to the apparent features, alignment, and lack of noise for performing auto refinement by Nyquist frequency (Zivanov et al. 2018). The best 3D classification metadata resolution was 5 Å and contained 204,000 particles. The mask was created with a particle size of 250 Å and low-passed-filtered to 20 Å. Then the post processed map was generated at 3 Å. This map was used to refine the CTF micrograph per particle by running three jobs: tilt aberrations, magnification, and defocus, respectively. The output of the last CTF refinement was trained about 10,000 particles to generate ideal parameters for Bayesian polishing, which led to a map at 2.5 Å resolution, estimated at the gold standard 0.143 FSC. The final post-process run was performed on the map and was revealed after B-factor sharpening. All parameters used in image processing are provided in (Table 2.4 in section 2.8).

2.5 Structure refinement and analysis

Due to the availability of high-quality atomic models of β-galactosidase in the Protein Data Bank (crystallographic or Cryo-EM), a suitable one was identified to perform a fit into the post processed masked map of both the apo protein and the complex. The chosen model was (PDB code: 5A1A) (Bartesaghi et al. 2015), also a single-particle Cryo-EM structure. Water and ligand molecules were removed from the atomic model before doing the fit, which was performed using UCSF Chimera (Pettersen et al. 2004). This provided a starting model to interpret the map. The side chains of the atomic model readily showed features inside the electron-potential map using *Coot*, (Emsley, Lohkamp, Scott, and Cowtan, 2010). However, not all side-chains fit the map correctly, hence rounds of refinement were done to achieve a good fit.

While the apo protein structure did not show any novel features, significant work was put into building and refining the covalent complex in preparation for deposition and publication. The atomic model of a single monomer (a homotetramer is generated via D2 symmetry) was refined with Servalcat v0.4.32, installed directly from the Python Package Index (PIP) and run on the command line using Bash scripts; local parts requiring minor refinements were adjusted manually by 'Sphere refine' and 'Sphere regularisation', which are tools in *Coot*, in order to improve model parameters.

The model fit was validated using *Coot* built-in tools, MolProbity (V. B. Chen et al. 2010) and by submission to the wwPDB validation web service. Use of these tools revealed an abnormally high (~3%) proportion of outliers as judged by the Ramachandran criterion. These criterion flags up any unusual conformations of the main chain; while outliers do occur in nature, they are rare and must always be supported by an unambiguous map – this was clearly not the case with a 2.7 Å resolution map. Therefore, a decision was made to try and reduce the proportion of outliers to bring it into the reported values for the highest-resolution structures (all crystallographic) of *E. coli* beta-galactosidase. One potential source of Ramachandran outliers is the presence of clashes with neighbouring residues that might force an awkward conformation on the main chain; these were checked thoroughly. Another obvious reason for unusual chemical geometry is the weak parameter to observation ratio typical of macromolecular structure refinement against low resolution data: this is certainly the case for our map. Geometry restraints (bond lengths, angles, torsions, chiral volumes and planes) are routinely used in macromolecular refinement, and the target function in Servalcat includes a weight for the experimental term, which may be decreased to give priority to geometry restraints. A few trial runs were done with lower alternative weights in Servalcat, but this failed to bring the proportion of outliers down from 3%.

One further way of increasing the number of observations and stabilising refinement is to use distance restraints derived from a comparison of the problem structure to a reference high-resolution target structure. The ProSMART software, distributed with both CCP-EM and CCP4, is able to produce such restraints. A high-resolution target was identified: (PDB code: 1JZ7) (Douglas H. Juers et al. 2001), with 0.3% Ramachandran outliers. ProSMART was used to derive distance restraints, which were then included in Servalcat refinement using external restraints. ProSMART produces restraints in the form of a text file where interatomic distances are specified, and these can be individually weighted depending on structural similarity with the Geman-McClure robust estimator. Servalcat then takes this text file as input and uses it to further restrain the model. Increasing the number of restraints also means increasing the number of refinement cycles to achieve convergence: whereas the refinements run initially would converge in 10-20 cycles, the inclusion of the external restraints pushed that figure to around 60. The final command that brought refinement to convergence looked like this:

```
# servalcat refine_spa --model ../beta-gal_deposit.pdb --
halfmaps half1.mrc half2.mrc --mask_for_fofc ../mask.mrc --pg
D2 --resolution 2.70 --cross_validation --ligand
UNL_GLU_link.cif --keyword_file ProSMART.txt --ncycle 60
```

Servalcat can perform cross-validation by doing separate refinements against the two half-maps produced by the reconstruction process. Overfitting can be avoided by monitoring the FSC between model and map, and model and the separate half maps. Wildly different values usually mean overfitting has been introduced.

A distinct volume of unmodelled map was observed next to glutamic acid 537 – this was also observed as positive density in the fo-fc difference map produced by Servalcat. The map emerged continuously from the carboxylate side-chain, and therefore it was hypothesised that the covalent ligand was present. An initial attempt was made to find a suitable ligand definition in the PDB Chemical Component Dictionary, but the entries describing reactions with 4-epi cyclophellitol contained chemical descriptions for the processed ligand where the epoxide is gone as opposed to the full chemical description (*e.g.* https://pubchem.ncbi.nlm.nih.gov/compound/71811328). Upon discussion with the Servalcat and AceDRG (the associated restraint dictionary generation software) teams, it was decided that AceDRG would be extended to be able to cope with the deletion of the epoxide, which would allow future users to download the chemical definition of cyclophellitol-style probes and create a link description that would process the ligand automatically, deleting the epoxide and bonding the anomeric carbon to 'OE2' in the glutamic acid side-chain. The modification of the software was completed in AceDRG version v277, which was tested to create the link description. The required instruction file contained:

```
LINK: RES-NAME-1 UNL ATOM-NAME-1 C2 FILE-1 AcedrgOut.cif DELETE
BOND O1 C2 1 RES-NAME-2 GLU ATOM-NAME-2 OE2 BOND-TYPE SINGLE
```

After that, Coot was used to model the ligand into the map, and a LINK record was created to instruct Servalcat to restrain the linkage using the provided dictionary file (see -- ligand in the

command above). Importantly, no torsion restraints were used in the refinement of the reacted probe, as the map was clear and the conformation of the ring did not deviate greatly from the expected $^4C_1$ chair.

While the identification of potential ions is perhaps easier in X-ray crystallography through the use of anomalous difference data, no such mechanism exists in single particle Cryo-EM; therefore, comparative studies were carried out in order to identify any remaining large blobs of electron density. Indeed, thesuperposition of the solved liganded structure to other available structures of the same protein revealed the presence of a magnesium ion near the ligand, Figure 2.5 shown the previous X-ray structure. This was evident in our map in the form of a large and strong positive difference density, suggesting that the magnesium ion was also present in our sample. Modelling the magnesium ion also required the introduction of three water molecules in order to complete the coordination, which is done by glutamic acid 416 ('OE2'), histidine 418 ('ND1') and glutamic acid 461 ('OE1'). No other water molecules were introduced.

Frustratingly, no difference map could be spotted to correspond to the expected sodium ion that is present in every crystallographic structure of this enzyme. As there was no evidence on the map, a decision was made not to include it in the final model.
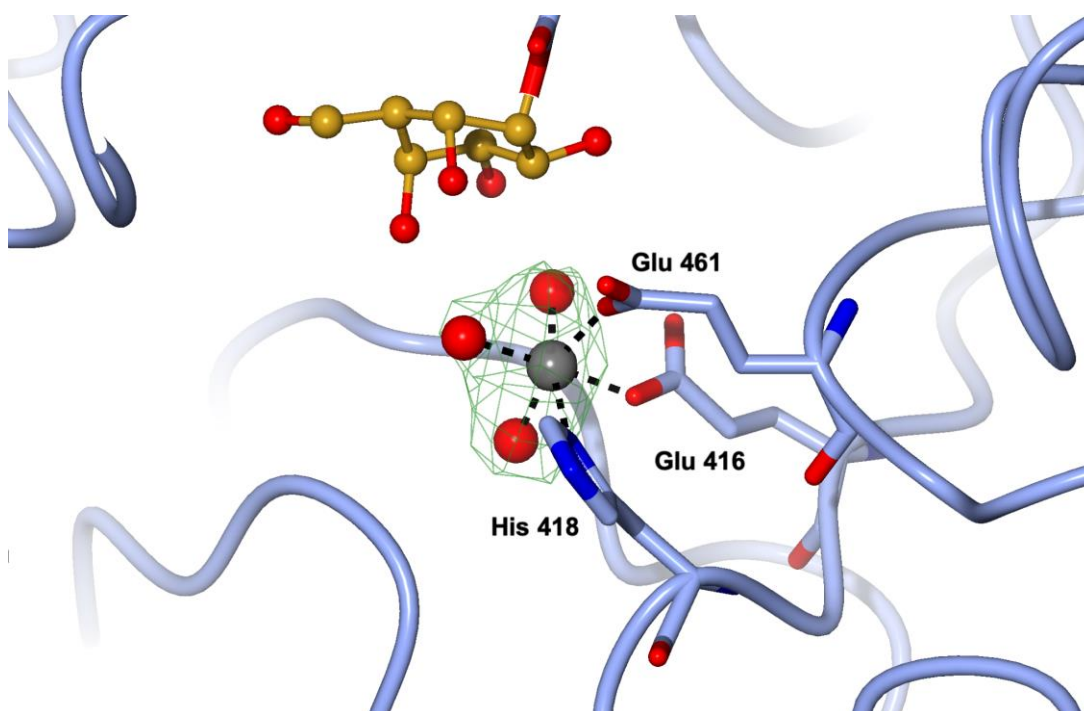
Figure 2.5 A close-up view of the magnesium ion and three waters that were modelled based on existing structure, (PDB code: 1JZ7) (Douglas H. Juers et al. 2001).
Difference map (green, contoured at +2.5 level) calculated by Servalcat at the previous refinement step, contains peak corresponding to the reacted probe.

Refinement statistics are summarised as follows (includes output from Servalcat):

RMSD from ideal
Bond lengths: 0. 008 A

Bond angles: 1.6 deg

Map-model FSCaverages (at 2.70 Å)

FSCaverage(full) = **0.8208**

Cross-validated map-model FSCaverages:

FSCaverage(half1) = **0.7661**

FSCaverage(half2) = **0.7354**

Atomic Displacement Parameter statistics

Chain A – protein (8218 atoms) min= 18.4 median= 53.8 max=354.9 A^2

Chain B – magnesium ion (1 atom) max= 50.0 A^2

Chain C – water molecules (3 atoms) min= 37.0 median= 43.4 max= 51.6 A^2

All combined (8,222 atoms) min= 18.4 median= 53.8 max=354.9 A^2

Weight used: 2.79

Ramachandran outliers (PDB validation report): 0

Side-chain outliers (PDB validation report): 3.8%

Clashscore (PDB validation report): 6

The map was deposited with the Electron Microscopy DataBase (EMDB) with identifier EMD-18488, and the coordinates were deposited at the Protein Data Bank (wwPDB) under identity code: 8QLZ. Valuable assistance was provided by Deborah Harrus (PDBe, EBI Cambridge), who monitored deposition, validation and annotated the entry.

Map, coordinates and final PDB validation report are available for pre-release inspection on Zenodo: 10.5281/zenodo.8384171

2.6 Human serum transferrin samples preparation procedures

2.6.1 Native Apo-hTF without purification

Apo Human transferrin from Sigma–Aldrich (cat no T4382-1G), synonym Human transferrin Siderophilin Lyophilised powder, was purchased from Sigma–Aldrich and solubilised in a 20 mM Tris-HCl, pH 7.5 and 200 mM NaCl following the protocol presented by (Noinaj et al. 2012). A decision was made to process the sample initially without any further purification and only to refine the process should problems arise down the line.

2.6.2 Second experiment: purification of the Apo-hTF and characterisation of yield

2.6.2.1 First step of purification using Ion exchange chromatography

A mass of 300 mg of apo human transferrin (cat no T4382-1G) was weighed and dissolved in 2 ml of buffer A (25 mM HEPES buffer, pH 7.5) and 5% NaCl. The purification instrument of Akta pure chromatography column was equilibrated with buffer A. Then, the protein was purified following published methodology (Pratt et al. 2018), using AKTA pure and Q-Sepharose FF ion exchange column in buffer A. The protein solution of 2 ml was injected into

Akta Pure in two separate runs. The protein was eluted using buffer B (25 mM HEPES, pH 7.5 and 1 M NaCl), a linear salt gradient (0–200 mM NaCl) and buffer exchanged into 10 mM HEPES buffer pH 7.5. The purity of apo-transferrin preparation was analysed using 10% Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). The SDS-PAGEs were then stained with 0.1% (w/v) Coomassie Brilliant Blue (CBB) to visualise protein bands (Figure 4.3).

2.6.2.2 Second step of purification size exclusion chromatography

All protein fractions resulting from ion exchange chromatography were pooled and concentrated into a volume of 2 ml and purified by Superdex 75, Hiload 16/600 column using SEC buffer (50 mM HEPES pH 7.5 150 mM NaCl). A 10% SDS-PAGE determined the purity of apo-transferrin preparation. The aliquots of protein fractions were stored at 4°C to prepare fresh grids and stored at -70 °C for longer-term purposes (Figure 4.4).

2.6.2.3 Analysis of the Apo-hTF MALDI-MS/MS

In order to analyse the peptide fingerprint of protein fractions, three bands were cut out of the SDS-PAGE gel and sent for Matrix-assisted laser desorption/ionisation mass spectrometry (MALDI-MS/MS). The analysis was done by the Metabolomics and Proteomics Technology Lab team at the University of York. Protein bands were at lane number 3 from the first purification step, Ion exchange chromatography, as shown in Figure 4.4.

Mass spectrometry sequence coverage of tryptic digests of three protein bands were shown. Identified peptides are shown in bold red. Sequence and numbering are for the Serotransferrin.

2.6.3 Formation holo forms of hTF- Ferric ammonium citrate and Ferrous ammonium sulphate

The iron-loaded (holo) forms of human serum transferrin were prepared as described in (Yang et al. 2012) approach, with modifications and technical support from Juliet Borgia in YSBL. The pure Apo-hTF fractions stored at -70 °C were defrosted by quickly bringing them to body temperature. The fractions were then pooled to yield approximately 6.5 ml of 4.44 mg/ml protein.

The sample was concentrated down to 194.59 mg/ml, equivalent to 2.53 mM. The concentration of 50 mM NaHCO3 was added to make (5 mM of final concentration) and 300mM of (Ferric ammonium citrate synonym (Ammonium Iron (III) citrate) CAS: 1185-57-5, cat F5879, Sigma Aldrich) and (Ferrous ammonium sulphate synonym (Ammonium Iron (II) sulphate) cat no A484053, Fisons), were added for making of the final concentration of 4 molar equivalents of Ferric salts was added before being incubated in the dark at room temperature overnight. Complexes were further purified by size exclusion columns using a Superdex 200, Hiload 16/600 column on Akta pure, using a buffer (50 mM HEPES pH 7.5 150 mM NaCl). The purity of proteins (Ferric ammonium citrate-hTF and Ferrous ammonium sulphate-hTF) was further verified using 10% SDS-PAGE.

2.6.4 Deglycosylation of hTF

Proteins are often glycosylated, meaning they contain carbohydrate molecules (glycans) attached to specific amino acid residues. Glycosylation is a common post-translational modification that can significantly affect protein structure, stability, function, and interactions. To improve the stability and homogeneity of the Cryo-EM sample, the glycoproteins were deglycosylated by digesting the polysaccharides and purified.

The aim was to release the sugars from *N*-linked glycosylation. For that, two deglycosylation enzymes were used, PNGase F (Peptide-N-Glycosidase F) and EndoH (Endoglycosidase H) deglycosylate the apo hTF. Deglycosylation was tested and used to remove sugars that may cause sample heterogeneity. Enzymatic reactions set up an overnight reaction with the target protein at room temperature then ran sample reactions on SDS-PAGE the next day to determine the best digestion.

The samples were treated and loaded on the gel according to the following Table 2.2.

Table 2.2 Reaction component for different samples treated with EndoH and PNGase F enzymes.

1	2	3	4	5	6	7	8	9	10
M	EndoH				PNGase F				
1	2	3	4	5	6	7	8	9	10
	Control of Apo hTF	Digestion Enzyme	Control follow manufacture steps	Follow the manufacture	Control of Apo hTF	Digestion Enzyme	Excluding N 40	Control follow manufacture steps	Follow manufacture steps
Denaturing Buffer			1 uL	1 uL				1 uL	1 uL
protein	8 uL	8 uL	8 uL	8 uL	8 uL	8 uL	8 uL	8 uL	8 uL
incubation			100 C for 10 min	100 C for 10 min				100 C for 10 min	100 C for 10 min
Buffer 10X	1 uL	1 uL	1 uL	1 uL	1.2 uL	1.2 uL	1.2 uL	1.2 uL	1.2 uL
Digestion Enzyme	------	1 uL	-----	1 uL	------	1 uL	1 uL	--------	1 uL
SEC Buffer	1 ul	-----	1 ul	-----	1 ul	----	1.8 uL	1 ul	------
N 40	_	_	_	_	1.8 uL	1.8 uL	----	1.8 uL	0.8 uL
total	10	10	11	11	12	12	12	12	12
Incubation overnight	Incubate at room temperature			Incubate one hour at 37 °C	Incubate at room temperature				Incubate one hour at 37 °C

2.7 Cryo-EM study of human serum transferrin

All grids of this study were discharged using Pelco Easiglow Discharge Cleaning System (Ted Pella Inc., USA) and vitrified by plunge-frozen first in liquid ethane and then cooled by liquid nitrogen on Thermo Fisher (FEI) Vitrobot device Mark IV (Thermo Fisher Scientific, Waltham, MA).

2.7.1 Grids preparation of native Apo-hTF

Fresh samples of hTF were titrated for concentration from 1.5 mg/ml to 0.1mg/ml. A volume of 2.5 µL of each protein titration was applied to the grids of Quantifoil R 2/2 200 mesh copper (Quantifoil Micro Tools GmbH, Jena, Germany). Local filter papers were used to dispense the excess of the specimen. The samples were blotted for 4 and 6 s at force - 3 and - 5 at a relatively 20 °C temperature and 100% humidity. The screening results showed the 4 s, -3, and 0.2 mg/ml blotted grid presented the best particle distribution conditions.

2.7.2 Grids preparation of Apo-hTF purified sample

A fresh sample was stored overnight at 4°C, and a concentration of 0.64 mg/ml was applied on a grid of Ultra Au Foil R 1.2/1.3 300 mesh gold support (Quantifoil Micro Tools GmbH, Jena, Germany). The protein blotting time for 6 s, at force - 4 at a relatively 20 °C temperature and 100% humidity. The following day the grid was screened, and data was collected.

2.7.3 Grids preparation of complexes of hTF- Ferric ammonium citrate and Ferrous ammonium sulphate

The complexes hTF- Ferric ammonium citrate and Ferrous ammonium sulphate were applied on grids of Ultra Au Foil R 1.2/1.3 300 mesh gold support (Quantifoil Micro Tools GmbH, Jena, Germany). After using 2.5 µL (~2 mg/ml) of each specimen, grids were vitrified in liquid ethane at blotting force -4 and time 6s.

2.8 hTF data acquisition

All datasets of this study were collected at Thermo Fisher Scientific Glacios transmission electron microscope (TEM) set at 200 keV with the direct electron detector Falcon IV in super-resolution mode at a standard magnification of 240/2 (KX) with the help and supervision from

Dr Jamie Blaza. All grids were clipped and loaded into the microscope by Dr Sam Hart and Dr Johan Turkenburg. The EPU program was used for screening and data collection.

2.8.1 First dataset collection: native Apo-hTF without purification

The day after preparing grids, the Cryo-EM data of hTF were screened and stocked up on the microscope. Three thousand two hundred fourteen micrographs were collected using a Falcon IV detector set on counting mode (0.574 Å pixel−1) employing a defocus range of -2.0 µm to -1.4 µm. Movies were gathered over a 5.36 s exposure and 50 (e-/Å^2) collection dose. Movies were collected in electron-event representation (EER) format, which captures electron events at a very high frequency (248 Hz) and typically does have a smaller file size than Mixed Raster Content (MRC) movie format (Guo et al. 2020). The number of EER frames in the movie was obtained by multiplying the fractions number of EPU frames184 by 7 (184×7=1,288). Then, the EER frames needed to be grouped, so each frame required a total dose of 1 (e-/Å^2), for 1,288/50=25.76 1 (e-/Å^2) per frame.

2.8.2 Second dataset collection: purified Apo-hTF

Cryo-EM data was screened and collected from a grid of a fresh purified sample of Apo-hTF that was stored overnight at 4°C.

Micrographs were collected using the detector set on counting mode (0.574 Å pixel−1) at a standard magnification of 240 (KX) employing a defocus range of -2.0, -1.8, -1.6, -1.4 µm. Movies were collected over a 4.66 s exposure, giving a total collection dose of 50 (e-/Å^2). The number of EER frames in the movie was obtained by multiplying the fractions number of EPU frames215 by 7 (215×7=1,288).

Then, the EER frames needed to be grouped, so each frame was required to have a total dose of 1(e-/Å^2), for 1,505/50=30.1 fractionation.

2.8.3 Third dataset collection: complex of hTF- Ferric ammonium citrate

After the last step of hTF- Ferric ammonium citrate purification, grids were prepared from overnight stored fresh samples at 4°C.

Cryo-EM data were collected using the detector set on counting mode (0.574 Å pixel−1), employing a defocus range of -2.0, -1.8, -1.6, -1.4 μm. Movies were collected over a 3.88 s exposure, giving a total collection dose of 50 (e-/Å^2). The number of EER frames in the movie was obtained by multiplying the fractions number of EPU frames133 by 7 (133×7=931). Then, the EER frames needed to be grouped, so each frame was required to have a total dose of 1e-/Å^2, for 931/50=18.62 fractionation.

2.8.4 Fourth dataset collection: complex of Ferrous ammonium sulphate

Data was screened the next day of preparing the grid from a snap-frozen sample of hTF-Ferrous ammonium sulphate that was stored at- 70°C.

Micrographs were collected by the detector set on counting mode (0.574 Å pixel−1) at a standard magnification of 240 (KX) employing a defocus range of -2.0, -1.8, -1.6, -1.4, -1.2, -1.0, -0.8, -0.6 μm. Movies were collected over a 3.88 s exposure, giving a total collection dose of 50 (e-/Å^2).

2.8.5 Fifth dataset collection: thinner ice grid

Grid ice thickness is a parameter that can affect the quality of the particles or the micrograph background. In particular, our small target protein was prepared under low humidity and higher blotting force to create a thinner, unbreakable amorphous ice layer, 4 mg/ml, 20 blotting force,70% humidity, 5s temperature 4C. Grids were prepared from a snap-frozen specimen that was previously purified by the SEC. They thawed quickly by body temperature and then used.

2.8.6 Sixth dataset collection: deglycosylated hTF by PNGase

Data was collected from the sample, which was incubated with PANGase for 16 H overnight and then purified with SEC. Micrographs were screened from different grid conditions, and the data was collected from the sample eluted at the highest chromatogram peak.

2.8.7 Seventh dataset collection: Volta phase plate (VPP)

Imaging using the Volta phase plate is an advanced technique used in electron cryo-microscopy (cryo-EM) to improve contrast and potentially the resolution of images of biological specimens. It allows to achieve better resolution by addressing some of the limitations of traditional phase contrast techniques in cryo-EM. For the seventh data set the VPP was inserted and exchanged, and the position adjusted every 60 exposures.

Table 2.3 The data collection of the Apo transferrin is incorporated in the acquisition.

Date	26/1/2022	20/4/2022
Sample	Apo Human transferrin	Purified Apo Human transferrin
Grid type	Quantifoil R 2/2 200 mesh copper grids.	Ultra Au Foil R 1.2/1.3 300 mesh gold grids.
Plunge freezer	FEI-Vitrobot Mark IV (Thermo Fisher Scientific)	
Sample freezing condition	0.2 mg/ml, blotting force -6 and blotting time 5s	0.64mg/ml, blotting force -6 and blotting time 4s
Data acquisition parameters		
Nominal magnification (KX)		240/2 KX
Spot size	4	4
Illuminated area (µm)	1.05	1.05
Intensity (decimal)	0.443	0.443
Calibrated pixel size (Å)	0.574 Å	
Dose		
Exposure time (s)	5.36 S	4.66 S
Total dose (e$^-$/Å^2)	50 e$^-$/Å^2	50 e$^-$/Å^2
Number of EPU frames	184 frames	215 frames
Number of EER frames	184×7 =1288	215×7 =1288
EPU parameters		
Defocus range (µm)	-2, -1.8, -1.6, -1.4	-2, -1.8, -1.6, -1.4
Autofocus cadence (µm)		10
Delay after stage shift (s)	20 s	5 s
Delay after image shift (s)	0.6 s	0.5 s
Exposures per hole	One number of shots	
Micrographs collected	3214	8,000

Ultra Au Foil R 1.2/1.3 300 mesh gold grid is a foil with ~2 µm circular holes and a spacing of ~1 µm between the holes. Quantifoil R 2/2 200 mesh copper grid is a foil with ~2 µm circular holes and a spacing of ~2 µm between the holes. All protein samples were prepared from the apo human transferrin (cat no T4382-1G).

Table 2.4 The data collection of the complexs transferrin is incorporated in the acquisition.

Date	23/5/2022	15/6/2022
Sample	Holo hTF (transferrin complex with ferric ammonium citrate)	Holo hTF (transferrin complex with ferrous ammonium sulphate)
Grid type	Ultra Au Foil R 1.2/1.3 300 mesh gold grids.	Ultra Au Foil R 1.2/1.3 300 mesh gold grids.
Plunge freezer	FEI-Vitrobot Mark IV (Thermo Fisher Scientific)	
Sample freezing condition	2 mg/mL, blotting force -6 and blotting time 4s	1.9mg/ml, blotting force -6 and blotting time 4s
Data acquisition parameters		
Nominal magnification (KX)	240/2 KX	
Spot size	3	3
Illuminated area (µm)	1.05	1
Intensity (decimal)	0.442	0.442
Calibrated pixel size (Å)	0.574 Å	
Dose		
Exposure time (s)	3.88 S	2.94 S
Total dose (e$^-$/Å^2)	50 e$^-$/Å^2	50.1 e$^-$/Å^2
Number of EPU frames	133 frames	707 frames
Number of EER frames	133×7=931	707
EPU parameters		
Defocus range (µm)	-2, -1.8, -1.6, -1.4	-2.0, -1.8, -1.6, -1.4, -1.2, -1.0, -0.8, -0.6
Autofocus cadence (µm)	10	
Delay after stage shift (s)	5 s	5 s
Delay after image shift (s)	0.5 s	0.5 s
Exposures per hole	One number of shots	
Micrographs collected	2600	3819

Ultra Au Foil R 1.2/1.3 300 mesh gold grid is a foil with ~2 µm circular holes and a spacing of ~1 µm between the holes. Quantifoil R 2/2 200 mesh copper grid is a foil with ~2 µm circular holes and a spacing of ~2 µm between the holes. All protein samples were prepared from the apo human transferrin (cat no T4382-1

Table 2.5 The data collection of thin ice, deglycosylated and VPP of transferrin is incorporated in the acquisition.

Date	30/8/2022	30/11 then 9/12/2022	4/5/2023
Sample	Thinner ice grid of Apo hTF	PNGas	VPP
Grid type	Ultra Au Foil R 1.2/1.3 300 mesh gold grids.	Ultra Au Foil R 1.2/1.3 300 mesh gold grids.	Ultra Au Foil R 1.2/1.3 300 mesh gold grids.
Plunge freezer	FEI-Vitrobot Mark IV (Thermo Fisher Scientific)		
Sample freezing condition	4mg/mL, blotting force 20+, 70 % humidity and blotting time 4s	2.7mg/mL, blotting force +1 and blotting time 2s	1.8 mg/mL blotting force -4 and blotting time 5s
Data Acquisition parameters			
Nominal magnification (KX)	240/2 KX		
Spot size	4	4	4
Illuminated area (µm)	1	1	1
Intensity (decimal)	0.443	0.443	0.443
Calibrated pixel size (Å)	0.574 Å		
Dose			
Exposure time (s)	5.71 s	5.64 s	5.64 s
Total dose (e$^-$/Å^2)	50 e$^-$/Å^2	50 e$^-$/Å^2	60 e$^-$/Å^2
Number of EPU frames	1372	1358	
Number of EER frames			
EPU parameters			
Defocus range (µm)	-2.0, -1.8, -1.6, -1.4	-2.0, -1.8, -1.6, -1.4, -1.2, -1.0, -0.8, -0.6	-2.0, -1.8, -1.6, -1.4, -1.2, -1.0, -0.8, -0.6
Autofocus cadence (µm)	10		
Delay after stage shift (s)	5 s	5 s	5 s
Delay after image shift (s)	0.5 s	0.5 s	0.5 s
Exposures per hole	One number of shots		
Micrographs collected	2000	2000	700

Ultra Au Foil R 1.2/1.3 300 mesh gold grid is a foil with ~2 µm circular holes and a spacing of ~1 µm between the holes. Quantifoil R 2/2 200 mesh copper grid is a foil with ~2 µm circular holes and a spacing of ~2 µm between the holes. All protein samples were prepared from the apo human transferrin (cat no T4382-1G).

2.9 hTF data processing

2.9.1 First datasets structure analysis: native apo-hTF without purification

Data was collected from micrographs of specimen 0.2 mg/ml. All movies were imported and processed on *RELION* (v. 3.1.2). Motion correction (*MotionCor2)* implementation in *RELION* was aligned. The beam-induced motion of all the micrograph frames was 3 x 3 tiled frames with a B-factor of 100 (Zhang et al, 2017). The contrast transfer function (CTF) parameters for each corrected micrograph were estimated using CTFFIND 4.1.10 (Rohou and Grigorieff 2015).

Ultimately, micrographs showed small spots without details or features indicating the clear transferrin molecule (Figure 4.20 ,1). Initial auto picking was performed using LoG (Laplacian-of-Gaussian) auto picking, producing 2D classes.
Using a Laplacian-of-Gaussian (LoG) filter of auto picking, particles initially extracted from 300 manually selected micrographs yielded 270,257 particles. Those particles were templates of 2D classification subjected to the second round to pick more particles and eventually were (602,180 particles).

The 2D classes box dimension was estimated by optimising the 2D classes from the available PDB atomic model for transferrin by "Calculate Distances" in *Coot* and adding about an extra 150 Å to most extended distances of 100 Å and diameter 150 Å mask, a micrograph after auto picking using the 2D class reference. Figure 4.20 ,2, A total of ten 2D classes, box size 216 pix and rescaled to 72 pix and mask diameter of 150 Å.

2D classes have generated De novo 3D models with insufficient features. Still, it was enough to predict from available atomic models that it might visualise the entire hTF molecule, N lobe or C lobe shown in Figure 3.6. The De novo 3D model was low-pass filtered without imposing any symmetry; 3D classes of 36th iterations and an angular sampling of 7.5 degrees obtained electron densities map range resolution between 11 and 10 Å, it shows in Figure 4.20, .

2.9.2 Second datasets results: purified Apo-hTF

Data was collected from micrographs of specimen 0.64mg/ml. All movies were imported, processed, and motion corrected on *RELION* (v. 3.1.2). Parameters of the contrast transfer function (CTF) for each corrected micrograph were estimated using CTFFIND 4.1.10 (Rohou and Grigorieff 2015).Particles were auto picked by performing LoG (Laplacian-of-Gaussian) Then, Particles extracted from 500 micrographs yielded particles. Those particles were subjected to 2D classification and eventually obtained only one initial model containing (226,731 particles) where two domains were apparently depicted.

Only one 2D class, box size 200 pixels and rescaled to 64 pixels and mask diameter of 150 Å viewing transferrin. The resulting 2D class was used as free 2D references for automated auto picking, which was produced in multi better-oriented classes (with a population of 602,180 particles).

Autopicked particles were noticed not overlapping and distributed neatly nearly 602,180 extracted particles from 22 2D classes were at a pixel size of 2.145 Angstrom/pixel Figure 4.21, (1).2D classes have generated a De novo 3D model class corresponding to the full apo-hTF PDB available atomic models features, shown in Figure 3.2. 200 box size was scaled to 64 for 3D refinement 240 *0.572/120= 2.145 for faster performance. 240 box size was scaled to 120 240 *0.572/120= 1.144 Å/pixel.

The 371,431 particles were 15 low-passes filtered and without imposing any symmetry; generated 3D classes of first 25th iterations and angular 7.5 degrees obtained 9.8 Å. Second 3D classification ran 32ed iterations and angular 3.7 degrees obtained an increased resolution to 8.4 Å. One best class of 3D classification was corresponding to the full apo-hTF and autorefined without a mask in the first round. Second round of refinement was performed using soft edge mask from the pass filtered refined map. Second round of regiment was performed by generating a better mask. This mask was created by fitting the atomic models of apo-hTF independently in the best 3D refined map to generate an mol map then create a mask on *RELION* out of this map. To create molmap the command "molmap #map resolution_value on Grid # 0, was applied on Chimera. The final 3D refined structure was subjected to obtain new 3D classifications. The best resulting 3D classification was subjected to 3D refinement.

2.9.3 Third dataset processing: complex of hTF- Ferric ammonium citrate

Data was collected from micrographs of specimen 2 mg/ml. All movies were imported and processed on *RELION* (v.4.0_beta). Motion correction (*MotionCor2*) implementation in *RELION* was aligned. The beam-induced motion of all the micrograph frames were 3 x 3 tiled frames with a B-factor of 100 (Zhang et al, 2017). Parameters of the contrast transfer function (CTF) for each corrected micrograph were estimated using CTFFIND ctffind-4.1.14 (Rohou and Grigorieff 2015).

Ultimately, micrographs showed spots without any details or features indicating transferrin. For that, Initial auto picking was performed using LoG (Laplacian-of-Gaussian) auto picking, particles initially extracted from 207 automatically selected micrographs yielded 83527 particles. Those particles that were extracted with a binned sized 1.79375 Å/pixel are subjected to 2D classification and eventually have obtained 2D classes contained (49,477 particles). The output of extraction particles subjected to 2D classification, box size 200 pixels and rescaled to 64 pixels and mask diameter of 150 Å. The resulting classes were selected according to the automatic scoring that was executed by python of *TOPAZ* warber which is a new implementation in *RELION* 4.0 for autopicking and de-noising. While the gui of scored 2D classes was opened, a second round for 2D selection was performed against the same extracted particles to select particles according to their distribution, for better manual selection. A second round of autopicking was performed against the 200 micrographs using the coordinates of (49,477 selected particles) to train the neural network in TOPAZ.

After satisfaction with autopicking parameters the same parameters were performed against all the corrected micrographs of hTF- Ferrous ammonium citrate. Also, 2D classification was performed against all the (2,790,085 extracted particles) of the datasets but using 200 hundred patches of gradient-driven algorithm (VDAM) that is new in *RELION* 4.0. which uses batches and it is faster than EM algorithm for such large datasets viewing transferrin.

2.9.4 Fourth dataset processing: complex of hTF- Ferrous ammonium sulphate

Data was collected from micrographs of specimen 1.9 mg/ml. All movies were imported and processed on *RELION* (v.4.0_beta). Motion correction (*MotionCor2*) implementation in *RELION* was aligned. The beam-induced motion of all the micrograph frames were 3 x 3 tiled frames with a B-factor of 100 (Zhang et al, 2017). Parameters of the contrast transfer function (CTF) for each corrected micrograph were estimated using CTFFIND ctffind-4.1.14 (Rohou and Grigorieff 2015).

For that, initial auto picking was performed using LoG (Laplacian-of-Gaussian) auto picking, particles initially extracted from 207 automatically selected micrographs yielded 83,527 particles. Those particles that were extracted with a binned size of 1.79375 Å/pixel are subjected to 2D classification and eventually have obtained 2D classes contained (49,477 particles). The output of extraction particles was subjected to 2D classification, box size 200 pixels and rescaled to 64 pixels and mask diameter of 150 Å. The resulting classes were selected according to the automatic scoring that was executed by python of the *TOPAZ* wrapper which is a new implementation in *RELION* 4.0 for autopicking and de-noising. While the GUI of scored 2D classes was opened, a second round for 2D selection was performed against the same extracted particles to select particles according to their distribution, for better manual selection as displayed in Figure 4.23. A second round of autopicking was performed against the 200 micrographs using the coordinates of (49,477 selected particles) to train the neural network in *TOPAZ*.

After satisfying autopicking parameters the same parameters were performed against all the corrected micrographs of hTF-Ferrous ammonium sulphate. Also, 2D classification was performed against all the (2,790,085 extracted particles) of the datasets but using 200 hundred patches of gradient-driven algorithm (VDAM) that is new in *RELION* 4.0. which uses batches and it is faster than EM algorithm for such large datasets viewing transferrin. That primer 2D class was used as a 2D free reference for the second round of auto picking with optimising the 2D classes by measuring the dimension of the PDB atomic model for transferrin by "Calculate Distances" in *Coot* and adding about an extra 150 Å to most extended distances of 70 Å and diameter 100 Å mask and considering following parameters.

The resulting 2D class was used as free 2D references for automated auto picking, which was produced in multi better-oriented classes (with a population of 602,180 particles). Autopicked particles were noticed not overlapping and distributed neatly and nearly 602,180 extracted particles from 22 2D classes were at a pixel size of 2.145 Angstrom/pixel.

2D classes have generated De novo 3D models with insufficient features, but it was enough to predict from available atomic models that it might visualise either of the half molecule, N lobe or C lobe. The 240 box size was scaled to 64 for 3D refinement 240 *0.574/120= 2.145 for faster performance. Then, the 240 box size was scaled to 120 240 *0.574/120= 1.144 Å/pixel.

The 371,431 particles were 30 low-pass filtered and without imposing any symmetry; generated 3D classes of first 25th iterations and angular 7.5 degrees obtained 9.8 Å. Second 3D classification ran 32ed iterations and angular 3.7 degrees obtained an increased resolution to 8.4 Å. Two different classes were used for the refinement of 3D classification. The atomic models of separate N and C lobes were fitted independently in the best 3D refined map to generate better-masked maps. This was performed using the molmap command "molmap #map number resolution on Grid # 0. Two maps for each specific lobe were subjected to obtain new 3D classifications. The best resulting 3D classification was subjected to 3D refinement.

2.9.5 Fifth dataset processing: thinner ice grid

Data was collected from micrographs of specimen 4mg/ml. All movies were imported and processed on *RELION* (v. 3.1.2) as mentioned in section (2.9.2).

2.9.6 Sixth dataset processing: deglycosylated hTF by PNGase

Data was collected from micrographs of specimen 2.7 mg/ml. All movies were imported and processed on *RELION* (v. 3.1.2) as mentioned in section (2.9.2).

2.9.7 Seventh dataset processing: Volta phase plate (VPP)

Datasets of 766 movies in MRC file format were collected using Volta phase plate (VPP) method (Danev et al. 2014). The movies were processed on *RELION* (v. 3.1.2). The motion correction job was skipped, and the CTF parameters were estimated using CTFFIND ctffind-4.1.14 (Rohou and Grigorieff 2015).Particles were picked first using LoG (Laplacian-of-Gaussian) to produce 2D classes, box size 300 pixels and rescaled to 150 pixels and mask diameter of 160 Å. These 2D classes were used as a reference for a second round of auto picking. Many rounds were run to collect more particles using 2D classification references.

The initial model was close to the dimensions of the PDB atomic model. Still, it was not centred in the middle of the voxel size, and it took one side of the box until it lost some of its edges, which could be because of the asymmetry of the transferrin biomolecule. Therefore, the initial model was replaced by employing an apo-transferrin map that I solved earlier at 6.8 Å to run the 3D classifications against the 188,644 particles of the 2D class averages. To obtain that replacement initial reference map, first, this reference map was low passed filtered to 20 Å and fixed its header (re-boxing) to be matched the voxel size of the selected particles using the following "vope" command in UCSF's Chimera: vop resample #1 onGrid #0. Finally, due to differences in the pixel size of particles (1.79375 Å/pixel) and the reference map (1.148 Å/pixel), the following "rescale" command was performed in the *RELION* argument tap: -angpix_ref 1.79375. The output of the best class was used to perform the second run of 3D classification.

The 3D refinement was run for all three output 3D classes. A second round of 3D refinement was performed for all the three 3D classifications by generating a mask by superimposing PDB atomic model on one of the 3D refined maps in UCSF's Chimera and using the following "molmap" command: molmap #1 15 onGrid #0.

Chapter 3   Determination of the Cryo-EM structure recombinant *Escherichia coli* β-Galactosidase in complex with 4-epi cyclophellitol

Objectives of this study

- To assess the suitability of a 200 keV cryo-microscope for structural studies of enzymatic inhibition, producing 3D reconstructions and refined atomic models of ligand-free and an enzyme-inhibitor covalent complex.
- To evaluate the current mechanisms for modelling of enzyme-cyclophellitol covalent linkages in atomic structures. Identify any issues, propose an alternative mechanism if needed.

Chapter summary

The *Escherichia coli* enzyme β-galactosidase, which is encoded by the *lacZ* gene, has been the subject of extensive and ongoing investigations since its discovery (Jacob and Monod 1961). Since its initial characterisation, *E. coli* β-galactosidase has been employed in a wide variety of applications in numerous fields. It is a hydrolytic transgalactosidase specific to the structure of β-D-galactopyranoside on lactose and shows both hydrolase and transglycosylase activities (Douglas H. Juers, Matthews, and Huber 2012). One primary function of this enzyme is hydrolytic transgalactosidase, which catalyses the hydrolysis of β-D-galactopyranoside-containing substrates and transfers galactosyl groups between molecules. The enzyme acts on substrates containing β-D-galactopyranoside moieties, cleaving the glycosidic bond between the galactose and the other moiety (usually a sugar or aglycone). This results in the release of galactose and the modification of the substrate molecule. This hydrolytic activity is widely used in biochemical, biotechnological, and analytical applications (Guerrero et al. 2015). β-Galactosidase helps *E. coli* metabolise lactose as a carbon source for energy, with its expression being dependent on the relative levels of glucose and lactose. Many breakthroughs have been achieved in understanding the structure of β-galactosidase the first was when its hydrolase activity was detected (Cohn 1957). In subsequent years, the first 3D structure was published (Jacobson et al. 1994), as well as the development of activity-based probes (ABPs) of enzyme action (Vocadlo and Bertozzi 2004).

Electron cryo-microscopy (Cryo-EM) has recently become ubiquitous in most structural biology laboratories worldwide. The resolution revolution, propelled by the development of new direct-electron detectors, better electron sources and efficient classification software methods, has put a new generation microscope in most laboratories. While high-end equipment is responsible for the highest resolution Cryo-EM structures in the Protein Data Bank (PDB), lower powered instruments have been reported to perform well with certain protein samples. In this study, we probed the 3D structure of recombinant β-galactosidase using ABPs to assess the suitability of a 200 keV cryo-microscope for structural studies of enzymatic inhibition, producing 3D reconstructions and refined atomic models of ligand-free and an enzyme-inhibitor covalent complex. The 3D structure of the complex with covalent 4-epi cyclophellitol at 2.7 Å revealed how the enzyme may accommodate these bulky groups. Our observations explain in part why β-galactosidase can tolerate substitutions in substrate

analogues and inhibitors in a manner other related enzyme cannot. We conclude that the use of 200 keV electron cryo-microscopy for imaging protein complexes of ABPs can yield maps at a resolution sufficient for building accurate atomic models, supporting the use of cryo-EM in producing next-generation specific inhibitors.

3.1 Introduction

Proteins, present and synthesised in eukaryotic, archaeal and prokaryotic organisms, perform all metabolic functions within the cellular realm. Proteins are a basic building block of cell membranes, subcellular compartments and cell walls. Moreover, proteins also function as hormones and enzymes that are responsible for almost all types of bodily functions (Brocchieri and Karlin 2005; Kesidis et al. 2020). Enzymes play a key role in the biochemical processes of a body by acting as catalysts. Their presence, on one hand, is essential for the normal physiological functioning of the body and, on the other hand, the excess and deficiency of certain enzymes might lead to disease (Ganem 1996). Studying anomalies affecting proteins often involves the creation of bioengineered species that can be tracked, a key feature of synthetic chemical biology. The detection of bioengineered enzymes includes following biological probes both *in vivo* and *in vitro*, imaging them using modern techniques such as fluorescence microscopy and mass spectrometry methods (Sadaghiani, Verhelst, and Bogyo 2007). Proteins, both wild and recombinant, are encoded by genes; genes are arranged into operons in prokaryotic organisms. The genes in an operon are regulated by a common upstream element and encode proteins that participate in a common metabolic pathway (Price, Arkin, and Alm 2006). One such system is the lac operon.

3.1.1 The lac operon

The lac operon is a classic example of prokaryotic gene regulation. This operon is a genetic cluster that encodes the proteins required to metabolise lactose (Jacob and Monod 1961). This operon is expressed only when lactose is present as the sole carbon source for growth – meaning no glucose can be harnessed. Lactose is metabolised in *Escherichia coli* (*E. coli*) by β-galactosidase, which will be the subject of this investigation. In contrast to eukaryotes, prokaryotic cells lack a nuclear membrane. This enables the messenger ribonucleic acid (mRNA) to translate genes directly into enzymes even as it is being transcribed. Consequently, gene expression and regulation govern enzyme production in these organisms (Jacob and

Monod 1961; Pelley 2011). Altering the substrate in the growing medium may induce the expression of new enzymes capable of metabolising that substrate. An example of this phenomenon appears when a culture of *E. coli* cells in a glucose-rich medium is transferred to a different medium containing lactose as the sole carbon source (Bren et al. 2016). Initially, the cells are quiescent and do not metabolise lactose. Subsequently, other metabolic activities decrease and eventually, cell division halts (Thomson et al. 2016). With time, the culture starts to grow rapidly utilising lactose as the energy source. During the quiescent period, the cells are induced to produce three enzymes that were not expressed in the presence of glucose. These enzymes are involved in the metabolism of lactose. Of these, one is lactose permease (LacY), encoded by the gene *lacY*; the lactose permease is a symporter, meaning that it uses a proton gradient towards the cell to transport lactose in the direction of the cell. The second enzyme is β-galactosidase (LacZ), encoded by the gene *lacZ*, which cleaves lactose into glucose and galactose. The third enzyme is galactoside acetyltransferase (LacA), encoded by *lacA*, which is known as non-metabolisable due to the fact that the product or substrate involved in the reaction catalysed by galactoside acetyltransferase (LacA) is not further processed or utilised by the cell in its metabolic pathways. (Romero-Campero and Pérez-Jiménez 2008; Clark, Pazdernik, and McGehee 2019),(Figure 3.1, A).
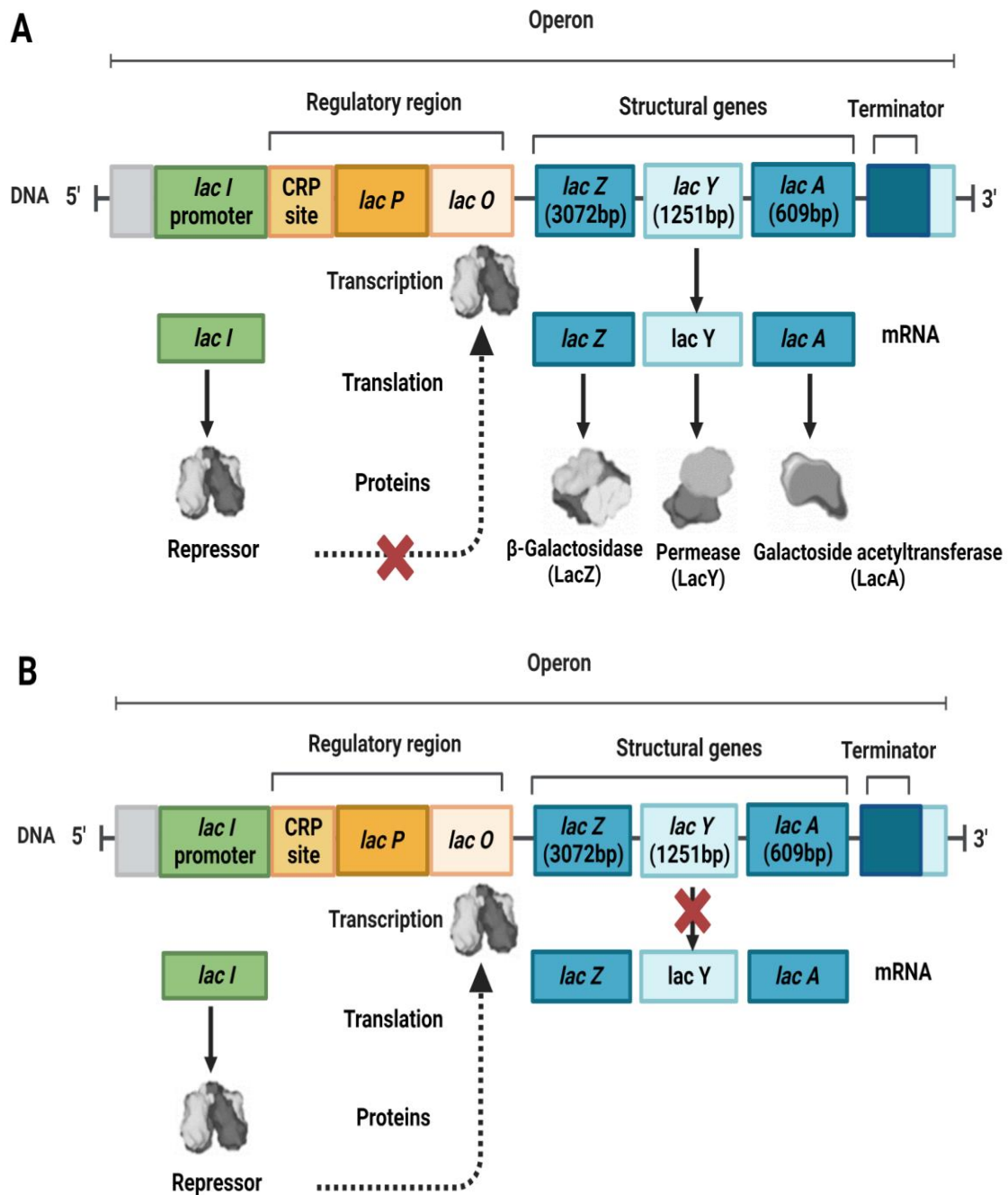
Figure 3.1 Schematic diagram of the Lac operon.
The Lac operon is made up of three structural genes, *lacZ*, *lacY* and *lacA*, that are all transcribed from a single promoter called *lacP*. The promoter is controlled by the repressor binding at the operator, *lacO*, and the CRP protein binding at the CRP site. (A) Operation in the presence of lactose. Lactose is converted into allolactose, which the repressor binds and then releases from the operator. The LacZ, LacY, and LacA proteins are produced by translating a single *lac* mRNA. (B) Operation in the absence of lactose. The repressor is constitutively expressed and binds to the *lacO* operator, thus preventing the binding of RNA polymerase and inhibiting the expression of the *lac* operon genes. It is maade with BioRender using data from (Romero-Campero and Pérez-Jiménez 2008; Clark, Pazdernik, and McGehee 2019).

The *E. coli* lac operon is a negative-inducible operon. Expression of the lac operon is controlled by the operator LacO, promoter LacP, and the LacI repressor encoded by the regulatory gene

lacI (Ptashne and Gann 2002; Ptashne 2004). The LacI repressor adopts its quaternary structure and binds to the LacO operator site to inhibit the expression of the structural genes in the lac operon, as shown in (Figure 3.1, B). The relative levels of lactose and glucose in the cell regulate the expression mechanism. Changes in these levels can result in one of the following scenarios:

1. In the absence of lactose, the repressor is constitutively expressed and binds to the LacO operator, thus preventing the binding of RNA polymerase, thereby inhibiting the expression of the lac operon genes. Notwithstanding, there is a residual level of expression of the Lac operon genes even when repressed.

2. When lactose (Figure 3.2, A) becomes available, it is converted to allolactose (Figure 3.2, B), a disaccharide of D-galactose and D-glucose with a glycosidic bond of β-1,6 instead of β-1,4 in lactose. Allolactose is an inducer of the lac operon. It binds and changes the structural conformation of the LacI repressor, thus preventing LacI binding to the lacO operator. The operon is thus transcribed by the RNA polymerase, and the proteins made can cleave lactose to produce further allolactose and metabolise lactose to glucose and galactose.
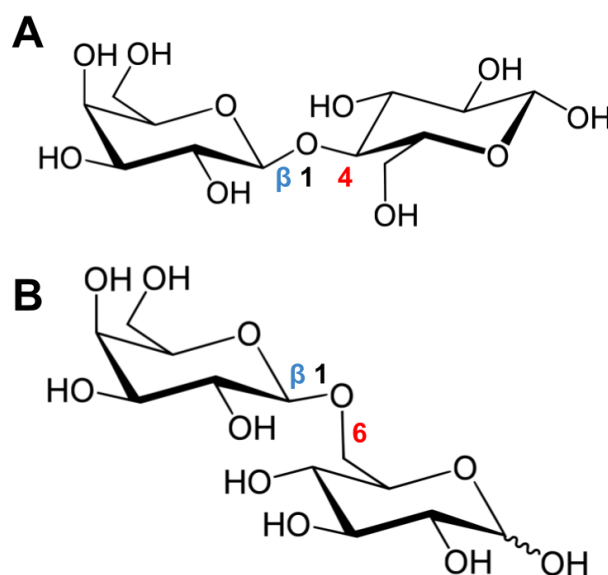
Figure 3.2 Chemical diagram of lactose and allolactose
(A) Lactose (B) Allolactose with their respective β1,4 and β1,6 links annotated. Figure generated with Google Draw using source material made in Chemdraw.

3. The bounded cAMP receptor protein (CRP); also known as catabolite activator protein (CAP) complex, is intricately tied to the polymerase bound to the *lacP* promoter.

Consequently, *lac* mRNA synthesis at high levels leads to the *lac* operon being transcribed extensively.

4. High glucose levels inhibit the enzyme adenylate cyclase. As a result, cAMP is not synthesised to activate the catabolite activator protein (CAP), leading to reduced transcription of the *lac* operon (Shimizu 2013). This regulatory pathway ensures the avoidance of futile expression of the *lac* operon (Pelley 2011). We can summarise these four cases in the following (Table 3.1).

Table 3.1 Summary of different possibilities leading to different levels of transcription.

Glucose	Lactose	CAP binds	Repressor binds	Level of transcription
Present	Absent	-	+	No
Present	Present	-	-	Low
Absent	Absent	+	+	No
Absent	Abundant	+	-	Strong

The lac operon genes cluster ultimately compromises the *lacZ* transcribed by mRNA polymerase to polycistronic mRNA (lac mRNA), which translates into three enzymes. The *lacZ* gene in *E. coli* translates to yield β-galactosidase (Douglas H. Juers, Matthews, and Huber 2012). The objectives of this study are to shed light on the suitability of a modern imaging technique such as Cryo-EM in the study of biological probes, previously only extensively characterised by high resolution X-ray crystallography.

3.1.2 Activity based probes (ABPs)

ABPs are reagents that facilitate the identification of enzyme activity and monitor the presence of an enzyme's active site (Tsai, Li, and Lo 2002; Sieber 2012; Kuo et al. 2018; L. Wu et al. 2019). Since enzymes are proteins in nature, ABPs are used with proteomic techniques to provide the proportion of total functional enzymes within an organism (Willems et al. 2014). ABPs have experienced substantial application in research studies since their labelling

properties contribute essential medicinal and manufacturing advantages. ABPs include reagents that have a strong binding potential and typically bind within the active site of enzymes. As a result of this binding, enzyme activity can be measured in a cell, tissue, or organism, *in vivo*, under specific conditions. Moreover, ABPs have also allowed us to study the binding capabilities of an enzyme inhibitor. The estimation of the inhibition potential of an enzyme inhibitor is beneficial in the treatment of diseases (Blum, Verhelst, and Ma 2021).

3.1.2.1 General catalytic mechanism of retaining glycosidases

Glycoside hydrolases (GH) enzymes are essential catalysts for carbohydrate energy extraction. Among the most abundant GH families are the retaining β-glycosidases GH2, characterised as the GH Clan A (Henrissat et al. 1995; Lee 2017). The *E. coli* (LacZ) β-galactosidase is classified among the broadest glycoside hydrolase family (EC 3.2.1.23). As previously introduced, it degrades lactose into glucose and galactose, which can then be oxidised to extract energy (Douglas H. Juers, Matthews, and Huber 2012). Glycoside hydrolases play a crucial role in metabolism. An in-depth understanding of their catalytic mechanism might lead to better developed biological and biomedical applications (Tailford et al. 2007; Lammens et al. 2009).

The active site of the *E. coli* β-galactosidase (LacZ) is located in the $(β/α)_8$ barrel fold region (a TIM, or triose-phosphate isomerase, barrel domain). The classification of all other glycosidase families is based on the location of the active site (Douglas H. Juers 2000; Douglas H.. More precisely, β-galactosidase is also classified as a retaining glycosidase, because it maintains the anomeric configuration of the anomeric substrate carbon (Brás, Fernandes, and Ramos 2010; Douglas H. Juers, Matthews, and Huber 2012). The active residues consist of a putative acid–base, Glu461 and an enzymatic nucleophile Glu537 (Gebler, Aebersold, and Withers 1992). The nucleophile Glu537 residue was identified through radiolabeled tracers in the High-Performance Liquid Chromatography method (HPLC). The nucleophilic position was initially mislabelled due to the inhibitor's promiscuity (Herrchen and Legler 1984), which was later corrected by the Stephen G. Withers group in Canada in 1992. Subsequently, a 3D structural analysis later fully supported the correct active site assumptions (Jacobson et al. 1994). Retaining β-glycosidase enzymes such as β-galactosidase hydrolyse and deglycosylate the substrate in two steps, called the Koshland double-displacement reaction mechanism

(Koshland 1953), as shown in Figure 3.3. The first catalytic step (glycosylation/galactosylation) describes the nucleophilic attack of the carboxylate (glutamic or aspartic acid, Glu537) residue of the enzyme at the anomeric carbon of the glycoside substrate with accompanying proton release from the acid residue or the base residue of the remaining aglycon. This is started by proton contribution from Glu461 to the specific oxygen of the glycosidic linkage to create a covalent intermediate, *i.e.*, glycosyl-enzyme.

In the following displacement (the deglycosylation step), a water molecule operates as the nucleophile to accomplish hydrolysis and deglycoslyate the covalent intermediate, deprotonated by the second carboxylic acid residue Glu461 (McCarter and Withers 1994). Both specific displacements occur within a planar oxocarbenium ion-like transition state, which places a strong constraint on what ring conformations the substrate can adopt. The acceptor reaction with glucose is essential for producing allolactose, the essential inducer of the lac operon.

Mechanism of retaining glycosidase (lacZ β-galactosidase)

R$_1$= Glucose(for lactose)

R$_2$= H (Water) for hydrolysis
R$_2$= another substrate (transglycoslation)
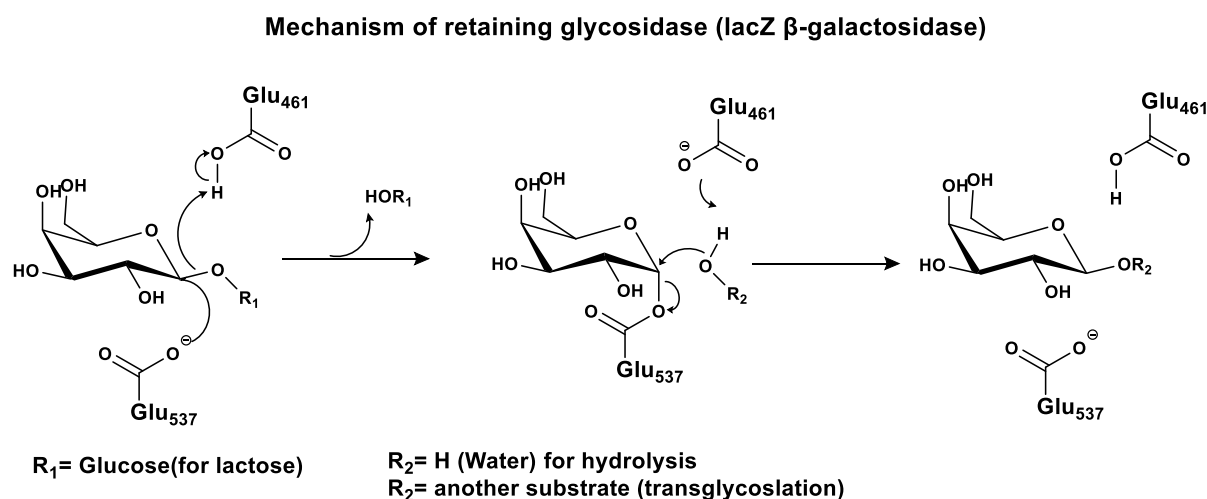
Figure 3.3 The double displacement reaction mechanism of LacZ β-galactosidase.
This figure shows two residues, Glu537 the nucleophilic residue, and Glu461 the acid/base residue. If R2 = H (water), the reaction is termed hydrolysis of the substrate; otherwise, it is known as a transglycosylation if another sugar is involved. The active centre nucleophile, Glu537, is the principal factor in the reactivity of the epoxides and aziridines discussed later in this chapter. Figure adapted from (McCarter and Withers 1994; Kjeldsen, Ardenkjær-Larsen, and Duus 2018).

3.1.2.2 Mechanism-based inhibitors of retaining glycosidases

All living organisms contain β-glycosidases, thus indicating the importance of these enzymes in various biochemical activities. However, certain environmental and genetic factors can alter or inhibit the function of these enzymes. For example, when a mechanism-based inhibitor comes into contact with a target enzyme, a covalent bond between the reactants

(enzyme and ligand) that is stable over time is created, thus leading to the permanent inactivation of the enzyme.

Mechanism-based enzyme inhibition can be performed most efficiently when an enzyme processes its substrate through the formation of a covalent intermediate (Shamsi Kazem Abadi et al. 2017). The covalent interaction between the enzyme and a functional group of the inactivator eliminates enzyme activity (Legler 1990). The bond is typically formed through a nucleophilic attack of an activated centre on the inhibitor causing the creation of a covalent complex. This class of inhibitors has attracted interest in probing β-glycosidase structure and function and identifying catalytic residues. β-glycosidase inhibitors can be categorised into several classes based on their interactional behaviour and impact (Rempel and Withers 2008).

Covalent inhibitors have gained significant importance in the fields of advanced therapeutics and biochemical diagnostic tools. The advantages of covalent enzyme inhibition in developing advanced therapeutics outweigh the initial concerns, such as potential off-target toxicity (Sutanto, Konstantinidou, and Dömling 2020). For example, therapeutic drugs that work on covalent enzyme inhibition are long acting in nature and, therefore, require a less frequent dosing regimen. Other advantages include enhanced efficiency in targeting binding sites that are impenetrable or shallow, efficient target ligand binding, and efficient elimination of competing ligands. Similarly, covalent enzyme inhibitors can help identify and target isolated ligand variants (Sutanto, Konstantinidou, and Dömling 2020).

3.1.2.3 Activity-based probes of retaining glycosidases

As discussed above, ABPs are reagents that facilitate the identification of enzyme activity and monitor the presence of the enzyme's active site (Tsai, Li, and Lo 2002; Sieber 2012; Kuo et al. 2018; L. Wu et al. 2019). A mechanism-based inhibitor allows for specific labelling of the desired enzyme function (Withers and Aebersold 1995).

ABPs typically contain three structural elements. First, a probe moiety (a reporter or tag) is commonly composed of a fluorescent group (for instance, BODIPY, rhodamine or fluorescein) for visualisation. Visualisation is done on a gel, living cells, or a biotin group for affinity enrichment, purification, and subsequent detection by mass spectrometry. Second, a reactive

moiety (warhead) is attached to the probe moiety through a third element, a linker, so that the probe does not interfere with the binding of the reactive moiety to the enzyme (Goddard-Borger, Wennekes, and Withers 2010).

ABPs have found various biological applications; α- and β-glycosidase ABPs have been synthesised based on this principle (Kallemeijn et al. 2012; Caner et al. 2016). Three major classes of ABPs have been created for retaining glycosidases: fluorinated glycosides, cyclophellitol epoxides (Stubbs 2014) and cyclophellitol aziridines. These three probes are chemically inert until the catalytic glycosidases activate them. Fluorinated glycosides, such as 2-deoxy-2-fluoro- (Vocadlo and Bertozzi 2004), (Figure 3.4), 5-fluoro- and 2-deoxy-2,2-difluoroglycosides, were first introduced by the Withers' laboratory as inhibitors of glycosidases (Gebler, Aebersold, and Withers 1992; Withers and Aebersold 1995).

Cyclophellitol epoxide is known for its selectivity towards the active site. It was isolated from the natural wood-rotting fungus *Phellinus sp.* (Atsumi et al. 1990). Afterwards, cyclophellitol aziridines were yielded by substituting the epoxide moiety in cyclophellitol with aziridine, substituting the oxygen with a three-membered nitrogen ring. The aziridine derivatives were reported to inhibit glucocerebrosidase (GBA1) (Li et al. 2014).
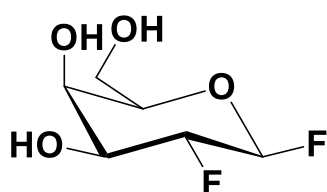
2-deoxy-2-fluoro-β-D-galactoside

Figure 3.4 The structure of fluorinated glycoside-derived ABPs.

3.1.2.3.1 Fluorinated inhibitors as ABP

Vocadlo and Bertozzi developed the strategy for detecting retaining glycosidases in complex mixtures based on extensive research by the Withers Laboratory on the inactivation of β-galactosidase (LacZ) by 2-flourogalactosides (Gebler, Aebersold, and Withers 1992). The primary hydroxyl group (C6-OH) of 2-deoxy-2-fluoro-b-D-galactosyl fluoride ABPs were replaced with an azide moiety, as shown in Figure 3.5, to detect labelled proteins through the

Staudinger–Bertozzi ligation or through azide-alkyne Huisgen cycloaddition (the 'click' reaction). Nevertheless, the tagged enzyme's potency was unstable compared to its non-probed enzyme. Subsequently, Chemo-selective Staudinger ligands were used to attach the FLAG tag, thus enhancing the stability.

The FLAG tag is a short, hydrophilic peptide used as an epitope tag recognized by specific antibodies, making it useful for detection and purification (Belardinelli and Pelleg 2012; Zhao, Li, and Liang 2013). When attached via the precise method of Staudinger ligation, it enhances the enzyme's stability (Bednarek et al. 2020). This increased stability could result from reduced structural disruption, preserving the enzyme's natural folding and function, and improved interaction with the surrounding environment, minimising denaturation or degradation (Köhn and Breinbauer 2004; Gentilucci et al. 2016).

The reaction of tagged proteins with a phosphine FLAG was conducted to visualise the active LacZ β-galactosidase on a Western blot. The ABP also enabled the labelling of this enzyme activity and several other recombinants retaining β-glycosidases in cell extracts (Fonović and Bogyo 2008). More galactosidases from various GH families have been labelled successfully using ABPs (Williams, Hekmat, and Withers 2006; Stubbs et al. 2008; Tsai et al. 2013). Figure 3.5 shows that the β-galactosidase has been labelled with FLAG epitope following by the 6-azido-2,6-dideoxy-2-fluoro-β-D-galactosyl fluoride (6Az2FGalF) treatment in the cell extracts.
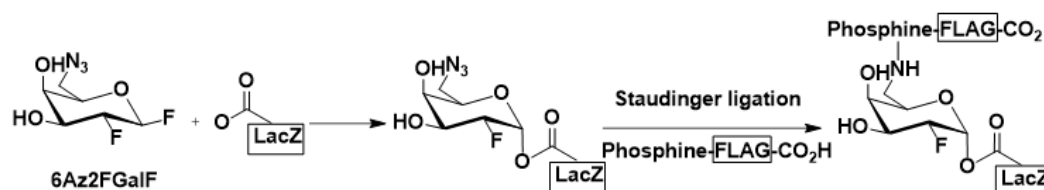
Figure 3.5 The activity-based labelling approach and subsequent Staudinger ligation.
Both LacZ β-galactosidase alone and in-cell extracts from *E. coli* were labelled by incubating with 6Az2FGalF, forming a covalent glycosidase adduct nucleophilic atom on the active site. Figures are taken from (Vocadlo and Bertozzi 2004).

3.1.2.3.2 Epoxide- and aziridine-based inactivators

These inactivators have found successful applications due to their high inactivation efficiency and stability. The mechanism-based inhibitor potential of cyclophellitol epoxide was tested on mice suffering from Gaucher disease symptoms, likely occurring from retaining β-glucosidase GBA1 inhibition (Atsumi et al. 1992). Subsequently, cyclophellitol (Figure 3.6, left

panel) acted as a permanent glycosidase inhibitor (Gloster et al. 2007). The Hermen Overkleeft group (Leiden University, The Netherlands) subsequently predicted that by combining multifunctional groups (biotin capture probes, azides for click chemistry and fluorescent groups) at the C6- position, β-glucosidase-specific ABPs could be created (Witte et al. 2010) (Figure 3.6, right panel).

Mechanism of a covalent inhibitor of natural product

Cyclophellitol   Cyclophellitol β-epoxide   Covalent intermediate
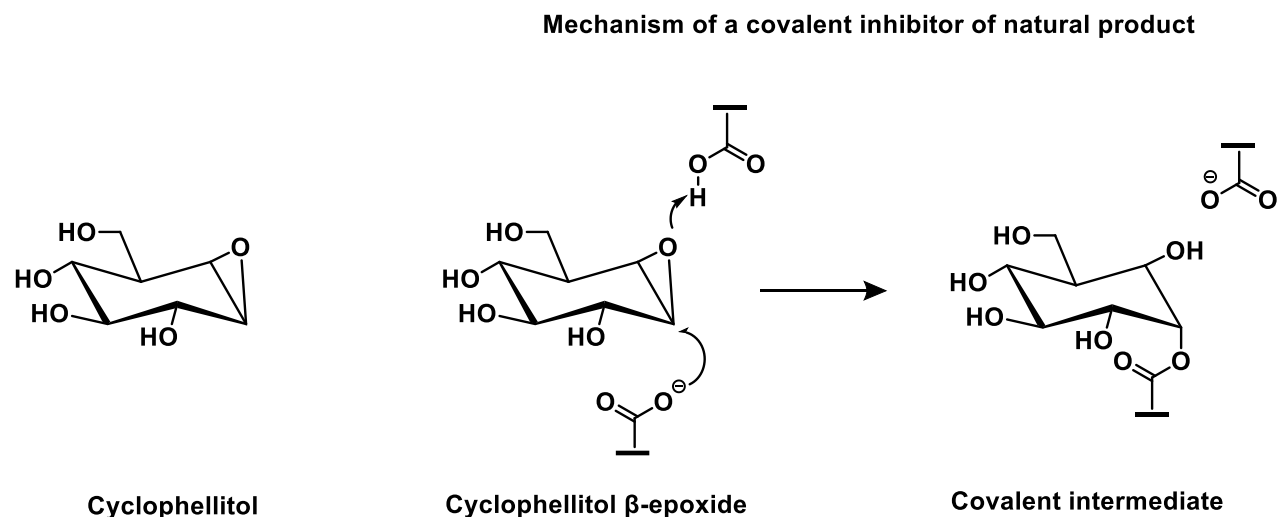
Figure 3.6 Retaining β-glucosidase mechanism-based inhibitors.
Structure of the natural product cyclophellitol (left) known β-epoxide labelling the nucleophile of a retaining β-glucosidase and their mechanism of reaction (right). The figure was adapted from (Ben Bdira et al. 2018).

Aziridines and epoxides are widely used to identify the catalytic residues of glycosidases as they exploit the first step of the catalytic enzyme mechanism. The inhibition mechanism requires attack of the epoxide or the aziridine by the enzymatic nucleophile, cleavage of the ring and formation of covalent bonding between the enzyme and the inhibitor, as shown in Figure 3.7. This bond emulates the intermediate state of the natural hydrolysis process. General acid or base residue protonated the inhibitor in this reaction to promote it (Withers and Aebersold 1995).

Mechanism covalent of based inhibitors

Acid/Base

Nucleophile

Covalent intermediate
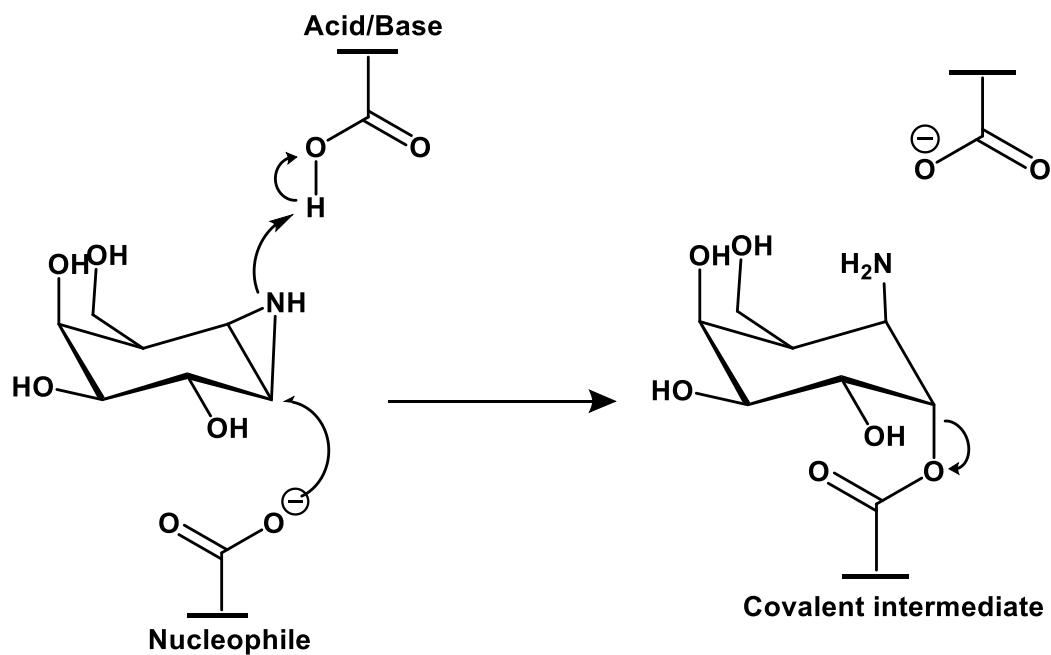
Mechanism covalent of activity based probes

Acid/Base

Nucleophile

Covalent intermediate
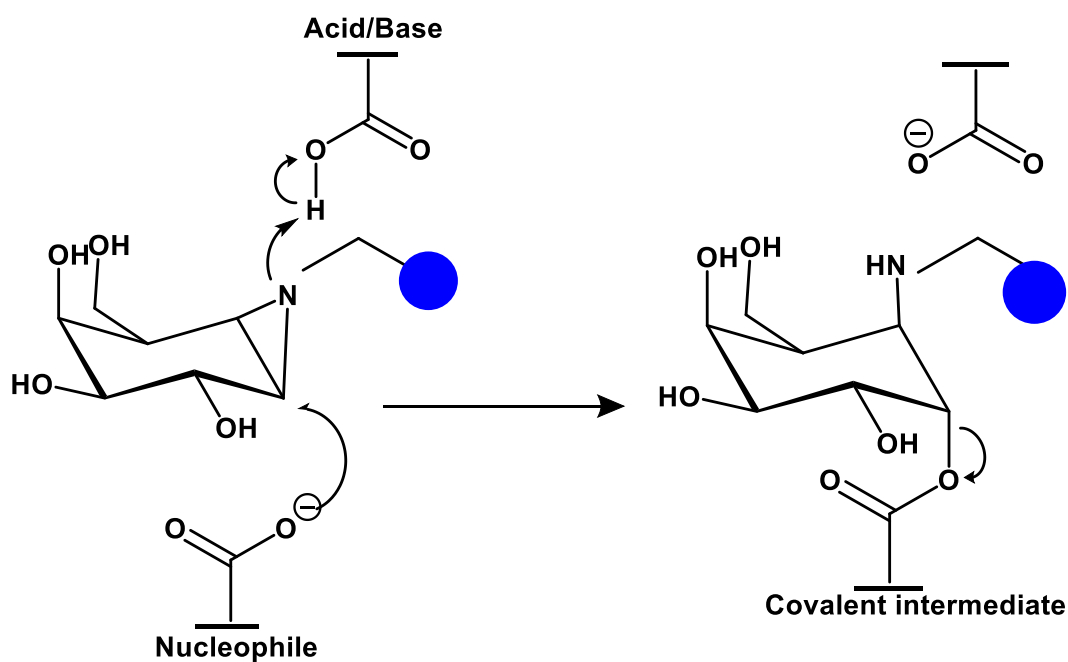
Figure 3.7 Mechanism of based inhibition.
Aziridine is a covalent inhibitor, the mechanism above, and Aziridines are ABPs of the enzyme active site, the mechanism underneath. The blue circle represents the reporter.

In 1974 Quaroni and colleagues used conduritol B epoxide (CBE) in sucrose and isomaltose

to identify the carboxylate, in Figure 3.8, left panel, (Quaroni, Gershon, and Semenza 1974).

Similarly, radioactive forms of CBE were also used to identify the catalytic residues of

glycosidases before bioinformatics emerged as a predictive tool that supported biologists in analysing and interpreting data (Rempel and Withers 2008). These tools are computational platforms developed by scientists, non-profits and open source. These enable comparing, assembling and annotating proteins and genes (Perello 2018). The structural symmetry of CBE allows it to label both α- and β-glycosidases (Herrchen and Legler 1984; Hermans et al. 1991). Moreover, CBE has been shown to interact not only with nucleophilic residues involved in the catalysis but, in some cases, with other active site residues. For instance, conduritol C *cis*-epoxide, illustrated in Figure 3.8, right panel, in β-galactosidase (LacZ), β-glucosylceramidase and almond β-glucosidase, the identification of the nucleophilic residues was initially erroneous due to the inhibitor's promiscuity, which was later corrected (Gebler, Aebersold, and Withers 1992; Miao et al. 1994).
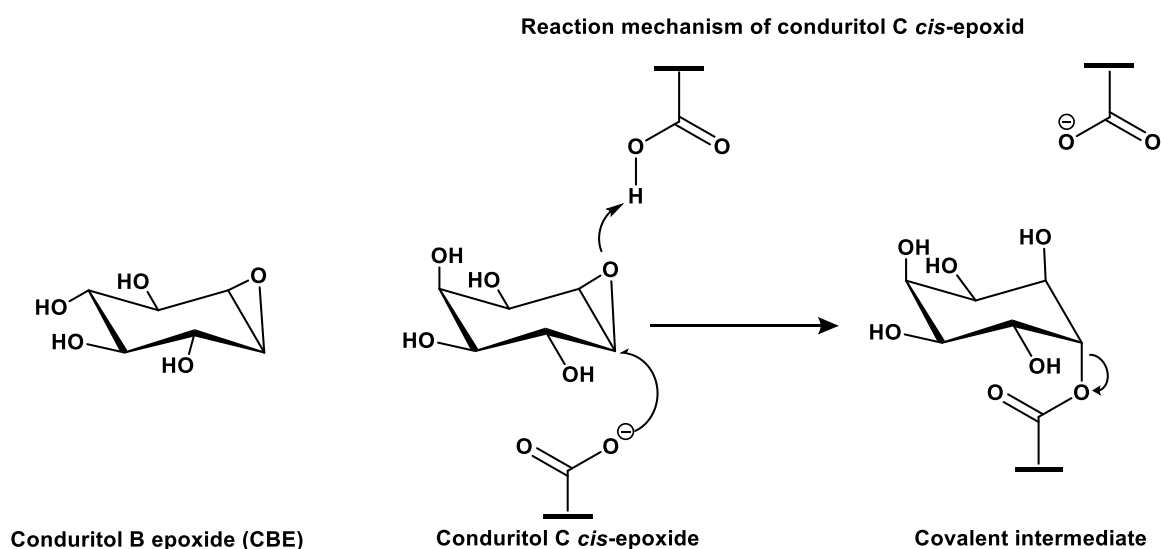
Figure 3.8 Potent irreversible epoxide-wielding probes of β-glucosidases.
 Conduritol B epoxide (CBE) (left) and Catalytic nucleophile-binding mechanism of conduritol C *cis*-epoxide (right). The mechanism of action was adapted from (Ben Bdira et al. 2018).

3.1.2.3.3 PDB examples of complex structures reacted with Cyclophellitol epoxide

The Protein Data Bank (PDB) (Berman et al. 2000) is a valuable resource for researchers to access three-dimensional structural information about biological macromolecules and their complexes with small molecules like inhibitors. A survey of structures of cyclophellitol epoxide inhibitors complexed with various glycosidase enzymes deposited in the PDB was conducted, and a summary is shown in Figure 3.9.

While some entries contained the intact ligand description in their ligand summary – *e.g.* PDB code 6IBR (Artola et al. 2019) – most of the entries surveyed displayed the probe in its reacted form, *i.e.,* with the epoxide gone. While this approach is clearly effective in modelling the reacted form, it does cause a mismatch between the entry's title (*e.g.* "beta-galactosidase in a covalent complex with 4-epi cyclophellitol") and the ligand definition shown in the relevant section. One reason for this mismatch could likely be that restraint dictionary generation software are currently not able to process the more complex reactions when generating a covalent link, offering essential functionality to link two atoms, delete leaving groups and, if required, change the order of bonds as part of their ligand processing mechanisms. This was the case for the main restraint generation software in the CCP4 and CCP-EM suites, AceDRG. Crucially for ligands such as 4-epi cyclophellitol, the epoxide must be opened meaning a bond must be deleted.
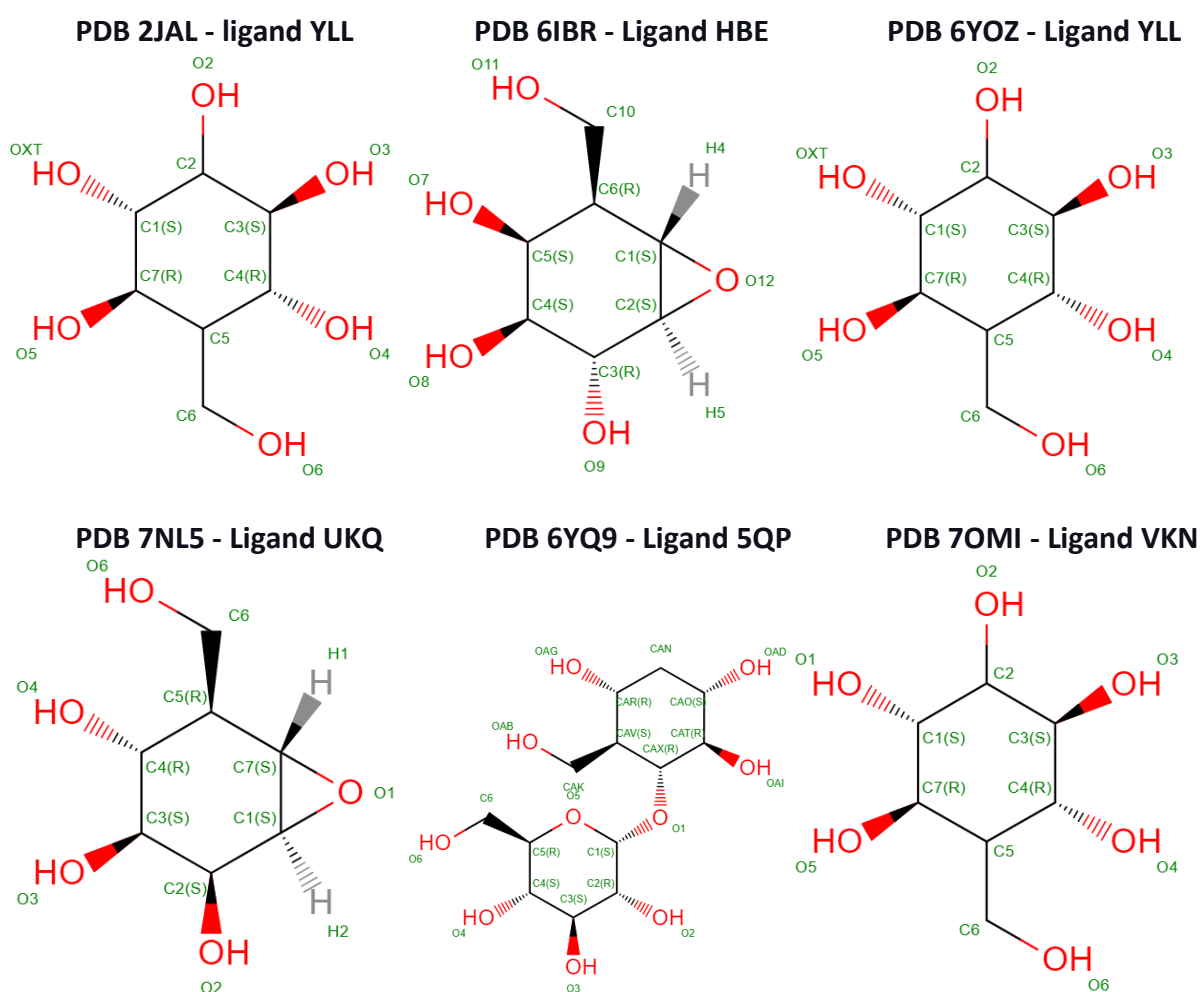
Figure 3.9 Structures of cyclophellitol epoxide inhibitors complexed with various glycosidase enzymes deposited in the PDB.

(PDB code: 2JAL) - ligand YLL β-glucosidase from Thermotoga maritima in complex with cyclophellitol (Gloster,

Madsen, and Davies 2007). (PDB code: 6IBR) - Ligand HBE Crystal structure of human α-galactosidase A in complex with α-galactose configured cyclophellitol epoxide LWA481 (Artola et al. 2019). PDB 6YOZ - Ligand YLL HiCel7B labelled with b-1,4-glucosyl cyclophellitol (de Boer et al. 2021). (PDB code: 7NL5) - Ligand UKQ Structure of the catalytic domain of the Bacillus circulans α-1,6 Mannanase in complex with an α-1,6-α-manno-cyclophellitol trisaccharide inhibitor (Schröder et al. 2021). (PDB code: 6YQ9) - Ligand 5QP Taka-amylase in complex with α-glucosyl epi-cyclophellitol epoxide inhibitor (Y. Chen et al. 2021). **(**PDB code: 7OMI) - Ligand VKN Bs164 in complex with mannocyclophellitol epoxide (McGregor et al. 2022). Structures adapted from PDB.

In discussions with the AceDRG developers (Fei Long, Robert Nicholls and Garib Murshudov) this lack of ability to delete bonds in the software was perceived as a deficiency which they sought to address. We coordinated efforts with the AceDRG team and annotators at the PDBe (Deborah Harrus and others) to upgrade the software to produce dictionaries that combine the intact description of an epoxide-containing probe with the instructions to produce the link, including the deletion of the relevant bond.

3.1.3 Structural studies of the bacterial retaining enzyme β-galactosidase

Treating disease with covalent inhibitors becomes an attractive target for many benefits – several advantages have been mentioned previously. Working out the structural basis of an inhibition mechanism requires structural elucidation in relatively high detail: molecular features must be identifiable at the atomic level, and therefore the resolution required for this sort of studies necessarily has to approach ~2 Å. The use of Cryo-EM techniques looks promising for the fast discovery and development of novel therapeutics, essentially due to the relative absence of bottlenecks (*e.g.* the requirement of crystals for doing X-ray crystallography) and increasing speeds of computational image reconstruction. However, achieving resolutions as low as 2 Å might be challenging. What makes it challenging is yet to be fully understood, but human and technical errors in sample preparation, microscope adjustments, inconsistencies during data collection and processing might be some factors that may deter. Cryo-EM techniques are more compatible with molecules of higher symmetry and larger molecular weight, such as ribosomes, viruses (Henderson 1995), and *E. coli* β-galactosidase (Agard et al. 2014). In this study, *E. coli* β-galactosidase (LacZ) was used as a model system because it has been successfully resolved in both liganded and unliganded forms at 2.2 Å (Bartesaghi et al. 2015; Saur et al. 2020) and 1.8 Å (Merk et al. 2020) respectively, using a microscope that is very similar to the entry-level 200 keV Cryo-EM microscopes installed throughout the UK in the present decade.

3.1.3.1 History of β-Galactosidase (LacZ) and applications

β-Galactosidase is not only readily available – can be isolated directly from bacteria – but also has a vital role in understanding gene expression control mechanisms and the operon structure. It also performs a crucial role as a living cell probe (Craig and Henderson 2013). Many functional properties of this enzyme have made it a desirable model in the development of biological imaging technology and software analysis procedures. The first property is the size of the structure of *E. coli* (LacZ) β-galactosidase, which has a large macromolecular weight for a bacterial protein (Wheatley et al. 2013). Another property of β-galactosidase is its a D2-symmetrical structure that, crucially, aids in image reconstruction and signal processing (Bartesaghi et al. 2014, 2015, 2018)**.**

The enzyme was evaluated for its specificity regarding the location of the galactosyl group and hydrolysis of D-galactopyranosides with various aglycones. It is tolerant to substrate-based colorimetric assays, *e.g.* o-nitrophenyl-b-D-galactopyranoside (ONPG) (Rouwenhorst, Pronk, and van Dijken 1989) along with 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal) having chromogenic aglycons (Douglas H. Juers 2000). The hydrolase reaction mechanism was first suggested by Yon and Sinnott in the early 1970s. They proposed a two-step kinetic mechanism that focused on the rate-limiting steps in prokaryotic gene expression by accumulating a trapped covalent glycosyl–enzyme intermediate through the use of 2-deoxy-2-fluorogalactose as inhibitor (Sinnott and Souchard 1973; Viratelle and Yon 1973). Later, they suggested a typical Koshland double-displacement mechanism for β-galactosidase (Figure 3.3), with Glu537 (identified by HPLC techniques using radiolabeled tracers) acting as the catalytic nucleophile (Gebler, Aebersold, and Withers 1992).

3.1.3.2 Early studies of β-Galactosidase structure

E. coli β-galactosidase is the most common enzyme used to investigate and monitor gene expression and regulation in biological systems (Tung et al. 2004; Wehrman et al. 2006; Kamiya et al. 2007). To date, over 60 X-ray crystallographic structures of β-galactosidase are available in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) (see http://www.cazy.org/GH2_structure.html). To date, two unliganded structures have been determined at relatively low resolutions. The first ever crystal structure of β-galactosidase was determined at 3.5 Å by (Jacobson et al. 1994). Afterwards, the same

structure was refined to reach 2.5 Å (Douglas H. Juers 2000). The Juers group managed the refinement of a single tetramer per asymmetric unit to push the resolution to 1.7 Å, which allowed a series of detailed atomic models to be built (Douglas H. Juers 2000). One year later, a LacZ–galactose complex was described at a higher resolution of 1.5 Å (Douglas H. Juers et al. 2001). This made it possible to extend the structure to probe a range of inhibitors and ligands (Matthews 2005).

The 3D structure of β-galactosidase is a D2-symmetric tetramer, with four identical subunits as the active form of the enzyme with a molecular weight of 465.4 kDa (Bartesaghi et al. 2014). It consists of four identical polypeptide monomers, each having a considerable molecular weight of approximately 116.3 kDa shown in Figure 3.10 chain D has an α/β or TIM barrel composition with the active site located at the barrel's C-terminal end that is critical of the active site (Douglas H. Juers 2000; Douglas H. Juers et al. 2001; Matthews 2005).
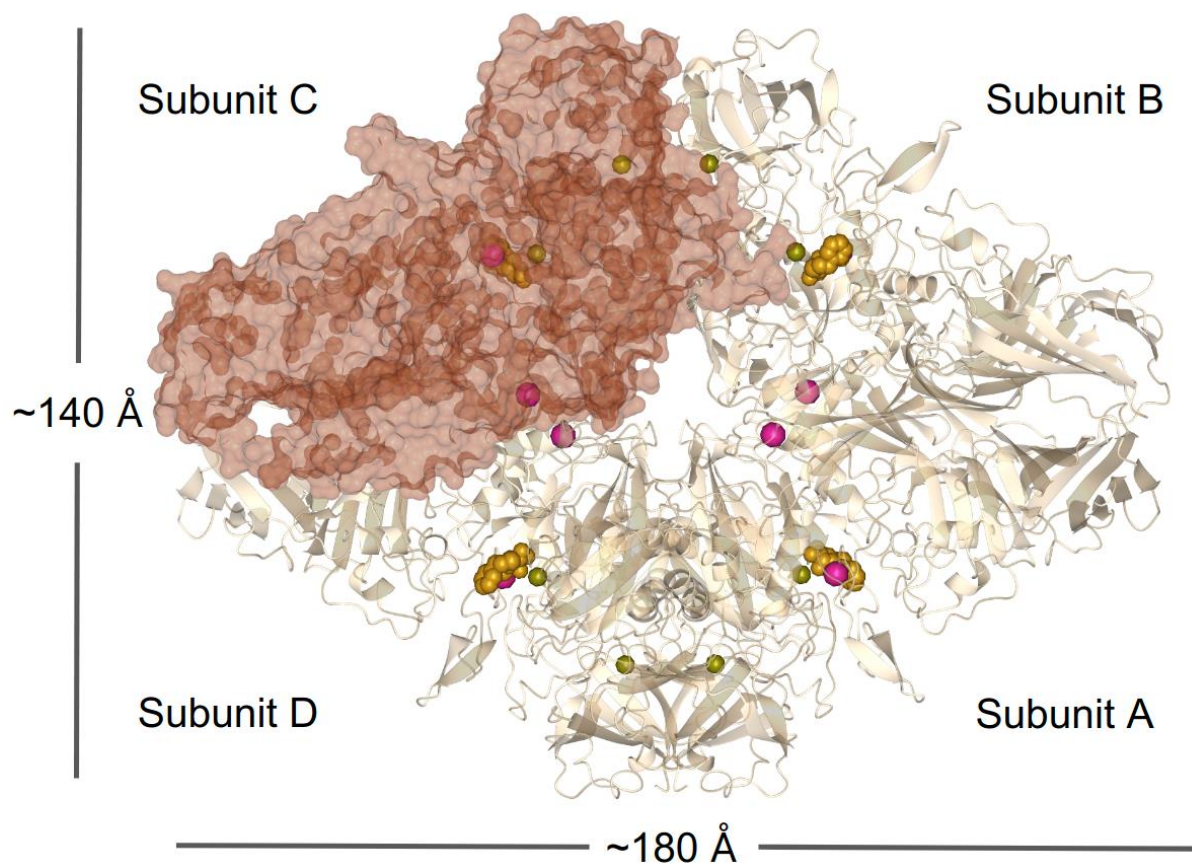
Figure 3.10 The D2 tetramer structure of the β-galactosidase from *E. coli* (PDB code: 5A1A).
Subunit C, shown in a transparent surface representation, is related to subunits A and D (shown as ribbons) by the 2-fold symmetry axis. The four active site subunits are bound with four gold spheres, each representing phenylethyl β-D-thiogalactopyranoside (PETG). The metal ions are depicted as eight small green spheres (Mg^{2+} ions) and eight large pink spheres (Na^+ ions). The figure was created using CCP4mg (McNicholas et al. 2011). The

four subunits of the D2-symmetrical molecule are related by three 2-fold symmetry axes: one horizontal, one vertical, and the third going through the centre of the figure perpendicular to the plane of the page.

The retaining β-galactosidase (EC 3.2.1.23) maintains the anomeric configuration of the substrate anomeric carbon (Brás, Fernandes, and Ramos 2010; Douglas H. Juers, Matthews, and Huber 2012; Lombard et al. 2014; Drula et al. 2022). It is a 1023-residue member of the glycoside hydrolases belonging to the CAZy (www.cazy.org) family 2 (GH2) (Drula et al. 2022). β-Galactosidase hydrolyses the non-reducing β-D-galactose residues present at the end and cleaves lactose into its two monosaccharide units (glucose and galactose) in bacteria (Figure 3.11); the non-reducing residue can also be reused in the bacterial cytoplasm. Furthermore, β-Galactosidase catalyses lactose's transgalactosylation to form allolactose, which acts as an inducer of the lac operon by binding to the lac repressor protein and preventing it from binding to the operator region, thereby relieving repression and allowing transcription of the *lac* genes to occur. Subsequently, allolactose can also be broken down into glucose and galactose (Huber, Kurz, and Wallenfels 1976; Douglas H. Juers, Matthews, and Huber 2012).
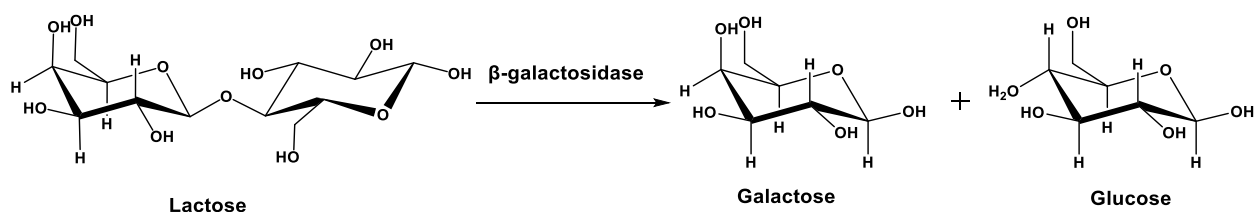
Figure 3.11 The reaction catalysed by β-galactosidase.
The hydrolysis of lactose with net preservation of the stereochemistry at the anomeric centre, where the glucoside group on the non-reducing β-D-galactose residue is substituted by a water molecule (the acceptor group). The figure is taken from (Douglas H. Juers, Matthews, and Huber 2012).

3.1.4 Chapter objectives

Macromolecular XRC relies on the formation of a physical protein crystal that can diffract X-rays. Cryo-EM, on the other hand, depends on the creation of a virtual crystal—an addition of views of a particle, computationally extracted from micrographs—which then leads to a 3D map. The objective of this study is to use laboratory standard Cryo-EM equipment to investigate β-galactosidase structure with ABPs in high resolution, which could open the doorway to explore the structures and functions of proteins in depth. The classification techniques, sample preparation protocols and imaging procedures used in Cryo-EM may be tuned to produce different or potentially better results. At least one sub-2 Å β-galactosidase structure has already been determined at 200 keV, which means Cryo-EM could provide

sufficient resolution to study ABPs. While in many cases XRC has the upper hand in resolution, Cryo-EM may provide a valuable niche to use techniques where only small quantities of the probe or target can be acquired. For example, with complicated or unstable probes and with enzymes that are isolated from the actual organism of interest rather than overexpressed. While these caveats do not apply to the β-galactosidase of *E. coli*, nonetheless, it provides a proof-of-concept platform for later studies.

To date and to the best of our knowledge, no systematic study exists on the influence of image classification and refinement parameters focused on ABPs bearing epoxide and aziridine scaffolding with β-galactosidase. This chapter described some variables to come up with a class method for the production and purification of the apo β-galactosidase following confirmation of the enzyme activity that was subjected to two ABPs. Subsequently, β-galactosidase was cryogenically treated, and a 3D structure was reconstructed to produce atomic models with an ABP selected to indicate the reactive groups and configurations, etc. Furthermore, this study would allow the comparison to similar studies in Cryo-EM, thus allowing the development of imaging preparation of selective inhibitors for β-galactosidase. Two complexes bearing the following small molecules were initially sought: β-galactose epoxide (Willems et al. 2014) and N-octyl norbornene-Cy5-galacto cyclophellitol aziridine (Beenakker et al. 2017), (Figure 2.5).

3.2 Results and discussion of β-galactosidase

The genes encoding for bacterial β-galactosidase expression in *E. coli* and protein purification was done successfully, and cryogenic in a form suitable for further structural study.

3.2.1 Overexpression and purification of β-galactosidase

3.2.1.1 Insert preparation

PCR of inserts have been carried out using primers and the genomic DNA template.

The *lacZ* genes were purified using the Montage PCR Clean Up Kit (Millipore) to yield DNA after a 15-minute spin; the DNA was stored at -20°C.

3.2.1.2 Transformation and overexpression of *lacZ* gene

Protein was His-tagged, and it was separated with a His-Trap column. Additionally, Tuner cells were used to transform the target gene instead of BL21 Gold cells. Transformation is the uptake and integration of single strands of foreign DNA as single strands into the chromosomes of bacteria through homologous recombination (plasmid) (Johnston, Martin, Fichant, Polard, Claverys, 2014). Consequently, the *lacZ* gene was cloned from BL21 Gold cells; therefore, if transformed in these cells, the bacterium's version of the protein will also be grown. The aim was to produce an inactive mutant of the protein with Tuner cells to ensure using a β-galactosidase-free strain so that there was no contamination from the host cell β-galactosidase.

Protein was expressed by producing soluble protein after inducing at 16°C using Tuner (DE3) cells, and the lysed cell pellet was resuspended in a buffer. Expression was conducted with 2 litre scale culture volumes of Tuner (DE3) cells containing pET-YSBLIC3C-β-Galactosidase. The pET-YSBLIC3C-β-Galactosidase strain showed protein expression at the correct size (116.3 kDa) and soluble.

SDS-PAGE was indicated that β-galactosidase expresses and is mainly found in a soluble state. However, it should be noted that in various cases, the target proteins were obtained in an insoluble form (Rosano and Ceccarelli, 2014). The SDS-PAGE shows that the protein was expressed solubly. The protein expressed soluble that was shown after cell lysis was sonicated

and then purified by Nickel column, which eluted the protein with imidazole buffer. It turned out that soluble protein was enough to be subjected to SEC.
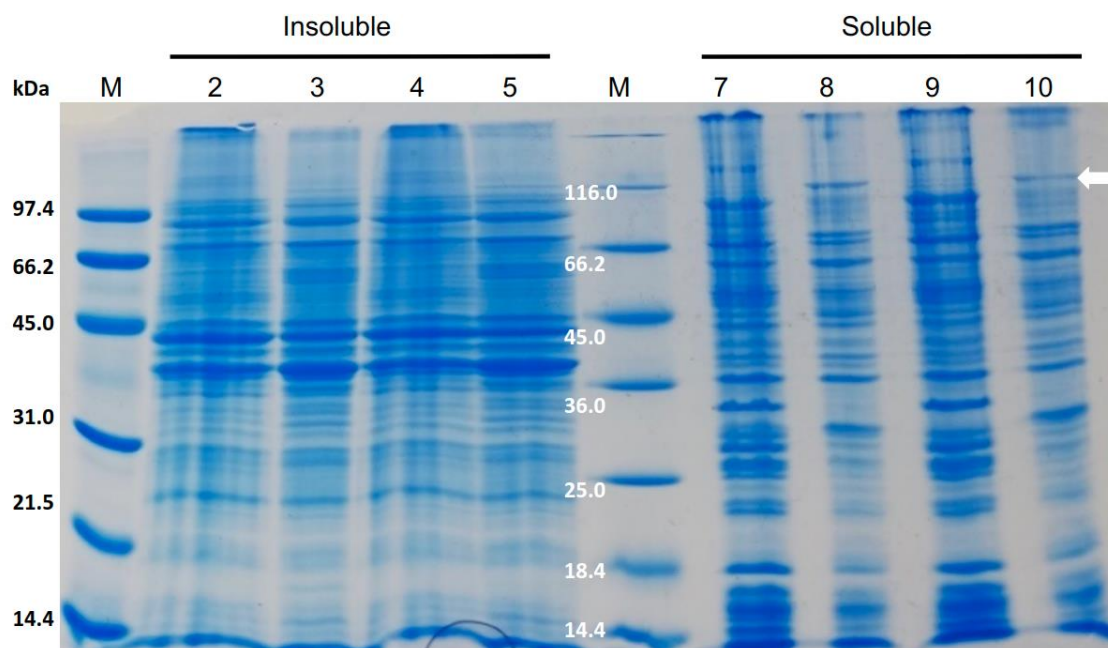
Figure 3.12 SDS-PAGE protein separation from *E. coli* Tuner-DE3 transformed with pET-YSBLIC3C plasmid. Lane 1 Bio-Rad Low Range SDS-PAGE standard molecular weight marker. Tuner-pET-YSBLIC3C-β-Galactosidase insoluble fractions 16°C Lanes 2-5, before induction Lane; 2 and 4 and after induction; Lane 3 and 5, lane 6 wide range protein molecular weight marker, unstained. Tuner- pET-YSBLIC3C-β-Galactosidase soluble fraction 16°C lanes 7-10, pre-induction lanes 7 and 9 and post-induction lanes 8 and 10. Lanes 8 and 10 (after induction) showed bands of the β-galactosidase molecular weight in 116.3 kDa, indicated by the white arrow.

Figure 3.12 showed that, while construct was expressed, Tuner (DE3) cells having pET-YSBLIC3C generated soluble protein following 0.5-1 mg/L IPTG induction. β-Galactosidase scaled up protein production. Protein acted as an active protein in the whole cell lysate and it was freely soluble on resuspension of the lysed cell pellet in a range of buffers.

3.2.1.3 Purification of LacZ β-Galactosidase

The fundamental requirement for a Cryo-EM project is a reproducible and steady supply of pure, stable, and fresh protein sample that behaves predictably when concentrated and placed on a Cryo-EM grid (Kampjut, Steiner, and Sazanov, 2021).

Therefore, β-Galactosidase was successfully purified using two purification steps, affinity column and SEC column.

Affinity chromatography column was utilised to yield the His6 tag β-galactosidase protein from the cell lysate after sonication (Figure 3.13). The His6 tag β-galactosidase protein was eluted as a major peak with an apparent molecular weight of around 116.3 kDa. The protein

was then further purified by size exclusion column (Figure 3.13) to give a final concentration

of ~ 8 mg/mL of cell culture.

Chromatogram of recombinant of β-Galactosidase
using HisTrap Nickel column
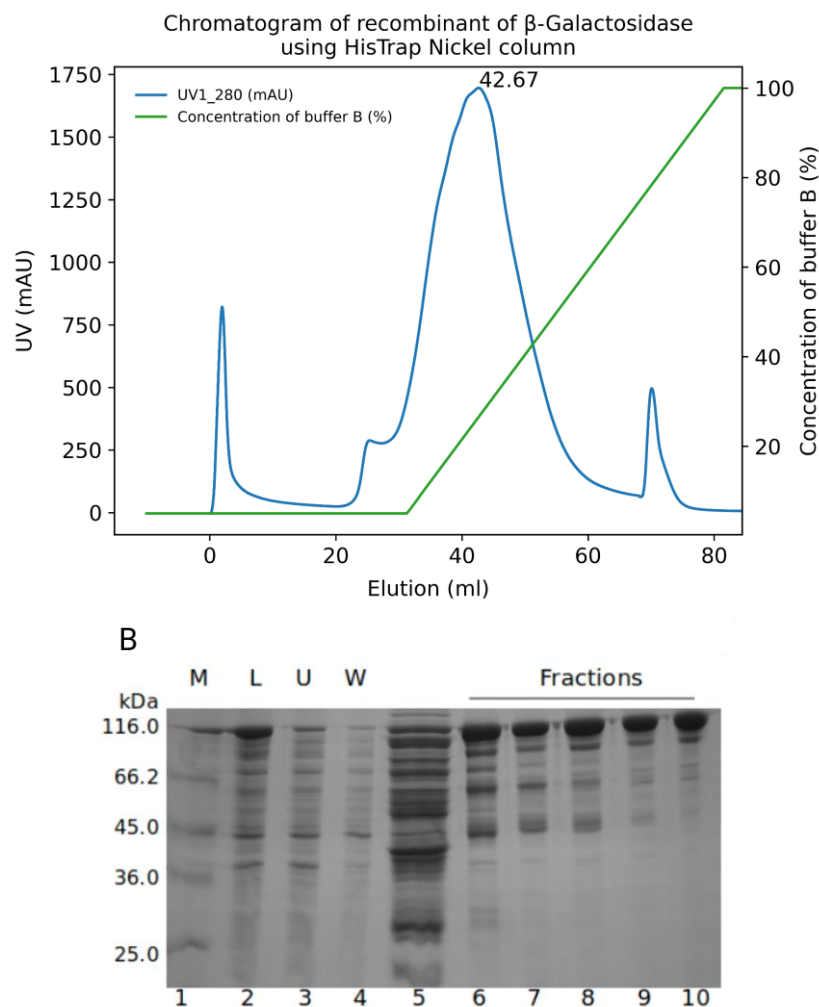
Figure 3.13 The first purification step of recombinant β-galactosidase.
(A) Chromatogram showing the elution of the β-galactosidase protein from a HisTrap Nickel column with a rising
concentration of buffer holding imidazole. Protein is determined by the UV absorbance at 280 nm (A280). (B)
SDS-PAGE separation of protein purified from *E. coli* Tuner-DE3 transformed with pET-YSBLIC3C-β-Galactosidase
using metal affinity-based chromatography. One of SDS-PAGE represents a part of fractions taken from the peak
between ~25-60 mL (lanes 5 -10) lane 5 fraction sample was excluded from the loading culture in SEC due to
massive contention; the rest of the beak fractions tested on a second of SDS-PAGE, not showing here. Lane 1 is
a low range molecular weight ladder Thermo Fisher), and lane 2 is a sample of the total cell lysate (protein that
was loaded onto the column), lane 3 is a sample of the unbound material that was removed buffer A(1-5mL),
and lane 4 is a sample of water was washed the peristaltic pump with 10 mL, β-Galactosidase has an expected
MW of 116.3 kDa. Lane 5 was not used for further purification due to contaminations.

The yield solution from the Nickel column was subjected to the SEC column. The SEC of His6

tag β-gGalactosidase provided two peaks (Figure 3.14, A). Nevertheless, fractions from the

largest peak were tested using SDS-PAGE; their bands revealed protein-like molecular weight

at the expected size of 116.3 kDa, with noticeable contaminants (Figure 3.14, B).
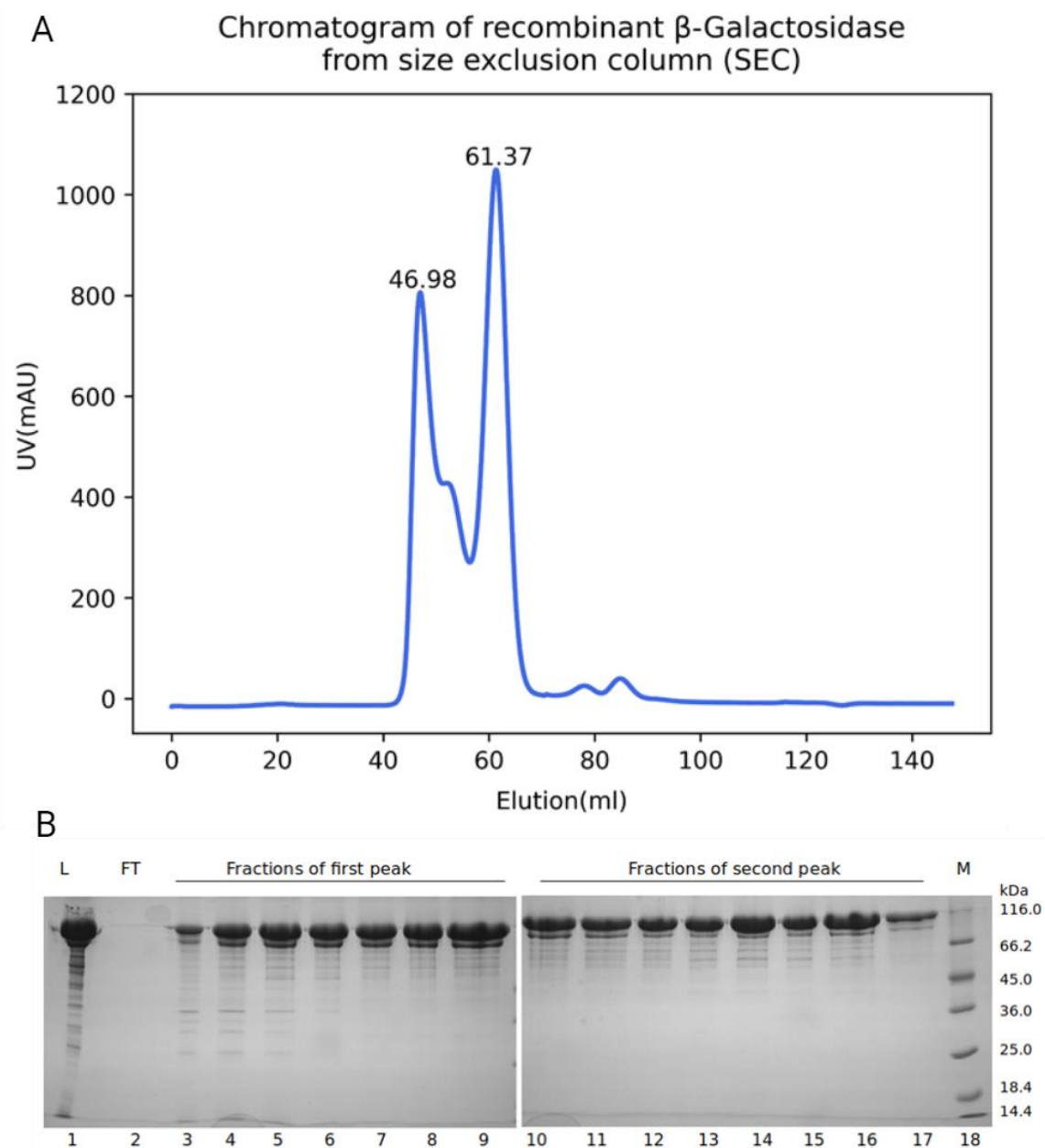
Figure 3.14 The second purification step of recombinant β-galactosidase.
(A) Chromatogram illustrates the elution recombinant of β-galactosidase protein from S200 size exclusion column. Protein is determined by the UV absorbance at 280 nm (A280). (B) SDS-PAGE separation of protein purified from *E. coli* Tuner-DE3 transformed with pET-YSBLIC3C-β-Galactosidase size exclusion chromatography. Two SDS-PAGEs of gel fractions were taken from two peaks between ~42-68 mL (lanes 3 -18). Lane 1 is a sample of the total protein that was loaded onto the column, and lane 2 is a sample of the flow-through. Lane 18 is a low range molecular weight ladder (Thermo Scientific); β-Galactosidase has an expected MW of 116.3 kDa.

3.2.1.4 Purification of commercial β-Galactosidase

The commercial sample was purified for the reasons explained in the previous section. The protein was eluted from the 120 mL HiLoad 16/600 Superdex 200 column after 50 mL (Figure 3.15, A). Protein peak fractions were analysed by SDS-PAGE to ensure protein purification (Figure 3.15, B).
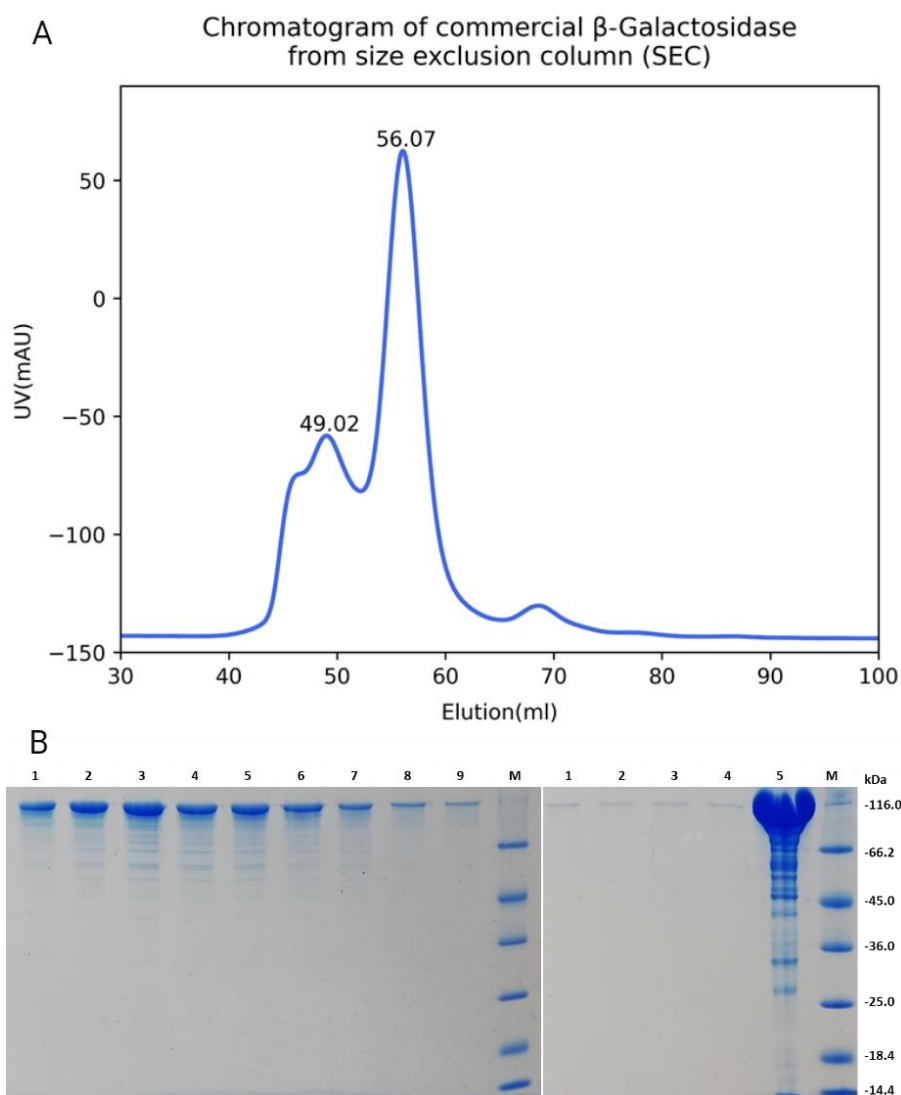
Figure 3.15 Purification of commercial β-galactosidase.
(A) Chromatogram showing the elution of the β-galactosidase monomer on HiLoad 16/600 Superdex 200 prep grade size exclusion column. Protein determined by the UV absorbance at 280 nm (A 280). (B) Two of 12% SDS-PAGE of size exclusion column fractions taken from the peak between ~50–70 mL (lanes 1-9, on the left gel, and lanes 1-4, on the right gel). Lane 10, on the left gel, and lanes 6, on the right gel) are Low range molecular weight ladder (Thermo Scientific), and lane 5, on the right gel, is a sample of the total protein that was loaded onto the column. β-Galactosidase has an expected MW of 116 kDa.

Protein sample fractions after size exclusion chromatography are rather diluted and need to be concentrated several folds before they can be used for grid preparation (Kampjut et al. 2021). In particular, after size-exclusion chromatography of β-galactosidase, the yield was not high; it had to be concentrated. After screening grids, the concentration of 1.8 mg/mL was suitable for grid preparation. All grids preparation of apo β-galactosidase was carried out from freshly size-exclusion chromatography specimens.

3.2.1.5 Preparing complex-covalent inhibitors and activity-based probes

The fluorescent β-galactose aziridine ABP (TB652) probe and the epoxide (LWA480) inhibitor have not been previously identified as readily binding to β-galactosidase in solution. The novelty of this study lies in the fact that covalent binding of β-galactosidase occurred even after 15 minutes of incubation.

The activity of the recombinant β-galactosidase was evaluated. Subsequently, the relationship between the ligand and the enzyme was investigated by Cryo-EM. The enzyme activity was studied, and irreversible inhibition by the activity-based probe was demonstrated.

Breen et al. 2017 demonstrated a similar approach using the configurational isomer cyclophellitol aziridine TB652 with the lysosomal enzyme acid β-glucocerebrosidase (GBA1). In 2022, Armstrong et al. solved the structure of human fucosidase, which has bound covalently with 'aziridine' ABP using single particle analysis (PDB code: 7PM4) (Armstrong et al. 2022). In 2023, another member of the same group has solved a complex crystal structure of Cellvibrio japonicus enzyme CjGH35 with LWA480, ligand ID YGX (PDB code: 8PEJ) (Kuo et al. 2023).

3.2.1.5.1 Determining enzyme activity and the efficiency of irreversible inhibition kinetics assays

3.2.1.5.1.1 4-MU-β-D-galactopyranoside

The fluorescent β-galactose aziridine TB652 probe was observed to undoubtedly bind to β-galactosidase in solution, with significant labelling taking place even after the 55-minute time point. Meanwhile, the galacto-configured epoxide LWA480 lost its inhibition potency and did not bind β-galactosidase after 55 minutes of incubation, as shownin Figure 3.16, D.
As a result, it has concluded for assay with 4-MU-β-D-galactopyranoside that 30 μM of inhibitors will inhibit β-galactosidase

A. Serial dilution of product -
MU Calibration curve

B. Serial dilution of enzyme
β-galactosidase

C. Kinetic of two compounds in different times

D. Zoom in on figure C (circled)
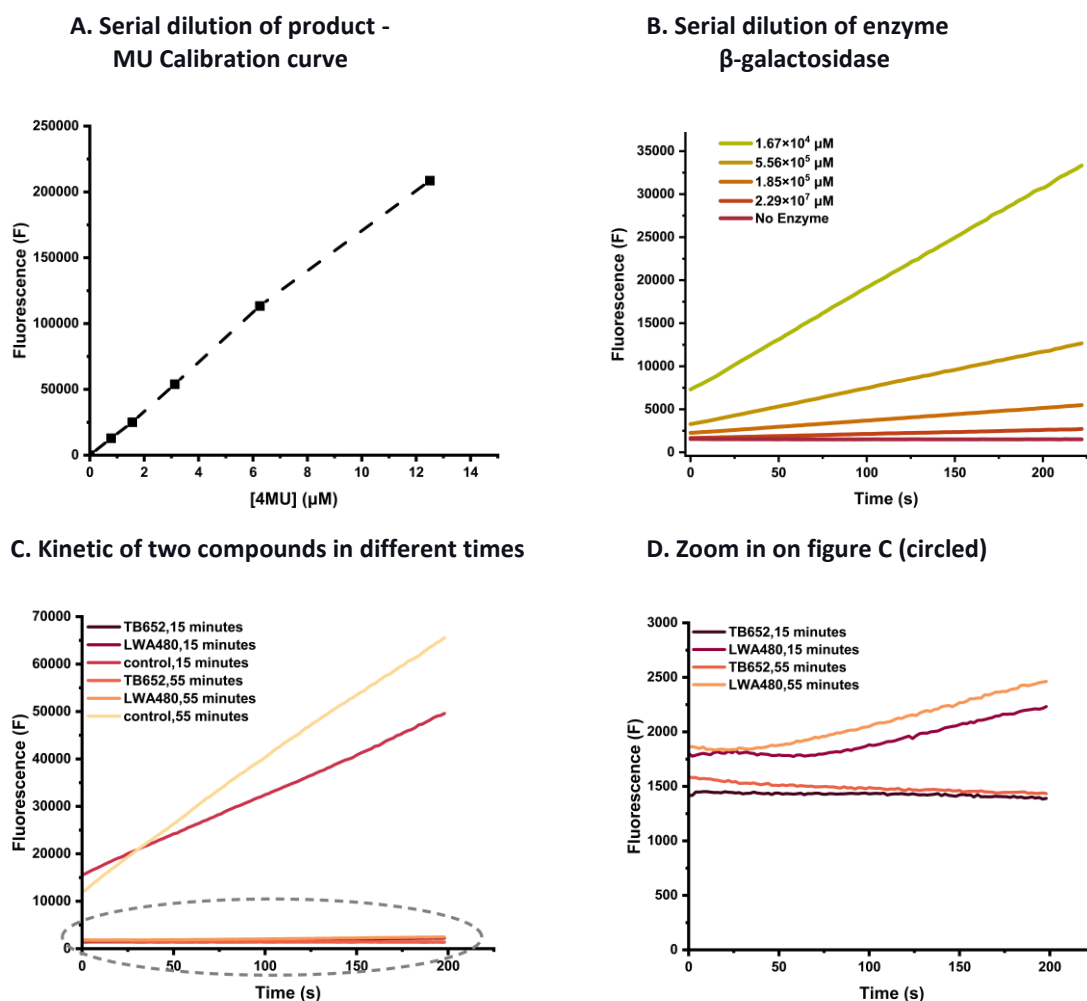
Figure 3.16 Kinetics assays plots of activity and inhibition using 4MU.
(A) Fluorescence (F) intensity vs [4MU] measured in 80 μL total volume of pH 7.5 HEPES buffer in a black 384-well plate using a range of (390-450) nm excitation wavelengths. A dashed black line linear best fit for F vs [4MU] from 0.78 to 12.5 μM at 360 nm excitation. (B) 4MU fluorescence intensity measured over 200 minutes with a 15-second cycle time using an excitation wavelength of 360 nm at 37°C for serial dilution ranging from $1.67{\times}10^4$ to $2.29{\times}10^7$ μM in 40 μL of pH 7.5 HEPES buffer in a black 384-well plate. C and D Plots of fluorescence vs time in the presence of two different times of two inhibitors of the aziridine (TB652) and epoxide (LWA480). (C) A large scale plot of the kinetic behaviour for the control enzyme and the complexes shows the kinetic inhibition of the complexes. (D) Zoom in of the circle dots large scale C plot that revealed complete inhibition of the enzyme with the aziridine probe and slight demolition of enzyme inhibition with the epoxide probe during the time.

3.2.1.5.1.2 4*p*NP-β-D-galactopyranoside

A standard curve was prepared by plotting absorbance values of known concentrations of 4-Nitrophenol [4*p*NP] ranging from 50 μM to 500 μM using a 40 μL volume. This standard curve was used to convert absorbance values to concentrations of 4-Nitrophenol [4*p*NP], Figure 3.17, A.

Plots in Figures 3.17, B, C, and D were generated to determine the product concentration produced by pNP from various catalytic enzyme concentrations, providing the initial reaction rates.

B panel in Figure 3.17 represents a serial dilution of β-galactosidase enzyme from the stock solution (10-fold in 1mL to give 1μg/mL (8.60 μM) 100-fold gives) are to be prepared. 1ug/mL was used to give a final concentration of 0.5 μg/mL (4.30 μM), Figure 3.17, B.

4-Nitrophenyl β-D-galactopyranoside (4pNP-β-D-galactopyranoside) was diluted in pH 7.0 buffer to final concentrations ranging from 0.01 μM to 10 μM and added 1μg/mL (8.60 μM) of the enzyme to give a final concentration of 0.5 μg/mL (4.30 μM) in each reaction (Figure 3.17, C).

Figures 3.17, B, C and D are plotted to determine the product concentration generated with pNP, generated from catalytic multi-enzyme concentrations as a function of pNP. These plots provide the initial reaction rate.

Then, multiple reaction rates at different substrate concentrations were measured to obtain a K_m value of 62 μM Figure 3.17, E. Michaelis-Menten constant K_m. It defines the attraction of an enzyme to a substrate. K_m represents the substrate concentration when the reaction velocity reaches half its maximum (Button, 1998). The test was run triplicate in order to improve the accuracy of the results.
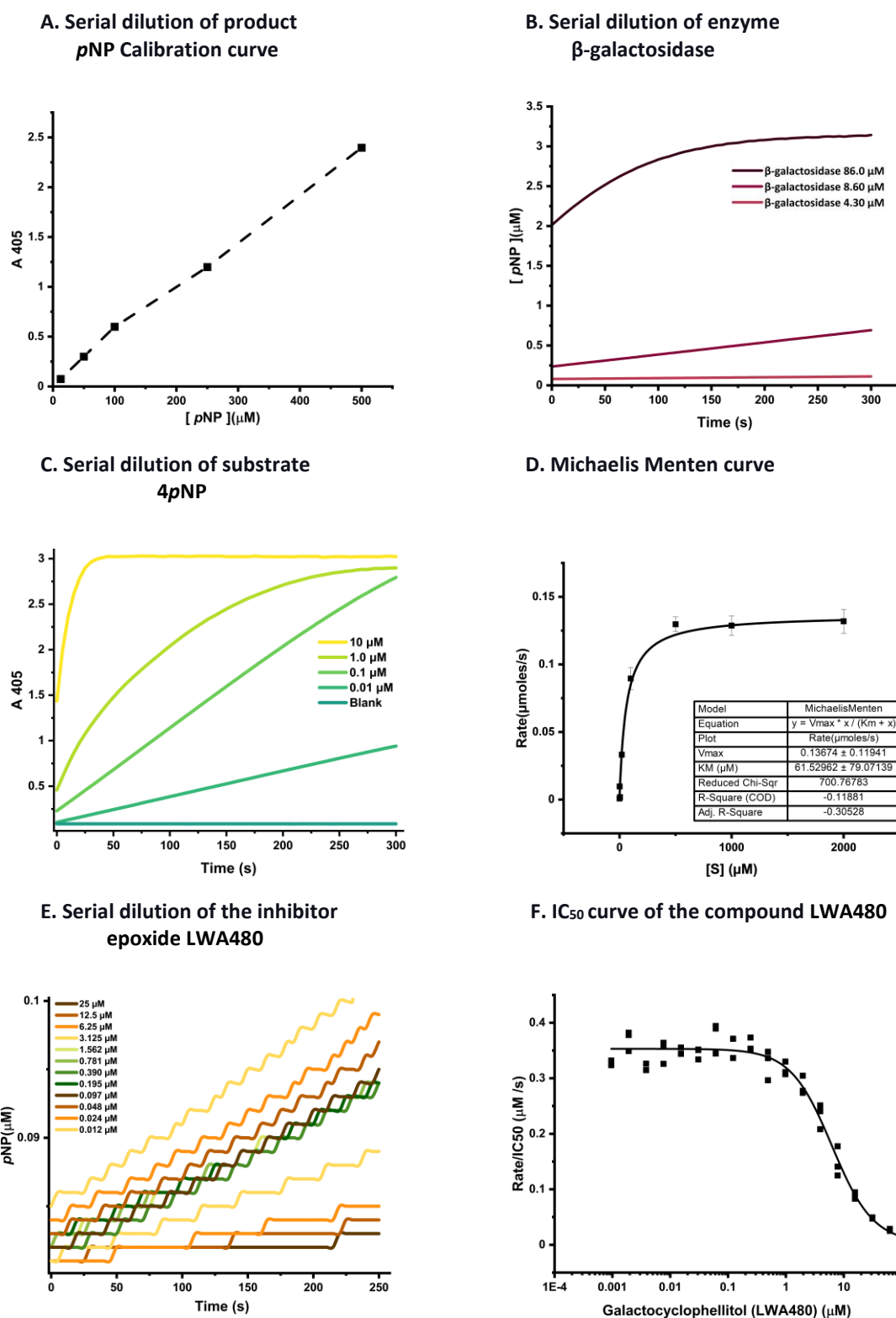
Figure 3.17 Catalytic enzyme and Inhibition kinetics using *p*NP.
(A) 4-Nitrophenol [4*p*NP] standard curve using colourimetric absorbance intensity vs [*p*NP] was measured in 40 µL total pH 7.5 HEPES buffer volume in a transparent background 384-well plate using a 405 nm excitation wavelength. A a dashed black line reveales a linear best fit for F vs [*p*NP] from 0.05 to 0.500 µM at 405 nm absorbance. (B) Serial dilution of the enzyme shows the *p*NP concentration vs time 10 µg/mL (86 µM) is the stock enzyme solution when diluted 10-fold in 1mL to give 1µg/mL (8.60 µM) 100-fold gives 0.5 µg/mL (4.30 µM). (C) Serial substrate titration from 10 to 10,000 nM (0.01 µM to 10 µM) 4-Nitrophenyl-β-D-galactopyranoside [4*p*NP] with the enzyme. (D) Michaelis Menten curve showing K_m value. Eight serial titrations of the substrate from 2000

to 0.1 µM. (E) Colourimetric plots of absorbance and pNP vs time for β-galactosidase in various concentrations of the inhibitor epoxide LWA480. (F) IC$_{50}$ curves of the LWA480 compound. The compound is slightly potent as the IC$_{50}$ values are in the low micromolar range.

The coefficients from both standard curves, Figure 3.16 A and 3.17 A, were used to calculate the initial reaction velocities by converting absorbance to concentrations. Experiments to gain the concentration of LWA480 that we need for the Cryo-EM sample, as was shown, result in enzyme inhibition, Figure 3.17, F. We determined its IC$_{50}$ value, which is half the maximal inhibitory concentration. Serial dilution of the substrate was determined. The IC$_{50}$ plot shows how much inhibitor is needed to inhibit 50% of the enzyme. Inhibitor dilutions from 25 to 0.012 µM are shown in Figure 3.17, G. At a concentration of 31 µM of the substrate 4pNP, which is below the K$_m$ because the binding was so tight the 62 µM value has adjusted inIC$_{50}$ was determined to be 6 µM, Figure 3.17, H. Attempts, the IC$_{50}$ against (LWA480 was run duplicate due to low quantity of the compound. IC$_{50}$ was not run for TB652 due to a lack of availability.

The IC50 value was reported to be 38 nM for the epoxide compound galactocerebrosidase (GALC), which was a very potent inhibitor (Marques et al. 2017). However, the non-tagged epoxide inhibitor activity was less potent than the inhibition with human enzymes.
The use of EM experimentation was a little concerning based on two factors: manufacturing time and the decreasing potency of the inhibitor after 30 minutes of incubation. The inhibitor was prepared six years ago, which might contribute to decreasing potency due to loss of activity from long-term storage. Furthermore, based on varying concentrations, the inhibitor was unable to bind irreversibly with the active site of the enzyme. Although using the Aziridine probe stabilised the irreversible inhibition and resulted in potent inhibition, very strong inhibition formed aggregation of proteins, as shown in section 3.3 figure 3.17. Moreover, this probe was dissolved in DMSO with a low freezing point. A low freezing point causes water to be crystalline, and denaturing of proteins occurs after a long incubation period. Nevertheless, two compounds (TB652) and (LWA480) showed inhibition, and TB652 was highly potent against β-galactosidase while (LWA480) showed potency when a fresh batch was used and so could be taken forward to the EM study.

A complex was generated of β-galactosidase with the inhibitor of β-galactose aziridine configured, N-octylnorbornene-Cy5-galactocyclophellitol aziridine (TB652). TB652 contained

fluorescent configured sulfonated cyanine 5 (Cy5) fluorophore and warhead aziridine. β-Galactosidase at a concentration of 2 mg/mL (equal to 17 µM) was added to 1 mM of ligand in the buffer of HEPES and 1 mM NaCl, pH 7.0, 10% DMSO, mechanism of the labelling shown in Figure 3.7, bottom panel. Another complex was obtained with the same protein concentration, and β-galactose epoxide was configured with Galactocyclophellitol (LWA480) with epoxide warhead (Willems et al. 2014)(Figure 2.4, below panel) The ligands were confirmed to effectively bound to β-galactosidase in solution with a brief incubation; thus, they were ideal targets for Cryo-EM incubation Proposed mechanism of β-galactosidase inhibition shown in Figure 3.18

Mechanism of based inhibition
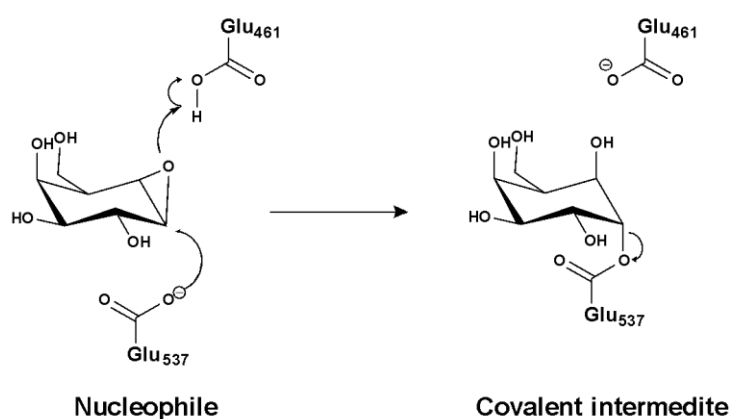
Covalent inhibitor
Mechanism of based inhibition
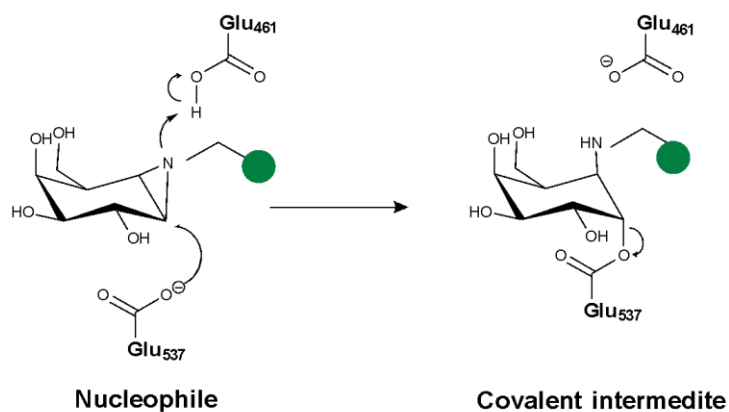
Activity based probe

Figure 3.18 Proposed mechanism of β-galactosidase inhibition.
Proposed mechanism of β-galactosidase inhibiting by epoxide compound (above). Mechanism of β-galactosidase labelling by the florescent aziridine compound (underneath).

3.2.2 Cryo-EM study of β-galactosidase

3.2.2.1 β-Galactosidase data acquisition

Four datasets were collected successfully in the electron-event representation (EER) format in a single 24 and 72 h session from a plunge-frozen two specimens of apo recombinant β-galactosidase in different concentrations, third specimen of apo commercial β-galactosidase and last specimen was the recombinant protein that incubated with ABP , using a Thermo Scientific Glacios operated at 200 keV microscope with a Falcon IV direct electron detector with optimal magnification pixel size, exposure time and dose rate.

The desired dose rate was achieved by altering the parameters of data acquisition and defining a template (Thompson et al. 2019). Data collection parameters of the datasets are given in the following in Table 3.2. A Cryo-EM data collection sample in Figure 3.19.

Table 3.2 Data collection and refinement parameters of the β-galactosidase datasets.

Sample	Recombinant apo β-gal	Recombinant apo β-gal	Commercial apo β-gal	Complex of recombinant β-gal
Sample concentration	2 mg/ml	3 mg/ml	1.8 mg/ml	2 mg/ml
Ligand concentration				0.33 mM
Defocus range(μm)	-2, -1,0.5, -1	-2	-2, -1.75, -1.5, -1.75, -1.0	-2, -1.75, -1.5, -1.75, -1.0
Fluence (e-/pix/s)	0.505	0.515	0.46	0.515
EER fractionation	24	23	18	18
Exposure time (s)	1, 2	3.73	3.88	5.12
Nominal magnification	240/2	240/2	240/2	240/2
Calibrated pixel size (Å)	0.574	0.574	0.574	0.574
Dose (e$^-$/Å^2)	50	49.98	50	50
No. of micrographs	1416	2094	1832	5931
No. of particles picked	136,013	124,686	146,239	736,882
No. of particles in final refinement	49,895	NA	73,037	204,567
Overall B-factor	- 136.4	NA	-37	- 74.9
Resolution (0.143; Å)	2.5	5	3.4	2.7

Atlas of the grid: Squares selected according to the colours. Each colour indicated the ice thickness. Blue and purple are thin ice, and orange is medium ice. At the same time, red is damaged.

Selected square: The holes measured, selected and acquired.

Final Image: It shows a micrograph containing particle distribution.

Selected hole: The green spot is the location of data collection, and the blue spot is autofocus.

Figure 3.19 Overview of Cryo-EM data collection and setting up data collection in EPU.

3.2.2.2 Image processing model building and refinement

All data were collected and processed using computer programs EPU and *RELION* 3.1.2. The norm to counterbalance the fogginess caused by the impact of stage shift and beam-induced movement is *MotionCor*2 (Brilot et al. 2012; X. Li et al. 2013) which filters the beam and the alignment of the movies (Zheng et al. 2017).

3.2.2.2.1 Apo recombinant β-Galactosidase

The first data of apo recombinant β-galactosidase was collected for the recombinant protein at a concentration of 3mg/ml. This data has generated 2D classes of similar top-view. The preferred orientations of these particles lead to limited resolution of 3D refinement maps at about 5 Å.

A.   Thin ice grid micrograph square both of 3 mg/mlLconcentration of protein

B.   2D classes specimen of 3 mg/mL concentration of recombinant β-galactosidase showing preferred orientation
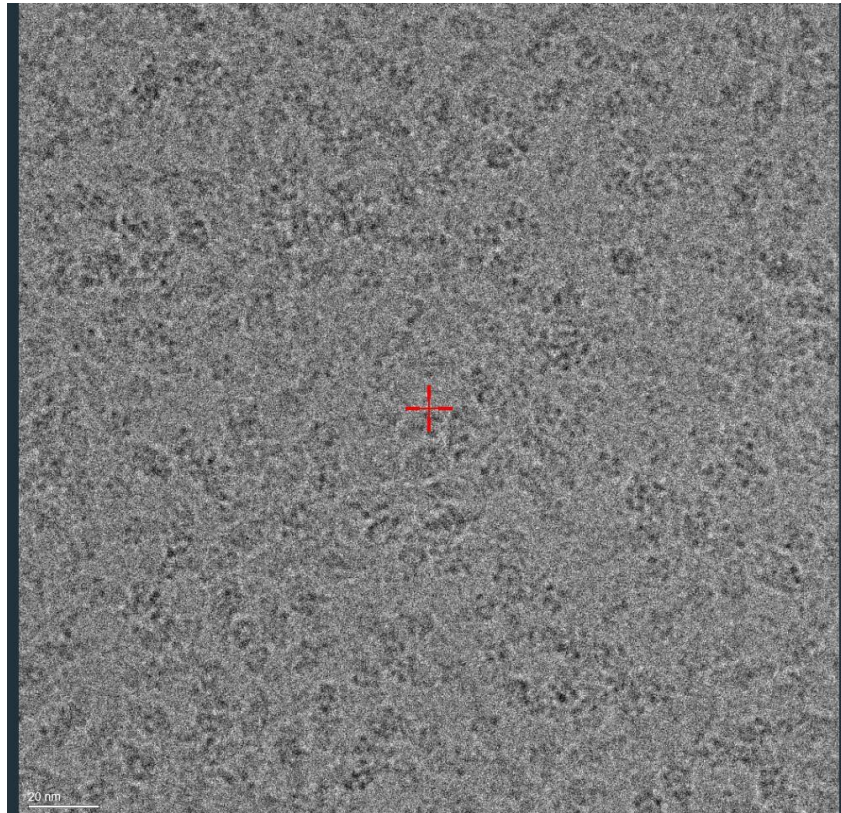
C. Thin ice grid micrograph square both of 3 mg/mL concentration of protein
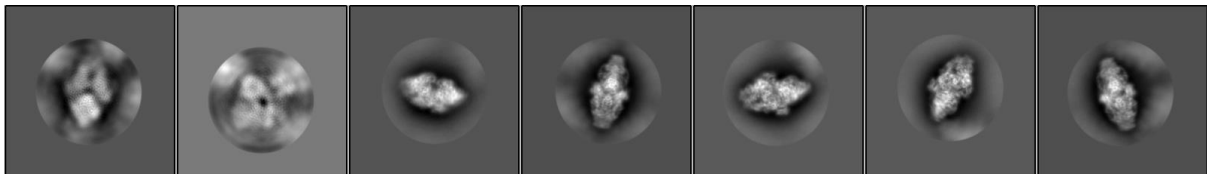
D. 2 mg/mL of recombinant β-galactosidase showing different views

Figure 3.20 Comparison of effecting the protein sample concentration on 2D classes.
(A) Thin ice grid micrograph square 3 mg/ml concentration of protein (B) The specimen of 3 mg/ml showed 2D free class average preferred orientation which contained about 5000 extracted particles. (C) Thin ice grid micrograph square2 mg/ml concentration of protein. (D)different views of 2D free class average of specimen at 2 mg/ml concentration.

The protein particles in the 3mg/ml concentration specimen were crowded in micrographs, as shown in Figure 3.20, A. However, the 2D classes lacked different views to build a complete 3D tetramer view in Figure 3.20, B. 2D classes suffered preferred orientations that led to low resolution. The processing was stopped when the 3D refinement was obtained at resolution 5 Å, in Figure 3.21, A.

These preferred orientations might be caused by the intensity of the particles, plot force, and air water interface (Glaeser 2018; Noble et al. 2018).
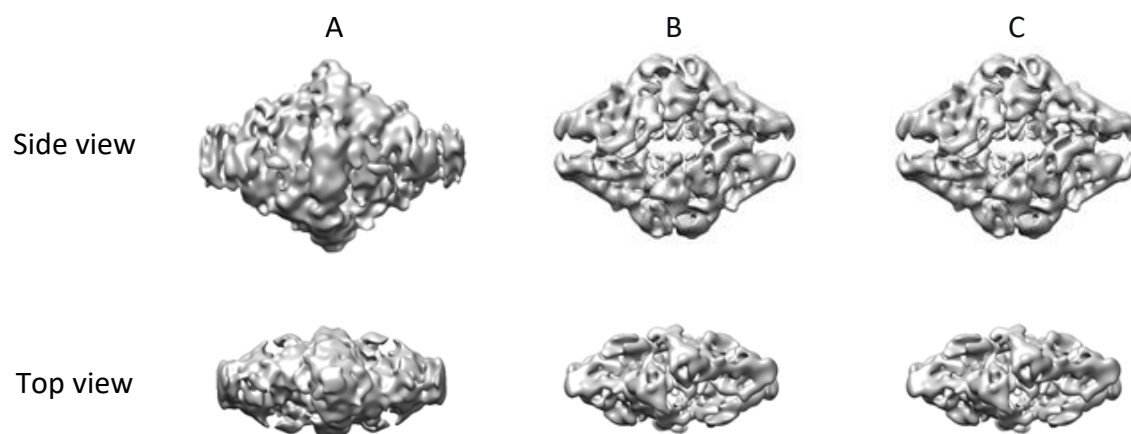
Figure 3.21 Comparison between 3D potential maps of apo proteins.

(A) Protein produced in house 3mg/ml. (B) Protein produced in house 2 mg/ml. (C) Commercial β-galactosidase. All above are side views while the bottom is top views of the maps.

A second new apoprotein dataset was collected by optimising the specimen preparation parameters to avoid the preferential orientation issue. For that, protein concentration was lowered from 3 mg/ml to 2 mg/ml and increased the plot force -5 instead of -10. This dataset showed multiple 2D class views in the second run. One of the best 3D classifications was used for 3D refinement. The low-pass filter of the 3D classification reference in the post-processing job was performed in low resolution to sort out particles and identify a homogenous group of particles in Figure 3.20, C.

The application of a low-pass filter is a useful method for improving particle contrast during the evaluation of particle picks. A low-pass filter only passes the frequencies lower than the defined cut off and excludes all other high-resolution frequencies. For example, only signals below 20 Å will pass through a 20 Å filter (Tegunov and Cramer, 2019).

The 3D refinement and post process were run several times between each job of CTF refinement per particles, estimate beam tilt, magnification and defocus. In addition, a final run of 3D refinement was performed after the Bayesian polishing job, which has increased the resolution from 2.6 Å to 2.5 Å, as in Figure 3.22.
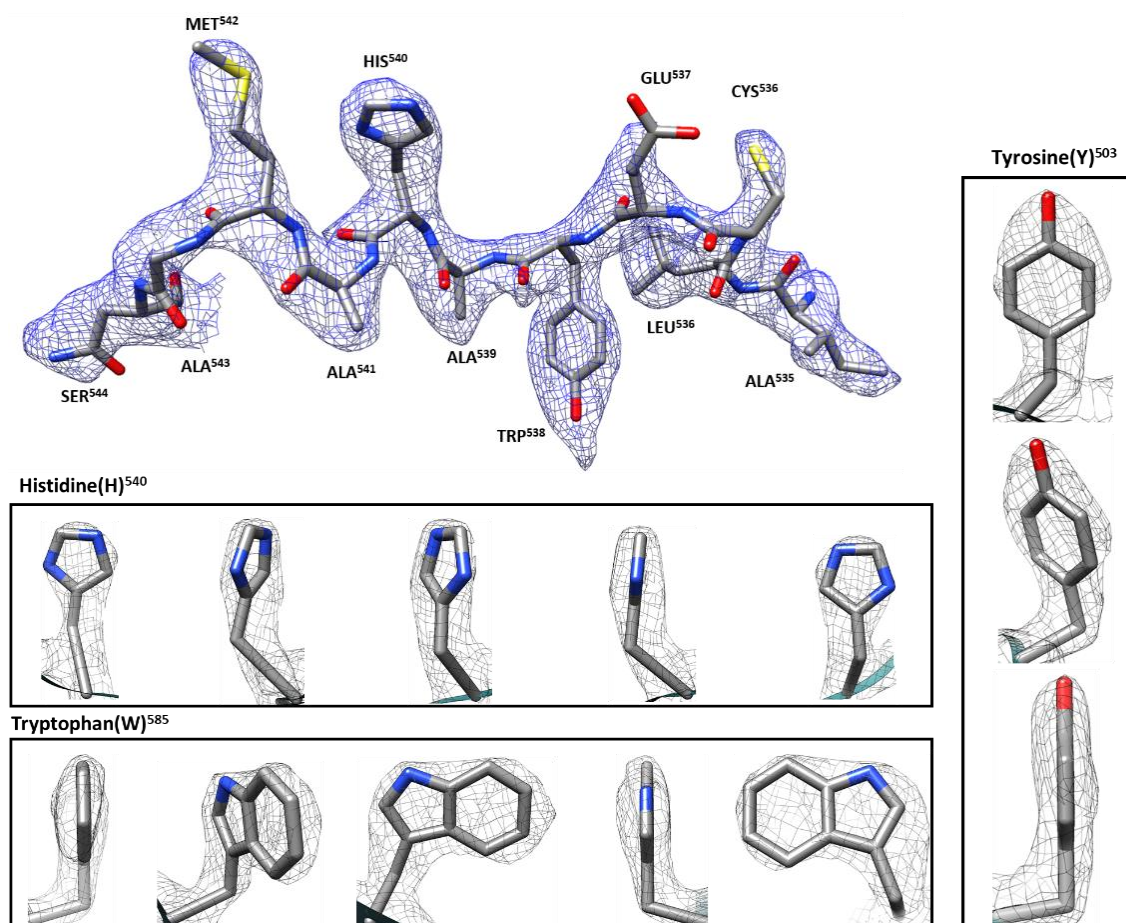
Figure 3.22 Cryo-EM 3D maps of the recombinant apo protein final at 2.5 Å.

3.2.2.2.2 Commercial apo β-Galactosidase

This EM structure was solved before we managed to express the protein in house, following the (Bartesaghi et al. 2015) protocol, which used the Tris buffer to elute the protein from the SEC.

It has been recommended that we avoid using a buffer that contains Tris HCL because the Tris molecule would occupy the active site, which is one reason we were not interested in imaging that protein with inhibitors in that buffer.

The interpretation of results in enzyme studies needs to be critically monitored, especially when using the Tris buffer. This is because Tris buffer acts as a competitive inhibitor owing to its strong binding abilities with the active site of the enzyme. This property of Tris buffer has also been suggested in many inhibition studies (Ghalanbor et al. 2008).

The majority of the particles did not suffer from preferred orientations and were randomly distributed. After data processing, both 2D classes and 3D reconstruction showed no problems with preferential views (Figure 3.23 A, B).
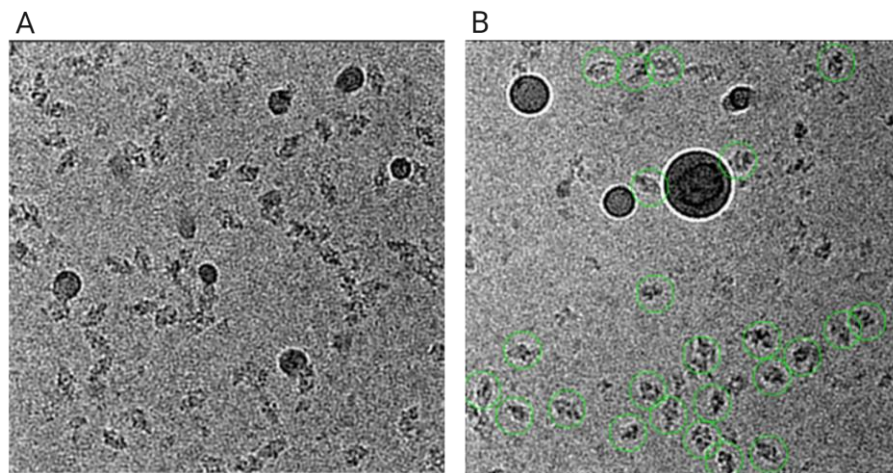
Figure 3.23 Micrographs of the commercial apo β-galactosidase data set.
A) A micrograph shows particle orientation distributions in thin ice. B) A representative micrograph which shows green circles representing manually picked particles.

2D class averages in (Figure 3.24, A) are a very good indication of how good the 3D map will become, and because we had few particles per some micrographs, we regrouped particles from multiple micrographs together.
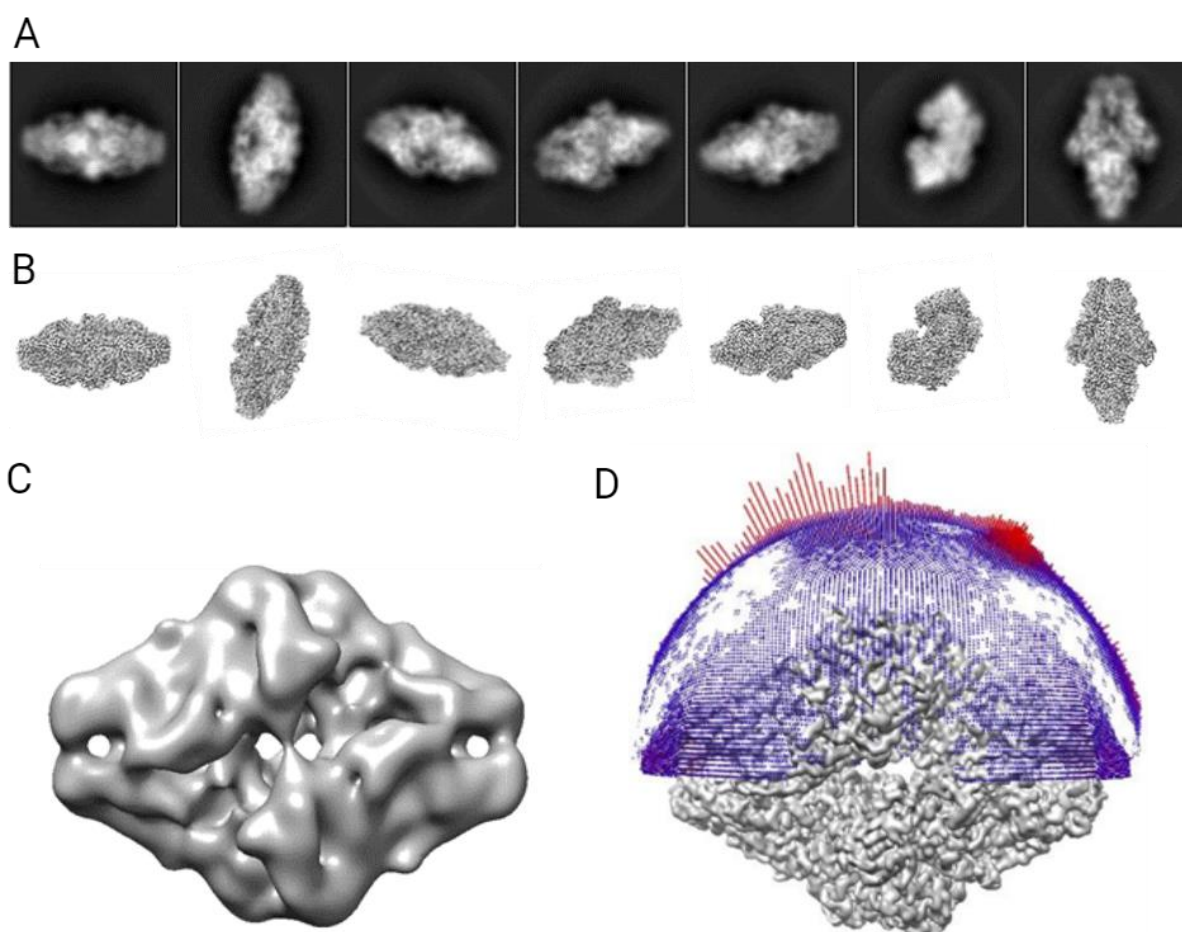
Figure 3.24 Analysis of the initial Cryo-EM data set of the apo β-galactosidase.
(A) 2D classes which contain different side views and show the correct shape of β-galactosidase (ordered from larger to smaller class). (B) Corresponding orientations of the 3D map to the 2D classes side view particles. (C) 3D initial model at 15 Å. (D) Angular distribution plot for the 3D refinement map at 3.4 Å.

Post-processing with mask yielded a 3.4 Å resolution after CTF refinement, and the local resolution estimations showed the core of the apo β-galactosidase was at a higher resolution (3.4 Å) as compared to the global average.

In theory, cryo-grids needed further optimisation to collect a data set which carries many particles. In this study, the final map resolution of the commercial achieved 3.4 Å with around 30,000 particles and 1832 movies. Local resolution and FSE resolution estimation plots were shown in Figures 3.25, respectively, compared with other studies that solve the final structure in higher resolution with a larger number of particles and fewer micrographs.

In most cases, 2 Å reached with around 40K particles of the final map (Bartesaghi et al. 2015), while around 3 Å was reached with 10 K particles (Bartesaghi et al. 2014). Three high

resolution complex structures were obtained from around 40 K to 70 K particles and around 500 micrographs (Saur et al. 2020). Bartesaghi et al. 2015 reported a complex structure solved at 2.2 Å from around 40 K particles and around 1487 micrographs (Bartesaghi et al. 2015).

Previous studies included fewer particles; about 11,726 out of 509 micrographs was generated at about 3.2 Å of the reconstructed map (Bartesaghi et al. 2014).

The motion-corrected micrographs have particularly good signal-to-noise ratios (SNRs). At the same time, the resolution projection for each particle is hard to evaluate because of the noise increase more than the signal; it has low (SNRs)(Sigworth 2016). Our data has many micrographs with few particles in each that could affect the contrast and lead to a lower resolution, for which collecting data with a more concentrated sample might improve the resolution. The quantity of picked particles with moderate distributions would increase the views of the 2D class average. Subsequently, it improves the angular orientation to generate better 3D reconstruction (Naydenova and Russo 2017; Nguyen et al. 2021). Cryo-EM structures data are shown in Table 3.3.

Table 3.3 The numerous Cryo-EM 3D structures of *E. coli* β-galactosidase have successfully been deposited into the RCSB PDB.

PDB	Protein concentration	Ligand	Å	Micrographs	Picked Particles	Final structure particles	TEM	Electron detector	Software	References
3J7H	2.3 mg/mL	Mg	3.2 Å	509	5,863	11,726	FEI Titan Krios	Gatan K2 (300 keV)	EMAN2and Frealign	(Bartesaghi et al. 2014)
4CKD	-	scFv13R4 and Mg	~5.4 Å	49	2,965	-	FEI Polara	FEI Falcon II (300 keV)	Frealign and MRC	(Vinothkumar, McMullan, and Henderson 2014)
5A1A	2.3 mg/mL	PETG	2.2 Å	1,487	93,686	41,123	FEI Titan Krios	Gatan K2 (300 keV)	EMAN 2and Frealign	(Bartesaghi et al. 2015)
6CVM	-	Mg, Na and PTQ	1.90 Å	-	298,715	-	FEI Titan Krios	Gatan K2 (300 keV)	Frealign 9.11	(Bartesaghi et al. 2018)
6DRV	-	-	2.20 Å	-	-	-	FEI Titan Krios	Gatan K2 Summit (300 keV)	Software platform	(Cianfrocco et al. 2018)
6X1Q	4.5 mg/mL	Mg and Na	1.8 Å	49,49	998,945	257,202	JEOL Cryo Arm	Gatan K3 (200 keV)	*RELION* 3.1and SerialEM	(Merk et al. 2020)
6TTE	0.17 mg/mL	PETG, PTQ and Mg	2.2 Å	562	136,013	49,895	FEI Titan Krios	FEI Falcon III (300 keV)	*RELION* 3.04	(Saur et al. 2020)
6TSH	0.17 mg/mL	DGN	2.3 Å	598	146,239	73,037	FEI Titan Krios	FEI Falcon III (300 keV)	*RELION* 3.04	(Saur et al. 2020)
6TSK	0.17 mg/mL	L-ribose	2.3 Å	517	124,686	40,086	FEI Titan Krios	FEI Falcon III (300 keV)	*RELION* 3.04	(Saur et al. 2020)

Resolution (Å), Fv antibody domains (scFv13R4), phenylethyl β-D-thiogalactopyranoside (PETG) Deoxygalacto-nojirimycin (2R,3S,4R,5S) -2- (hydroxymethyl) piperidine-3,4,5-triol (DGJ), beta-L-ribopyranose (L-ribose), and Acceleration Voltage (keV).

Through the employment of cryo-EM, four structures of apo β-galactosidase and five liganded β-galactosidase structures have been solved (Table 3.3). The evolved software to obtain structures in the early days was EMAN, then Frealign and now *RELION* (Merk et al. 2020) and (Saur et al. 2020). The Subramanian group has enhanced the resolution of 3D structures of β-galactosidase in the PDB (PDB code: 6CVM) to 1.9 Å (Bartesaghi et al. 2018). Even though they were using the same Titan Krios transmission electron microscope (FEI Company, Hillsboro, OR) equipped with the same Gatan K2 detector as in previous studies (Bartesaghi et al. 2014, 2015), the improvement came from the 3D reconstruction software, Frealign, which was rewritten in 2016 (Grigorieff 2016). Frealign's imaging strategy that combines local defocus, local drift, and dose weighting leads to an increase in the number of selected particles and, therefore, to an improved final resolution.

The highest resolution of all cryo-EM structures of β-galactosidase in the PDB (PDB code: 6X1Q) is 1.8 Å (Merk et al. 2020). Interestingly, the researchers in this study used an electron microscope operated at a lower acceleration voltage, the 200 keV CRYO ARM (JEOL Inc. Tokyo, Japan) than electron microscopes that had been used in all previous structures (Bartesaghi et al. 2014; Vinothkumar, McMullan, and Henderson 2014; Bartesaghi et al. 2015, 2018; Cianfrocco et al. 2018). Consequently, the 200 keV acceleration voltage was chosen for the present study since it would allow us to see the probes bound to β-galactosidase. It should also be noted that the number of particles used by Merk and co workers to refine the final structure is much higher than in most of the other cases. This can be explained by many advances such as specimen quality, protein concentration and purity, detector technology, electron source, and the use of the open source *RELION* 3.04 software enables auto picking of particles and aberration correction.

Apo β-galactosidase at 2.5 Å   Apo β-galactosidase at 3.4 Å
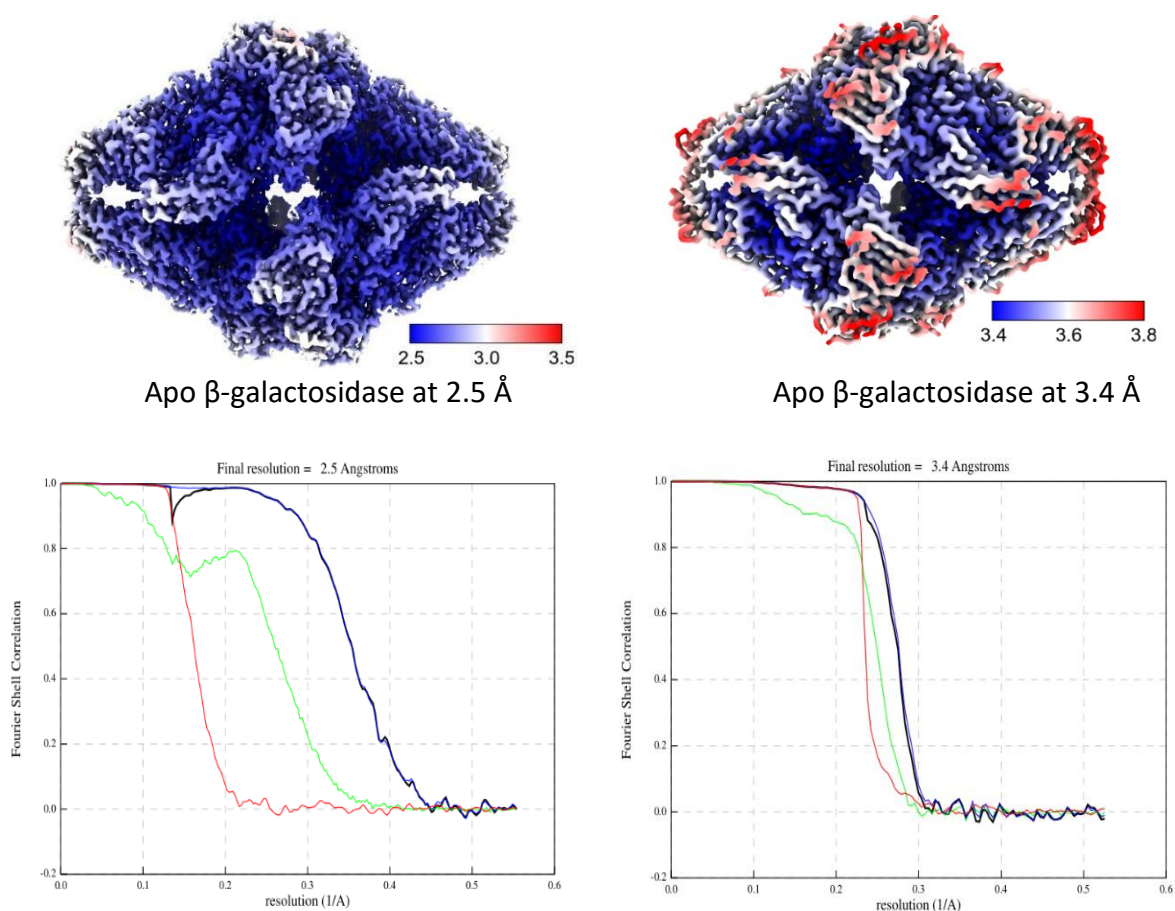
Figure 3.25 Comparison between in-house and commercial protein resolution of this study.
Local resolution maps of apoprotein lab production (left) and the commercial (right)l, respectively, showing the core of β-galactosidase at 2.5 and ~3.4 Å, and Local resolution maps coloured to the different scales.
Fourier Shell Correlation (FSC) curves demonstrate an estimated resolution of 2.5 (below left) and 3.4 Å (below right). The final resolution of the data sets was generated using the 0.143 gold standard FSC between two independent half maps. In black: Fourier Shell Correlation Corrected; in green: Fourier shell correlation unmasked Maps; in blue: Fourier shell correlation masked maps; in red: Corrected Fourier shell correlation phase randomised masked maps.

3.2.2.2.3 Enzyme-ligand complex

Applying single particle analysis of Cryo-EM, the recombinant β-galactosidase, bound to aziridine moiety, was not determined because when the grids were screened, the majority of the particles were aggregated Figure 3.26. In addition, the compound was not available in a sufficient amount to repeat the experiment.

The reason might be because the incubation time was longer than 30 min, and this inhibitor has tightly covalent bonds to the protein and needs further purification after making the complex solution, or the DMSO could denature the protein.
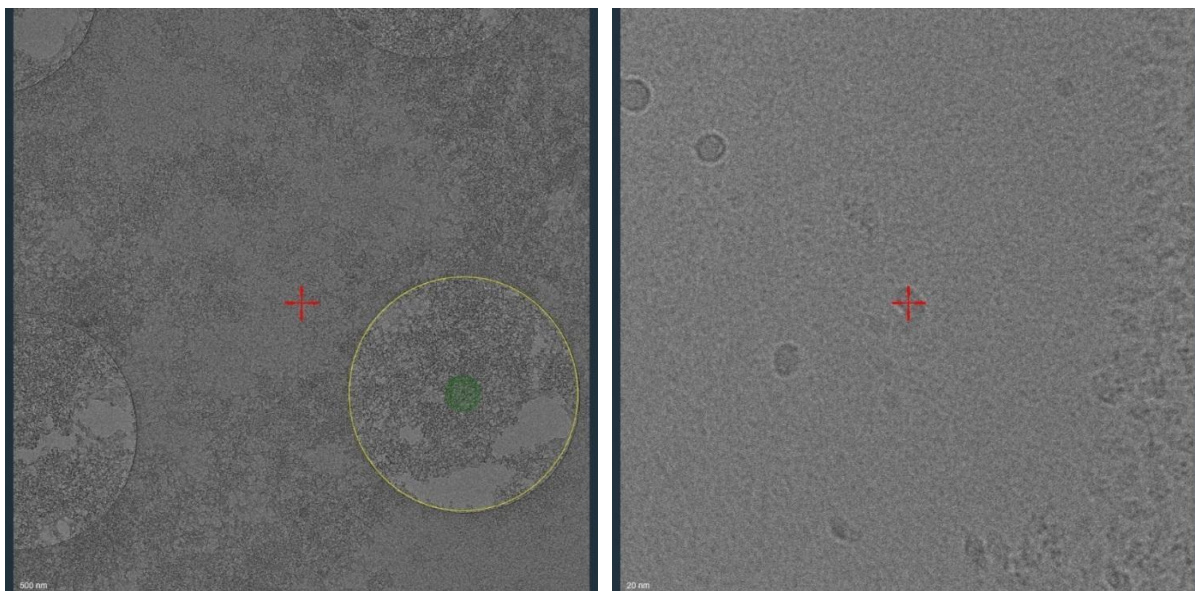
Figure 3.26 Screened micrographs of the protein complex with aziridine probe TB652.
An overly figure of Copper grid holes shows a clear aggregation of the protein complex. The scale bar for the micrographs is 500 nm (left). A micrograph shows massively aggregated particles, the scale bar for the micrographs is 20 nm (right).

Applying single particle analysis of Cryo-EM, the recombinant β-galactosidase, bound to epoxide moiety, was determined. Particles in the micrographs were homogenous (Figure 3.27, A); they showed multi-views and random distribution.

This compound was suitable for the EM study because it was in HEPES pH 7.5, 0.5 M NaCl buffer, 7.5 pH same as the SEC buffer for our protein of interest.

For auto picking particles, the high-resolution map of apoprotein was used directly and optimised the auto picking parameters as mentioned in 2.4.1.

Therefore, the EM de novo determination of the initial model was obtained in 15 Å by extracting the reference-free 2D classes at (0.889778 Angstrom/pixel, 450-pixel box) and low-pass filtered at 20 Å (Figure 3.27, B). The first run of 2D classification free reference was used as a template model because a sufficient number of particles were selected and included a variety of views. The initial map was used to obtain 3D classification (Figure 3.27, C). The best-represented 3D class particles of the homotetramer subunit were selected and then used for 3D refinement by using a gold-standard criterion.

Postprocessing the tetramer with a mask corresponding to the particles' dimension, removing the noisy solvent in the particle's edges, and imposing D2 symmetry led to a 2.7 Å resolution map. Angular distribution analysis demonstrated that there was a weak preferred orientation for the particles (viewed along with the top view of the tetramer, road shape) (Figure 3.27, D)., overall balanced distribution of particles over different directions (Figure 3.27, A, B). The conclusive map reveals that a subunit of tetramer binds to the epoxide inhibitor. It recognises the covalent bond forms a tight binding with active domain (in *Coot* map) and clearly determined density.
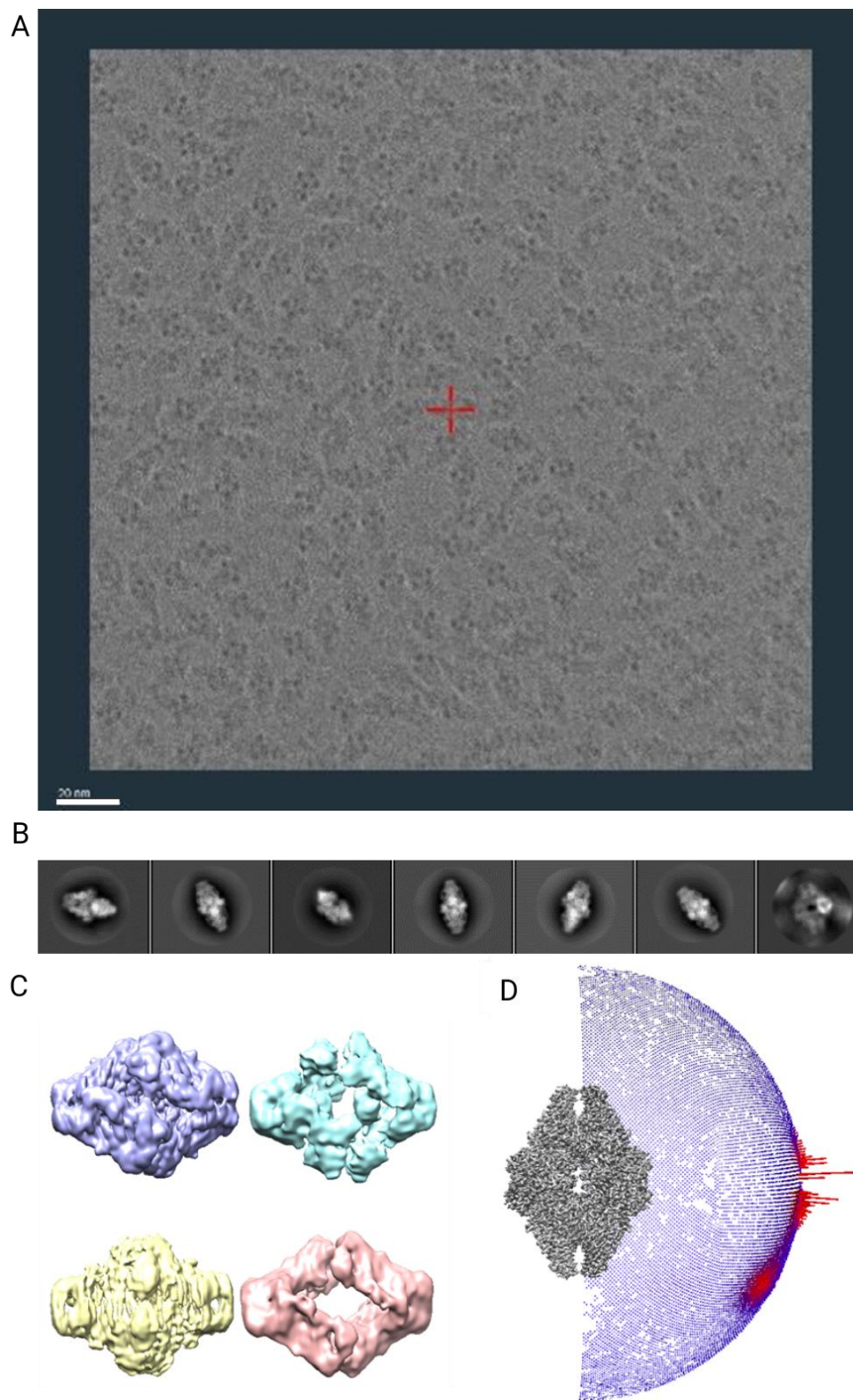
Figure 3.27 The Cryo-EM analysis of the ligand β-galactosidase.
(A) An illustrative micrograph of the β-galactosidase sample by the EPU Cryo-EM. The scale bar represents 20 nm. (B) Representative 2D class averages of β-galactosidase particle images. C) 3D classification of four different classes. (D) Angular distribution of ligand protein with LWA480 map at 2.7 Å.

3.2.3 The structure of E. coli β-Galactosidase in a covalent complex with 4-epi cyclophellitol

3.2.3.1 Model building and refinement of apo- β-Galactosidase

Two apoprotein maps were built and one complex structure was generated. The electron density map for our apo β-galactosidase structure was determined in-house using a single particle electron cryo-microscopy. Nine β-galactosidase Cryo-EM structures had been solved before this study. In 2020, Merk and colleagues used a 200 keV microscope: this is the voltage that we employed in the present work (Merk et al. 2020) . Also, note that their final map has the highest number of particles as compared to most of the other cases. This can be explained by many advances such as specimen quality (R. F. Thompson et al. 2019), protein concentration and purity, detector technology, electron source and programs that produce the reconstructions and enable auto picking of particles and correcting aberrations.

3.2.3.2 Model building and refinement of β-Galactosidase covalent complex

The objective of this chapter was to assess the suitability of a 200 keV cryo-microscope for structural studies of enzymatic inhibition, producing 3D reconstructions and refined atomic models of enzyme-inhibitor covalent complexes. Here, β-Galactosidase was incubated with epoxide LWA480 and data processed to 2.7 Å as shown in Figure 3.28.
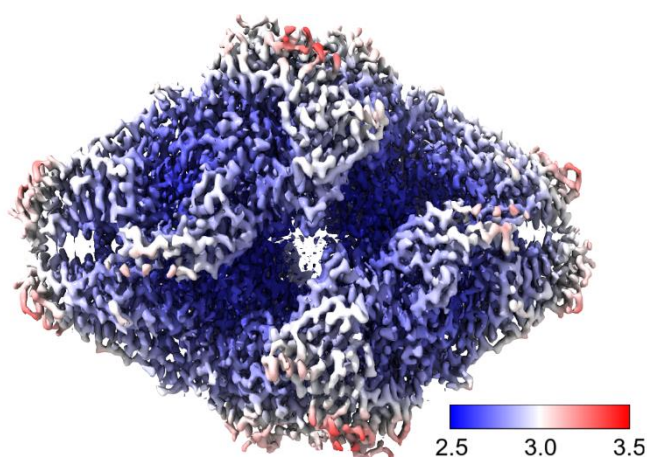
Figure 3.28 Local resolution of β-galactosidase with epoxide inhibitor at 2.7 Å.
As expected, the core of the molecule showed higher resolution than the outer ends of the structure.

The structure of the covalent complex was obtained as previously described (see *Materials and methods*). The 3D map was deposited on the Electron Microscopy DataBase (EMDB) with identifier EMD-18488, and cartesian coordinates were deposited on the worldwide Protein Data Bank (wwPDB) under code: 8QLZ.

The key interest of the complex is obviously the covalent ligand. As illustrated below, the epoxide ring needs to open and be protonated to give the OH group, the stereochemistry at the anomeric carbon (which is attacked by the enzymatic nucleophile) will flip, and that will then be covalently bound to the catalytic nucleophile. This expected chemistry is visible in the finished structure (Figure 3.29). The structure of the ligand was refined without restraining torsion angles, allowing the shape and gradients of the map to accommodate the chair conformation – indeed, saturated rings in the pyranose form easily deviate from the expected minimal-energy conformation whenever there is either a modelling mistake (*e.g.* a misinterpretation of the map) or the gradients in the map are weak.

Glu$_{461}$

β-Galactosidase acid residue

HO

HO

HO

HO

OH

O

H

Galactocyclophellitol or
4-epi cyclophellitol

O

O$^{\ominus}$

Glu$_{537}$

β-Galactosidase nucleophile residue

HO

HO

HO

OH

OH

C2

EO2

EO1

O

O

Glu$_{537}$

Reacted Galactocyclophellitol
or 4-epi cyclophellitol

β-Galactosidase in covalent state

Glu 416

EO2

EO1

Glu 537
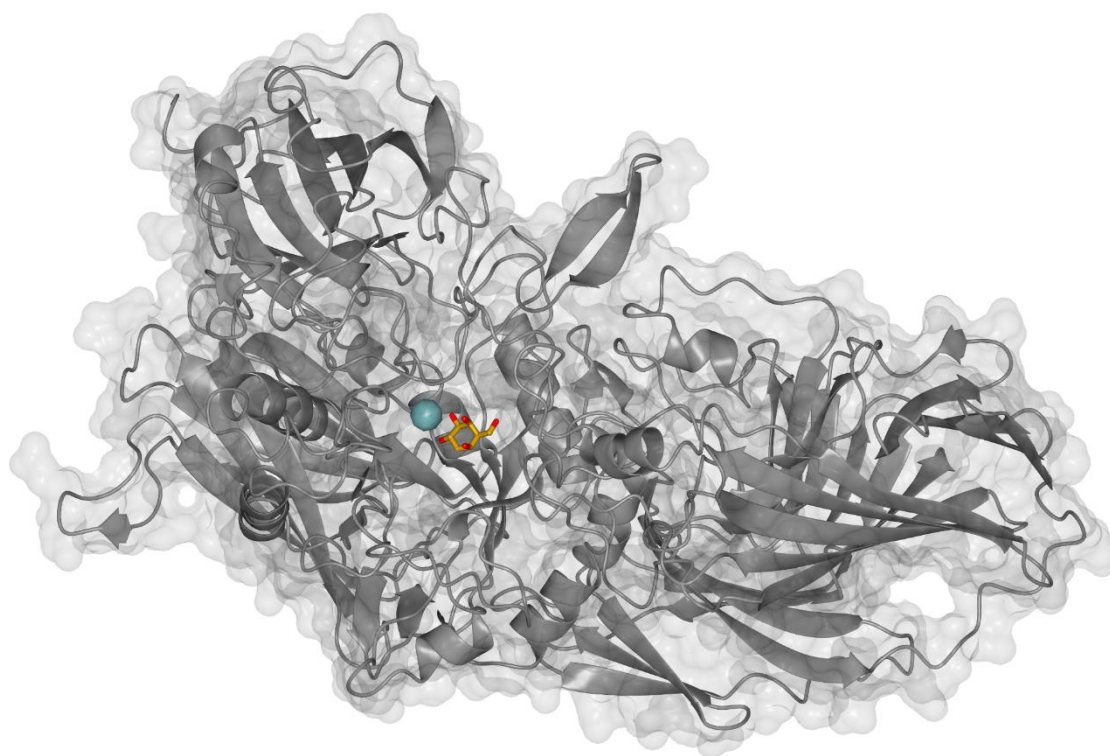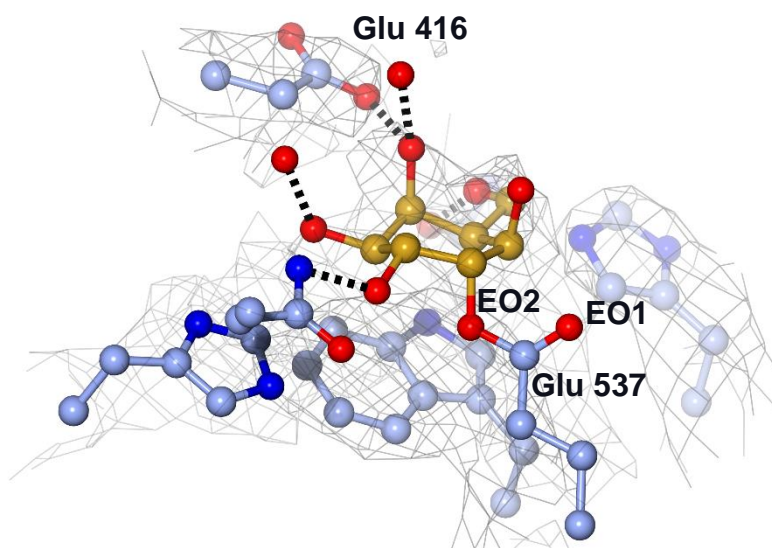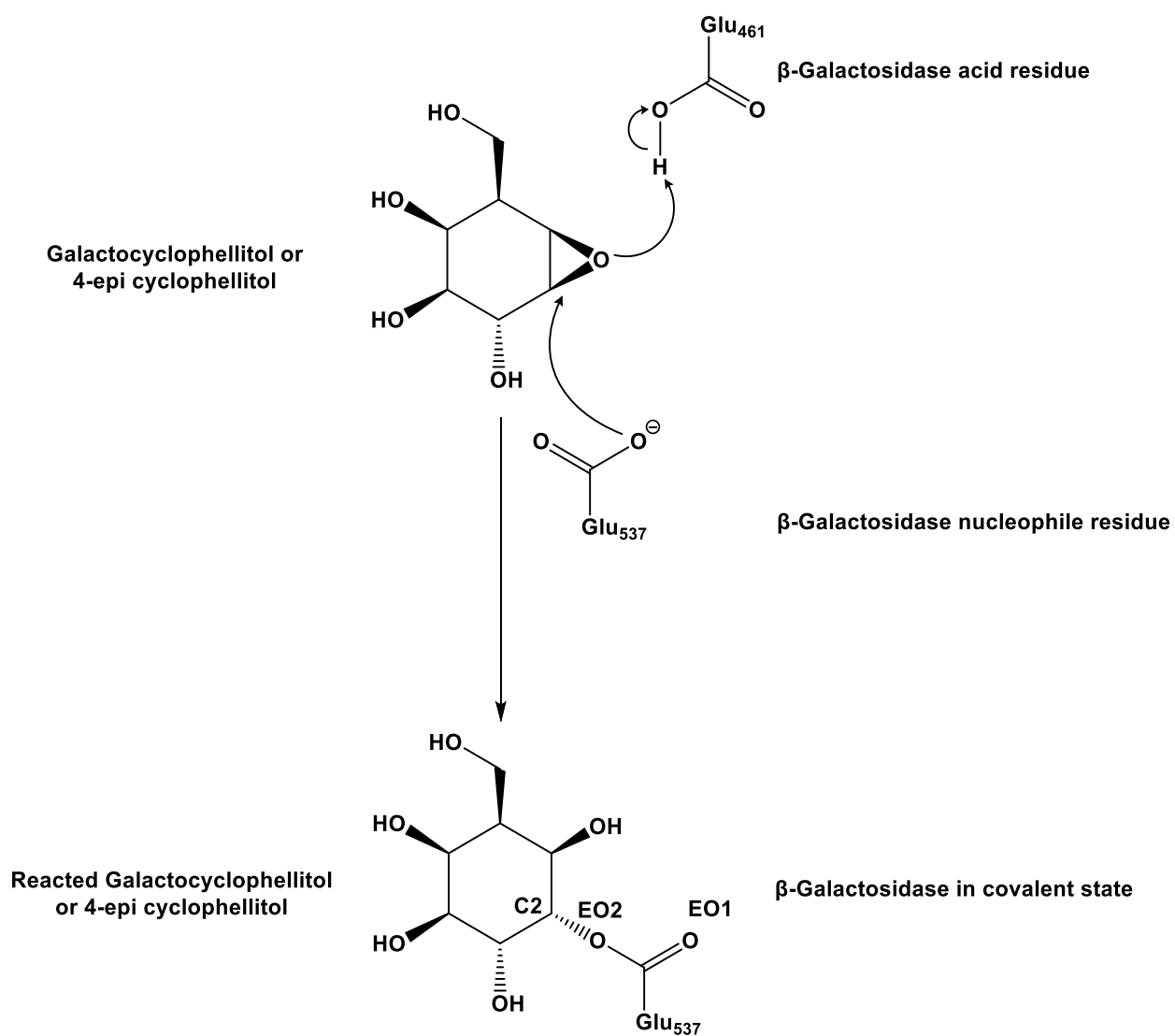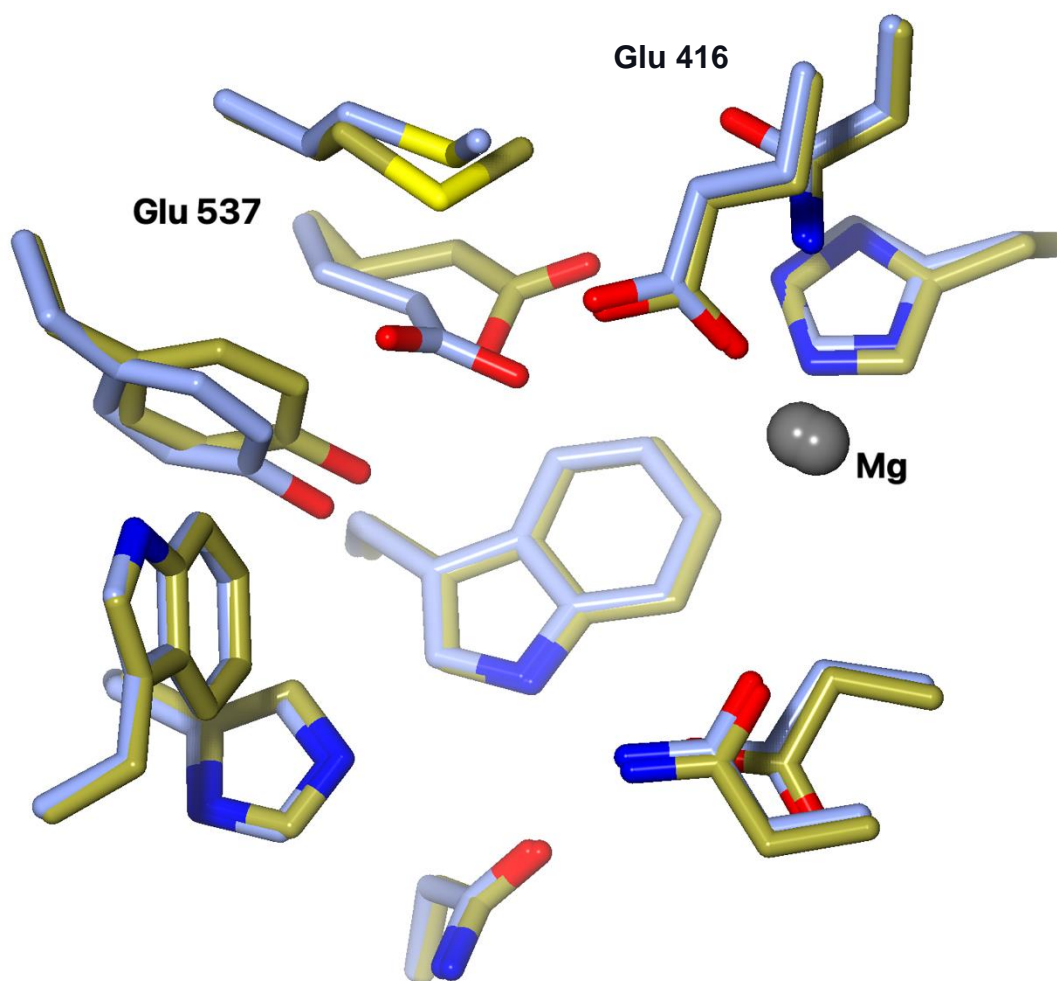
Figure 3.29 The covalent bond and binding pocket of β-galactosidase.
The first figure of β-galactosidase subunit (grey) shows the binding pocket of the ligand (gold cylinder) and(Mg^{2+} ion) (blue sphere). The second figure is the mechanism of covalently bound. The third figure, modelling of the reacted 4-epi cyclophellitol into the map around the active residues, Glu537 and Glu461. The map is contoured at 0.2 V (the *depositor recommended value* on EMDB). The fourth figure is an overlay of active site residues; yellow is the high-resolution XRD structure of a complex with ligand beta-galactose (PDB code: 1JZ7); blue is our probed β-galactosidase (PDB code: 8QLZ). The third panel shows a rotated view with respect to the second panel in order to provide an unobstructed view of the different residues.

The active site with the bound probe (Figure 3.29) has a very similar shape to that of the high-resolution XRD structure of a complex with ligand beta-galactose (Douglas H. Juers et al. 2001). Indeed, the shape of the map and the geometry of the covalently attached probe is comparable to that in the structure of a beta-glucosidase from *Thermotoga maritima* – (PDB code: 2JAL) (Gloster, Madsen, and Davies 2007).– determined to 1.9 Å resolution, with the exception of the galacto-configuration (4-epi). This can be seen in Figure 3.30:
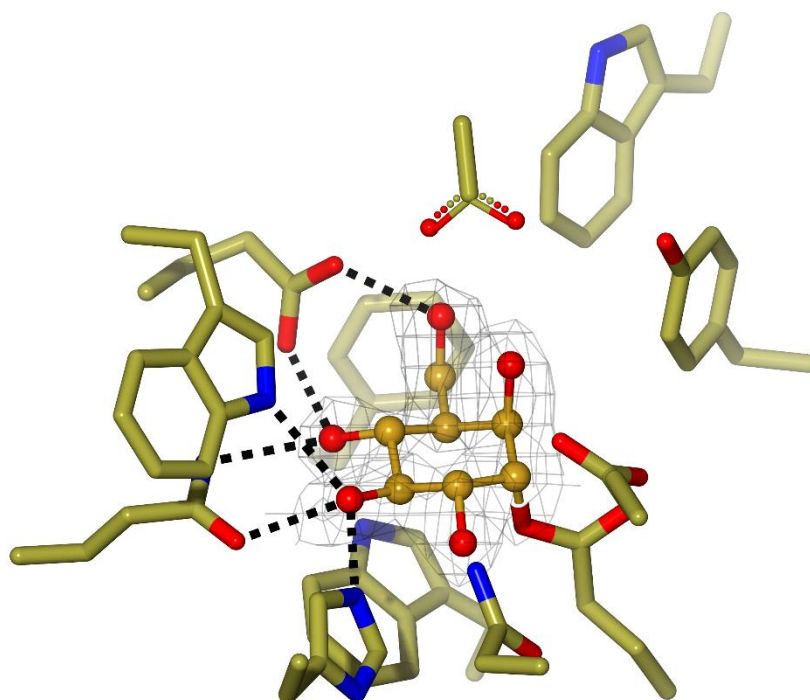
Figure 3.30 The covalent bond and pocket of a cyclophellitol inhibited β-glucosidase from Thermotoga maritima (PDB code: 2JAL).

The figure shows the conformation and overall shape of the reacted cyclophellitol linked to a beta-glucosidase.

This is a crystallographic structure: electron density is contoured at 1 sigma.

Similar tetramer-shaped assemblies were observed for the apo commercial β-galactosidase, recombinant β-galactosidase of complex with 4-epi cyclophellitol and the Apo (Figure 3.30)

EM map of the Apo commercial β-Galactosidase at 3.4 Å

EM map of the recombinant β-Galactosidase in complex with 4-epi cyclophellitol at 2.7 Å

EM map of the Apo recombinant β-Galactosidase at 2.5 Å
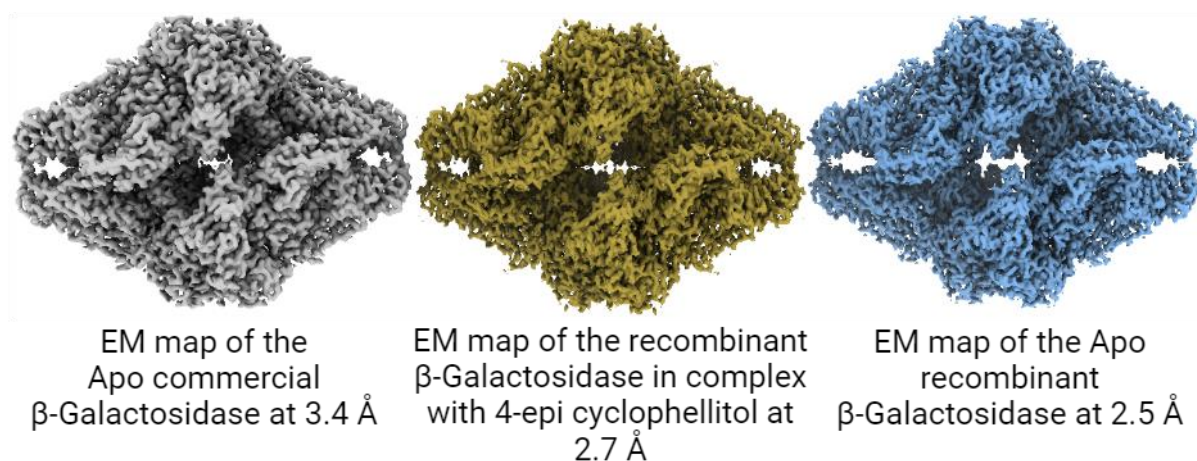
Figure 3.31 Final β-galactosidase potential maps
Apo commercial β-galactosidase at 3.4 Å; grey, recombinant β-galactosidase in complex with 4-epi cyclophellitol at 2.7 Å; yellow and apo recombinant β-galactosidase at 2.5 Å; blue. Figures were generated with Chimera X (Pettersen et al. 2021).

The overall quality of the structure according to the wwPDB validation scales is roughly on par or better than all structures – perhaps an expected result taking into account the gradual improvements in software methods – with slight but not qualitatively differences when compared to EM structures determined at similar estimated resolution (Figure 3.32).
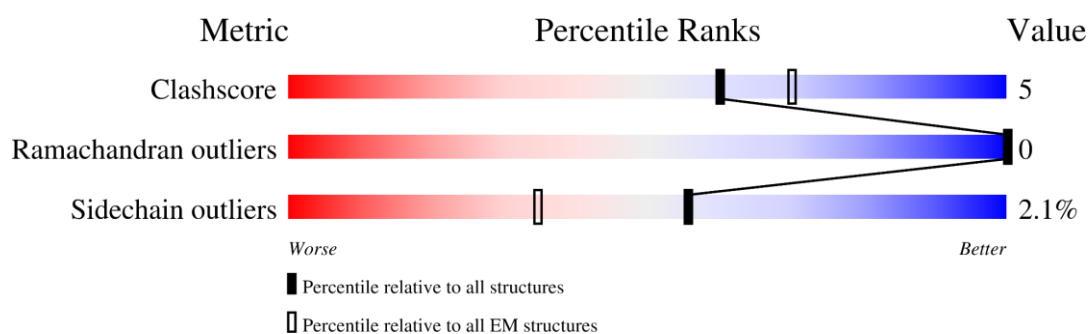
Figure 3.32 The wwPDB validation sliders for our deposition, (PDB code: 8QLZ).
Most issues concern side-chains that are not well defined in the EM map. The map for the main chain was excellent, and this is likely reflected in the result for the Ramachandran analysis, which matches that of many other higher resolution structures of the same enzyme.

A number of issues were flagged up upon deposition, but these mostly concern non-planarity of the guanidinium group in arginines. This is issue highlights an ongoing problem concerning sometimes outdated prior chemical knowledge used in the wwPDB validation reports: modern restraint dictionary generators (*e.g.* phenix.elbow, or AceDRG in the case of this study) have relaxed the planarity restraints on arginine according to the latest knowledge (Moriarty et al. 2020), and the wwPDB continues to check for strict planarity, highlighting non-planar guanidinium groups as outliers.

The map was deposited on the Electron Microscopy Data Base (EMDB) with identifier EMD-18488, and the coordinates were deposited on the worldwide Protein Data Bank (wwPDB) under code 8QLZ. Valuable assistance was provided by Deborah Harrus (PDBe, EBI Cambridge), who monitored deposition, validation and annotated the entry.

Map, coordinates and final PDB validation report are available for pre-release inspection on Zenodo: 10.5281/zenodo.8384171

3.3 Discussion

This study reports the final maps of commercial and recombinant β-galactosidase imaged at 1.8 mg/mL and 2 mg/mL concentrations using Quantifoil Cu grids on a 200 keV microscope. It's worth noting that the majority of previous β-galactosidase Cryo-EM structures determined by other groups were generated using high-voltage instruments such as 300 keV (listed in Table 3.3).

The use of a 200 keV instrument in this study is significant. While an electron microscope with a power of 300 keV is generally used for recording higher resolution structures, the 200 keV instrument is typically used for screening (Mills 2021), due to somewhat limited resolution of such instruments (Egerton 2011) as mentioned in section 1.1.2.1.

During my PhD, a super high-resolution structure of a noncovalent complex of β-galactosidase with the inhibitor PETG, determined at 1.1 Å, was published (Chakraborty 2023). This study underscored the importance of the optimum quality of the specimen, which allowed to generate an ultra-resolution EM structure using an old-generation Krios G1 300 keV.

The previous highest-resolution map of β-galactosidase was deposited by (Merk et al. 2020). It was solved using a 200 keV power microscope. The specimen quality was important for deriving this structure at high resolution, as mentioned in sections 3.1.3 and 3.2.3.1. The grids used in this study were prepared from fresh purified commercial apo protein using HEPES buffer. Furthermore, protein sample concentration and the number of the final particles were higher than those used in our study or previous EM maps.

In our troubleshooting experiments, when the grids were prepared at a slightly higher concentration of 3 mg/mL to produce 2D classes, the samples suffered from preferred orientations (a side view) of the molecule, as described in section 3.2.2.2.1.

The atomic model of β-galactosidase has been derived many times and is available in complex with other compounds from XRC (Douglas H. Juers et al. 2001; Jancewicz et al. 2012; Jancewicz et al. 2012; Wheatley et al. 2013), and cryo-Em structures (Vinothkumar, McMullan, and Henderson 2014; Bartesaghi et al. 2015; Bartesaghi et al. 2018; Saur et al. 2020). The E. coli retaining enzyme produced in-house was blocked with the compound 4-epi

cyclophellitol, and its inactivation reaction was confirmed in the solution. Subsequent inhibition analysis experiments were performed to generate a high-resolution 3D structure of β-galactosidase covalently bound to an inhibitor for the first time via cryo-EM. The obtained density map was used to build an atomic model. Several models in complex with cyclophellitol epoxide compounds, derived from the current study, have been deposited with the PDB.

AceDRG has generated a dictionary of the compound of interest to us. We have considered conformational changes by deleting atoms and/or forming bonds using CCP4. Servalcat in CCPEM was performed for refinement. We sought to preserve the ligand during module building and refinement. Figure 3.9. shows the number of cyclophellitol epoxide inhibitors with the deletion of the main ingredient, epoxide.

3.4 Conclusions

Cryo-EM is becoming a mainstay in structural biology laboratories around the world. The proliferation of more affordable 200 keV instruments with lower electron energy (compared to the 300 keV instruments like Titan Krios), coupled with improvements in software and standard practices are bringing about a revolution in solving the structure of bigger complexes. However, only few examples are available of detailed studies of enzymatic reactions done using the lower energy 200 keV instrument. Thus, the main aim of this project was successfully accomplished in that we show that this kind of routine work is feasible and that ligands can be confidently resolved even for cases and at resolutions where crystallographic data -were ambiguous. Indeed, the refinement of the equivalent ligand conformations in XRD would have required additional torsion restraints, which was not necessary using in our case.

Chapter 4   Single particle microscopy of a small glycoprotein target: human serum transferrin

Objectives of this study

- Push the technical limits of Cryo-EM to develop a glycosylated reference system for chemical glycobiology projects by determining the structure of a small and flexible target: human serum transferrin.

Chapter summary

The electron cryo-microscopy (Cryo-EM) community has long sought high-resolution structures of biological glycoproteins weighing less than 100 kilodaltons (kDa). While improvements in sample preparation have allowed for the visualisation of glycosylated specimens in this size range, the reconstruction process in this size range is still very problematic and might fail for certain samples. In this work, we show that by employing a transmission electron microscope operating at 200 keV paired with a modern direct electron detector, traditional defocus-based Cryo-EM techniques may be utilised to identify structures of glycosylated specimens weighing less than 100 kDa, focusing on the fully glycosylated human transferrin.

Our structures of the glycosylated transferrin were purified, and the 3D apo state structure was revealed at 6.8 Å, which showed a "fully-opened" conformation of N lobe.

Furthermore, for enhancing transferrin stability, it was further explored with two iron compounds that bind strongly yet reversibly, then purified, showing transferrin's conformational alterations.

The two structures of holotransferrin show that bound ligands may be detected through conformation change and subsequently examined in metal interactions. These structures show unique binding mechanism states with "fully-closed" conformation of the N lobe. That suggested Cryo-EM may discover the conformation changes of glycoproteins (≤ 80 kDa) within a dataset at 8 Å.

Despite the purification or strong binding to the iron salts, 2D and 3D classification observations revealed flexibility and compositional variability.

The structural studies on human serum transferrin provided here offer mechanistic insight into the structure of apo and its complexes. They will direct future efforts to improve and tackle sample preparation and processing of small glycoproteins.

4.1 Introduction

This chapter documents our structure determination studies of the Human serum transferrin (hTF) protein in its natively glycosylated apo state, which we intend to use as a model for studying glycan mobility and microheterogeneity.

Although many hTF crystal structures are now available, only some structures of the natively glycosylated states have been elucidated. Human serum transferrin (hTF) is a monomeric glycoprotein (~80 kDa), synthesised in Hepatocytes and secreted into the blood plasma, where it transports iron (Luck and Mason 2012). Transferrin (TF) shows similarity to iron transporters family members lactoferrin (Lf) and melanotransferrin (MTf) (Hao et al. 2018). These Tf proteins have qualities in common with the ability to bind iron tightly and reversibly. The TF family's primary function is to provide transportation, binding, and solubility of the ferric ion (Lambert 2012).

Transferrin (Tf) has been broadly studied in biomedicine due to its essential role as an iron transporter from the liver parenchymal cells, intestine, and reticuloendothelial system, to the proliferation of body cells. Apart from that, its physiologic role as granulocyte/pollen-binding protein (GPBP) is removing particulate organic matter and allergens from serum (Sass-Kuhn et al. 1984). It plays a crucial role in preventing iron from participating in redox reactions and acting as the major defence of the body's immune system by uptaking iron from pathogens.

Human Serum Transferrin is reported to be primarily *N*-glycosylated (Sturiale *et al.* 2005; Pascreau, Auditeau, and Borgel 2023) and partially O-glycosylated (Coudert *et al.* 2023). Changes in this glycosylation give rise to two populations of different molecular weights. Due to the difference in weight, they may be separated by chromatography. Typically, it has two biantennary, *N*-linked, fully sialylated glycans with core-linked fucose and very little heterogeneity.

4.1.1 Transferrin Structure

hTF has a bilobal fold into the N terminal and C terminal lobe, each lobe cleft bound to a molecule of Fe3+. The main lobes are separated into two subdomains, N1 and N2 and C1 and

C2, as illustrated in Figure 4.1. hTF about 679 residues. The N lobe (331 residues) and C lobe (341 residues) are joined by a bridge of seven amino acids (residues 332–338) and are ~40% similar in sequence. Presumably, this occurred due to gene duplication and fusion (Park et al. 1985). In primary sequence, the subdomains N1 (residues 1–93) and C1 (residues 339–425 and 573–679) are discontinuous, while the subdomains N2 (residues 94–246) and C2 (residues 426–572) are continuous (Wally et al. 2006b; Silva et al. 2021).

Figure 4.1, left panel, represents the model after zooming out and separating transferrin. The amino terminal lobe of transferrin represents the N lobe, and the carboxylate terminal lobe represents the C lobe. Both lobes have residues; the N lobe has residues running from 1–331, while the C lobe has residues from 339–679. A hinge connects the subdomains in the lobes, forming a cleft that comprises the iron-binding site (Wally et al. 2006).

Both lobes of the molecule are joined via a linker (7 amino acid residue), the interface of which holds a hydrophobic patch. Furthermore, two salt bridges cause the interaction among the separated subdomains. In contrast to other proteins present in plasma, the TF structure contains about 19 disulphide bonds that stabilise the iron-binding sites. The N protein lobe contains eight disulphide bonds, and the C protein lobe has 11 disulphide bonds (Benjamín-Rivera et al. 2020).
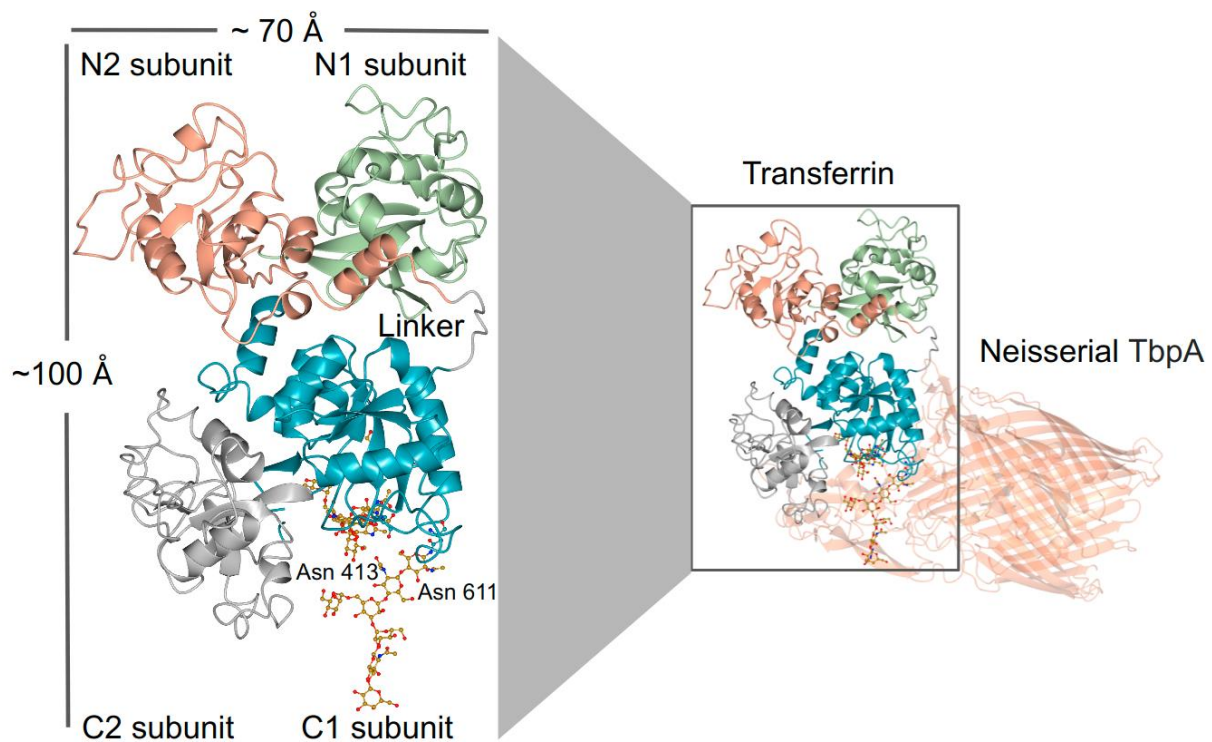
Figure 4.1 A complex of human serum transferrin with Neisseria meningitidis serogroup B transferrin binding protein A.
Crystal structure of hTF bound to Neisseria meningitidis serogroup B TbpA, (PDB code: 3V8X) (right). The ribbon model of separated complete apo hTF lobes depicts subunits C1 in turquoise, C2 in grey, N1 in green, and N2 in orange (left). The peptide linker between the N and C lobes is in grey: Asn 413 and 611 of glycosylated sites. Figures were generated with CCP4mg (McNicholas et al. 2011).

4.1.2 Pathway of human transferrin

Transferrin functionis to transport iron in the body. In the apo state, transferrin does not bind to iron in plasma, whilst transferrin bound to iron forms the holo-state, revealing conformation change (Benjamín-Rivera et al. 2020). After iron binding in the transferrin molecule, the transportation to bone marrow occurs, synthesising haemoglobin and fragments of erythrocytes. The iron from the body is lost by desquamation of epithelial cells, menstruation and sweating, thus making it an obligatory phenomenon and leading to no way of its regulation. On that account, strict regulation of intestinal absorption remains the major control for iron homeostasis (Anderson and Shah 2013).

The pathway of apo (iron-free) and holo (iron-bound) human transferrin involving the transferrin receptor 1 (TFR1) includes multiple specific steps, such as clathrin-mediated endocytosis, pH-dependent iron release, and intracellular trafficking cells (Brock 1994; Templeton 2002). Figure 4.2 illustrates the cellular iron uptake by the TF cycle (Levina and Lay 2019a; Deffieu et al. 2021; Gammella et al. 2017).

The clathrin-coated vesicles pinch off from the membrane, becoming uncoated endocytic vesicles. These vesicles mature into early endosomes, which are initially neutral in pH. Early endosomes mature into late endosomes through acidification driven by ATP-dependent proton pumps (*e.g.,* ATPase pumps). The ATP-dependent proton pumps transport protons (H^+ ions) actively into the endosome, lowering its pH. This acidification is crucial for subsequent steps in iron metabolism. Iron is released from transferrin in the acidic environment when the endosomal pH drops below 5.6. The low pH induces a conformational change in the transferrin-TFR1 complex, facilitating the release of iron ions. Apo-transferrin is then recycled back to the cell surface for further iron uptake (Silva et al. 2021).

Within the late endosome, the ferrireductase enzyme STEAP3 reduces Fe(III) ions to Fe(II). This reduction step is essential for efficient iron transport out of the endosome (McKie 2005; Ohgami et al. 2005). The reduced iron ions (Fe^{2+}) are transported out of the late endosome into the cytoplasm by DMT1 (divalent metal transporter 1), also known as DCT1 (divalent cation transporter 1) (McKie 2005). DMT1 is responsible for transporting various divalent metal ions, including Fe^{2+}, across cellular membranes. Once in the cytosol, iron ions participate in various cellular processes, including haemoglobin synthesis, DNA replication, and mitochondrial respiration.

The pH of endosomes plays a crucial role in iron binding and release from transferrin. The acidic environment of late endosomes facilitates the release of iron ions from transferrin by inducing a conformational change in the transferrin-TFR1 complex. Additionally, the acidic pH is necessary for the ferrireductase activity of STEAP3, which reduces Fe(III) to Fe(II) for efficient iron transport out of the endosome (Yanatori and Kishi 2019).
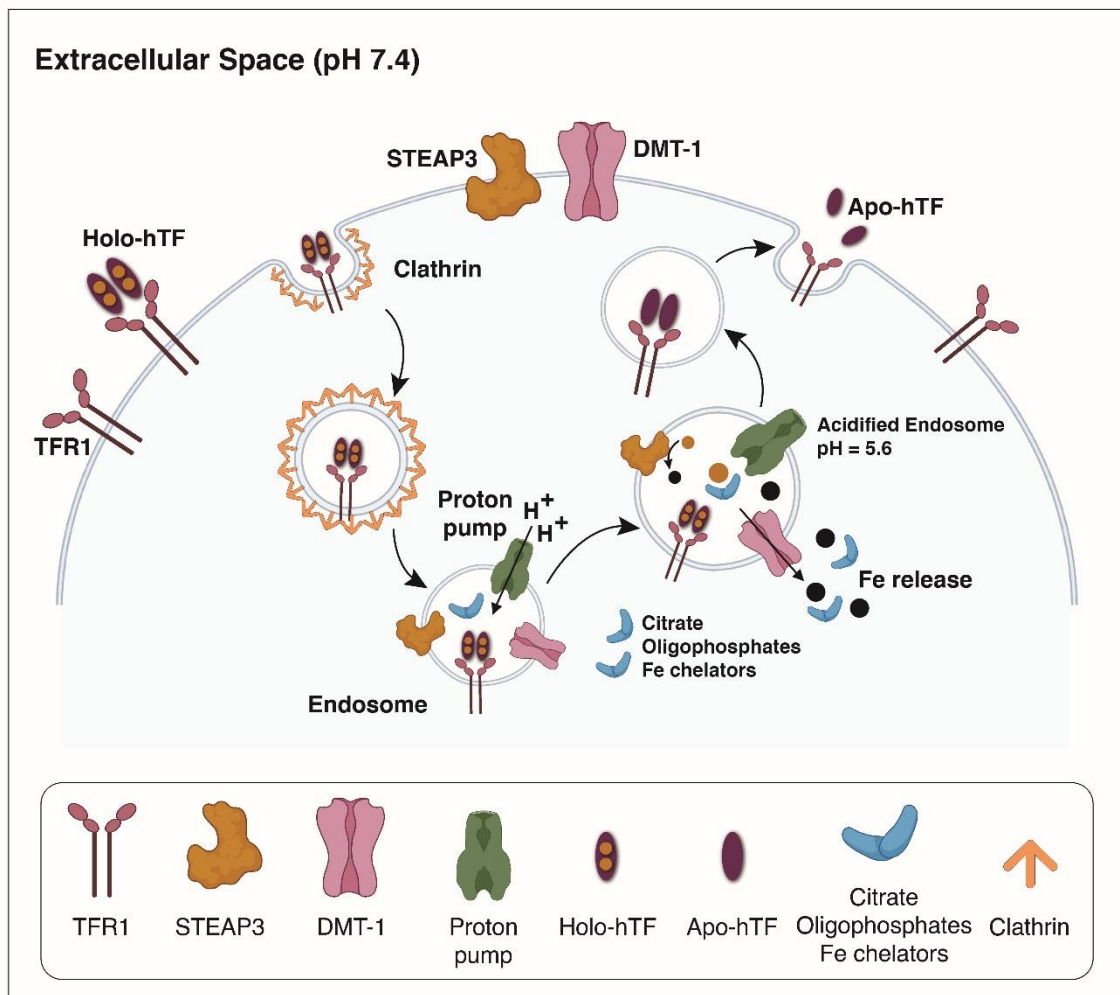
Figure 4.2 Illustration of transferrin cycle.
Iron-loaded Tf attaches to TfR1 on the cell surface, clathrin-mediated endocytosis occurs causing complex internalisation. Dissociation of Fe(III) from Tf occurs in the acidic endosome. Fe(III) gets reduced to Fe(II) by ferrireductase (STEAP3) and transported into the cytosol divalent metal transporter 1 (DMT1). The apo-TF/TFR1 complex is transported to the cell surface and apo-TF is freed into the bloodstream. The figure was created with BioRender.

4.1.3 Binding mechanism effect

hTF has a conformational transition in both N and C lobes during the binding or releasing iron mechanism (Hamilton et al. 2004); (Benjamín-Rivera et al. 2020).
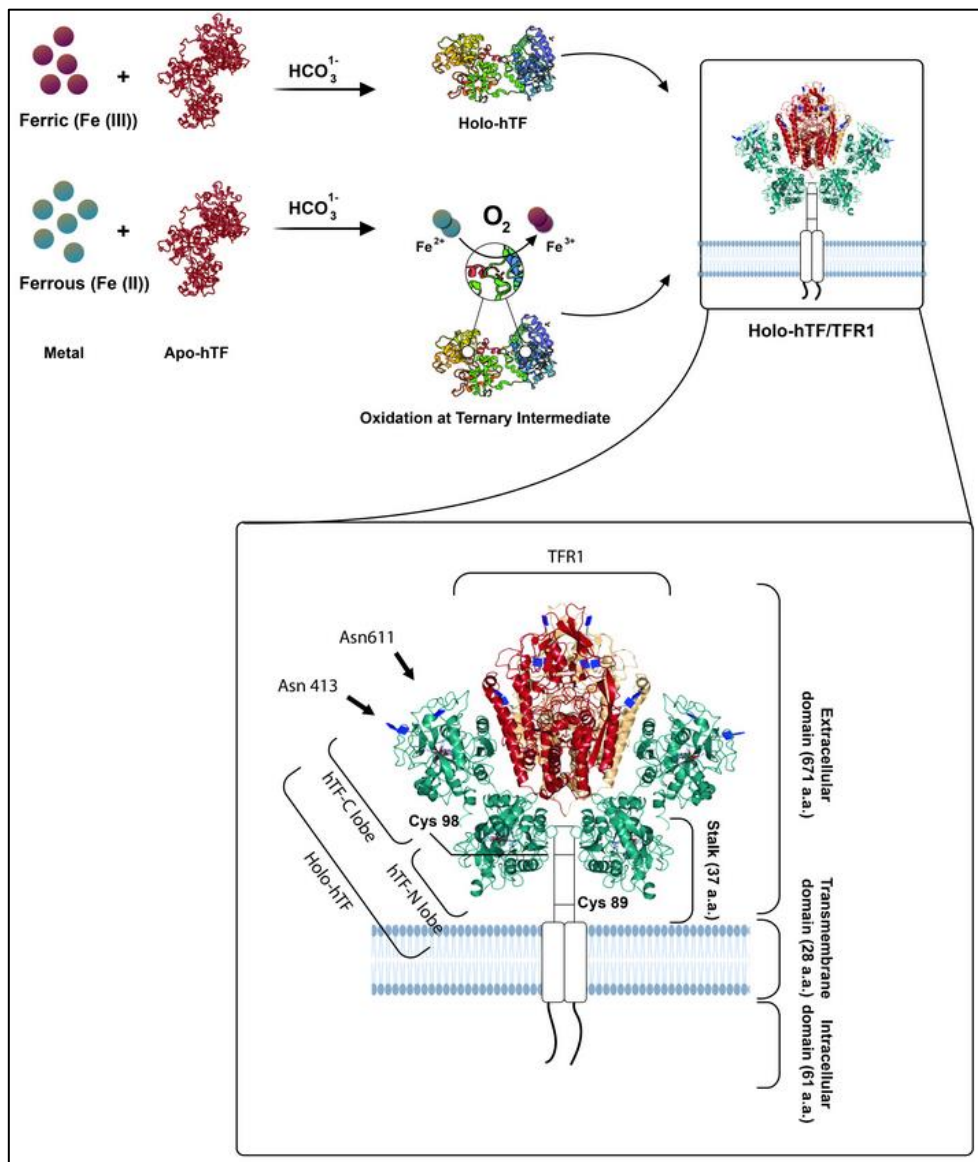
Figure 4.3 The processes of formation of Holo-hTF complex by hTF binding with Fe.
In the first and more common process, Fe(III) released in the bloodstream immediately binds by serum hTF in presence of bicarbonate anion. B) In the second process, Fe(II) that leaches into the blood is bound by hTF by bicarbonate synergism, hTF acting as ferroxidase and oxidase. The figure's layout was inspired by (Benjamín-Rivera et al. 2020) ,and created with BioRender, UCSF Chimera (Pettersen et al. 2004) and CCP4mg (S. McNicholas et al. 2011).

Figure 4.3 depicts the binding of iron to the serum transferrin. Two possible states for iron ions are ferrous or ferric ions. A distinctive characteristic of Tfs is the necessity of an anion having a synergistic action, *e.g.* bicarbonate, for stronger iron binding. Iron-binding of each lobe is carried out by similar existing residues, two tyrosine, one aspartate, and one histidine residue, and the bicarbonate achieves the coordination globe. There are six coordinates in an iron atom oriented in a pseudo octahedral layout. The bond lengths in the lobes of TF, as well as different forms, are almost similar. In the iron-binding process, both domains play a role and get together, closing down the binding cleft.

A crystallographic illustration of the ''locked'' conformation and ''extended'' protein conformation has been established for ovotransferrin (Agarwal et al. 1992; Al-Refaie et al. 1995). The 'locked' protein conformation inhibits the chelators from joining the iron-binding site; for this reason, the conformation plays a significant role in the iron release which occurs on the opening of the cleft (Hall et al. 2002; Wally et al. 2006).

4.1.4 PDB data bank of transferrin, glycosylated structure and structures of sub-100 kDa 3D structure

A search of the Protein Data Bank showed that over forty 3D structures of transferrin have been determined by X-ray crystallography, NMR spectromicroscopy and Cryo-EM.
The first cryo-EM low-resolution human transferrin map was determined in a complex with receptor-transferrin at 7.5 Å (PDB code: 1SUV) (Cheng et al. 2005). A few years later, hTF EM structures were determined in a complex of malaria parasites at sub-3 Å (PDB codes: 6D03, 6D04 and 6D05), as shown in Figure 4.29 (Gruszczyk et al. 2018).

The complete amino acid sequence of transferrin was determined in the 1980s (MacGillivray et al. 1982). N lobe characteristic was the first component noted in the human transferrin (MacGillivray et al. 1998; Bewley et al. 1999; H.-W. Yang et al. 2000; Adams et al. 2003)), the N lobe was determined independently in high resolution by crystallographers multiple times (MacGillivray et al. 1998; Bewley et al. 1999; H.-W. Yang et al. 2000; Adams et al. 2003), while half of the hTF of C lobe with iron remained constructed in an unpublished model at 3.3 Å (Hamilton et al. 2004).

The complete structure of hTF remained elusive to crystallise. Until Wally and their team solved the structure of the full-length hTF at 2.70 Å. The 3D of hTF was revealed in two crystal forms. The first model is for apo glycosylated transferrin (PDB code: 2HAV), and the second is for recombinant apo unglycosylated transferrin (PDB code: 2HAU). Interestingly, the structure of the native glycosylated hTF, which is similar to this study target, was deposited without data showing glycans due to the lack of density surroundings Asn413 and Asn611, suggesting that the glycan components are stretchy and exist in numerous arrangements. Moreover, the unglycosylated recombinant human TF showed absences of glycans (Wally et al. 2006).

The available C1 subunit of holo rabbit transferrin was used for molecular docking for the complete hTF structure due to the high sequence similarity of rabbit TF. Transferrin (Tf) and lactoferrin (Lf) from the family iron transporter consist of an identical composition of amino acids as well as similar secondary (also their disulphide linkages) and tertiary structures.

The apo-hTF structure by (Wally et al. 2006)'s group can be compared to other structures such as those of di-ferric rabbit TF (2.6 Å, 79% alike), porcine (2.15 Å, 72% similar) and Human and Bovine lactoferrin (2.2 Å, 2.8 Å 60% correspondingly) respectively (Hall et al. 2002). Figure 4.4 shows a section of the sequence alignment obtained by T-coffe, which shows the structural layout of the amino acid sequences of different transferrin proteins from various species. Each line in the alignment represents the amino acid sequence of a transferrin protein from a specific species. Human Serum transferrin (PDB code: 3QYT), (Yang et al. 2012), human lactoferrin (PDB code: 1B0L), (X. L. Sun et al. 1999), Rabbit serum transferrin (PDB code: 1JNF), (Hall et al. 2002), bovine lactoferrin (PDB code: 1BLF), (Moore et al. 1997). Porcine serum transferrin (PDB code: 1H76), (Hall et al. 2002) is not in the figure. The figure was inspired by (Vogel 2012); (Kell, Heyden, and Pretorius 2020).
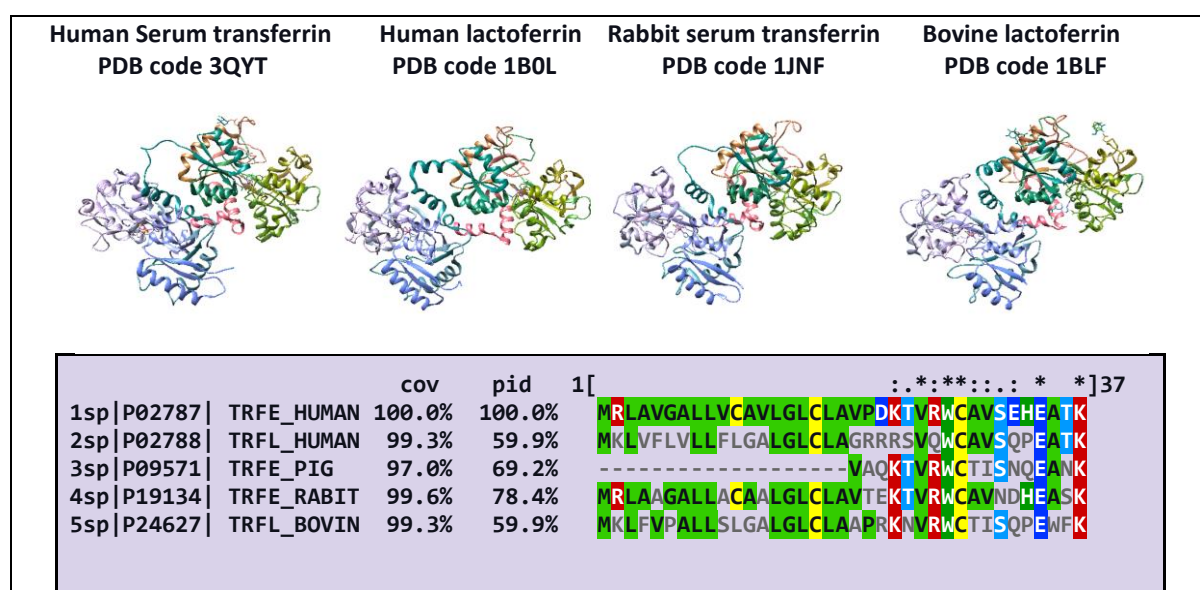
Figure 4.4 Comparison of human serum transferrin with models of several homologues.
Ribbon diagrams generated by UCSF Chimera (Pettersen et al. 2004). A representative extract of their T-coffee sequence alignment is shown at the bottom; five protein sequences correspond to different transferrin proteins from various species, including human (TRFE_HUMAN, TRFL_HUMAN), pig (TRFE_PIG), rabbit (TRFE_RABIT), and bovine (TRFL_BOVIN). Sequences were aligned with hTF in T-coffee (Notredame, Higgins, and Heringa 2000). The conserved residues are highlighted by using the ClustalW colour scheme (J. D. Thompson, Higgins, and Gibson 1994). The conservation percentage (%cov) indicates the degree of conservation of amino acid residues across the aligned sequences. The percent identity (%pid) represents the percentage of identical residues between the aligned sequences. The dots (.) indicate identical residues among the sequences. The colons (:) indicate conserved substitutions. The asterisks (*) indicate semi-conserved substitutions. Gaps (-) are inserted to maximise the alignment between sequences.

Generally, protein sequence alignment is a fundamental bioinformatics technique used to compare and analyse the similarities and differences between two or more protein sequences (Nute, Saleh, and Warnow 2019). It serves several essential purposes in biological and biomedical research, such as homology detection, functional annotation, functional domains, conserved motifs, comparative studies, structural prediction and drug discovery (Guex et al. 2009).

However, despite sharing significant sequence similarity, these transferrin proteins may have functional differences due to variations in their amino acid sequences. Transferrin is a multifunctional protein involved in iron binding and transport, as well as immune regulation and cell growth (Baker, Baker, and Kidd 2002). Variations in the sequence may reflect species-specific adaptations or functional differences in transferrin proteins (Lambert, 2012).

Sequence alignment typically involves aligning the entire length of protein sequences (Lipman et al. 1989). However, there are cases where researchers might focus on a specific region. Aligning the N-terminal region of proteins like lactoferrin and transferrin, especially when comparing isoforms within the same species or homologous proteins across species, can be a helpful procedure for several explanations (Albar et al. 2014). The most intriguing biological activities reside in the protein's uniquely positively charged N-terminal region (Vogel 2012). Aligning this region can help identify conserved structural elements, such as alpha-helices or beta-sheets, which are important for maintaining the protein's tertiary structure (Sitbon and Pietrokovski 2007). Also, it can help trace the origins and divergence of these proteins, it helps identify the key residues or binding sites responsible for the interactions (Rausell et al. 2010) and the N-terminal region may have epitopes targeted for diagnostics or drug development in the case of transferrin receptors, which interact with transferrin and lactoferrin (Macedo and de Sousa 2008).

4.1.5 Cryo-EM of small targets

Since our target is a small glycoprotein, and because of its restrictive resolution at the nanometer to subnanometer level, electron cryo-microscopy (Cryo-EM) was not always the first option for structural biologists until a few years ago. Proteins with low molecular weights

(less than 100 kDa) have traditionally been thought to be too small to resolve their structures using single particle Cryo-EM. Regarding resolution, this approach now competes with X-ray crystallography. It can be employed to discover the atomic structures of macromolecules that are either resistant to crystallisation or challenging to form in certain biological functions.

For the first attempt, the small specimens allowed analysis without using a phase plate, but this was only sometimes the case. Evidence from Electron Microscopy Data Bank (EMDB) entries suggests that nowadays, it is possible to resolve small samples through conventional methods, but only if they are very stable (M. Wu and Lander 2020).

Slightly larger (in kDa) molecules have been solved by Cryo-EM. Some other Cryo-EM structures of the sub-150 have been obtained, illustrated in Figure 4.5 ATP-binding exporter Tmr from *Thermus thermophilus*, with a molecular weight of 151 kDa, was reconstructed at 2.8 Å resolution fairly recently (Hofmann et al. 2019). Another study identified a comparatively more minor structure of about 131 kDa. Safarian's group has studied the structure of *E. coli* cytochrome bd-I oxidase using single particle Cryo-EM at a resolution of 2.7 Å (Safarian et al. 2019).
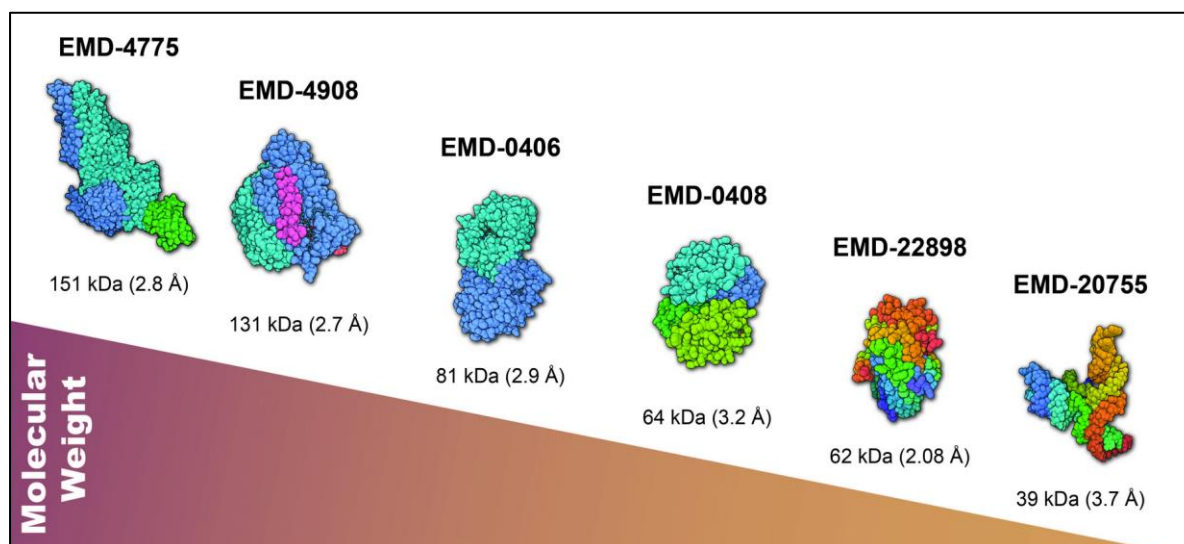
Figure 4.5 High-resolution sub-150 kDa protein structures determined by Cryo-EM analysis. EMD-4775 ATP-binding Exporter, EMD-4908 Cytochrome Bd-Oxidase State 1, EMD-0406 Horse liver alcohol Dehydrogenase, EMD-0408 Human methemoglobin, EMD-22898 SARS-CoV-2 protein 3A in lipid nanodiscs, EMD-20755 Apo SAM-IV Riboswitch RNA. The figure was adapted and updated by (M. Wu and Lander 2020) and created with BioRender.

There are noteworthy examples of high resolution with 200 keV operating microscopes, such as Herzik and colleagues successfully determining the structure of sub-100 kDa at the heterotetrameric and homodimeric structure. In comparison, it was impossible to solve heterodimer at ~32-kDa or unsymmetric structure isolable kinase at ~43 kDa (Herzik, Wu, and Lander 2019). Moreover, a unique small structure was carried out during the recent global pandemic and tackled the drug design that advanced imaging. SARS-CoV-2 protein 3A in lipid nanodiscs was ~62kDa, which was solved at 2.08 Å using an operating system of 300 keV (Kern et al. 2021). The smallest RNA molecule was determined using Cryo-Em are Apo and S-adenosylmethionine (SAM) riboswitches at 3.7 Å and 4.1 Å resolution, respectively (Kaiming Zhang et al. 2019). Figure 4.6 describes the details of various protein molecules obtained by the Cryo-EM method. Horse liver alcohol dehydrogenase (HLAD)is a homodimer protein ~82kDa, Human methemoglobin is a heterotetramer ~64 kDa (Herzik, Wu, and Lander 2019) SARS-CoV-2 protein 3A in lipid nanodiscs is a dimer ~62kD (Kern et al. 2021) Apo and S-adenosylmethionine (SAM) riboswitches RNA is a monomer ~39kDa(Kaiming Zhang et al. 2019).
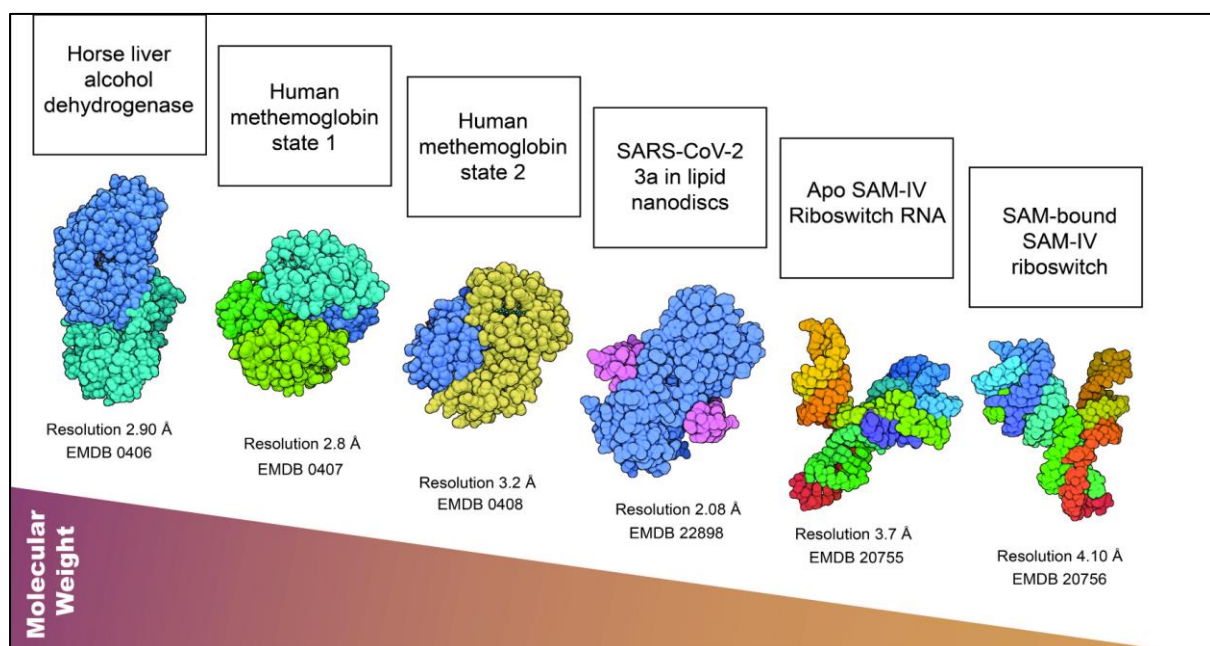
Figure 4.6 Low molecular weight biomolecules, less than 100 kDa, solved by single particle Cryo-EM at 200 keV and 300keV.

4.1.6 Protein glycosylation and glycans

Since human serum transferrin is a glycosylated protein, it is essential to understand what glycosylation and glycans are. Glycosylation is a process that generates a plethora of cellular glycans that are commonly connected to proteins and lipids.

One of the four fundamental constituents of cells, glycans may also be the most common and diversified type of biopolymer found in nature. Glycans make up a sizable portion of the mass and structural variety in biological systems, existing as covalent connections of saccharides frequently coupled to proteins and lipids. The study of glycans and their derivatives' structure, chemistry, production, and biological function is the main goal of the science of glycobiology (Ohtsubo and Marth 2006).

The diverse glycan compositions found in biological molecules and the absence of practical instruments for investigating the glycosylation process and its function have severely hampered glycoscience research (Krasnova and Wong 2016). One crucial cellular process controlling several physiological and pathological activities is glycosylation (Pinho and Reis 2015). A large, varied, and tightly controlled array of cellular glycans typically joined to proteins and lipids are produced through glycosylation. The glycosyl transferases and glycoside hydrolases, which are responsible for transferring and processing glycans, are crucial for the growth and physiology of living creatures, according to studies on the role of glycans during the past ten years. Numerous critical biological functions, such as cell adhesion, molecular trafficking and clearance, receptor activation, signal transduction, and endocytosis, are mediated by glycans (Ohtsubo and Marth 2006).

4.1.6.1 The glycoprotein hTF

Protein glycosylation is defined as a non-templated process driven by enzymes that add complex carbohydrate components to amino acid residue side chains (Moremen, Tiemeyer, and Nairn 2012). This is a ubiquitous protein modification; mainly, Transferrin molecules are glycosylated proteins, excluding the fish TF (Hongzhe Sun, Li, and Sadler 1999). The TFs present in mammals have *N*-glycan chains linked to the asparagine residues and is partially *O*-glycosylated. However, the *N*-glycosylation process in this protein class is highly variable (Hongzhe Sun, Li, and Sadler 1999).

hTF is the most common glycoprotein in human serum (Karimian Amroabadi et al. 2018). It does not always have *O*-glycans attached (Moremen, Tiemeyer, and Nairn 2012). *O*-glycans were shown for the first time in human amniotic fluid transferrin (van Rooijen et al. 1998). Human serum transferrin primarily undergoes *N*-linked glycosylation, where sugar molecules are attached to specific asparagine (Asn) residues within the protein structure.

It contains two full-length *N*-glycans; changes in this glycosylation give rise to two populations of different molecular weights. Due to the difference in weight, they may be separated. Typically, it has two biantennary fully sialylated glycans with fucose and very little heterogeneity. For example, the complete monomer of hTF in complex with *Neisseria meningitidis* Figure 4.7 displays the glycan present in the human transferrin.
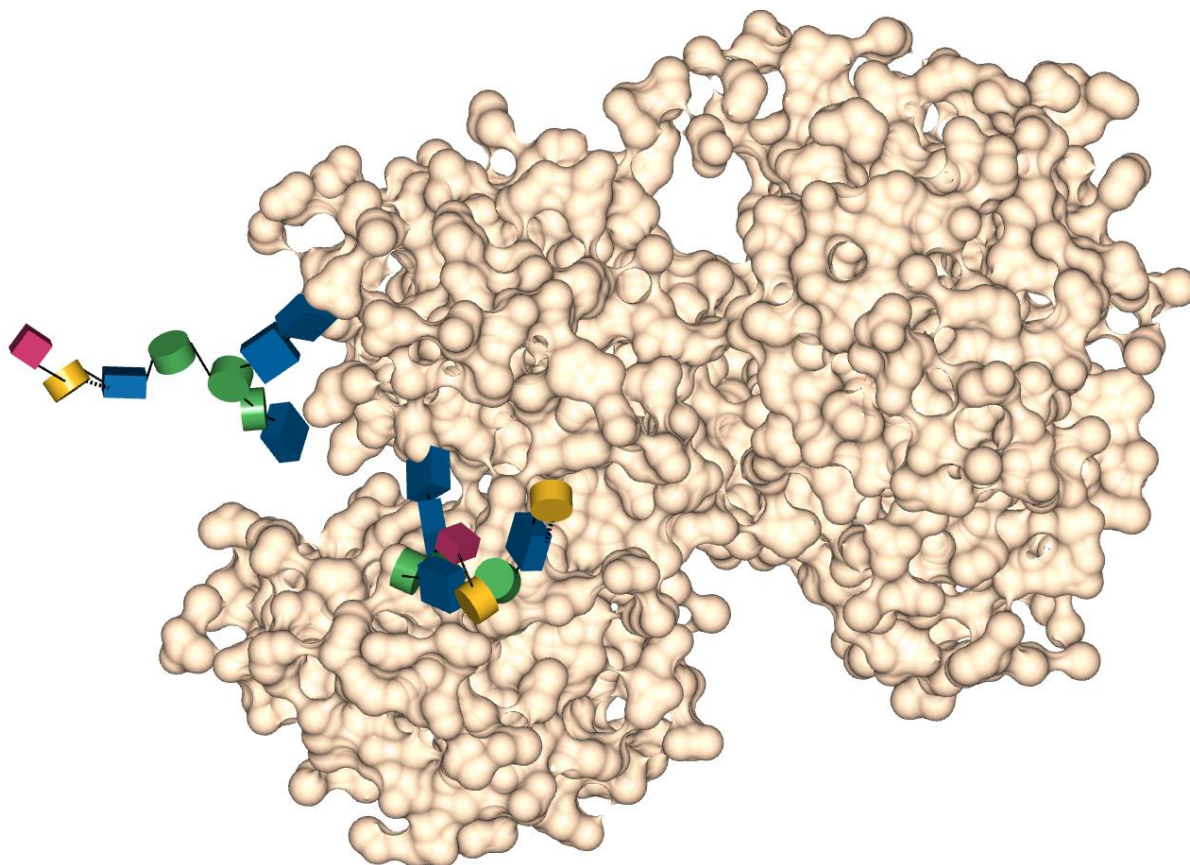
Figure 4.7 Human serum transferrin (hTF) glycan complex (figure: monomer imported in (PDB code: 3V8X), chain B.
hTF with glycans *N*-Linked shows as Glycoblocks (Stuart McNicholas and Agirre 2017). The figure was generated using CCP4mg (S. McNicholas et al. 2011). The observed glycans block *N*-acetyl glucosamine depicted in the Blue square, mannose depicted in the green circle, galactose shown in the yellow circle and fuschia diamond: *N*-acetyl neuraminic acid (sialic acid).

Carbohydrates make up about 6% of the human TF, containing two types of complex oligosaccharide chains having various degrees of branching at Asn413 and Asn611 in the

carboxy-terminal domain of the mature protein (MacGillivray et al. 1983); (Dorland et al. 1977; Satomi et al. 2004); (Spik et al. 1975). At the Asn472 position, a low 2% relative molar abundance of glycosylation has been seen. Glycosylation was recognised in Ser51, Asn432and Asn630 (Kristiansen et al. 2004), Asn432 (Asn-Lys-Ser) and Asn630 (Asn-Val-Thr) (Satomi, Shimonishi, and Takao 2004).

Four branched glycans have also been recognised; however, the most chain of carbohydrates in the hTF has a bi-antennary arrangement and terminates in the sialic acids (Dorland et al. 1977; Satomi et al. 2004). The crystal hTF structures show that carbohydrates are not only flexible but can occur in multiple conformations (Wally et al. 2006b; Yang et al. 2012). However, interestingly it was found that desialylation did not seemingly increase the Tf clearance (Ashwell and Morell 1974). In the blood plasma, the desialylated isoforms can be recognised (de Jong and van Eijk 1988; del Castillo Busto et al. 2005). Murakami and its group observed a sialo-transferrin isoform that lacked the sialyl-galactose component in the brain (Murakami et al. 2019). Generally, this isoform is employed as a clinical cerebrospinal fluid marker (Warnecke et al. 2004).

Lack of glycans seems to have no impact on the function and ability of hTF binding to iron; this is because the primary function of transferrin is to bind and transport iron, which is mediated by specific binding sites on the protein rather than the attached glycans. The binding of iron to transferrin occurs primarily through interactions with specific amino acid residues in the protein structure, particularly in the N- and C-terminal lobes of the molecule, rather than interactions with glycans. The sugars provide slight protection for proteolysis, which is the degradation of proteins by proteolytic enzymes. The presence of glycans can sterically hinder access to susceptible peptide bonds, reducing the susceptibility of the protein to proteolytic cleavage and controlling the hydrophilicity of hTF, which affects its solubility and interactions with other molecules. Glycans increase hydrophilicity by introducing polar groups, such as hydroxyl (-OH) and amino (-NH2) groups, which enhance water solubility and prevent aggregation (Zhao and Enns 2013; van Veen et al. 2004). Carbohydrates in hTF are necessary for interacting with the human transferrin receptor (TFR), also known as TFR1. The glycans on hTF serve as recognition sites for TFR, facilitating the binding and internalisation of hTF-TFR complexes into cells through receptor-mediated endocytosis (Zhao and Enns 2013). However, despite its biological function, hTF

glycans are employed as biomarkers for analysing and indicating alcohol consumption (Arndt, 2001) and carbohydrate deficiency syndromes (Jaeken 1995; Gordon 2000). Human Serum Transferrin was reported to be an extensively glycosylated protein in different locations listed in Table 4.1 the position of the glycans group is in red font, and its background is in yellow.

4.1.6.1.1 Analysis, validation and remodelling of (PDB code: 3V8X)

The *Privateer* software was used to assess the quality of the glycan structures on (PDB code: 3V8X), which is the transferrin structure with the longest deposited *N*-glycans. It was found that the structure contained numerous pyranosides in high-energy conformations (orange backgrounds behind SNFG blocks in Figure 4.8), which are usually associated with modelling errors, particularly in *N*-glycans where the sugars are expected to remain in their chair conformation.
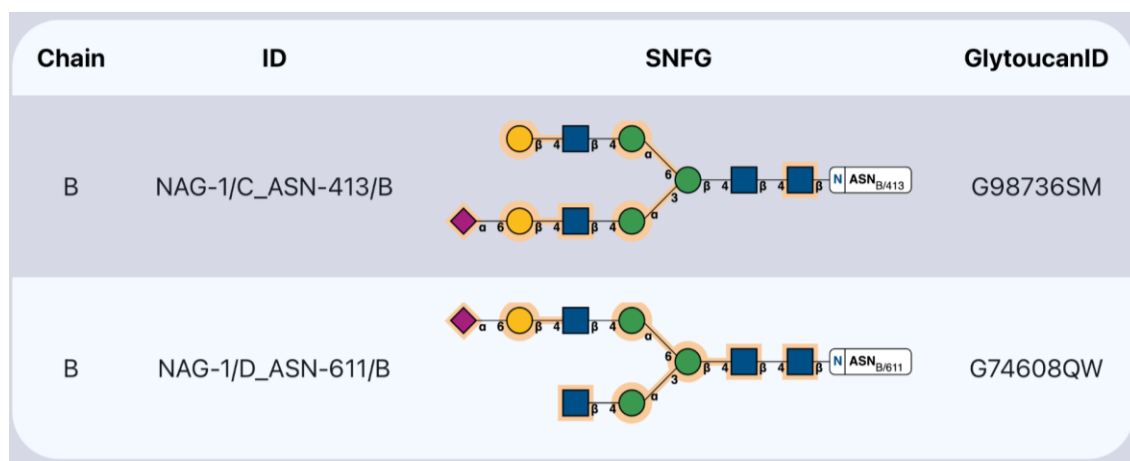
Figure 4.8 Validation of the original 3V8X glycans.
Glycans illustrated using *Privateer* (Agirre, Iglesias-Fernández, et al. 2015) as run through the new web server (Dialpuri et al. 2024). Orange backgrounds denote unusual features in the carbohydrate structures, usually correlated with modelling errors.

The structure was loaded into *Coot*, and manually inspected. A strong discrepancy was found in the glycan attached Asn 611, where the beta-mannose was flipped, cascading errors down the glycan chain due to the different ring-to-ring distances involved in 1-3 an 1-6 linkages. The glycans were deleted, and a new model was produced by hand. The results can be seen in Figure 4.9 (validation) and 4.10 (3D structure).
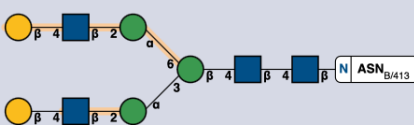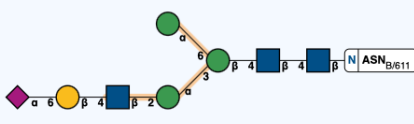
Chain	ID	SNFG	GlytoucanID
B	NAG-870: 32/B_ASN-413: 32/B		G36191CD
B	NAG-863: 32/B_ASN-611: 32/B		G06209KS

Figure 4.9 Validation of the remodelled 3V8X glycans.

Glycans illustrated using as run through the new web server (Dialpuri et al. 2024). While some of the glycosidic linkages continue to be in unusual conformations according to *Privateer*'s internal torsion database, all the issues affecting the pyranosides have been resolved owing to the changes in glycan structure providing a better match to the electron density map. Work done in collaboration with Phuong Thao Pham (YSBL, University of York), who advised on suitable changes to the glycan structures.

The output structure was refined using 20 cycles of Refmac5 with default regular restraints and no torsion angle restraints and re-validated using the *Privateer* web server (Dialpuri et al. 2024).

Asn 413

Asn 611

Original & Remodlled
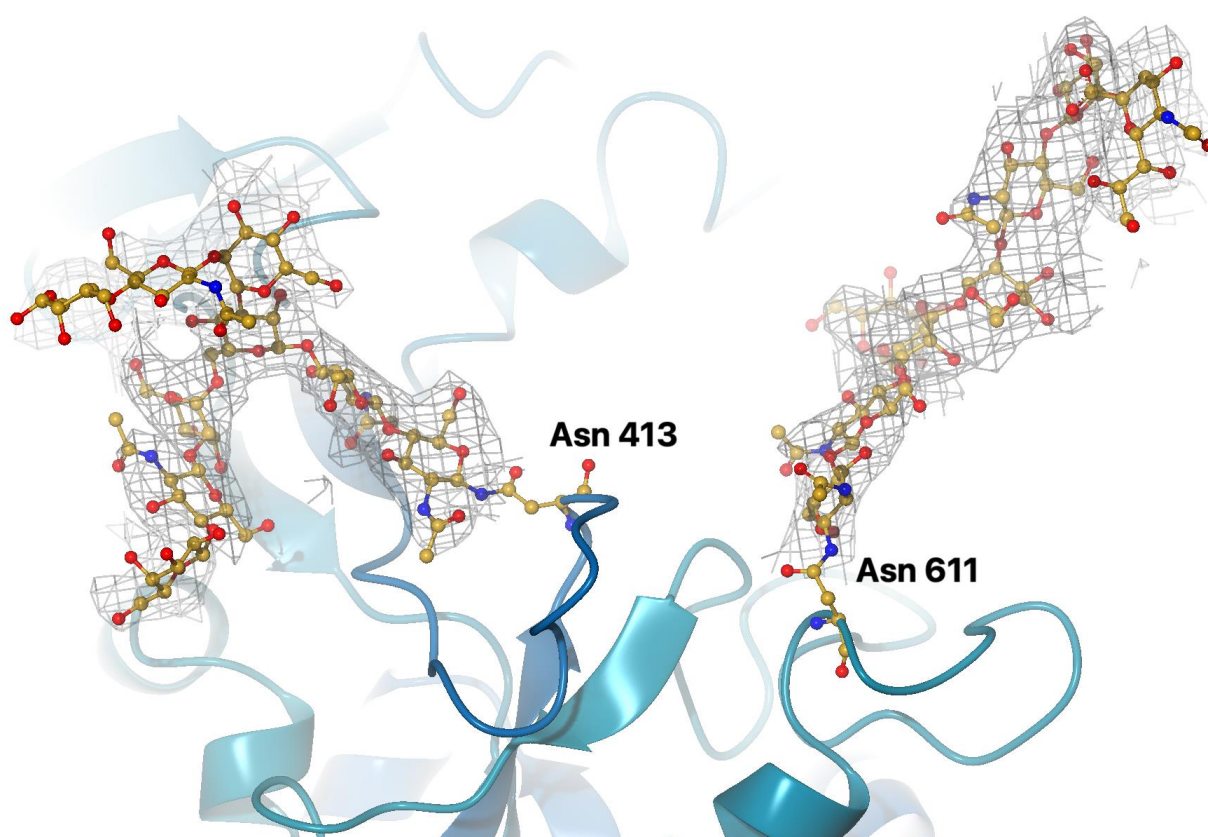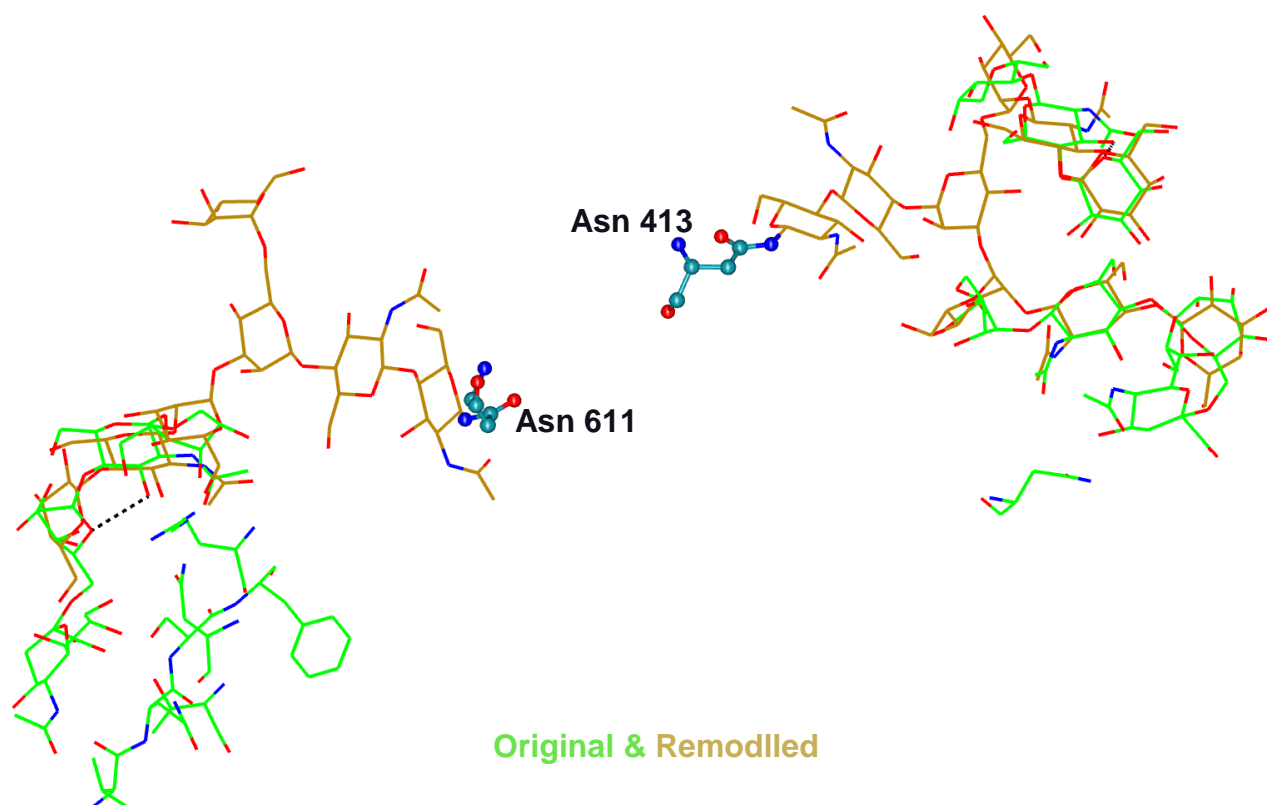
Figure 4.10 Top: the remodelled 3D structure of the *N*-glycans in (PDB code: 3V8X). 2mFo-DFc electron density coming out of the refinement process was contoured at 1 sigma. Bottom: overlay of original vs remodelled.

The remodelled structure, validation reports and torsional analyses have been uploaded to Zenodo along with the rest of the supplementary data contained in this thesis: 10.5281/zenodo.8384171

Table 4.1 List of some hTF Glycanase positions, sequence and method of their analysing methods.

Glycosylated residue	Analysing method	Description	Sequence	Reference
Ser51		*O*-linked (GalNAc...) serine	SDGPSVACVK	Uniprot (Coudert et al. 2023)
Asn432	Liquid chromatography and tandem mass spectrometric (LC-MS/MS)	*N*-linked (GlcNAc...) (complex) asparagine	YNKSDNCEDT	(Kristiansen et al. 2004)
Asn630	(LC-MS/MS)	*N*-linked (GlcNAc...) (complex) asparagine	RQQQHLFGSN	(Kristiansen et al. 2004)
Asn413, Asn611	Nuclear magnetic resonance (NMR) spectroscopy	Carboxy-terminal domain of the mature protein asparagine		(Dorland et al. 1977; Satomi et al. 2004)
Asn413, Asn611	Electron density of crystal	Asn611 in the C lobe of FeNFeC-hTF, indicative of the connection of two glycan moieties (NAG, *N*-acetylglucosamine) to these residues		(Noinaj et al. 2012)

The following sequence, Figure 4.11, is amino acid single letter codes of hTF. N (Asn432, Asn630) and S (Ser51) are amino acid single-letter codes that refer to the three-letter amino acid coloured in red font and its background in yellow, which identifies the glycan positions. N lobe sequence is represented in blue and C lobe in black.

MRLAVGALLVCAVLGLCLAVPDKTVRWCAVSEHEATKCQSFRDHMKSVIP**S**DGPSVACVKKASYLDCIR
AIAANEADAVTLDAGLVYDAYLAPNNLKPVVAEFYGSKEDPQTFYYAVAVVKKDSGFQMNQLRGKKSCH
TGLGRSAGWNIPIGLLYCDLPEPRKPLEKAVANFFSGSCAPCADGTDFPQLCQLCPGCGCSTLNQYFGYS
GAFKCLKDGAGDVAFVKHSTIFENLANKADRDQYELLCLDNTRKPVDEYKDCHLAQVPSHTVVARSMGG
KEDLIWELLNQAQEHFGKDKSKEFQLFSSPHGKDLLFKDSAHGFLKEVPPRMDAKMYLGYEYVTAIRNLR
EGTCPEAPTDECKPVKWCALSHHERLKCDEWSVNSVGKIECVSAETTEDCIAKIMNGEADAMSLDGGFV
YIAGKCGLVPVLAENY**N**KSDNCEDTPEAGYFAIAVVKKSASDLTWDNLKGKKSCHTAVGRTAGWNIPMG
LLYNKINHCRFDEFFSEGCAPGSKKDSSLCKLCMGSGLNLCEPNNKEGYYGYTGAFRCLVEKGDVAFVKH
QTVPQNTGGKNPDPWAKNLNEKDYELLCLDGTRKPVEEYANCHLARAPNHAVVTRKDKEACVHKILRQ
QQHLFGS**N**VTDCSGNFCLFRSETKDLLFRDDTVCLAKLHDRNTYEKYLGEEYVKAVGNLRKCSTSSLLEAC
TFRRP

Figure 4.11 The complete sequence of human serotransferrin, it was exported from the UniProt platform (P02787).

4.1.7 Transferrin's iron complexes

Complexes are the subject of our investigation since they may change the protein's overall structure and possibly stabilise it.

Transferrins are a group of bi-lobed iron-binding proteins essential for binding ferric iron and maintaining it in solution, which regulates the body's requirement of this crucial metal (Wally et al. 2006).

Human serum transferrin is made in the liver and released into the bloodstream; it absorbs Fe(III) from the gastrointestinal tract and distributes it to cells that need iron by connecting to specific transferrin receptors (TFR) there. The endosome releases iron once the hTf-TFR complex is picked up by receptor-mediated endocytosis (Klausner et al. 1983). Iron attachment and release cause significant conformational changes, in which two subdomains in each lobe rigidly twist around a hinge to shut or open (Wally et al. 2006).

Iron-free hTF (apo-hTF) remains linked to the TFR at low pH, which is necessary for hTF's re-utilisation. Apo-hTF is released to take up new iron when the apo-hTF TFR complex is brought back to the cell surface.

Due to its dual ability to serve as an electron acceptor and donor, iron is a necessary component for all living things and is utilised in several physiological and metabolic processes. Iron may be found in many oxidation states (ranging from 2 to 6), with +2 and +3 being the most prevalent. Iron is vital in many biological processes, including biological oxidations, oxygen transport, and energy generation, because of the simplicity of interconversion between the ferrous ($Fe2+$) and ferric ($Fe3+$) forms. While reduced ($Fe2+$) or free iron in the cell is poisonous because of its propensity to produce detrimental reactive oxygen species through Fenton chemistry, which can damage cellular components such as lipids, nucleic acids, and proteins, $Fe3+$ is not soluble in water in the unbound form. To survive, each cell must be able to obtain and maintain the required quantity of iron, which can be poisonous in excess. Thus, living organisms have evolved sophisticated processes to control secure and regulated iron storage and transportation (Abdizadeh et al. 2017).

There are several ways to take iron supplements medically, but they all have the same basic design: an iron core encased in a carbohydrate covering. The reticuloendothelial system must

digest these nanoparticles to release iron, which is then taken up by the iron-binding protein transferrin and dispersed throughout the body, with most of the iron going to the bone marrow. Due to its vigorous redox activity, free iron, which may be harmful to cells and tissues, is in danger of exposure during this process. In ferric pyrophosphate citrate (FPC) one iron atom is complexed with one pyrophosphate and two citrate anions. This new parenteral iron creation varies from traditional intravenous iron formulations (Pratt et al. 2018).

4.1.7.1 Ferric pyrophosphate citrate ($C_{18}H_{24}Fe_4O_{42}P_6$) bonds with hTF

Iron may be transferred from ferric pyrophosphate (FPC) to apo transferrin. According to kinetic studies, FPC contributes iron to apo-transferrin quickly (Feraheme, 2022).

A product that replaces iron in soluble form is ferric pyrophosphate citrate. Free iron can stimulate the generation of free radicals, lipid peroxidation, and interactions with other forms of iron in plasma, all of which have negative consequences. Pyrophosphate and citrate have a substantial complexing effect on the ferric ion (Gupta et al. 1999).
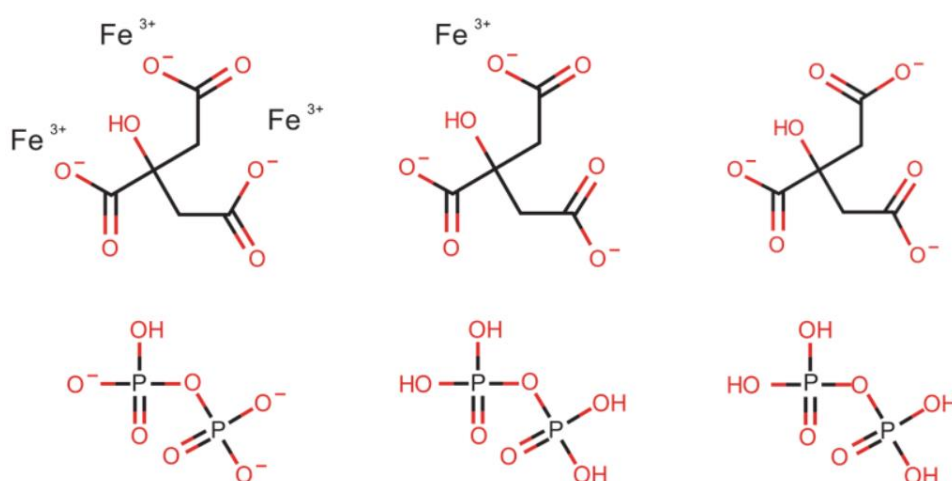
Figure 4.12 A chemical diagram based on a crystal structure of Ferric Pyrophosphate Citrate.
(National Center for Biotechnology Information ,2024a).

The oligomeric complex of Iron (III)-citrate-pyrophosphate, which lacks carbohydrate shell, transports iron to plasma transferrin effectively. Unique iron (III)-citrate-pyrophosphate ternary complex oligomeric structure makes up ferric pyrophosphate citrate (FPC), which is stable in aqueous solution for a considerable amount of time (Yang et al. 2012).

4.1.7.1.1 Crystal structure of human transferrin bound to FPC (6CTC)

In 2018 a publication, ferric pyrophosphate citrate (FPC), was studied. Iron atoms are complex, with one pyrophosphate and two citrate anions making up ferric pyrophosphate citrate (FPC). As a result of this investigation, they demonstrated that 1) FPC can directly transfer iron to apo transferrin. 2) According to kinetic investigations, FPC contributes iron to apo-transferrin with quick binding kinetics. 3) FPC can also contribute iron to the transferrin structure's two iron-binding sites. Examining the iron-binding sites reveals that the iron atoms in both locations are completely encapsulated and have formed connections with the protein's side chains as well as the pyrophosphate and carbonate anions. In conclusion, FPC may directly and quickly contribute iron to transferrin without endangering cells and tissues by exposing them to the harmful effects of free, redox-active iron (Pratt et al. 2018).

Human transferrin coupled to FPC (PDB code: 6CTC) is shown at a 2.6 Å resolution co-crystal structure, with the N- and C-terminal lobes denoted by grey and green ribbons, respectively, shown in Figure 4.13. As both the N and C lobe iron-binding sites contain an iron atom, it is possible that FPC may provide iron in both lobes of human apo-transferrin. The structure of FPC-bound transferrin, as shown in Figure.4.13 A, B, more closely resembles that of iron-sulphate-bound transferrin Figure 4.15, B,C, and bismuth-NTA- Figure 4.15,E (Yang et al. 2012). than either the fully open, apo-hTF form or the fully closed ferric-hTF form (Noinaj et al. 2012). On the electron density map, close to the site where iron binds in the N lobe, there was a noticeable differential density peak that suggested the presence of an iron atom.
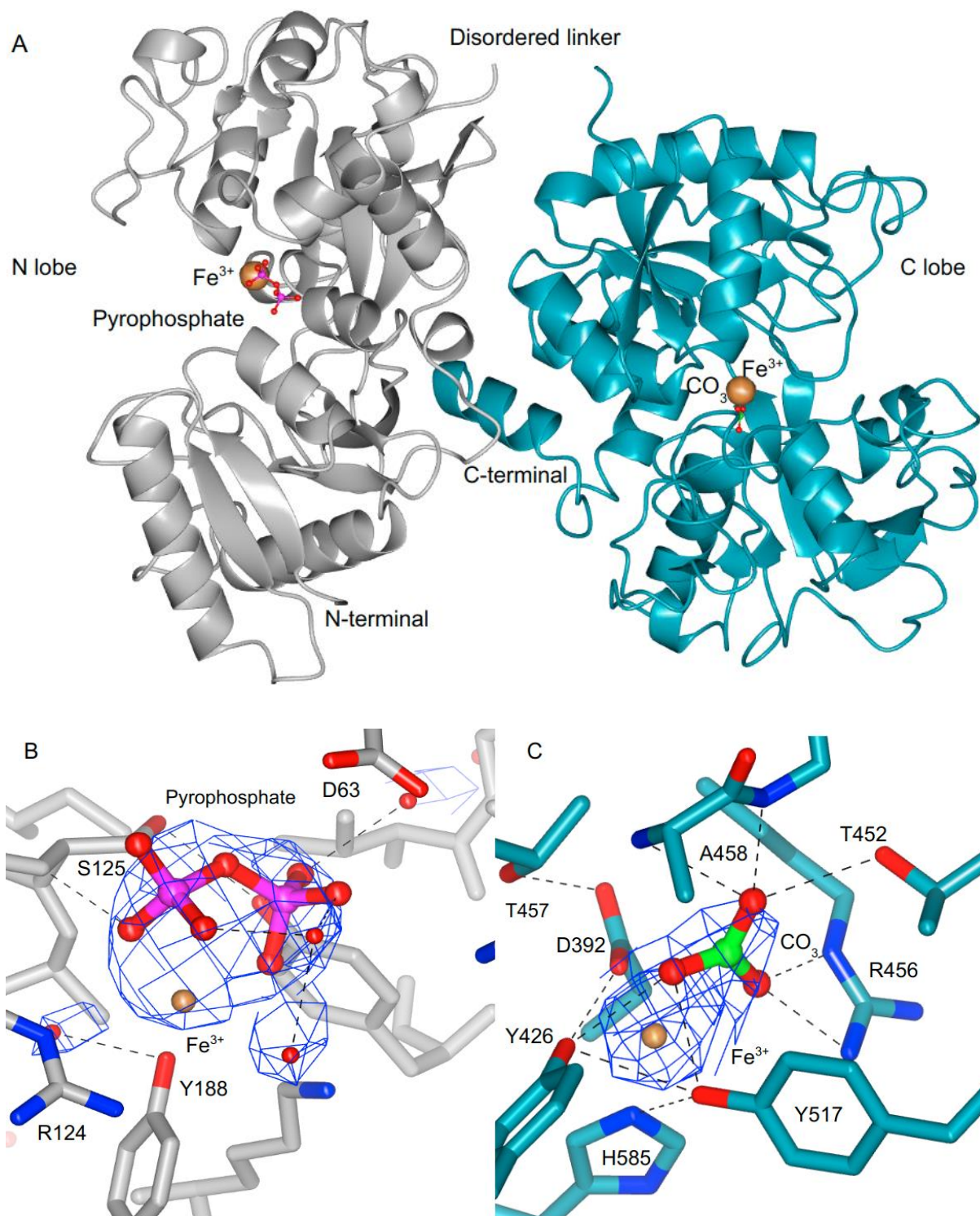
Figure 4.13 Crystal structure of FPC-bound transferrin.
(A) The N- and C-terminal lobes are depicted by grey and dark cyan ribbons, respectively, in a co-crystal structure of human transferrin coupled to FPC at 2.6- Å resolution. Orange spheres represent iron atoms. Pyrophosphate and carbonate are shown in sticks rendering. The linker between the N lobe and C lobe was disordered and is labelled; N and C termini are also labelled. (B) The N lobe is bound to iron and pyrophosphate. C) The C lobe bonded to iron and individual amino acids are labelled (Pratt et al. 2018). The figures were generated using CCP4mg (S. McNicholas et al. 2011)

Figure 4.13, B shows the N lobe bound to iron and pyrophosphate. The pyrophosphate molecule establishes three coordinate linkages to the iron atom in addition to direct hydrogen bonds with the backbone amide of S125, the side chain of Y95, water-mediated hydrogen bonds with the backbone amide of A64, and the carbonyl of P247 (Pratt et al. 2018).

Figure 4.13, C represents the C lobe bound to iron and carbonate. The position of pyrophosphate efficiently blocks D63 and H249 in the N lobe from reaching the iron atom, allowing for the construction of an encapsulated iron. The C lobe contains a well-encapsulated iron atom that is coordinated by D392, Y426, Y517, H585, and carbonate. One of the strongest electron density characteristics is the N-lobe iron density (3.5 in the 2mFobs-DFc map), although the C lobe has a stronger affinity for iron. The most robust electron density feature (10 in the 2mFobs-DFc map) is the C lobe iron density (Cannon and Chasteen 1975; Princiotto and Zapolski 1975).

4.1.7.1.2 Partially opened structures of hTF

It is widely known that hTF transfers metal through an endocytosis process that is controlled by the hTF receptor (TfR) and that the tertiary structure of hTF is essential for its identification by the transferrin receptor. Metal binding and release accompany the modification of hTF's tertiary structure; specifically, hTF occludes the lobe cleft following metal binding and opens up with metal dissociation. Due to TfR's relatively high binding affinity for ferric-bound hTF (holo-hTF) than it is for the non-ferric-bound form (apo-hTF) in the extracellular region (pH 7.5), which results in the internalisation of ferric-bound hTF and the discharge of apo-hTF on the cell surface, thus a metal-triggered conformational change is also crucial for hTF's biological turnover. TfR binds to apo-hTF more preferentially than holo-hTF when in the endosome (pH 5.6), which causes apo-hTF to be transported with TfR8 from the endosome to the extracellular membrane. In such a procedure, hTF is recycled while iron is given to the cells (Pratt et al. 2018).

4.1.7.1.3 Ferric ammonium citrate ($C_6H_{11}FeNO_7$)

Having a mild ammonia scent, ferric ammonium citrate is a solid that ranges in colour from yellowish brown to crimson. In water, it is soluble. Environment-related risk is the main danger. To stop it from contaminating the environment, immediate action must be taken. It is employed as an additive in feed as well as in medicine and the creation of blueprints.

Influenza A, HIV, Zika, and enterovirus 71 (EV71) infections were prevented by the iron salt ferric ammonium citrate (FAC). Other iron salts and citrates did not show viral suppression, indicating that both the iron ion and citrate ion were necessary for the antiviral properties of FAC. FAC prevented viral release from endosomes and induced viral fusion to prevent viral infection. Distinct iron-dependent cell death was linked to FAC-induced liposome aggregation and intracellular vesicle fusion. Research demonstrating the unique antiviral action of FAC and suggesting the therapeutic potential of iron in managing viral infections was published in 2018 (Wang et al. 2018).
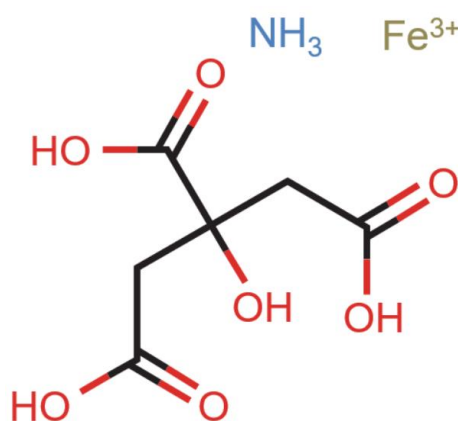
Figure 4.14 A chemical structure of Ferric ammonium citrate.
(National Center for Biotechnology Information, 2024b).

In 1996 a paper was published to investigate the syncytiotrophoblast ability to control transferrin receptor (TfR) production in response to changes in the maternal iron supply. The primary trophoblast cell culture served as the model. Desferrioxamine (DFO), diferric transferrin (hTf-2Fe), and ferric ammonium citrate (FAC) mediums were used to cultivate the trophoblast cells that were isolated from term human placentas. TfR synthesis was not altered by FAC. TfR synthesis was obviously stimulated by iron restriction caused by DFO. These findings demonstrate that the rate of TfR synthesis is altered by the developing trophoblast cells in response to perturbations in the iron supply (mediated by transferrin). Their findings revealed that the placenta had the ability to alter its ability to absorb iron temporarily (Kroos et al. 1996).

4.1.7.1.4 Ferrous Ammonium Sulphate

When lobes are loaded with Fe^{3+}, they are more selective when they are loaded with Fe^{2+} (which binds as Fe^{3+}), Al^{3+}, Ga^{3+}, and Bi^{3+}. These metal ions show comparable shift

modifications in the Met residues, indicating that they cause comparable protein structural changes (H. Sun et al. 1998). Transferrin exhibits higher affinity and selectivity for Fe^{3+} ions due to the specific coordination chemistry between the protein and the metal ion (Baker, 1994; Li, Sadler and Sun, 1996). This selectivity ensures that iron transport remains efficient and specific, maintaining iron homeostasis within the body (Knutson 2017).

4.1.7.1.5 Crystal structures of diferric hTF (3QYT) and bismuth bound hTF (4H0W)

It is well known that proteins constantly undergo dynamic changes to sustain their unique biological roles. However, the absence of structural information on intermediate conformers makes it unclear how TF proteins undergo conformational changes following metal binding and separation. Although a structure for the TF proteins would be crucial for understanding how they alter conformation, they exist in an in-between state where they assume a "partially-opened" shape. To this purpose, scientists provide the crystal structures of bismuth-bound hTF (PDB code: 4H0W) and diferric hTF (PDB code: 3QYT). The two hTF structures, representing two crucial protein conformers throughout the metal release process, prominently display distinct "partially-opened" conformations in the N lobe for the first time. A structural foundation for the possible use of transferrin as a metallodrug delivery "vehicle" in healthcare is provided by the structure of bismuth-hTF (PDB code: 4H0W), which is also significant since it gives direct information on Bi (III) coordination (Yang et al. 2012).

4.7.2.2 Conformational changes in the N-lobe of hTF

It has traditionally been thought that the uptake and release of metal ions by hTF are pH-dependent processes and accompanied by the transposition of lobe opening and closure, *i.e.*, from "fully-opened" to "fully-closed" state upon iron binding at extracellular pH (pH 7.5) and from "fully-closed" to "fully-opened" conformation upon iron release at endosomal pH (pH 5.6). Surprisingly, the N lobes of the diferric and mono-bismuth bound forms of hTF adopt a specific "partially-opened" conformation, which is distinct from both the apo-hTF with a "fully-opened" conformation and the isolated iron-bound recombinant N lobe of hTF (Fe-hTF/2N) and a recently reported diferric hTF with a "fully-closed" (Yang et al. 2012).

4.7.2.3 Interlobe communication in diferric hTF (3QYT) and bismuth bound hTF (4H0W)

The interlobe interactions also have an impact on the tertiary structure of hTF. The N and C lobes of hTF are connected by a peptide (C[331]PEAPTNEC[339]), where two disulfide bonds are formed, serving as the two distinct iron-binding domains (Cys331-Cys137 and Cys339-Cys596).

In conclusion, the first crystal structures of differic and bismuth-bound transferrin, with their N lobes displaying distinctive "partially-opened" conformations. In contrast to the C-four lobe's highly conserved binding residues, the N-N2-subdomain lobe only has operated or two tyrosines that might be involved in metal binding. hTF structures provide a foundation for comprehending the process of metal/metallodrug absorption and release in transferrin and constitute an essential protein conformer (Yang et al. 2012).
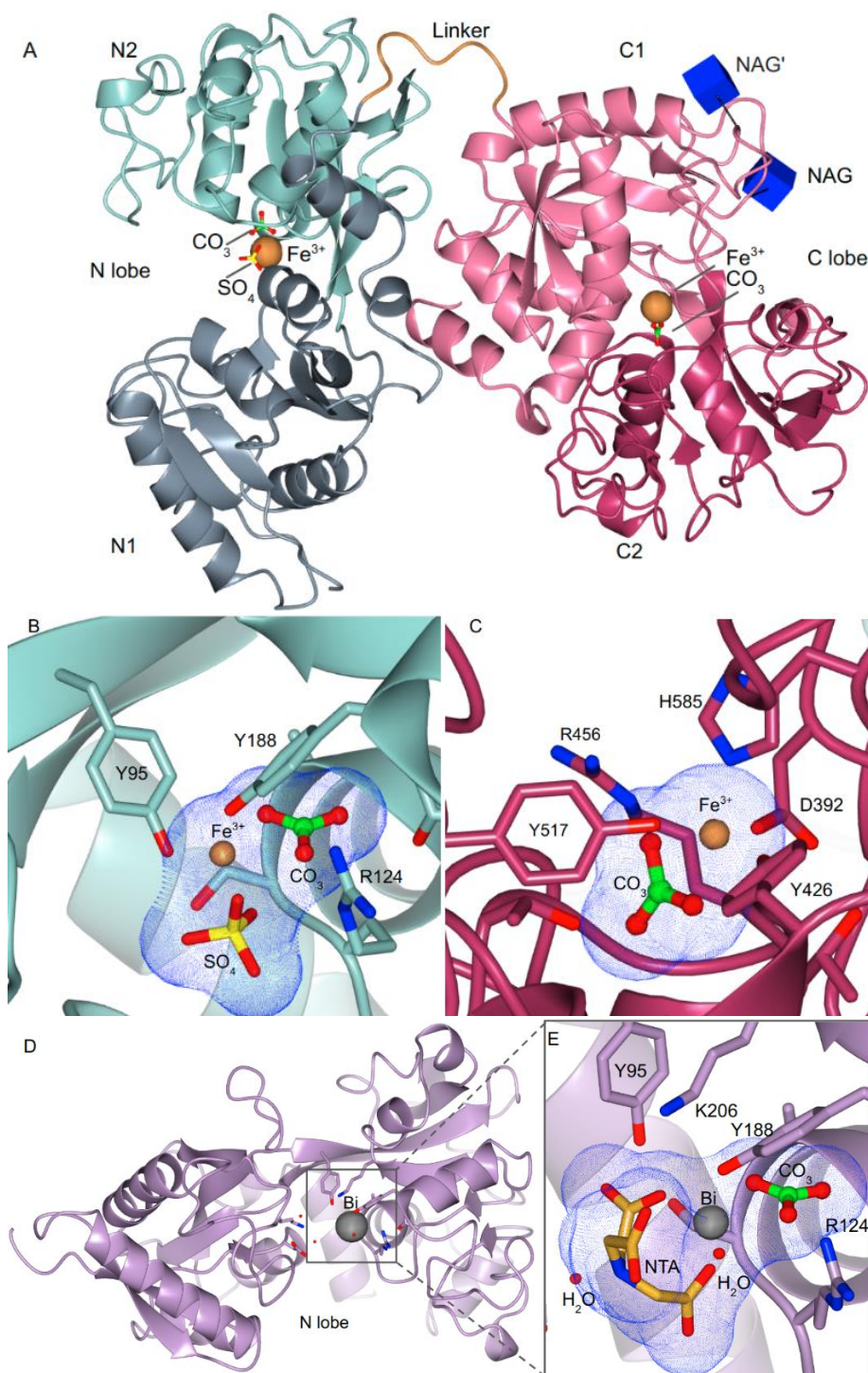
Figure 4.15 Interlobe communication in diferric hTF (3QYT) and bismuth bound hTF (4H0W).

(A) hTF (PDB code: 3QYT), FeNFeC-overall hTF's structure is shown as a ribbon, with the subdomains N1 and N2 coloured slate grey and sea green, respectively. The peptide linker is peru, C1 is pale violet-red, and C2 is maroon. Peru sphere representations of Fe (III) ions are shown. The NAG and NAG' *N*-acetylglucosamine moieties are shown as blue Glycoblocks representations (Stuart McNicholas and Agirre 2017). (B) Coordination of the Fe (III) in FeNFeC-N lobe. A surface including the Fe (III), CO_3^{2-} and SO_4^{2-} were modelled by normal blue dots. (C) The C lobe of FeNFeC-hTF contains an iron-binding core. A normal blue dots surface applied on the Fe (III) and CO_3^{2-}.

(D) hTF (PDB code: 4H0W), BiNFeC-anomalous hTF's atomic model. Tyr188 and Bi (III) residues are shown as stick and spherical models, respectively, while the hTF backbone is depicted as a lilac ribbon. E) Bi (III) coordination in the N lobe of BiNFeC-hTF and a normal blue dots surface applied on the Fe (III), NTA and CO_3^{2-}. The Tyr95 side chain is 9.9 Å distant from the Bi(III) (Yang et al. 2012). The figures were generated using CCP4mg (S. McNicholas et al. 2011).

4.2 Results and discussion of human serum transferrin

4.2.1 Purification of the apo-hTF and characterization of the yield

4.2.1.1 First step of purification of apo-hTF using Ion exchange chromatography (IEX)

Human transferrin was obtained from Sigma-Aldrich (cat no T4382-1G) as a quick way to obtain workable quantities for what was to become a model system; alas, the purity of the protein needed to be assessed in detail in order to evaluate the suitability of the source. The sample was soluble in the HEPES buffer at room temperature. HEPES is a suitable buffer system for preserving the structure and conformation of transferrin due to its superior buffering capacity and the pK_a value of 7.3 being close to the physiological pH. It is beneficial for maintaining physiological pH conditions, which is important for preserving the structure and conformation of transferrin. The protein was successfully purified using two purification steps: The Ion exchange chromatography (IEX) method (Pratt et al. 2018). The hTF was eluted as a major peak with an apparent molecular weight greater than 70 kDa from a Q-Sepharose FF ion exchange column (Figure 4.16). Eluates of IEX were pooled, concentrated and stored overnight in the fridge for the second purification step.

Further purification was performed by size exclusion chromatography (SEC) because purification by IEX yielded material that contained impurities with molecular weights around 55kDa, in addition to the ~80 KDa target protein.

Figure 4.16 First step of purification of the apo-hTF protein.
(A) Chromatogram illustrating the elution of apo-hTF from Ion exchange chromatography (IEX) using a Q-Sepharose FF ion exchange column; eluting proteins are monitored by the UV absorbance at 280 nm (A280). (B) Two 10 % SDS-PAGEs of gel profiles of hTF elution, retention volume of fractions was taken from a high peak between ~40-50 mL, hTF is indicated in both gels. Lane 1, low range molecular weight marker; in left gel, in lanes 2-5 eluted protein fractions, right gel in lanes 2-4 eluted protein fractions and LS lane is the loading sample. All eluted fractions of hTF have an expected MW of 77kDa and have been used for further investigation and the second purification step.

4.2.1.1.1 Results MALDI-MS/MS gel

Protein after purification was expected to become homogeneous but Coomassie Brilliant Blue (CBB), stained SDS-PAGE shows purified hTF at the expected molecular weight of ~ 80 kDa and two bands were migrated around 55 and 50 kDa. The MALDI-MS/MS protein identification procedure has been completed for three bands the spectra were searched against a combined human subset of SwissProt (20242 sequences;11289677 residues). Mass spectrometry analysis indicated all protein bands (80 kDa, 55 kDa and 50 kDa). All bands give a match to Serotransferrin, suggesting that the lower bands result from a breakdown of the higher band. This result has highlighted further heterogeneity due to (unpredictable) cut fragments.
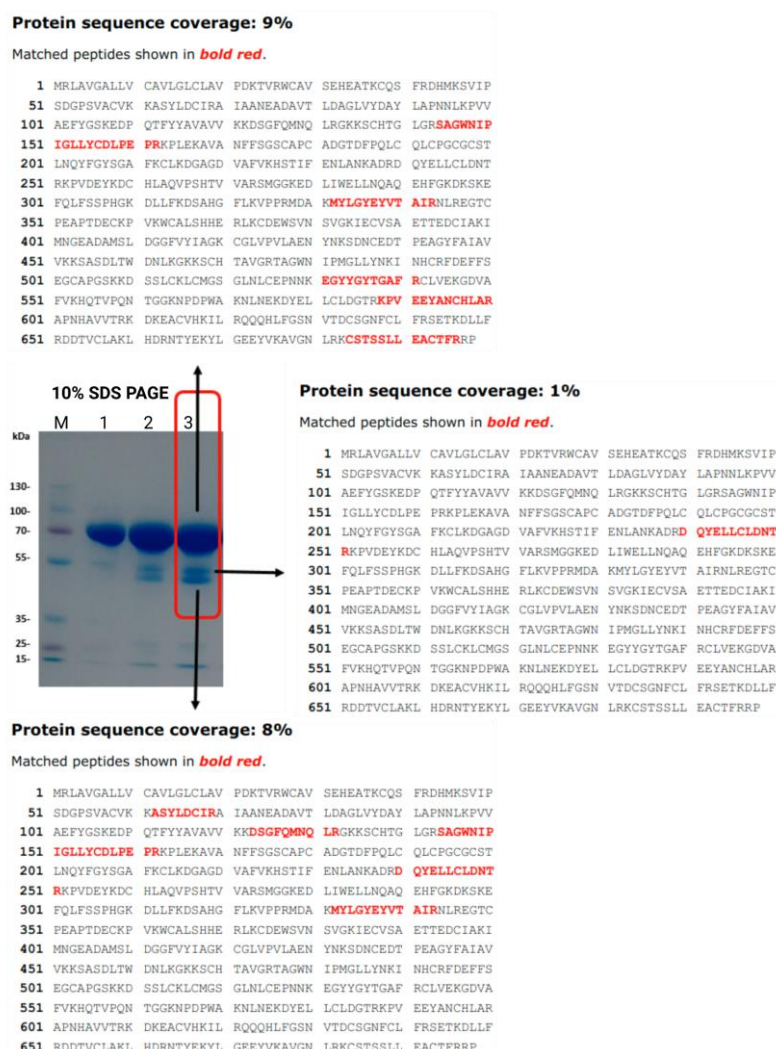
Figure 4.17 Protein bands of the purified apo-hTF protein.
SDS-PAGE of the hTF protein which were purified using Ion exchange chromatography. Protein bands were sent for peptide fingerprinting using MALDI-MS/MS. Mass spectrometry sequence coverage of tryptic digests of three protein bands were shown. Identified peptides are shown in bold red. Sequence and numbering are for the wild type of Serotransferrin, human serum transferrin.

The exact molecular weight of an *N*-glycosylation site attached to human transferrin, such as an entire oligosaccharides chain branch at Asn413, would be approximate.

4 *N*-Acetylglucosamine: 4 × ~221.21 Da = ~884.84 Da

3 Mannose: 3 × ~180.16 Da = ~540.48 Da

2 Galactose: 2 × ~180.16 Da = ~360.32 Da

1 Sialic acid: 1 × ~309.26 Da = ~309.26 Da

Total molecular weight for one glycan = 884.84 Da + 540.48 Da + 360.32 Da + 309.26 Da = ~2094.9 Da ~2kDa (2 *N*-glycans, total: 4 kDa)

Unlike *N*-glycosylation, which often involves a common core structure, *O*-glycosylation can vary more widely. However, a typical *O*-glycan structure includes

Total molecular weight for one typical monosaccharides *O*-glycan

N-Acetylgalactosamine: 1 × 203.19 Da = 203.19 Da

Galactose: 1 × 180.16 Da = 180.16 Da

Sialic acid: 1 × 309.26 Da = 309.26 Da

Total molecular weight = 203.19 Da + 180.16 Da + 309.26 Da = 692.61 Da ~0.692 kDa (one *O*-glycan, ~0.7 kDa)

4.2.1.2 Second step of purification apo-hTF using size exclusion chromatography (SEC)

The second purification step was conducted to produce a homogeneous sample. The UV absorbance readings for SEC showed an A280 of ~1, indicating the presence of the protein within the purified sample (Figure 4.18, A). Protein was eluted, and fractions were collected. Elution of apo-hTF (Figure 4.18, B) was confirmed by SDS-PAGE, and fractions of lanes 2-6 corresponding to the peak, were used for imaging with a transmission electron microscope (TEM), which showed that the purified hTF molecules were too small to visualise their specific shape, (Figure 4.18, C).
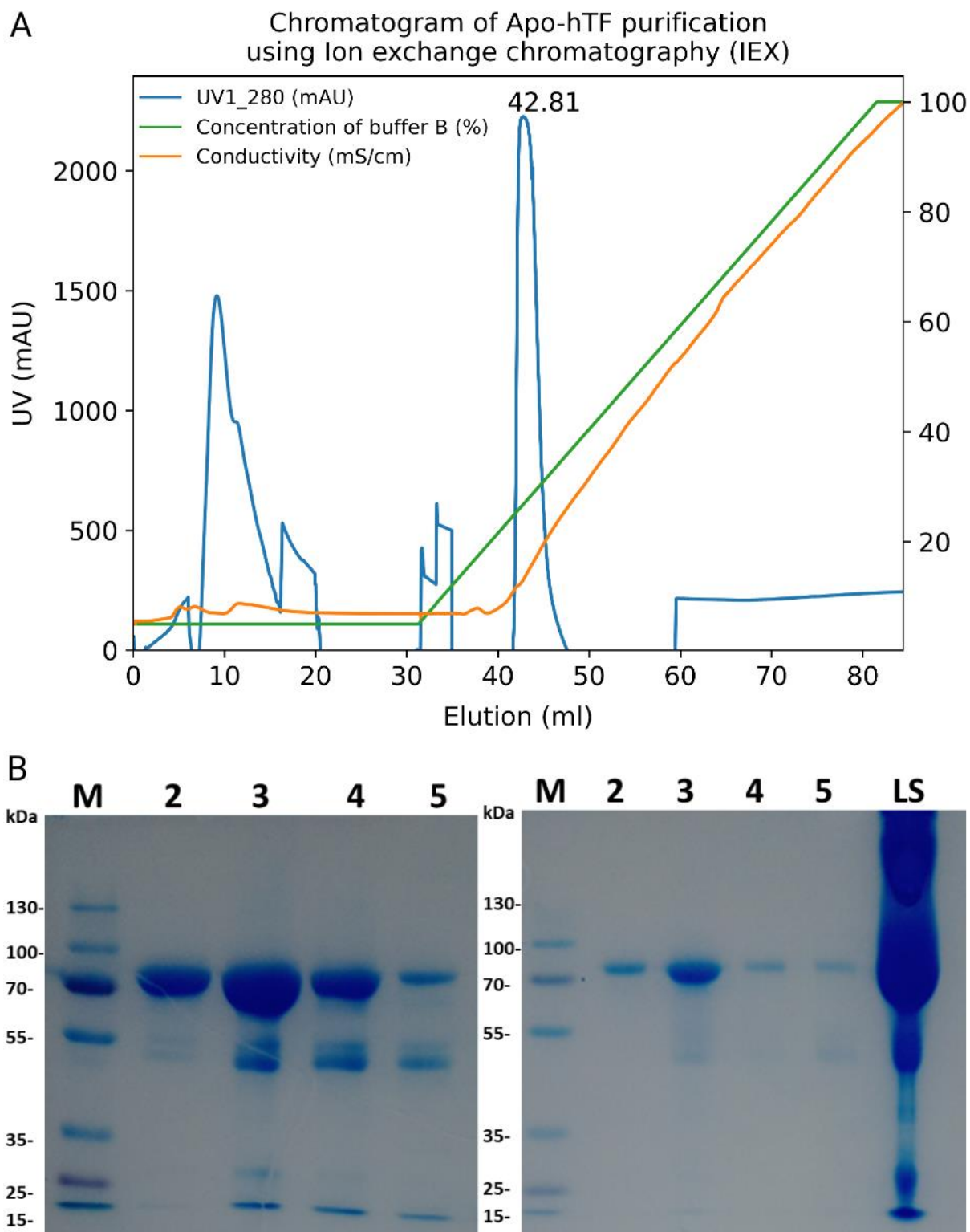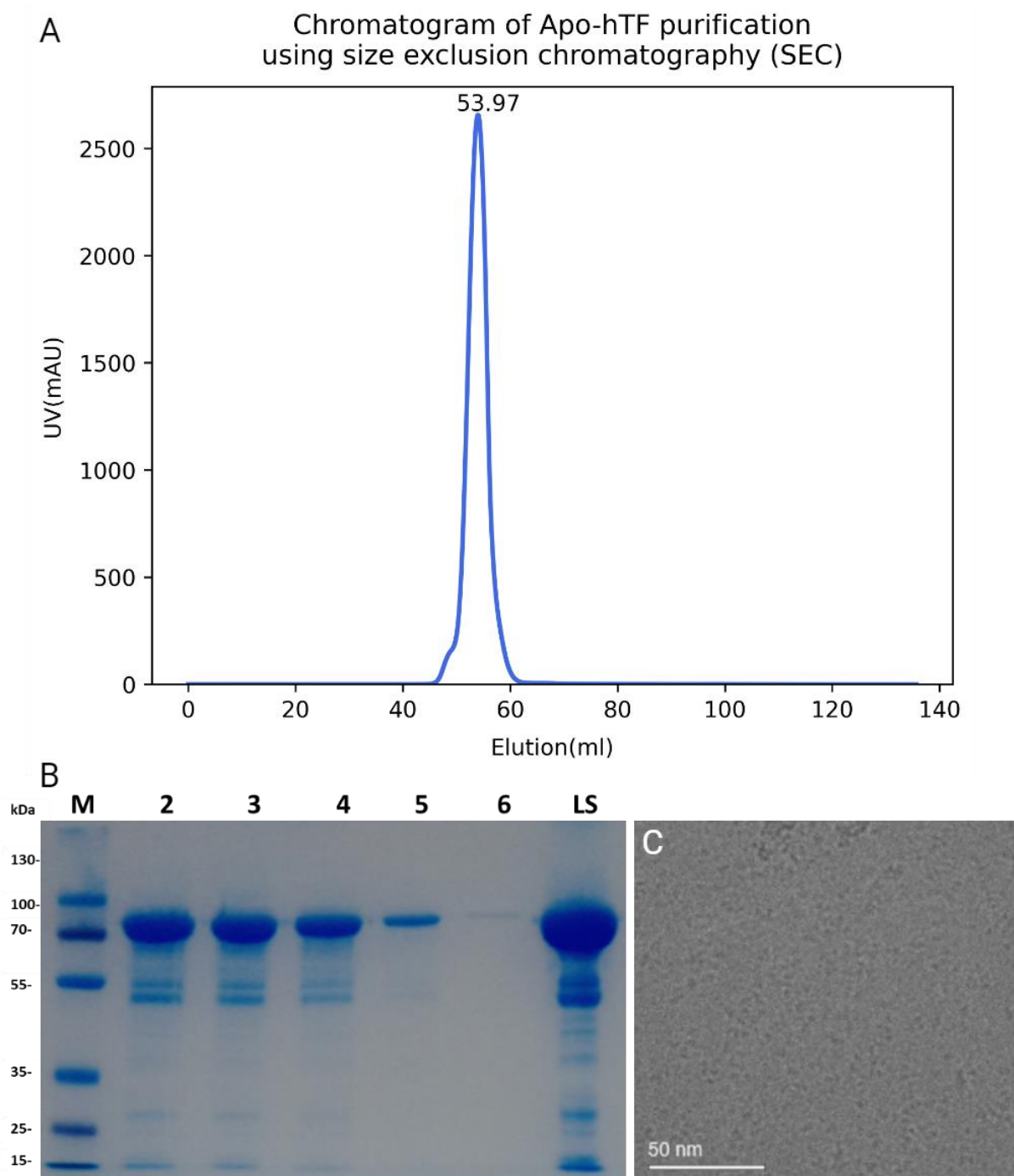
Figure 4.18 Second step of purification of the apo-hTF protein.
(A) Chromatogram illustrating the elution of apo-hTF from size exclusion chromatography (SEC) using a Superdex 75; Hiload 16/600 column; the UV absorbance determines protein at 280 nm (A280). (B) 10 % SDS-PAGE profile of some protein fractions collected and corresponding to the highest protein peak after SEC elution between ~42-60 mL. Lane 1, low range molecular weight marker; lanes 2-6 eluted protein fractions and lane 7 LS the loading sample. (C) A screened TEM micrograph image of the SEC purified hTF protein of defocus set at -2 µm and scale bar, 50 nm.

4.2.2 Formation and purification complexes of hTF-Ferric ammonium citrate and Ferrous ammonium sulphate

4.2.2.1 Reconstitution of hTF iron complexes in vitro

In order to obtain a stable structure of the glycosylated hTF that connects to the iron salt complexes, an in vitro assembly procedure was carried out by binding the purified protein to Ferric ammonium citrate and Ferrous ammonium sulphate, respectively.

4.2.2.1.1 Purification the hTF iron complexes

To obtain hTF protein complex suitable for structural determination using cryoEM, additional purification was done using SEC. To improve sample homogeneity and remove unbound iron, the hTF complexes were purified on Akta pure utilising Superdex 200, Hiload 16/600 columns. SDS-PAGE analysis revealed that the hTF complexes were segregated based on molecular weight (80kDa) as the apo-hTF dose (Figure 4.19) and (Figure 4.20).

The protein concentration of the hTF ferric ammonium citrate and ferrous ammonium sulphate eluted from the SEC was about 4.7 mg/mL and 3.5 mg/mL, respectively. However, reducing the protein solution volume of each of the different fractions using a centrifugal protein concentrator increased the protein concentration to reach the desired concentration before preparing grids.

A Chromatogram of the complex the hTF-Ferric ammonium citrate using size exclusion chromatography (SEC)

B

C
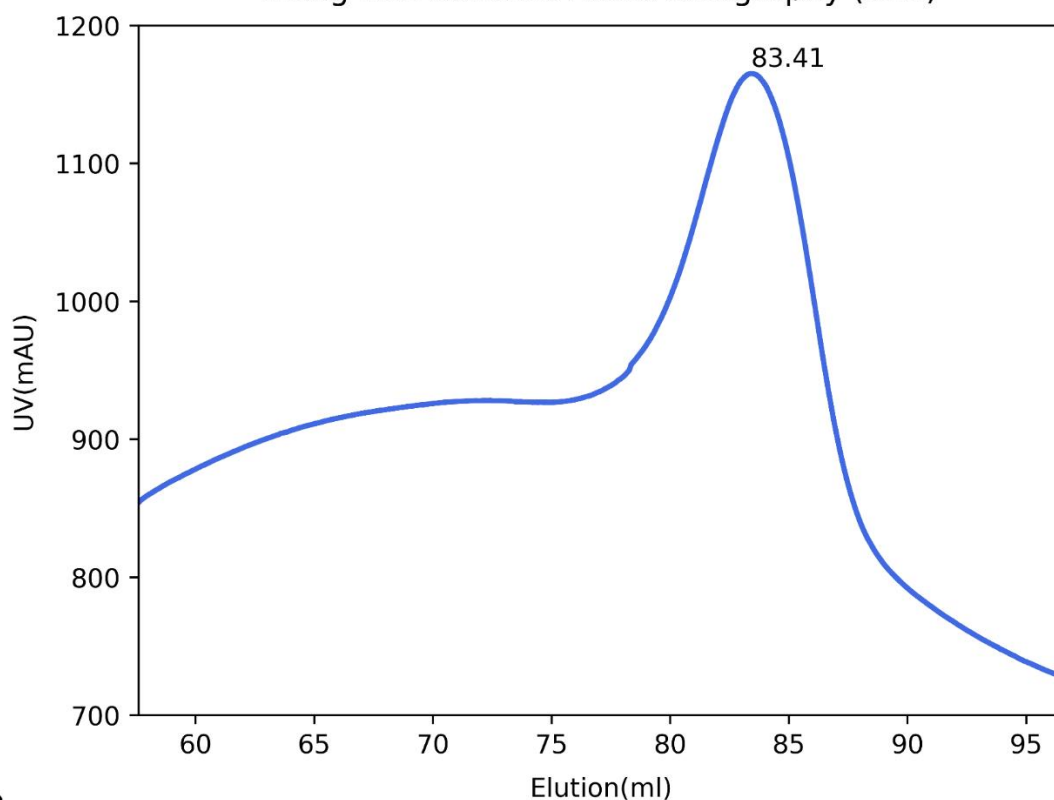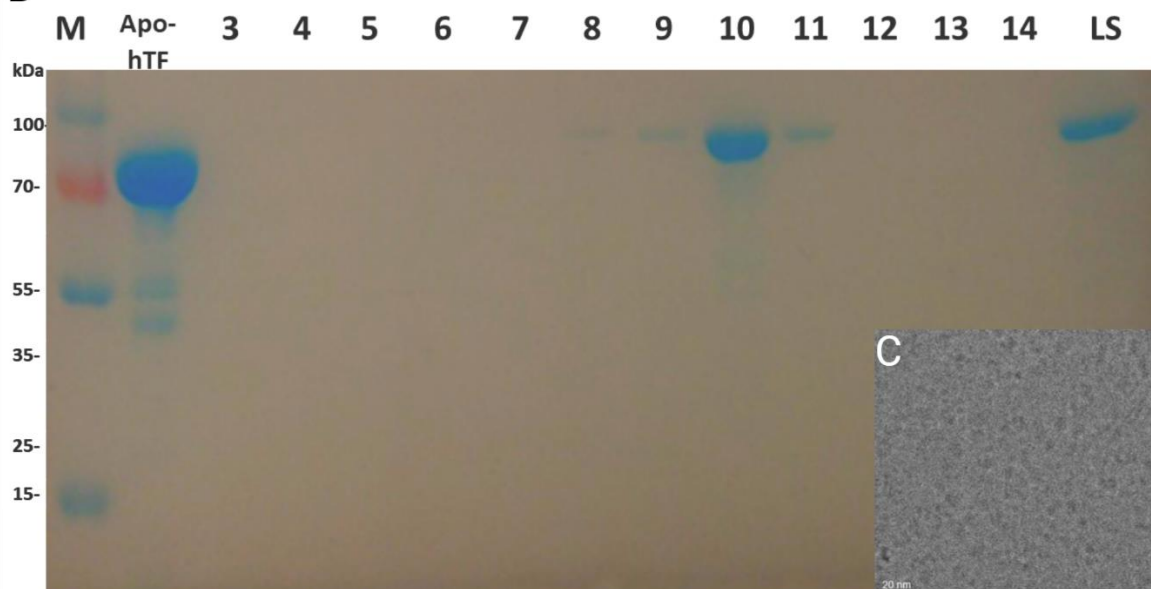
Figure 4.19 Purification of the hTF-Ferric ammonium citrate protein.
(A) Chromatogram illustrating the elution profile of the protein-ligand complex from SEC using a Superdex 200, Hiload 16/600 column. (B) 12% SDS PAGE profile of the protein complex fractions collected and shown lanes 3 to 14, which correspond to the first peak of the chromatogram of elution between ~58 - 97 mL; lane 1 is the low molecular weight marker, lane 2 is the purified apo-hTF control sample, and LS is the complex loading sample. hTF complex has an expected MW of ~70 kDa. (C) A screened TEM micrograph image of the purified hTF-Ferric ammonium citrate of defocus set at -2 μm and scale bar, 20 nm.

A Chromatogram of the complex the hTF-Ferrous ammonium sulphate using size exclusion chromatography (SEC)

B

C
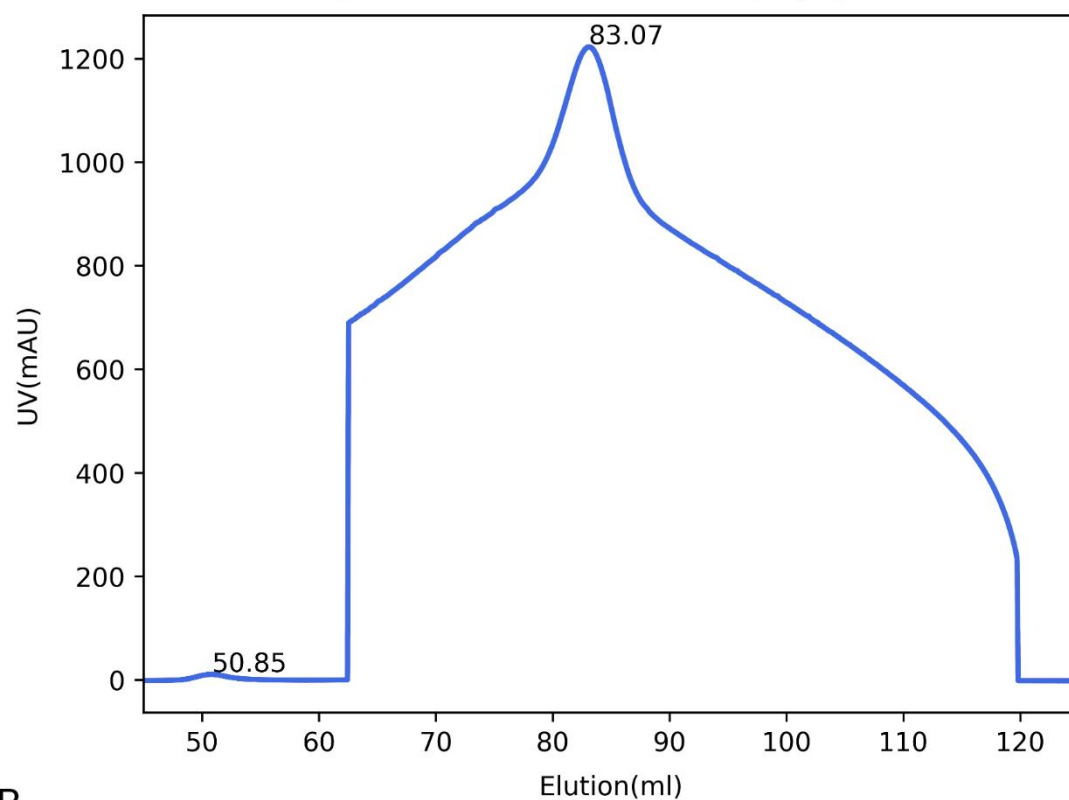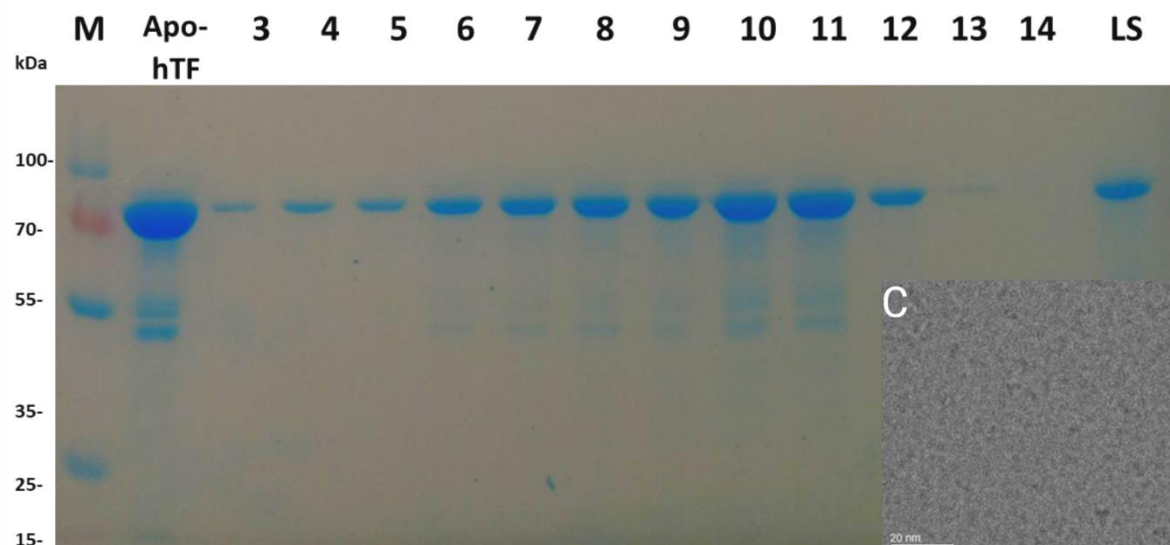
Figure 4.20 Purification of the hTF-Ferrous ammonium sulphate protein.
(A) Chromatogram illustrating the elution profile from SEC using a Superdex 200, Hiload 16/600 column. (B) 12% SDS PAGE profile of the protein complex fractions collected and shown lanes 3 to 14, which correspond to the high peak of the chromatogram; lane 1 is the low molecular weight marker, lane 2 is the purified apo-hTF control sample, and LS is the complex loading sample. hTF complex has an expected MW of ~70kDa. (C) A screened TEM micrograph image of the purified hTF-Ferrous ammonium sulphate of defocus set at -2 µm and scale bar, 20 nm.

4.2.3 Deglycosylation of hTF

Cryo-EM has only partially elucidated glycosylated protein structures, due to their dynamics (Masiulis et al.2019; Yang et al. 2021). However, fascinating large glycoproteins such as HIV-1 envelope trimer (PDB code: 5FUU) have been derived at 4.19 Å (Lee et al. 2016). Glycan flexibility can introduce heterogeneity into the EM datasets because vitrified sugars are captured at varying conformations (Tsai et al. 2024). The mix of molecular structures within samples complicates the analysis, meaning that different protein molecules may have different sugar configurations attached, which makes it challenging to separate particles based on structural similarities accurately (Sobti et al. 2021).

Distinctions of 2D and 3D classes of glycosylated protein may cause problematic blurs due to the complexity and heterogeneity of models (Cheng 2018), which leads to removing a significant amount of data during processing corresponding to the noise of glycan shields or branches of EM volume maps (Tsai et al. 2024). A deglycosylation technique releases N-glycans enzymatically or *O*-glycans chemically from glycoproteins, which could improve sample preparation and reduce heterogeneity caused by glycosylated proteins with flexible glycoproteins (Yang et al. 2017).

Both deglycosylation enzymes Endoglycosidase H (Endo H) and Peptide-N-Glycosidase F (PNGase F, EC 3.5.1.52) were tested and used to remove the sugars that might cause the sample heterogeneity. These enzymes perform digestion in different metabolic pathways. PNGase F is the most effective method of removing nearly all N-linked oligosaccharides from glycoprotein or asparagine-linked hybrid or complex, While Endo H cleaves N-linked glycans between the two *N*-acetylglucosamine (GlcNAc) residues. To Prepare an optimal deglycosylated protein sample for CryoEM, I followed specific manufacturer instructions, supplied with EndoH or PNGase enzymes provided by New England Biolabs along with appropriate buffers provided in the kit.
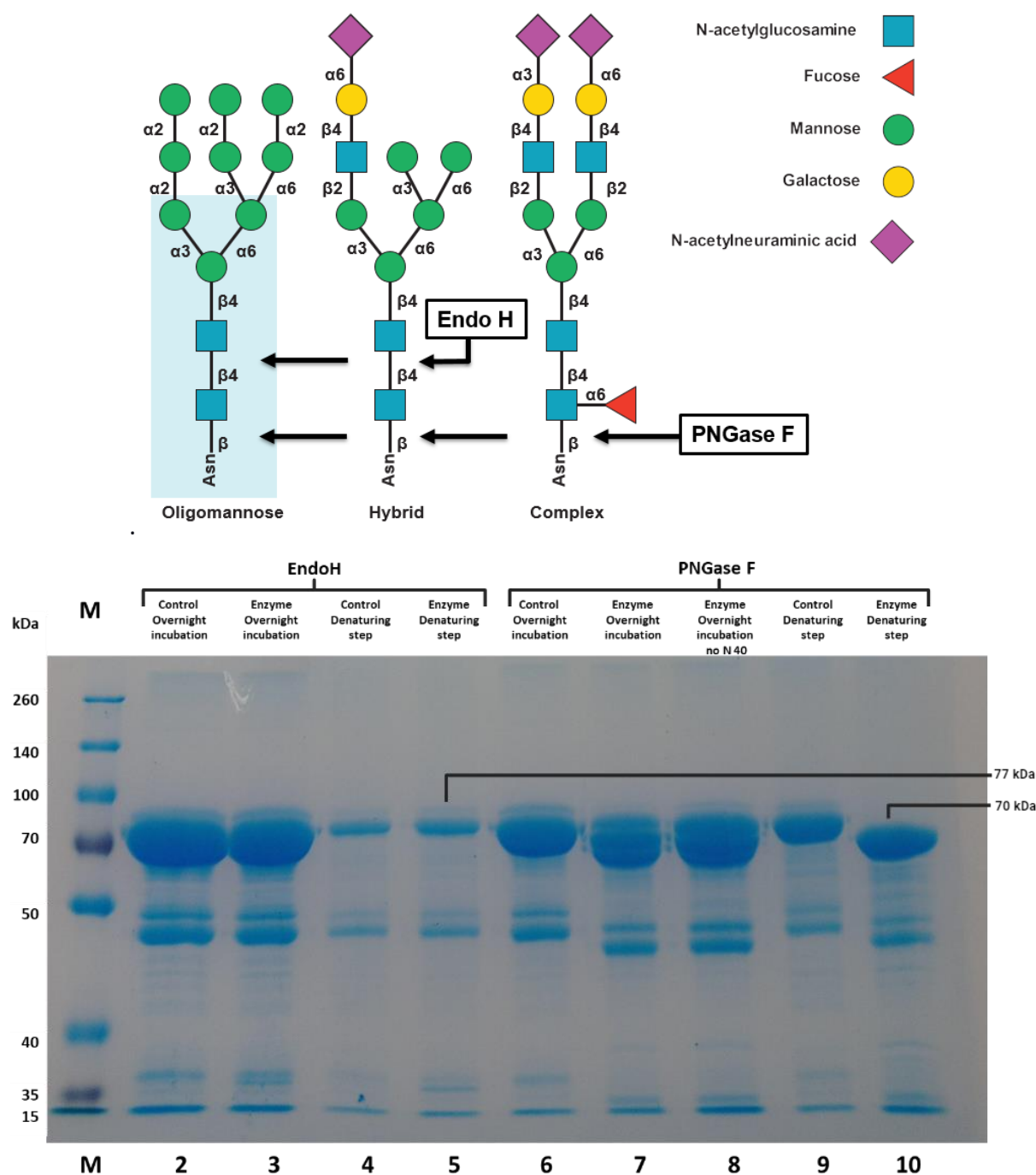
Figure 4.21 EndoH and PNGase digestion sites and their hTF digestion reaction analysis using 10% SDS-PAGE. Sample reactions were loaded at 3 µg/uL each.

The optimal concentration and incubation conditions were determined for the reaction mixture. Experiments have run at the recommended period and temperature specified in the manufacturer's instructions. Incubation times were used for 1 hour at 37°C, and some directions were incubated overnight for 16 hours.

When using EndoH and PNGase F enzymes specified in the manufacturer's instructions, proteins are typically in their folded state. These enzymes are used to remove N-linked

glycans from glycoproteins, and for them to effectively recognize and cleave the glycan moieties, the proteins usually need to be correctly folded. This ensures that the glycosylation sites are accessible to the enzymes. If the proteins were unfolded, it might hinder the enzymatic activity as the structural conformation necessary for recognition would be disrupted (Briggs 2017; Bresciani et al. 2018).

SDS-PAGE gel analysis indicates that samples that are treated with EndoH do not observe a distinct lower band while lower bands appear at lanes 7,8 and 10 for samples treated with PNGase (Figure 4.21)

The hTF sample treated with PNGase (lane 10) showed clearer evidence of deglycosylation occurring on SDS-PAGE gel than the hTF sample subjected to Endo H (lane 5). PNGase treatment has resulted in the appearance of a lower molecular weight band, while EndoH treatment did not produce a lower band (Figure 4.21). For that, it was decided to incubate a large amount of the enzyme overnight with the digestion enzyme and then purify and prepare the grids (Figure 4.22).

A Chromatogram of the deglycosylated hTF by Peptide-N-GlycosidaseF (PNGas F) using size exclusion chromatography (SEC)
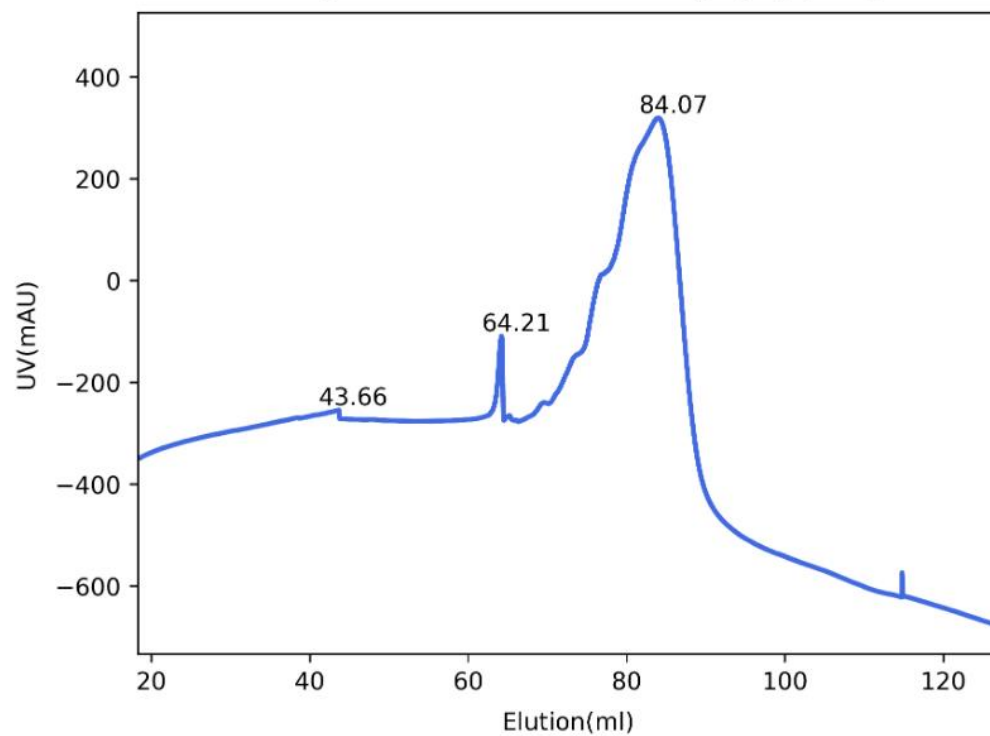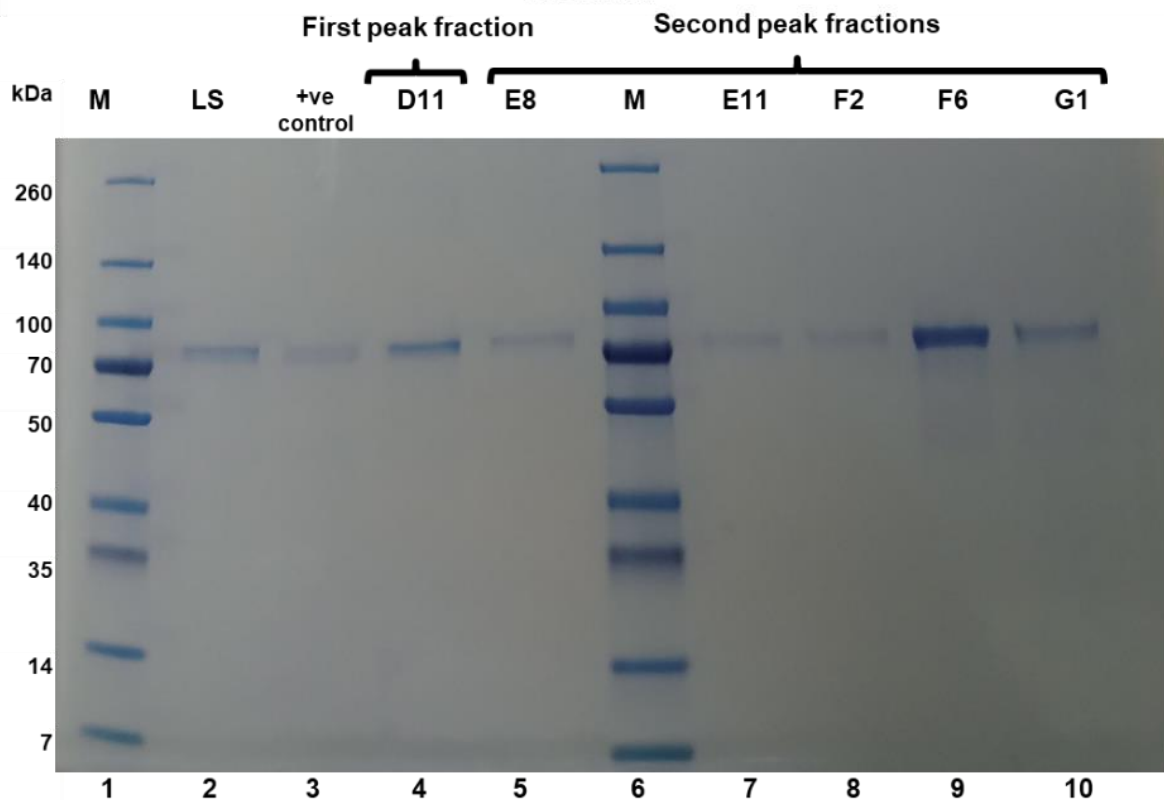
Figure 4.22 Purification of the hTF treated with PNGase.

(A) A chromatogram of apo-hTF digested by PNGase and purified using SEC.(B) 12% SDS-PAGE analysis of some eluted fractions corresponded to the two chromatographic peaks. Sample reactions were loaded at one µg/ul each.

4.2.4 Cryo-EM study of human serum transferrin

The study focuses on addressing challenges associated with imaging small, flexible, and glycosylated molecules of human transferrin (hTF) using cryo-electron microscopy (cryo-EM). To optimise sample preparation, plunge freezing parameters such as sample concentration and blotting conditions were adjusted to enhance ice quality and achieve uniform particle distribution (Table 2.3, 2.4 and 2.5) However, grids prepared from these molecules exhibited varying ice thickness and quality before screening due to the behaviour of small, flexible molecules during plunge freezing, presenting challenges in achieving uniform and vitreous ice. The presence of crystalline ice or ice artefacts indicated imperfect vitrification, potentially obscuring or distorting the small hTF molecules. Subsequent micrographs obtained after screening displayed limited or poorly defined hTF particles, compromising particle visibility and clarity due to their small size and distribution. Additionally, acquiring a diverse range of particle views and orientations for 3D reconstruction proved challenging with these difficult-to-image molecules, impacting the accuracy and completeness of structural analysis. Micrographs also exhibited higher levels of background noise and unwanted artefacts, further complicating the visualisation of small hTF molecules and subsequent data interpretation.

4.2.4.1 First datasets structure analysis: native apo-hTF without purification

A decision was made to initially process the sample without any further purification and refine the process should problems arise down the line.

These data sets were processed on *RELION* (v. 3.1.2).

The first round of 3D classification was in limited resolution to distinguish the details, whether an N or C Lobe map. Moreover, fitting N and C atomic module lobes individually to maps indicated the similarity of map dimensions shows in Figure 4.23, (5). The second run of 3D classification was obtained using the identified 3D class map with green. In contrast, the entire hTF atomic molecule fits into the three 3D classes EM map as a single rigid body. It was obtained as an acceptable fit.
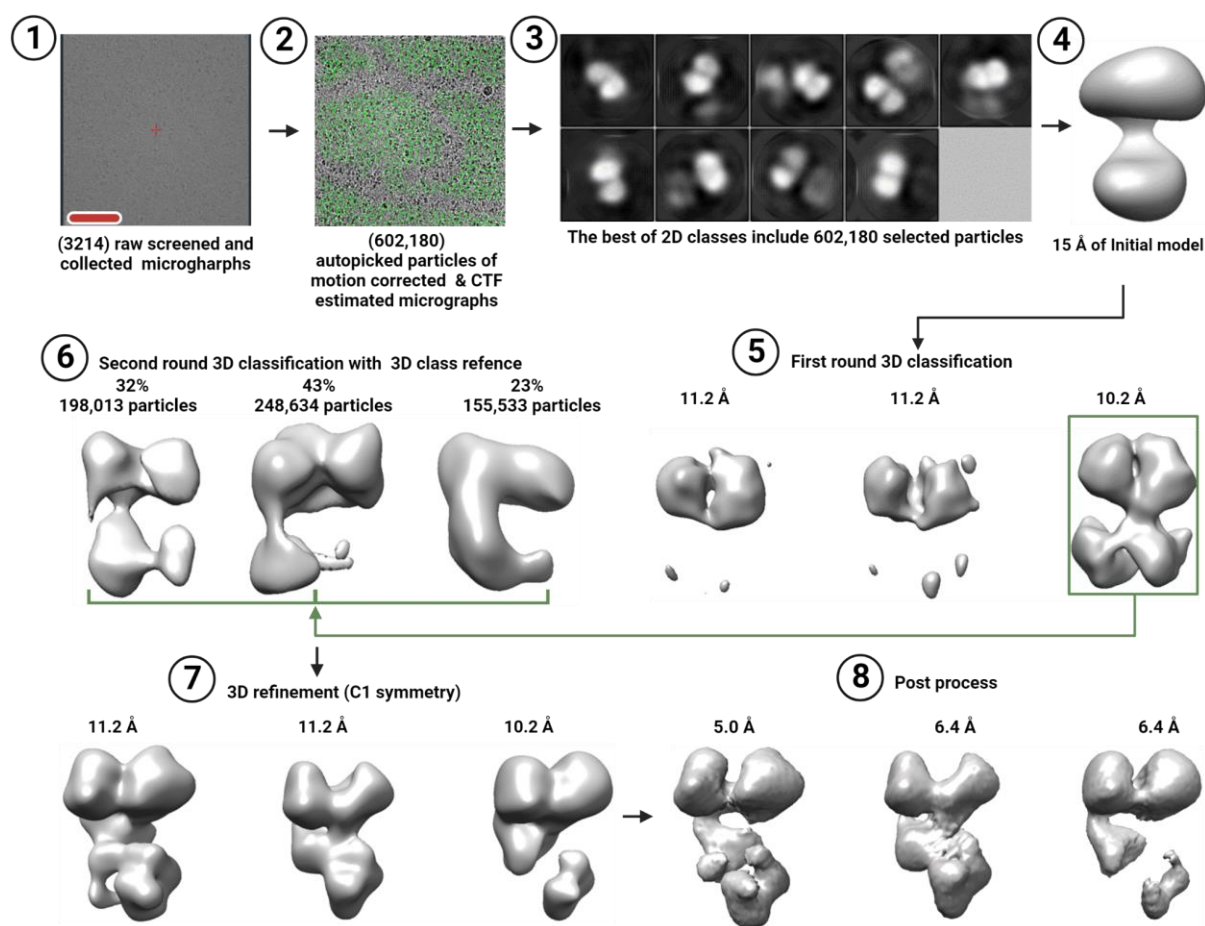
Figure 4.23 CryoEM data processing workflow of native apo-hTF.
(1) Represent a screened micrograph at -2 μm defocus and a 0.2 mg/mL sample concentration. The red scale bar indicates 20 nm. (2) A micrograph showed picked particles automatically. These particles have generated (3) different views of 2D classes that generated (4) the initial model, then (5) and (6) two rounds of 3D classification, (7) 3D refinement and (8) final post process maps.

4.2.4.2 Second dataset structure analysis: purified apo-hTF

After the 2D classification, 371,431 particles were initially sorted by 3D classification using an initial model generated from the original data lowpass filtered to 15 Å. The best 3D class (21%) was used as a reference for 3D refinement, with a total of 171,431 particles. This 3D refined map was sorted into amps for the N and C lobes. Two new runs of 3D classification were performed separately to inspect the selected particles and 2D class averages for each type of lobe. For both lobes, the resolution of 3D classes led to a resolution of 6.8 Å, shown in Figure 4.24.
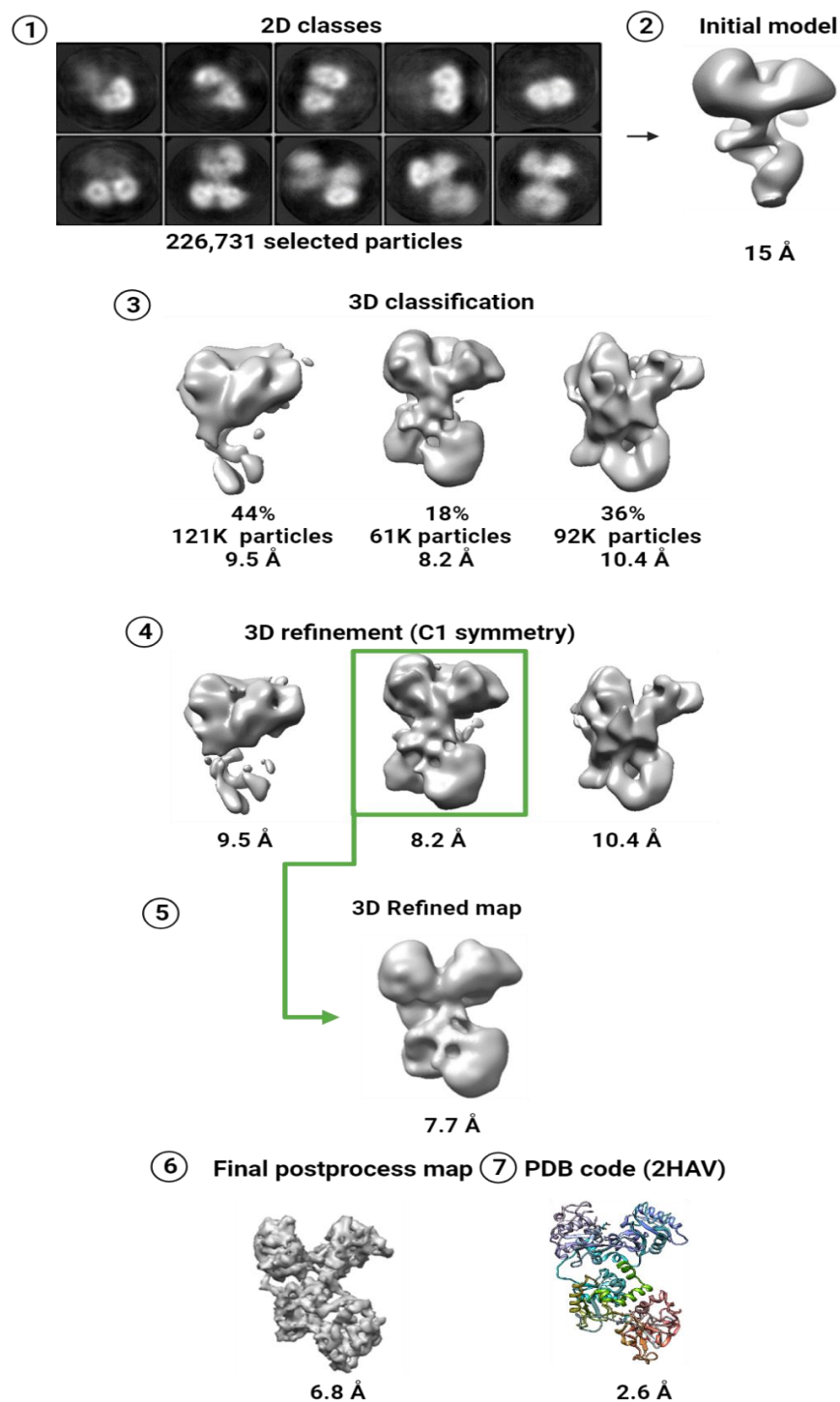
Figure 4.24 CryoEM data processing workflow representative of purified hTF.
(1)2D classes generated from purified hTF sample grids (2) An initial model (3) 3D classification (4) 3D refinement (5) Second run of the best-refined map using 3D mask map and (6) final post process potential map (7). (PDB code: 2HAV) of the atomic model was superimposed to the final map.

4.2.4.3 Third datasets structure analysis: complex of hTF- Ferric ammonium citrate

In order to increase our protein stability, apo-hTF was incubated with two iron compounds, Ferric ammonium citrate and Ferric ammonium sulphate. These protein iron complexes did not improve the resolution but generated two holo-structures with "fully-closed" conformation of the N lobes, shown in Figures 4.25 and 4.26, respectively.
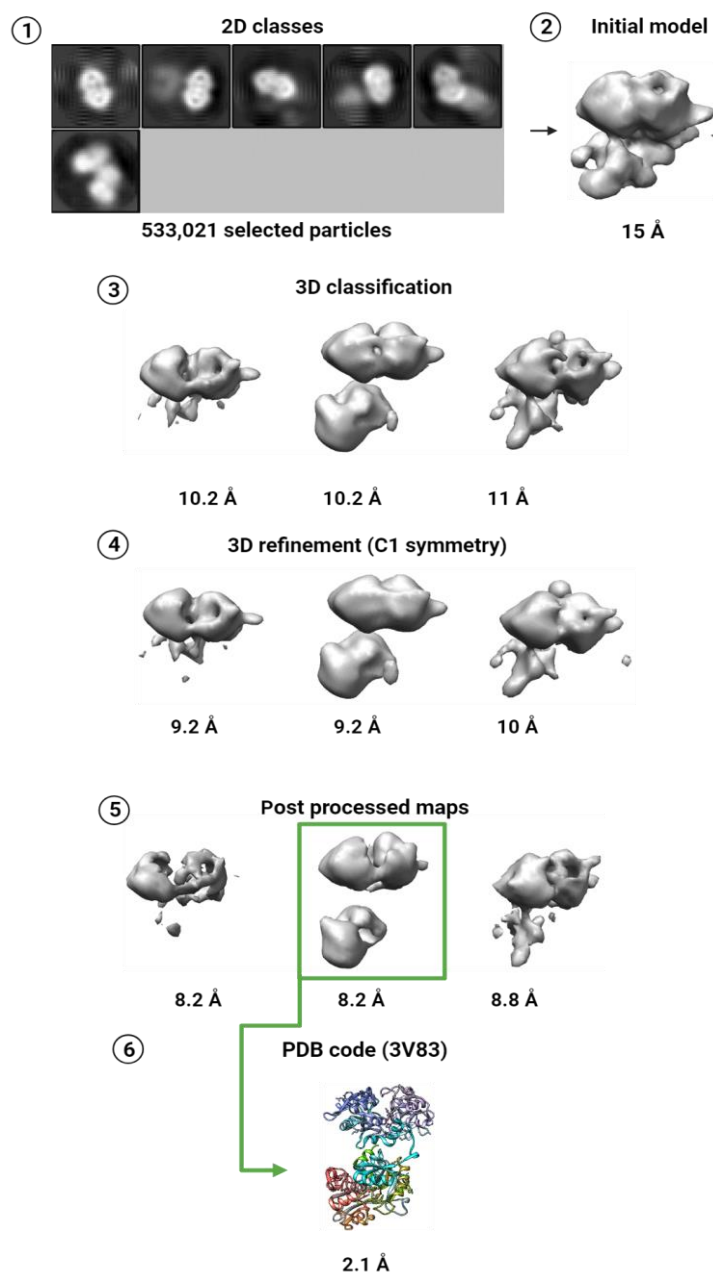
Figure 4.25 CryoEM data processing workflow represents a purified complex of hTF -Ferric ammonium citrate. (1) 2D classes generated from a purified complex of hTF -Ferric ammonium citrate sample grids (2) An initial model (3) 3D classification (4) 3D refinement (5) final post process potential maps of different 3D classes (6). (PDB code: 3V83) of the atomic model was superimposed to the final map to show the fully closed conformation N lobe.

4.2.4.4 Fourth datasets structure analysis: complexes of hTF- Ferrous ammonium sulphate
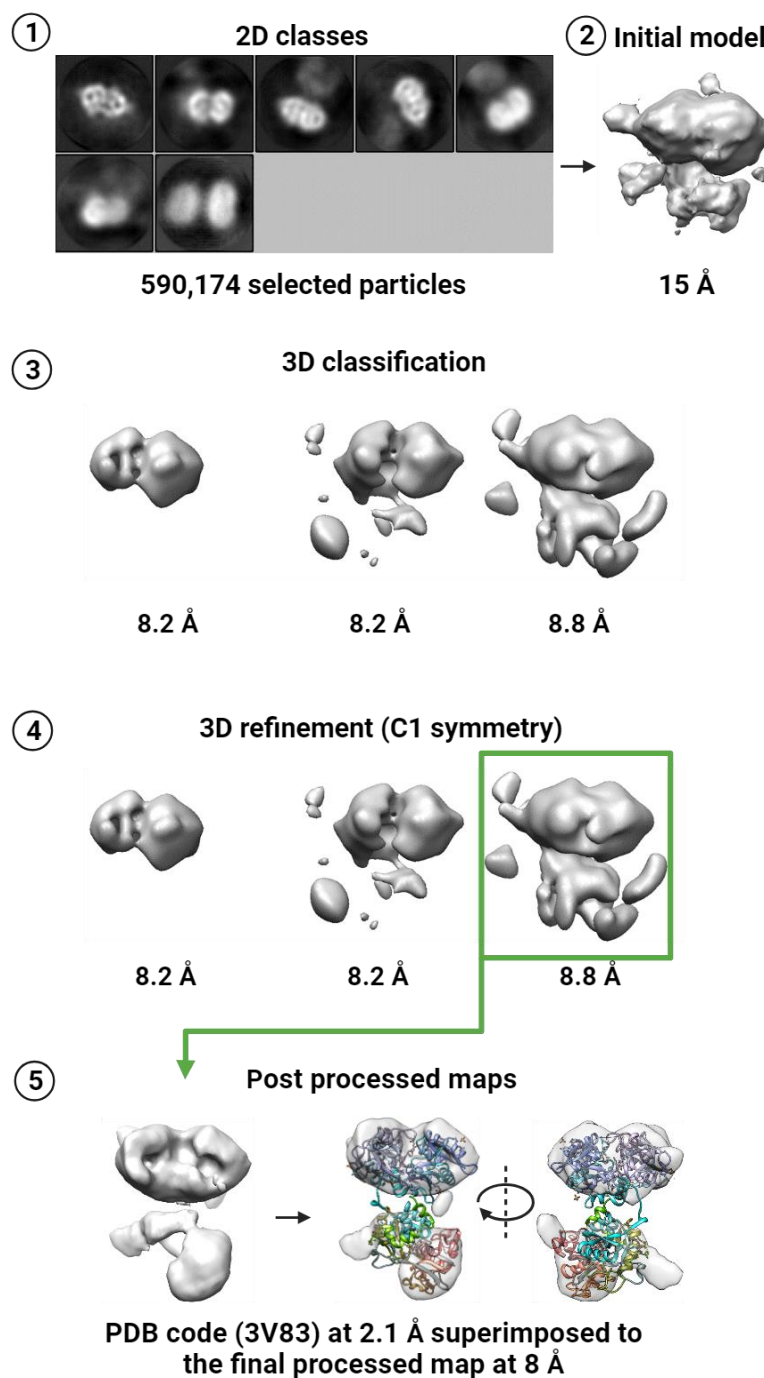
Figure 4.26 CryoEM data processing workflow representative from a purified 3hTF -Ferric ammonium sulphate. (1) 2D classes generated from a purified complex of hTF -Ferric ammonium sulphate sample grids (2) An initial model (3) 3D classification (4) 3D refinement (5) A Post process potential map of the best of the 3D classes with two sides of fitting angles for the atomic model (PDB code: (3V83), which is entirely closed conformation N lobe.

4.2.4.5 Fifth datasets structure analysis: Thinner ice grid protocol

In order to get thinner, unbreakable ice hTF grid sample was prepared at low humidity and higher blotting force by applying sample 4 mg/ml, 20 blotting force ,70% humidity, 5s

Grids were prepared from a snap-frozen specimen previously purified by SEC. They were thawed quickly at body temperature and then used. These proteins prepared with thinner ice generated a map that did not improve the resolution, shown in Figure 4.27.
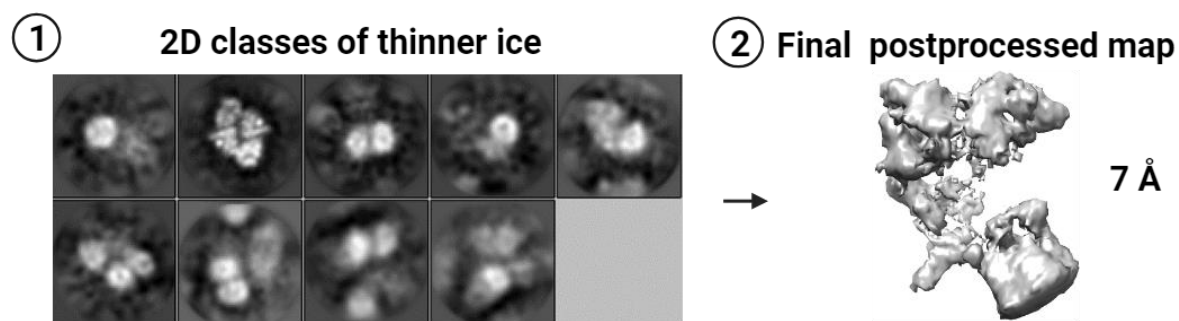
Figure 4.27 CryoEM data processing workflow representative from an apo-hTF thinner ice grids.
(1) 2D classes generated from purified hTF thinner ice grids (contained 250,019 particles) (2) Final post-processed map with limited resolution.

4.2.4.6 Sixth datasets structure analysis: deglycosylated hTF by PNGase

To produce a homogenous and decreasing protein motion hTF grid by removing suggestions with PNGase, these proteins prepared without glycans generated a map that did not improve the resolution, shown in Figure 4.28.
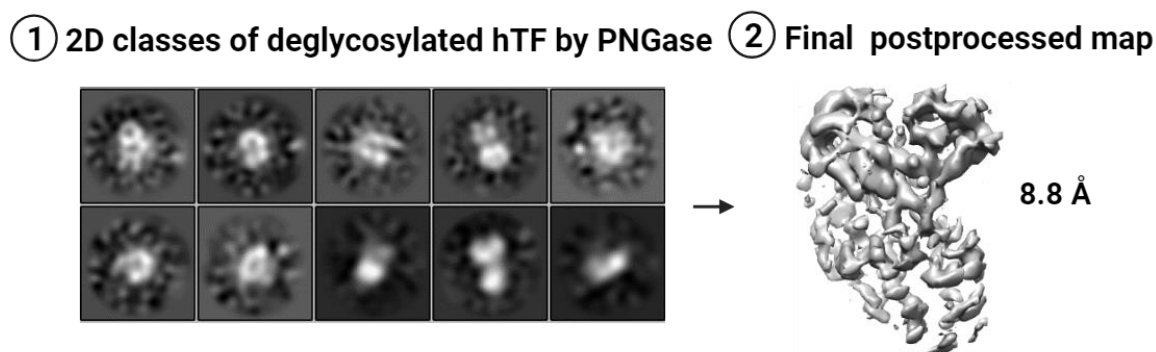
Figure 4.28 CryoEM data processing workflow representative from samples treated with PNGase.
(1) 2D classes generated from a purified hTF which was deglycosylated overnight by the PNGase. 2d classes were generated based on 188,644 particles (2) Final post-processed map with limited resolution.

4.2.4.7 Seventh datasets structure analysis: Volta phase plate (VPP)

The VPP device technology was applied to the microscope Glacios to collect datasets of apo transferrin in an attempt to improve the final map quality and resolution shown in Figure 4.29.
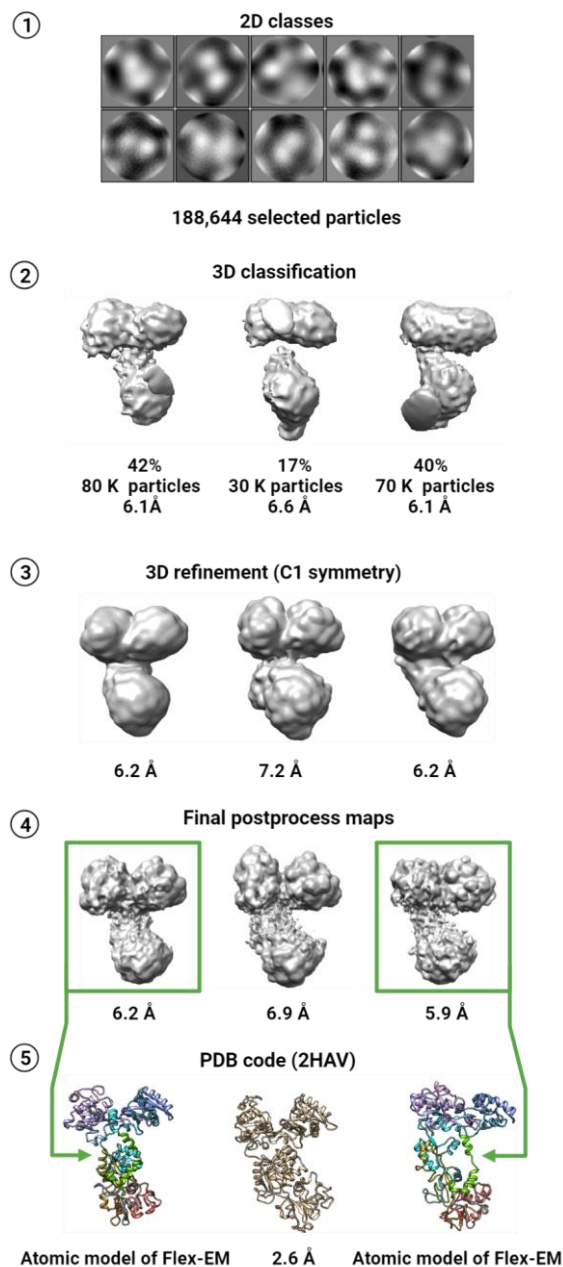
Figure 4.29 CryoEM data processing workflow representative hTF data collected using VPP method.
(1) 2D classes generated from a previous purified and frozen hTF sample grids (2) 3D classification (3) 3D refinement (4) final post process potential maps (5) Atomic modules were generated by using flexible fitting of Flex-EM for two of best resolution final maps relatively.

4.3 Discussion

The stability and functionality of the mature protein depend on glycans. Among other biological activities, glycans have an impact on receptor binding and the half-life of enzyme function. The two main kinds of protein glycosylation are glycans attached to the nitrogen atom (N-linked) of asparagine side chains or to the oxygen atom (*O*-linked) of serine and threonine side chains. While *O*-glycans are frequently broken by a chemical reaction, *N*-glycans can be liberated enzymatically by glycosidases (Yang et al. 2017). *N*-glycans are post-translational protein modifications that have been linked to important molecular and physiological processes and pathologies such as cancer, including tumour cell invasion, cell motility and metastasis development, angiogenesis, and immune system regulation(Pinho and Reis 2015). *O*-glycans are often short chains consisting of only a few sugars or they can be extended by the addition of different monosaccharides resulting in a longer chain.

Different stages of complex *N*-glycans

N-glycans can also be complex, with the mannose residues only present at the core and additional monosaccharides, including glucosamine or galactose, sialic acid, and fucose, being used to expand the branches. Based on their structural characteristics, *N*-linked glycans are divided into oligomannose, hybrid, and complex type glycans (Figure 4.30) (Fisher and Ungar 2016).
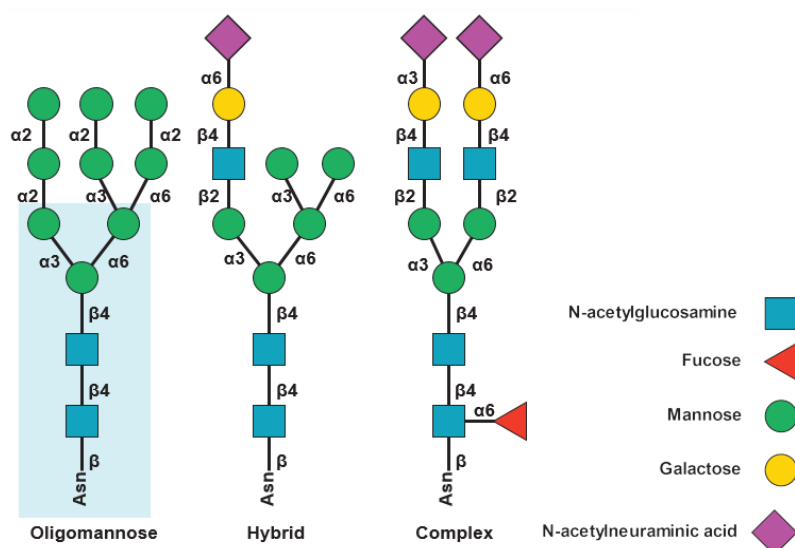
Figure 4.30 Different stages of complex *N*-glycans: oligomannose, hybrid, and complex.
The core of all three types is marked with a blue rectangle adapted from (Ivan Gudelj and Lauc, 2018).

The bulk of *N*-glycans seen on cell surfaces and in secretions are complex glycans, which come in bi-, tri-, and tetraantennary forms. Sialic acid residues are a typical termination point for complex glycans. Additional alterations are also feasible, including the inclusion of a fucosyl residue on the innermost GlcNAc and/or a bisecting GlcNAc at the mannosyl core. The human serum has a prominent transferrin glycoform, the tetra sialo-transferrin isoform (n = 4; 75%-80%)normally, which is projected to have two biantennary *N*-glycans on the 413 and 611 asparagine residues, for a total of four sialic acid residues(Gonzalo, Radenne, and Gonzalo 2014).

A previous study used electron microscopy and small-angle X-ray scattering to characterise the TbpB-transferrin complex and solved its crystal structure. The extra density that was found during the latter stages of refining was traced to the hTF residues N413 and N611, that are thought to be potential *N*-linked glycosylation sites. So, for these two residues, N-linked glycans have been built. The resulting structure had a R/R$_{free}$ value of 0.22/0.28 and was solved to 2.60 Å. Using Phaser-CCP435, the crystal structure of the diferric hTF was solved. PHENIX36 was used for refinement, and the structure was solved to a resolution of 2.1 Å, yielding final R/R$_{free}$ values of 0.19/0.23 (Noinaj et al. 2012).

Also, another model size 260-kDa of iron import complex from *N. meningitidis* strain K454 (serogroup B). These proteins were purified under identical circumstances. In the publication, the conformational changes of iron-free (apo) and iron-loaded (holo) transferrin in crystal and solution states, respectively, at three different pH values of physiological significance, were examined using two potent X-ray techniques, such as Macromolecular X-ray Crystallography (Merogroup B) that was modelled using a combination of X-ray crystallography, small-angle X-ray scattering, and electron microscopy (Noinaj et al. 2012).

In another study, both the iron-free (apo) and iron-loaded (holo) versions of human serum transferrin were purchased commercially (Sigma-Aldrich, St. Louis, MO, X) were subject to Small Angle X-ray Scattering (SAXS). Despite several attempts to increase crystal quality, a 3.0 Å resolution crystallographic model of glycosylated apo-Tf could not be improved (Campos-Escamilla et al. 2021). The revised structure and the X-ray data were uploaded to the Protein Data Bank and designated (PDB code: 7Q1L). SAXS data were uploaded to SASBDB (https://www.sasbdb.org) and verified there. The position of two NAG moieties bound to

residue Asn413 could be seen thanks to the apo-Tf model's data, which was sufficient for displaying the whole protein structure. After the second NAG element, additional density was visible, suggesting lengthening of the glycosyl chain, which may have been the main obstacle to improving crystal quality. Thus, glycosylation may even aid in crystallisation, although the resolution limit and diffraction quality may be compromised (Mesters and Hilgenfeld 2007). Glycosylated and non-glycosylated apo versions of human transferrin (PDB codes: 2HAV and 2HAU) were superposed with chain A and had RMSD values of 0.56 and 0.70, respectively. The resolution's capacity to shed light on the specifics of how things work is still somewhat constrained (Campos-Escamilla et al. 2021).

In a 2018 study, human apo-Tf samples (pro-325) (ProSpec) and human apo-Tf (T2036) (Sigma Aldrich) were obtained from commercial sources. An iron(III)-citrate-pyrophosphate oligomeric complex called ferric pyrophosphate citrate effectively transfers iron to plasma transferrin without a carbohydrate shell. FPC has a distinct iron(III)-the citrate-pyrophosphate ternary complex oligomeric structure that is stable for a long time in aqueous solution. Since the intended medication does not include carbohydrates or sugar, Moreover, the linker could not be modelled either (Pratt et al. 2018). As both the N and C lobe iron-binding sites contain an iron atom, it is possible that FPC may provide both lobes of human apo-transferrin with iron. On the electron density map, close to the site where iron binds in the N-lobe, there was a noticeable differential density peak that suggested the presence of an iron atom (Pratt et al. 2018).

Preparation and crystallisation of FeNFeC-hTF (PDB code:3QYT)

A 100 mg/ml stock of lyophilized apo-hTF was purchased from Sigma and reconstituted in 10 mM HEPES at pH 7.4. The sitting drop vapour diffusion technique was utilised at 271 K to create the crystals needed for data acquisition. Brifly a reservoir solution made up of 10 mM HEPES pH 7.4 and 14–15 percent PEG 3000 was combined with diferric hTF at a concentration of 30 mg/ml. After a few days, orange-red crystals began to form. After being cryoprotected in a solution of 10 mM HEPES pH 7.4, 30% PEG 3000, and 15% ethylene glycol, the crystals were flash-frozen in liquid nitrogen (Yang et al. 2012).

Determination of structure of FeNFeC-hTF

The diffraction data were gathered at 100 K at BL17U and analysed with HKL200050 at the Shanghai Synchrotron Radiation Facility. Using the N lobe of apo-hTF (PDB code: 2HAV) and the C lobe of iron-bound rabbit serum transferrin (PDB code: 1JNF) as search models, the structure was solved using the molecular replacement approach using the program Phaser51 from the CCP4 suite52 (Yang et al. 2012).

BiNFeC-hTF (4H0W)

Apart for being collected immediately above bismuth's L-III absorption edge (0.92000 Å), the diffraction data were gathered identically to those for FeNFeC-hTF and analysed with HKL2000. Using the structure of FeNFeC-hTF as a search model, the structure was solved using the molecular replacement technique with the program Phaser52 from the CCP4 suite30. The Protein Data Bank now contains the coordinates and structural factors (PDB code: 4H0W) (Yang et al. 2012).

Additionally, electron densities are seen close to the Asn413 and Asn611 residues in the C lobe of FeNFeC-hTF, indicating that two glycan moieties (NAG and *N*-acetylglucosamine) are linked to these residues. The first linked glycan moiety is easily discernible in the structure of FeNFeC-hTF, even though inadequate electron densities prevented the molecular replacement of full glycan chains in the model (Yang et al. 2012). Interestingly, the two hTF structures display distinct "partially opened" conformations in the N lobe with different extents of opening for the first time, which reflects two crucial protein conformers throughout the metal release process. Moreover, the structure of BiNFeC-hTF offers support for the prospective use of transferrin as a metallodrug delivery "vehicle" in healthcare. This information on Bi(III) coordination is directly provided by the structure (Yang et al. 2012), shown in Figure 4.31
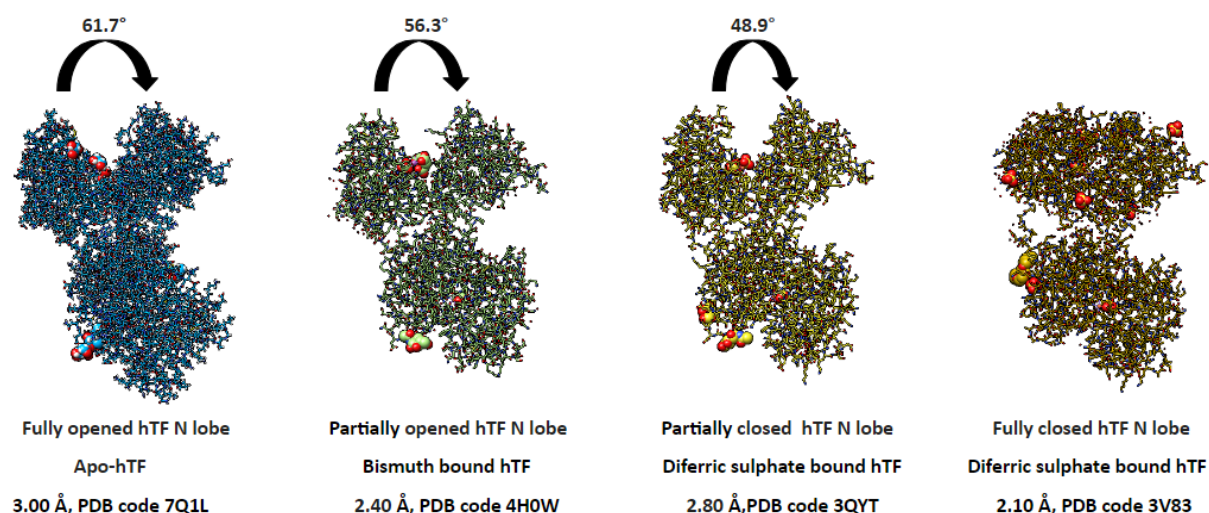
Figure 4.31 Representation of the N lobes on the molecular surfaces of apo-hTF, BiNFeC-hTF, FeNFeC-hTF, and Fe-hTF/2N.
The opening extent of the N lobe in the FeNFeC-hTF (red), BiNFeC-hTF (purple), and apo-hTF (green) structures in relation to the Fe-hTF/2N is depicted schematically in the diagram (blue)(Gruszczyk et al. 2018) .

In 2004, scientists mutated the structure to its non-glycosylated form. Residues 122 to 760 of recombinant human TfR ectodomains were produced in CHO cells and purified in accordance with the instructions in (Lawrence, 1999.). Non-glycosylated full-length Tf was produced, in which the two natural Asn 413 and Asn 611 glycosylation sites were ablated by mutation to Asp (Zak et al. 1997)Cheng et al. 2004). In 2018, Sigma-Aldrich published apo-transferrin isolated from human serum (Catalog Number T4382). The Cryo-EM structure PvRBP2b-TfR1-Tf ternary complex was established with an overall resolution of 3.7Å. The homodimeric TfR1 (residues 120–760) and two molecules of iron-bound Tf (residues 1-679) make up the PvRBP2b–TfR1–Tf complex, which has two molecules of PvRBP2b attached on either side (Figure 4.32)(Gruszczyk et al. 2018).
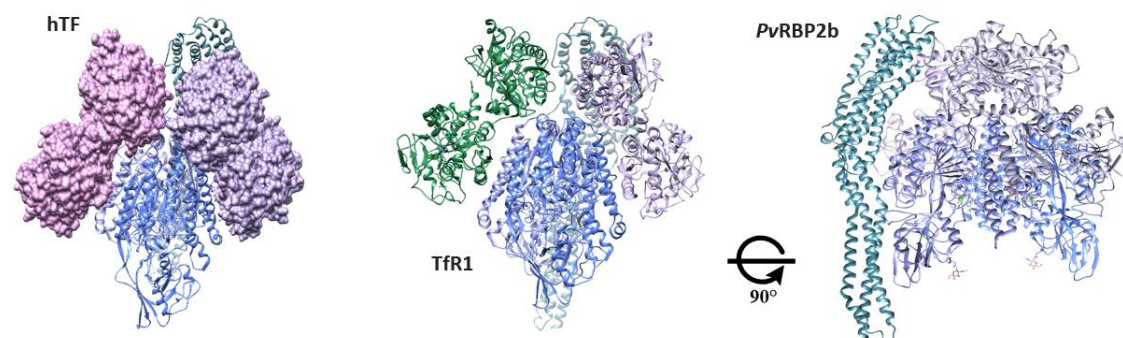
Figure 4.32 The 3.7 Å resolution Cryo-EM structure of the ternary complex of PvRBP2b-TFR1-Tf.
Labelling the secondary structure with two orthogonal images of the PvRBP2b structure. Two orthogonal views of the PvRBP2b-TFR1-Tf ternary complex with two ligands. Ferric ions (Fe^{3+}), red spheres; TFR1, Blue; hTf, purple; and PvRBP2b, green (Gruszczyk et al. 2018).

Mutational studies reveal that PvRBP2b residues that are crucial to complex formation are conserved; This raises the possibility of creating antigens that function in all *P. vivax* strains. Functional investigations of TfR1, a necessary housekeeping protein, reveal that this protein is hijacked by *P. vivax* by attaching to locations that control host specificity, without impairing its cellular function of transporting iron. The inhibitory epitopes were characterised from the crystal and solution structures of PvRBP2b upon interaction with antibody fragments. Their findings provide a structural foundation for understanding the molecular mechanism of inhibitory monoclonal antibodies as well as how the *P. vivax* reticulocyte-binding protein interacts with its receptor., the Cryo-EM maps have been added to the Electron Microscopy Data Bank with the accession codes EMD-7783, EMD-7784, and EMD-7785. The PvRBP2b-Tf-TfR1 one-ligand complex, two-ligand complex subclass 1, and two-ligand complex subclass 2 each have their atomic model coordinates deposited in the Protein Data Bank with codes 6D03, 6D04, and 6D05 (Gruszczyk et al. 2018).

The hTF protein target was previously considered exclusive to solve by crystallography as a monomer. Our results demonstrate that the ability of single particle analysis of Cryo-EM data to resolve the HTF in its native state at a resolution of structures can show the conformation change of the N-lobe between apo and holo forms.

Thinner ice grids

Thinner ice provides better contrast in cryo-EM images because electrons can pass through it more easily. This can be advantageous for visualising fine details in specimens. However, extremely thin ice can also lead to a reduction in image resolution because fewer electrons interact with the specimen.

Sample preparation of small proteins requires a special method to produce thinner ice film that embeds our target protein in layers to ensure least background noise with best dynamic signals.

Deglycosylation

PNGase cleaves the *N*-linked glycans from glycoproteins, leaving behind the deglycosylated protein. This cleavage involves the removal of glycan moiety attached to the asparagine (N) residue in the protein. The removal of the glycan results in a decrease in the molecular weight

of the glycoprotein because the glycan is relatively large compared to the protein itself. As a result, the protein migrates faster through the SDS-PAGE gel, leading to the appearance of a lower band on the gel corresponding to the deglycosylated protein. The extent of deglycosylation is often reflected in the intensity of this lower band.

EndoH, on the other hand, specifically cleaves high-mannose and some hybrid *N*-linked glycans but does not cleave complex glycans. High-mannose glycans are a particular type of *N*-linked glycan. In cases where the glycoprotein contains only high-mannose glycans, EndoH treatment will result in the removal of these glycans, but this may not significantly alter the molecular weight of the protein core. Therefore, when EndoH is used to deglycosylate glycoproteins with high-mannose glycans, a distinct lower band on the SDS-PAGE gel may not appear because the change in molecular weight may be relatively minor.

Data processing of highly glycosylated protein can be improved by using advanced classification algorithms that can better handle flexible structures or increase the amount of training data to mitigate overfitting. Additionally, techniques like sub-classification, where particles are classified into smaller, more homogeneous groups before final classification, can also help improve the accuracy of structural analysis in the presence of flexible sugars.

Volta phase plate VPP

The Volta phase plate VPP technique (Danev et al. 2014), has been used to increase the contrast of the target protein transferrin images. The smallest biomolecule has been reported to solve by VPP according to EMDB, the flexible RNA fragment of apo SAM-IV riboswitch (~40 kDa), revealing a final 3D map of ~7 Å. A later study explored the more accessible data collection without VPP, and obtained the same apo RNA 3D map with a relatively high resolution of 3.7Å, mentioned in section 2.1.5. RNA provides a higher electron scattering of phosphate backbones to supply adequate contrast for visualising this small structure (Kaiming Zhang et al. 2019). VPP improves visual contrast and enables in-focus data capture, enhancing the signal-to-noise ratio (Danev, Tegunov, and Baumeister 2017).

The 2D averages of the apo hTF with VPP exhibit low quality compared to the 2D averages of the apo hTF in Cryo-EM independently.

4.4 Conclusions

At the start of this study, structures for the human serum transferrin in apo and complex forms were known. However, the majority of these structures were so far solved by crystallography, and very few were solved by Cryo-EM with hTF as part of a larger protein complex (Cheng et al. 2004; Gruszczyk et al. 2018). Very little was known about glycosylated apo hTF. The relatively low resolution of the hTF structure derived in this study was mainly due to the heterogeneity and flexibility of the molecule.

However, post processed maps of our Cryo-EM data have clearly shown a strong iron interaction and close-structured conformation at different complexes. Static snapshots that we attempted to derive from single particle analysis data were hampered by intrinsically low-resolution maps and/or the molecule size. We conclude that glycosylation is perhaps the main reason for unsuccessful efforts to improve map quality, along with the high flexibility of the emolecule, as seen in cryo-EM maps. From this study, we can identify conformational details of the mechanism of the iron interacting with transferrin, including its apo state, which will have important biostructural research impacts on inform further studies involving the glycosylated sample preparation and processing of Cryo-EM data.

5 Concluding Remarks

This study, despite using a 200 keV electron microscope typically employed for screening, achieved significant results. This demonstrates the feasibility of routine work with cryo-EM using such instruments, a finding that should inform and update our colleagues in the field. The first study also showed that high-resolution structures could be obtained even with a 200 keV microscope, challenging the notion that higher-voltage microscopes are necessary for such work. This is particularly relevant as previous cryo-EM studies of β-galactosidase were primarily conducted using higher voltage instruments, yet this study achieved comparable results using a 200 keV microscope.

Troubleshooting experiments showed that preparing grids at slightly higher concentrations could lead to preferred orientations of the molecules, affecting image quality and data interpretation. The quality of the specimen, including sample concentration and grid preparation methods, significantly influenced the resolution and quality of the obtained cryo-EM structures.

In the second study, which focused on transferring structure, glycosylation played a significant role. It affected the conformational flexibility and heterogeneity of the molecule, contributing to challenges in obtaining high-resolution cryo-EM structures. Deglycosylation methods using enzymes like PNGase and EndoH were discussed as strategies to improve the resolution of highly glycosylated proteins. However, glycosylation remained a significant challenge in achieving high-quality cryo-EM maps. The heterogeneity and flexibility induced by glycosylation were identified as the main reasons for the difficulty in improving map quality, suggesting that studies on such proteins should focus on glycosylated protein sample preparation and data processing optimisation. Advanced classification algorithms and sub-classification methods were mentioned as potential approaches to handle flexible structures and improve the accuracy of structural analysis, especially in the presence of flexible sugars. The Volta phase plate (VPP) technique was discussed to increase contrast in cryo-EM images. However, it was noted that the 2D averages of apo hTF with VPP exhibited lower quality compared to Cryo-EM without VPP.

The present study aimed to determine the structure of two molecules, β-galactosidase and transferrin, using cryo-electron microscopy (Cryo-EM). The Cryo-EM maps of β-galactosidase were obtained at different concentrations using a 200 keV microscope, which is significant given that most previous studies utilised higher-voltage instruments. Despite challenges in sample preparation, including preferred orientations at higher concentrations, the study successfully obtained high-resolution structures of β-galactosidase. Regarding transferrin, the study focused on glycosylated apo-hTF, a relatively understudied area compared to crystallography-based structures. Cryo-EM maps revealed insights into the iron interaction and conformational dynamics of transferrin. However, the resolution was limited due to the molecule's heterogeneity, flexibility, and glycosylation. Despite these challenges, the study provides valuable information on iron transferrin interactions and highlights the importance of further research on glycosylated transferrin and Cryo-EM data processing techniques.

Advanced Cryo-EM techniques, such as the Volta phase plate (VPP), were explored to improve contrast and resolution. While VPP enhanced the signal-to-noise ratio, it did not significantly improve the quality of transferrin density maps compared to traditional Cryo-EM. Overall, the study underscores the potential of Cryo-EM in elucidating complex molecular structures and interactions and the feasibility of obtaining high-resolution cryo-EM structures of β-galactosidase while also acknowledging the challenges posed by glycosylation and flexibility. Further research is needed to address these challenges and enhance the resolution of Cryo-EM structures, particularly for glycosylated proteins like transferrin. Transferrin highlights the challenges posed by glycosylation and suggests avenues for further research to improve sample preparation and data processing techniques.

References

Abdizadeh, H., A. R. Atilgan, C. Atilgan, and B. Dedeoglu. 2017. "Computational Approaches for Deciphering the Equilibrium and Kinetic Properties of Iron Transport Proteins." *Metallomics: Integrated Biometal Science* 9 (11): 1513–33.

Adams, Ty E., Anne B. Mason, Qing-Yu He, Peter J. Halbrooks, Sara K. Briggs, Valerie C. Smith, Ross T. A. MacGillivray, and Stephen J. Everse. 2003. "The Position of Arginine 124 Controls the Rate of Iron Release from the N-Lobe of Human Serum Transferrin. A Structural Study." *The Journal of Biological Chemistry* 278 (8): 6027–33.

Afonine, Pavel V., Bruno P. Klaholz, Nigel W. Moriarty, Billy K. Poon, Oleg V. Sobolev, Thomas C. Terwilliger, Paul D. Adams, and Alexandre Urzhumtsev. 2018. "New Tools for the Analysis and Validation of Cryo-EM Maps and Atomic Models." *Acta Crystallographica. Section D, Structural Biology* 74 (Pt 9): 814–40.

Agard, David, Yifan Cheng, Robert M. Glaeser, and Sriram Subramaniam. 2014. "Chapter Two - Single-Particle Cryo-Electron Microscopy (Cryo-EM): Progress, Challenges, and Perspectives for Further Improvement." In *Advances in Imaging and Electron Physics*, edited by Peter W. Hawkes, 185:113–37. Elsevier.

Agarwal, M. B., S. S. Gupte, C. Viswanathan, D. Vasandani, J. Ramanathan, Neena Desai, R. R. Puniyani, and A. T. Chhablani. 1992. "Long-Term Assessment of Efficacy and Safety of L1, an Oral Iron Chelator, in Transfusion Dependent Thalassaemia: Indian Trial." *British Journal of Haematology*. https://doi.org/10.1111/j.1365-2141.1992.tb06445.x.

Agirre, Jon. 2017. "Strategies for Carbohydrate Model Building, Refinement and Validation." *Acta Crystallographica. Section D, Structural Biology* 73 (Pt 2): 171–86.

Agirre, Jon, Mihaela Atanasova, Haroldas Bagdonas, Charles B. Ballard, Arnaud Baslé, James Beilsten-Edmands, Rafael J. Borges, et al. 2023. "The CCP4 Suite: Integrative Software for Macromolecular Crystallography." *Acta Crystallographica. Section D, Structural Biology* 79 (Pt 6): 449–61.

Agirre, Jon, Gideon Davies, Keith Wilson, and Kevin Cowtan. 2015. "Carbohydrate Anomalies in the PDB." *Nature Chemical Biology* 11 (5): 303.

Agirre, Jon, Javier Iglesias-Fernández, Carme Rovira, Gideon J. Davies, Keith S. Wilson, and Kevin D. Cowtan. 2015. "*Privateer*: Software for the Conformational Validation of Carbohydrate Structures." *Nature Structural and Molecular Biology* 22 (11): 833–34.

Al-Refaie, F. N., C. Hershko, A. V. Hoffbrand, M. Kosaryan, N. F. Olivieri, P. Tondury, and B. Wonke. 1995. "Results of Long-Term Deferiprone (L1) Therapy: A Report by the International Study Group on Oral Iron Chelators." *British Journal of Haematology*. https://doi.org/10.1111/j.1365-2141.1995.tb05274.x.

Anderson, Erik R., and Yatrik M. Shah. 2013. "Iron Homeostasis in the Liver." *Comprehensive Physiology*. https://doi.org/10.1002/cphy.c120016.

Arndt, T. 2001. "Carbohydrate-Deficient Transferrin as a Marker of Chronic Alcohol Abuse: A Critical Review of Preanalysis, Analysis, and Interpretation." *Clinical Chemistry* 47 (1): 13–27.

Artola, Marta, Christinne Hedberg, Rhianna J. Rowland, Lluís Raich, Kassiani Kytidou, Liang Wu, Amanda Schaaf, et al. 2019. "α- D -Gal-Cyclophellitol Cyclosulfamidate Is a Michaelis Complex Analog That Stabilizes Therapeutic Lysosomal α-Galactosidase A in Fabry Disease." *Chemical Science* 10 (40): 9233–43.

Ashwell, G., and A. G. Morell. 1974. "The Role of Surface Carbohydrates in the Hepatic Recognition and Transport of Circulating Glycoproteins." *Advances in Enzymology and Related Areas of Molecular Biology* 41 (0): 99–128.

Aslanidis, C., and P. J. de Jong. 1990. "Ligation-Independent Cloning of PCR Products (LIC-PCR)." *Nucleic Acids Research* 18 (20): 6069–74.

Atanasova, Mihaela, Haroldas Bagdonas, and Jon Agirre. 2020. "Structural Glycobiology in the Age of Electron Cryo-Microscopy." *Current Opinion in Structural Biology* 62 (June): 70–78.

Atanasova, Mihaela, Robert A. Nicholls, Robbie P. Joosten, and Jon Agirre. 2022. "Updated Restraint Dictionaries for Carbohydrates in the Pyranose Form." *Acta Crystallographica. Section D, Structural Biology* 78 (Pt 4): 455–65.

Atsumi, S., H. Iinuma, C. Nosaka, and K. Umezawa. 1990. "Biological Activities of Cyclophellitol." *The Journal of Antibiotics* 43 (12): 1579–85.

Atsumi, S., C. Nosaka, H. Iinuma, and K. Umezawa. 1992. "Inhibition of Glucocerebrosidase and Induction of Neural Abnormality by Cyclophellitol in Mice." *Archives of Biochemistry and Biophysics* 297 (2): 362–67.

Baker, E.N. 1994. "Structure and Reactivity of Transferrins", in A.G. Sykes (ed.) *Advances in Inorganic Chemistry*. Academic Press, pp. 389–463.

Baker, E.N., Baker, H.M. and Kidd, R.D. 2002. 'Lactoferrin and transferrin: functional variations on a common structural framework', *Biochemistry and cell biology = Biochimie et biologie cellulaire*, 80(1), pp. 27–34.

Bartesaghi, Alberto, Cecilia Aguerrebere, Veronica Falconieri, Soojay Banerjee, Lesley A. Earl, Xing Zhu, Nikolaus Grigorieff, et al. 2018. "Atomic Resolution Cryo-EM Structure of β-Galactosidase." *Structure* 26 (6): 848–56.e3.

Bartesaghi, Alberto, Doreen Matthies, Soojay Banerjee, Alan Merk, and Sriram Subramaniam. 2014. "Structure of β-Galactosidase at 3.2-Å Resolution Obtained by Cryo-Electron Microscopy." *Proceedings of the National Academy of Sciences of the United States of America* 111 (32): 11709–14.

Bartesaghi, Alberto, Alan Merk, Soojay Banerjee, Doreen Matthies, Xiongwu Wu, Jacqueline L. S. Milne, and Sriram Subramaniam. 2015. "2.2 Å Resolution Cryo-EM Structure of β-Galactosidase in Complex with a Cell-Permeant Inhibitor." *Science* 348 (6239): 1147–51.

Bauer, E. 1994. "Low Energy Electron Microscopy." *Reports on Progress in Physics*. https://doi.org/10.1088/0034-4885/57/9/002.

Bednarek, Christin, Ilona Wehl, Nicole Jung, Ute Schepers, and Stefan Bräse. 2020. "The Staudinger Ligation." *Chemical Reviews* 120 (10): 4301–54.

Beenakker, Thomas J. M., Dennis P. A. Wander, Wendy A. Offen, Marta Artola, Lluís Raich, Maria J. Ferraz, Kah-Yee Li, et al. 2017. "Carba-Cyclophellitols Are Neutral Retaining-Glucosidase Inhibitors." *Journal of the American Chemical Society* 139 (19): 6534–37.

Belardinelli, Luiz, and A. Pelleg. 2012. *Adenosine and Adenine Nucleotides: From Molecular Biology to Integrative Physiology*. Springer Science & Business Media.

Ben Bdira, Fredj, Marta Artola, Herman S. Overkleeft, Marcellus Ubbink, and Johannes M. F. G. Aerts. 2018. "Distinguishing the Differences in β-Glycosylceramidase Folds, Dynamics, and Actions Informs Therapeutic Uses." *Journal of Lipid Research* 59 (12): 2262–76.

Benjamín-Rivera, Josué A., Andrés E. Cardona-Rivera, Ángel L. Vázquez-Maldonado, Christian Y. Dones-Lassalle, Héctor L. Pabón-Colon, Héctor M. Rodríguez-Rivera, Israel Rodríguez, et al. 2020a. "Exploring Serum Transferrin Regulation of Nonferric Metal Therapeutic Function and Toxicity." *Inorganics*. https://doi.org/10.3390/inorganics8090048.

Bewley, M. C., B. M. Tam, J. Grewal, S. He, S. Shewry, M. E. Murphy, A. B. Mason, R. C.

Woodworth, E. N. Baker, and R. T. MacGillivray. 1999. "X-Ray Crystallography and Mass Spectroscopy Reveal That the N-Lobe of Human Transferrin Expressed in Pichia Pastoris Is Folded Correctly but Is Glycosylated on Serine-32." *Biochemistry* 38 (8): 2535–41.

Blum, Galia, Steven H. L. Verhelst, and Xiaowei Ma. 2021. "Editorial: Development and Applications of New Activity-Based Probes." *Frontiers in Chemistry* 9 (September): 754294.

Boer, Casper de, Nicholas G. S. McGregor, Evert Peterse, Sybrin P. Schröder, Bogdan I. Florea, Jianbing Jiang, Jos Reijngoud, et al. 2021. "Correction: Glycosylated Cyclophellitol-Derived Activity-Based Probes and Inhibitors for Cellulases." *RSC Chemical Biology* 2 (6): 1701.

Bonsor, Daniel, Stephanie F. Butz, Julianna Solomons, Stephanie Grant, Ian J. S. Fairlamb, Mark J. Fogg, and Gideon Grogan. 2006. "Ligation Independent Cloning (LIC) as a Rapid Route to Families of Recombinant Biocatalysts from Sequenced Prokaryotic Genomes." *Organic and Biomolecular Chemistry* 4 (7): 1252–60.

Brás, Natércia F., Pedro A. Fernandes, and Maria J. Ramos. 2010. "QM/MM Studies on the β-Galactosidase Catalytic Mechanism: Hydrolysis and Transglycosylation Reactions." *Journal of Chemical Theory and Computation*. https://doi.org/10.1021/ct900530f.

Bren, Anat, Junyoung O. Park, Benjamin D. Towbin, Erez Dekel, Joshua D. Rabinowitz, and Uri Alon. 2016. "Glucose Becomes One of the Worst Carbon Sources for E.coli on Poor Nitrogen Sources due to Suboptimal Levels of cAMP." *Scientific Reports* 6 (April): 24834.

Bresciani, Alberto, Ottavia Cecchetti, Antonino Missineo, Pier Giorgio Pacifici, Licia Tomei, and Steven Rodems. 2018. "Development of a Broadly Applicable Assay for Measurement of Glycan-Directed Enzymatic Activity." SLAS Discovery : Advancing Life Sciences R & D 23 (9): 941–50.

Briggs, John B. 2017. "Glycan Characterization." In Analytical Characterization of Biotherapeutics, 257–332. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Brilot, Axel F., James Z. Chen, Anchi Cheng, Junhua Pan, Stephen C. Harrison, Clinton S. Potter, Bridget Carragher, Richard Henderson, and Nikolaus Grigorieff. 2012. "Beam-Induced Motion of Vitrified Specimen on Holey Carbon Film." *Journal of Structural Biology*. https://doi.org/10.1016/j.jsb.2012.02.003.

Brocchieri, Luciano, and Samuel Karlin. 2005. "Protein Length in Eukaryotic and Prokaryotic Proteomes." *Nucleic Acids Research* 33 (10): 3390–3400.

Brock, Jeremy H. 1994. *Iron Metabolism in Health and Disease*. W B Saunders Company.

Brooks, Harold B., Sandaruwan Geeganage, Steven D. Kahl, Chahrzad Montrose, Sitta Sittampalam, Michelle C. Smith, and Jeffrey R. Weidner. 2012. "Basics of Enzymatic Assays for HTS." In *Assay Guidance Manual*, edited by Sarine Markossian, Abigail Grossman, Kyle Brimacombe, Michelle Arkin, Douglas Auld, Chris Austin, Jonathan Baell, et al. Bethesda (MD): Eli Lilly and Company and the National Center for Advancing Translational Sciences.

Cabra, Vanessa, and Montserrat Samsó. 2015. "Do's and Don'ts of Cryo-Electron Microscopy: A Primer on Sample Preparation and High Quality Data Collection for Macromolecular 3D Reconstruction." *Journal of Visualized Experiments*. https://doi.org/10.3791/52311-v.

Campbell, Melody G., Anchi Cheng, Axel F. Brilot, Arne Moeller, Dmitry Lyumkis, David Veesler, Junhua Pan, et al. 2012. "Movies of Ice-Embedded Particles Enhance Resolution in Electron Cryo-Microscopy." *Structure* 20 (11): 1823–28.

Campos-Escamilla, Camila, Dritan Siliqi, Luis A. Gonzalez-Ramirez, Carmen Lopez-Sanchez,

Jose Antonio Gavira, and Abel Moreno. 2021. "X-Ray Characterization of Conformational Changes of Human Apo- and Holo-Transferrin." *International Journal of Molecular Sciences* 22 (24). https://doi.org/10.3390/ijms222413392.

Caner, Sami, Xiaohua Zhang, Jianbing Jiang, Hong-Ming Chen, Nham T. Nguyen, Hermen Overkleeft, Gary D. Brayer, and Stephen G. Withers. 2016. "Glucosyl Epi-Cyclophellitol Allows Mechanism-Based Inactivation and Structural Analysis of Human Pancreatic α-Amylase." *FEBS Letters* 590 (8): 1143–51.

Cannon, J. C., and N. D. Chasteen. 1975. "Nonequivalence of the Metal Binding Sites in Vanadyl-Labeled Human Serum Transferrin." *Biochemistry* 14 (21): 4573–77.

Carragher, B., Y. Cheng, A. Frost, R. M. Glaeser, G. C. Lander, E. Nogales, and H. -w. Wang. 2019. "Current Outcomes When Optimizing 'standard' Sample Preparation for Single-particle cryo-EM." *Journal of Microscopy*. https://doi.org/10.1111/jmi.12834.

Casañal, Ana, Bernhard Lohkamp, and Paul Emsley. 2020. "Current Developments in Coot for Macromolecular Model Building of Electron Cryo-Microscopy and Crystallographic Data." *Protein Science: A Publication of the Protein Society* 29 (4): 1069–78.

Castillo Busto, M. Estela del, Maria Montes-Bayón, Elisa Blanco-González, Juris Meija, and Alfredo Sanz-Medel. 2005. "Strategies to Study Human Serum Transferrin Isoforms Using Integrated Liquid Chromatography ICPMS, MALDI-TOF, and ESI-Q-TOF Detection: Application to Chronic Alcohol Abuse." *Analytical Chemistry* 77 (17): 5615–21.

Cheng, Yifan, Robert M. Glaeser, and Eva Nogales. 2017. "How Cryo-EM Became so Hot." *Cell* 171 (6): 1229–31.

Cheng, Yifan, Olga Zak, Philip Aisen, Stephen C. Harrison, and Thomas Walz. 2004. "Structure of the Human Transferrin Receptor-Transferrin Complex." *Cell* 116 (4): 565–76.

Cheng, Yifan, Olga Zak, Philip Aisen, Stephen C. Harrison, and Thomas Walz. 2005. "Single Particle Reconstruction of the Human Apo-Transferrin-Transferrin Receptor Complex." *Journal of Structural Biology* 152 (3): 204–10.

Chen, Vincent B., W. Bryan Arendall 3rd, Jeffrey J. Headd, Daniel A. Keedy, Robert M. Immormino, Gary J. Kapral, Laura W. Murray, Jane S. Richardson, and David C. Richardson. 2010. "MolProbity: All-Atom Structure Validation for Macromolecular Crystallography." *Acta Crystallographica. Section D, Biological Crystallography* 66 (Pt 1): 12–21.

Chen, Yurong, Zachary Armstrong, Marta Artola, Bogdan I. Florea, Chi-Lin Kuo, Casper de Boer, Mikkel S. Rasmussen, et al. 2021. "Activity-Based Protein Profiling of Retaining α-Amylases in Complex Biological Samples." *Journal of the American Chemical Society* 143 (5): 2423–32.

Cianfrocco, Michael A., Indrajit Lahiri, Frank DiMaio, and Andres E. Leschziner. 2018. "Cryoem-Cloud-Tools: A Software Platform to Deploy and Manage Cryo-EM Jobs in the Cloud." *Journal of Structural Biology* 203 (3): 230–35.

Clark, David P., Nanette J. Pazdernik, and Michelle R. McGehee. 2019. "Regulation of Transcription in Prokaryotes." *Molecular Biology*. https://doi.org/10.1016/b978-0-12-813288-3.00016-1.

Cohn, M. 1957. "Contributions of Studies on the Beta-Galactosidase of Escherichia Coli to Our Understanding of Enzyme Synthesis." *Bacteriological Reviews* 21 (3): 140–68.

Coudert, Elisabeth, Sebastien Gehant, Edouard de Castro, Monica Pozzato, Delphine Baratin, Teresa Neto, Christian J. A. Sigrist, Nicole Redaschi, Alan Bridge, and UniProt Consortium. 2023. "Annotation of Biologically Relevant Ligands in UniProtKB Using ChEBI." *Bioinformatics* 39 (1). https://doi.org/10.1093/bioinformatics/btac793.

Guerrero, Cecilia, Carlos Vera, Raúl Conejeros, and Andrés Illanes. 2015. "Transgalactosylation

and Hydrolytic Activities of Commercial Preparations of β-Galactosidase for the Synthesis of Prebiotic Carbohydrates." *Enzyme and Microbial Technology* 70 (March): 9–17.

Craig, Douglas B., and Anna Henderson. 2013. "Electrophoretic Heterogeneity Limits the Utility of Streptavidin-β-Galactosidase as a Probe in Free Zone Capillary Electrophoresis Separations." *The Protein Journal* 32 (2): 81–88.

Crispin, Max, A. Radu Aricescu, Veronica T. Chang, E. Yvonne Jones, David I. Stuart, Raymond A. Dwek, Simon J. Davis, and David J. Harvey. 2007. "Disruption of α-Mannosidase Processing Induces Non-Canonical Hybrid-Type Glycosylation." *FEBS Letters*. https://doi.org/10.1016/j.febslet.2007.04.020.

Danev, Radostin, Bart Buijsse, Maryam Khoshouei, Jürgen M. Plitzko, and Wolfgang Baumeister. 2014. "Volta Potential Phase Plate for in-Focus Phase Contrast Transmission Electron Microscopy." *Proceedings of the National Academy of Sciences of the United States of America* 111 (44): 15635–40.

Danev, Radostin, Dimitry Tegunov, and Wolfgang Baumeister. 2017. "Using the Volta Phase Plate with Defocus for Cryo-EM Single Particle Analysis." *eLife* 6 (January). https://doi.org/10.7554/eLife.23006.

Deffieu, Maika S., Ieva Cesonyte, François Delalande, Gaelle Boncompain, Cristina Dorobantu, Eli Song, Vincent Lucansky, et al. 2021a. "Rab7-Harboring Vesicles Are Carriers of the Transferrin Receptor through the Biosynthetic Secretory Pathway." *Science Advances* 7 (2). https://doi.org/10.1126/sciadv.aba7803.

Dialpuri, Jordan S., Haroldas Bagdonas, Lucy C. Schofield, Phuong Thao Pham, Lou Holland, Paul S. Bond, Filomeno Sánchez Rodríguez, Stuart J. McNicholas, and Jon Agirre. 2024. "Online Carbohydrate 3D Structure Validation with the Privateer Web App." Acta Crystallographica. Section F, Structural Biology and Crystallization Communications 80 (Pt 2): 30–35.

D'Imprima, Edoardo, Davide Floris, Mirko Joppe, Ricardo Sánchez, Martin Grininger, and Werner Kühlbrandt. 2019. "Protein Denaturation at the Air-Water Interface and How to Prevent It." *eLife*. https://doi.org/10.7554/elife.42747.

Doerr, Allison. 2015. "Cryo-EM Goes High-Resolution." *Nature Methods*. https://doi.org/10.1038/nmeth.3469.

Dorland, L., J. Haverkamp, B. L. Schut, and J. F. Vliegenthart. 1977. "The Structure of the Asialo-Carbohydrate Units of Human Serotransferrin as Proven by 360 MHz Proton Magnetic Resonance Spectroscopy." *FEBS Letters* 77 (1): 15–20.

Drula, Elodie, Marie-Line Garron, Suzan Dogan, Vincent Lombard, Bernard Henrissat, and Nicolas Terrapon. 2022. "The Carbohydrate-Active Enzyme Database: Functions and Literature." Nucleic Acids Research 50 (D1): D571–77.

Dubach, Victor R. A., and Albert Guskov. 2020. "The Resolution in X-Ray Crystallography and Single-Particle Cryogenic Electron Microscopy." *Crystals*. https://doi.org/10.3390/cryst10070580.

Echlin, Patrick. 2011. *Handbook of Sample Preparation for Scanning Electron Microscopy and X-Ray Microanalysis*. Springer Science and Business Media.

Egerton, R. F. 2011. "Electron Energy-Loss Spectroscopy in the Electron Microscope." https://doi.org/10.1007/978-1-4419-9583-4.

Eisenthal, Robert, Michael J. Danson, and David W. Hough. 2007. "Catalytic Efficiency and kcat/KM: A Useful Comparator?" *Trends in Biotechnology* 25 (6): 247–49.

Evans, Philip R. 2007. "An Introduction to Stereochemical Restraints." *Acta Crystallographica. Section D, Biological Crystallography* 63 (Pt 1): 58–61.

Evans, Philip R., and Garib N. Murshudov. 2013. "How Good Are My Data and What Is the

Resolution?" *Acta Crystallographica Section D Biological Crystallography*.
https://doi.org/10.1107/s0907444913000061.

Faruqi, Abdul Raffey, Richard Henderson, and Greg McMullan. 2014. "Direct Electron
Detectors for Electron Microscopy." *Proceedings of 22nd International Workshop on
Vertex Detectors — PoS(Vertex2013)*. https://doi.org/10.22323/1.198.0044.

Feja, B., and U. Aebi. 1999. "Determination of the Inelastic Mean Free Path of Electrons in
Vitrified Ice Layers for on-Line Thickness Measurements by Zero-Loss Imaging." *Journal
of Microscopy* 193 (1): 15–19.

Fernández, José-Jesús, Josér Sanjurjo, and José-María Carazo. 1997. "A Spectral Estimation
Approach to Contrast Transfer Function Detection in Electron Microscopy."
Ultramicroscopy 68 (4): 267–95.

Fisher, Peter, and Daniel Ungar. 2016. "Bridging the Gap between Glycosylation and Vesicle
Traffic." *Frontiers in Cell and Developmental Biology* 4 (March): 15.

Fogg, Mark J., and Anthony J. Wilkinson. 2008. "Higher-Throughput Approaches to
Crystallization and Crystal Structure Determination." *Biochemical Society Transactions*
36 (Pt 4): 771–75.

Fonović, Marko, and Matthew Bogyo. 2008. "Activity-Based Probes as a Tool for Functional
Proteomic Analysis of Proteases." *Expert Review of Proteomics* 5 (5): 721–30.

Fukuda, Yoshiyuki, Ulrike Laugks, Vladan Lučić, Wolfgang Baumeister, and Radostin Danev.
2015. "Electron Cryotomography of Vitrified Cells with a Volta Phase Plate." *Journal of
Structural Biology* 190 (2): 143–54.

Gammella, Elena, Paolo Buratti, Gaetano Cairo, and Stefania Recalcati. 2017. "The
Transferrin Receptor: The Cellular Iron Gate." *Metallomics*.
https://doi.org/10.1039/c7mt00143f.

Ganem, Bruce. 1996. "Inhibitors of Carbohydrate-Processing Enzymes: Design and Synthesis
of Sugar-Shaped Heterocycles." *Accounts of Chemical Research*.
https://doi.org/10.1021/ar9502184.

Gebler, J. C., R. Aebersold, and S. G. Withers. 1992. "Glu-537, Not Glu-461, Is the
Nucleophile in the Active Site of (lac Z) Beta-Galactosidase from Escherichia Coli." *The
Journal of Biological Chemistry* 267 (16): 11126–30.

Gentilucci, Luca, Pierluigi Tosi, Adriano Bauer, and Rossella De Marco. 2016. "Modern Tools
for the Chemical Ligation and Synthesis of Modified Peptides and Proteins." Future
Medicinal Chemistry 8 (18): 2287–2304.

Ghalanbor, Zahra, Nasser Ghaemi, Sayed-Amir Marashi, Massoud Amanlou, Mehran Habibi-
Rezaei, Khosro Khajeh, and Bijan Ranjbar. 2008. "Binding of Tris to Bacillus
Licheniformis Alpha-Amylase Can Affect Its Starch Hydrolysis Activity." *Protein and
Peptide Letters* 15 (2): 212–14.

Glaeser, Robert M. 2018. "PROTEINS, INTERFACES, AND CRYO-EM GRIDS." *Current Opinion
in Colloid and Interface Science* 34 (March): 1–8.

Gloster, Tracey M., Robert Madsen, and Gideon J. Davies. 2007. "Structural Basis for
Cyclophellitol Inhibition of a Beta-Glucosidase." *Organic and Biomolecular Chemistry* 5
(3): 444–46.

Gloster, Tracey M., Peter Meloncelli, Robert V. Stick, David Zechel, Andrea Vasella, and
Gideon J. Davies. 2007. "Glycosidase Inhibition: An Assessment of the Binding of 18
Putative Transition-State Mimics." *Journal of the American Chemical Society* 129 (8):
2345–54.

Goddard-Borger, Ethan D., Tom Wennekes, and Stephen G. Withers. 2010. "Molecular
Probes: Getting Lucky in the Lysosome." *Nature Chemical Biology*.

Gonzalo, Philippe, Sylvie Radenne, and Sylvie Gonzalo. 2014. "Biomarkers of Chronic Alcohol Misuse." *Current Biomarker Findings*, January, 9.

Gordon, N. 2000. "Classic Diseases Revisited: Carbohydrate-Deficient Glycoprotein Syndromes." *Postgraduate Medical Journal*. https://doi.org/10.1136/pmj.76.893.145.

Grigorieff, N. 2016. "Frealign: An Exploratory Tool for Single-Particle Cryo-EM." *Methods in Enzymology* 579 (June): 191–226.

Grimm, R., D. Typke, M. Bärmann, and W. Baumeister. 1996. "Determination of the Inelastic Mean Free Path in Ice by Examination of Tilted Vesicles and Automated Most Probable Loss Imaging." *Ultramicroscopy* 63 (3-4): 169–79.

Gruszczyk, Jakub, Rick K. Huang, Li-Jin Chan, Sébastien Menant, Chuan Hong, James M. Murphy, Yee-Foong Mok, et al. 2018. "Cryo-EM Structure of an Essential Plasmodium Vivax Invasion Complex." *Nature* 559 (7712): 135–39.

Guo, Hui, Erik Franken, Yuchen Deng, Samir Benlekbir, Garbi Singla Lezcano, Bart Janssen, Lingbo Yu, Zev A. Ripstein, Yong Zi Tan, and John L. Rubinstein. 2020. "Electron-Event Representation Data Enable Efficient cryoEM File Storage with Full Preservation of Spatial and Temporal Resolution." *IUCrJ* 7 (Pt 5): 860–69.

Gupta, A., N. B. Amin, A. Besarab, S. E. Vogel, G. W. Divine, J. Yee, and J. V. Anandan. 1999. "Dialysate Iron Therapy: Infusion of Soluble Ferric Pyrophosphate via the Dialysate during Hemodialysis." *Kidney International* 55 (5): 1891–98.

Hall, D. R., J. M. Hadden, G. A. Leonard, S. Bailey, M. Neu, M. Winn, and P. F. Lindley. 2002. "The Crystal and Molecular Structures of Diferric Porcine and Rabbit Serum Transferrins at Resolutions of 2.15 and 2.60 A, Respectively." *Acta Crystallographica. Section D, Biological Crystallography* 58 (Pt 1): 70–80.

Hamilton, David H., Isabelle Turcot, Alain Stintzi, and Kenneth N. Raymond. 2004. "Large Cooperativity in the Removal of Iron from Transferrin at Physiological Temperature and Chloride Ion Concentration." *Journal of Biological Inorganic Chemistry: JBIC: A Publication of the Society of Biological Inorganic Chemistry* 9 (8): 936–44.

Hao, Liyuan, Qiang Shan, Jingya Wei, Fengtao Ma, and Peng Sun. 2018. "Lactoferrin: Major Physiological Functions and Applications." *Current Protein and Peptide Science*. https://doi.org/10.2174/1389203719666180514150921.

Henderson, R. 1995. "The Potential and Limitations of Neutrons, Electrons and X-Rays for Atomic Resolution Microscopy of Unstained Biological Molecules." *Quarterly Reviews of Biophysics* 28 (2): 171–93.

Henrissat, B., I. Callebaut, S. Fabrega, P. Lehn, J. P. Mornon, and G. Davies. 1995. "Conserved Catalytic Machinery and the Prediction of a Common Fold for Several Families of Glycosyl Hydrolases." *Proceedings of the National Academy of Sciences of the United States of America* 92 (15): 7090–94.

Hermans, M. M., M. A. Kroos, J. van Beeumen, B. A. Oostra, and A. J. Reuser. 1991. "Human Lysosomal Alpha-Glucosidase. Characterization of the Catalytic Site." *The Journal of Biological Chemistry* 266 (21): 13507–12.

Herrchen, Monika, and Gunter Legler. 1984. "Identification of an Essential Carboxylate Group at the Active Site of lacZ Beta-Galactosidase from Escherichia Coli." *European Journal of Biochemistry*. https://doi.org/10.1111/j.1432-1033.1984.tb07947.x.

Herzik, Mark A., Jr, Mengyu Wu, and Gabriel C. Lander. 2019. "High-Resolution Structure Determination of Sub-100 kDa Complexes Using Conventional Cryo-EM." *Nature Communications* 10 (1): 1032.

Hofmann, Susanne, Dovile Januliene, Ahmad R. Mehdipour, Christoph Thomas, Erich Stefan, Stefan Brüchert, Benedikt T. Kuhn, et al. 2019. "Conformation Space of a Heterodimeric

ABC Exporter under Turnover Conditions." *Nature* 571 (7766): 580–83.

Hoh, Soon Wen, Tom Burnley, and Kevin Cowtan. 2020. "Current Approaches for Automated Model Building into Cryo-EM Maps Using Buccaneer with CCP-EM." *Acta Crystallographica. Section D, Structural Biology* 76 (Pt 6): 531–41.

Huber, R. E., G. Kurz, and K. Wallenfels. 1976. "A Quantitation of the Factors Which Affect the Hydrolase and Transgalactosylase Activities of Beta-Galactosidase (*E. coli*) on Lactose." *Biochemistry* 15 (9): 1994–2001.

Inkson, B. J. 2016. "Scanning Electron Microscopy (SEM) and Transmission Electron Microscopy (TEM) for Materials Characterization." *Materials Characterization Using Nondestructive Evaluation (NDE) Methods*. https://doi.org/10.1016/b978-0-08-100040-3.00002-x.

Jacob, F., and J. Monod. 1961. "Genetic Regulatory Mechanisms in the Synthesis of Proteins." *Journal of Molecular Biology* 3 (June): 318–56.

Jacobson, R. H., X. J. Zhang, R. F. DuBose, and B. W. Matthews. 1994. "Three-Dimensional Structure of Beta-Galactosidase from *E. coli*." *Nature* 369 (6483): 761–66.

Jaeger, Joachim. 2005. "Macromolecular Structure Determination by X-ray Crystallography." *eLS*. https://doi.org/10.1038/npg.els.0002723.

Jaeken, J. 1995. "Carbohydrate-Deficient Glycoprotein Syndromes." *Inborn Metabolic Diseases*. https://doi.org/10.1007/978-3-662-03147-6_38.

Jancewicz, Larisa J., Robert W. Wheatley, Gopinath Sutendra, Michael Lee, Marie E. Fraser, and Reuben E. Huber. 2012. "Ser-796 of β-Galactosidase (Escherichia Coli) Plays a Key Role in Maintaining a Balance between the Opened and Closed Conformations of the Catalytically Important Active Site Loop." *Archives of Biochemistry and Biophysics*. https://doi.org/10.1016/j.abb.2011.11.017.

Jong, G. de, and H. G. van Eijk. 1988. "Microheterogeneity of Human Serum Transferrin: A Biological Phenomenon Studied by Isoelectric Focusing in Immobilized pH Gradients." *Electrophoresis* 9 (9): 589–98.

Juers, Douglas H, T. D. Heightman, A. Vasella, J. D. McCarter, L. Mackenzie, S. G. Withers, and B. W. Matthews. 2001. "A Structural View of the Action of Escherichia Coli (lacZ) Beta-Galactosidase." *Biochemistry* 40 (49): 14781–94.

Juers, Douglas H, R. H. Jacobson, D. Wigley, X. J. Zhang, R. E. Huber, D. E. Tronrud, and B. W. Matthews. 2000. "High Resolution Refinement of Beta-Galactosidase in a New Crystal Form Reveals Multiple Metal-Binding Sites and Provides a Structural Basis for Alpha-Complementation." *Protein Science: A Publication of the Protein Society* 9 (9): 1685–99.

Juers, Douglas H. 2000. *A Structural View of Beta-Galactosidase in Action*.

Juers, Douglas H., Brian W. Matthews, and Reuben E. Huber. 2012. "*LacZ*β-Galactosidase: Structure and Function of an Enzyme of Historical and Molecular Biological Importance." *Protein Science* 21 (12): 1792–1807.

Kallemeijn, Wouter W., Kah-Yee Li, Martin D. Witte, André R. A. Marques, Jan Aten, Saskia Scheij, Jianbing Jiang, et al. 2012. "Novel Activity-Based Probes for Broad-Spectrum Profiling of Retaining β-Exoglucosidases in Situ and in Vivo." *Angewandte Chemie* 51 (50): 12529–33.

Kamiya, Mako, Hisataka Kobayashi, Yukihiro Hama, Yoshinori Koyama, Marcelino Bernardo, Tetsuo Nagano, Peter L. Choyke, and Yasuteru Urano. 2007. "An Enzymatically Activated Fluorescence Probe for Targeted Tumor Imaging." *Journal of the American Chemical Society* 129 (13): 3918–29.

Karimian Amroabadi, Marzieh, Asghar Taheri-Kafrani, Leily Heidarpoor Saremi, and Ali Asghar Rastegari. 2018. "Spectroscopic Studies of the Interaction between Alprazolam

and Apo-Human Serum Transferrin as a Drug Carrier Protein." *International Journal of Biological Macromolecules* 108 (March): 263–71.

Kato, Takayuki, Fumiaki Makino, Takanori Nakane, Naoya Terahara, Takeshi Kaneko, Yuko Shimizu, Sohei Motoki, Isamu Ishikawa, Koji Yonekura, and Keiichi Namba. 2019. "CryoTEM with a Cold Field Emission Gun That Moves Structural Biology into a New Stage." *Microscopy and Microanalysis*. https://doi.org/10.1017/s1431927619005725.

Kell, Douglas B., Eugene L. Heyden, and Etheresia Pretorius. 2020. "The Biology of Lactoferrin, an Iron-Binding Protein That Can Help Defend Against Viruses and Bacteria." *Frontiers in Immunology* 11 (May): 1221.

Kern, David M., Ben Sorum, Sonali S. Mali, Christopher M. Hoel, Savitha Sridharan, Jonathan P. Remis, Daniel B. Toso, Abhay Kotecha, Diana M. Bautista, and Stephen G. Brohawn. 2021. "Author Correction: Cryo-EM Structure of SARS-CoV-2 ORF3a in Lipid Nanodiscs." *Nature Structural and Molecular Biology* 28 (8): 702.

Kesidis, Athanasios, Peer Depping, Alexis Lodé, Afroditi Vaitsopoulou, Roslyn M. Bill, Alan D. Goddard, and Alice J. Rothnie. 2020. "Expression of Eukaryotic Membrane Proteins in Eukaryotic and Prokaryotic Hosts." *Methods* 180 (August): 3–18.

Kjeldsen, Christian, Jan Henrik Ardenkjær-Larsen, and Jens Ø. Duus. 2018. "Discovery of Intermediates of lacZ β-Galactosidase Catalyzed Hydrolysis Using dDNP NMR." *Journal of the American Chemical Society* 140 (8): 3030–34.

Klausner, R. D., G. Ashwell, J. van Renswoude, J. B. Harford, and K. R. Bridges. 1983. "Binding of Apotransferrin to K562 Cells: Explanation of the Transferrin Cycle." *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.80.8.2263.

Knoll, M., and E. Ruska. 1932. "Das Elektronenmikroskop." *Zeitschrift F�r Physik*. https://doi.org/10.1007/bf01330526.

Köhn, Maja, and Rolf Breinbauer. 2004. "The Staudinger Ligation-a Gift to Chemical Biology." *Angewandte Chemie* 43 (24): 3106–16.

Kontziampasis, Dimitrios, David P. Klebl, Matthew G. Iadanza, Charlotte A. Scarff, Florian Kopf, Frank Sobott, Diana C. F. Monteiro, Martin Trebbin, Stephen P. Muench, and Howard D. White. 2019. "A Cryo-EM Grid Preparation Device for Time-Resolved Structural Studies." *IUCrJ* 6 (Pt 6): 1024–31.

Koshland, D. E. 1953. "STEREOCHEMISTRY AND THE MECHANISM OF ENZYMATIC REACTIONS." *Biological Reviews*. https://doi.org/10.1111/j.1469-185x.1953.tb01386.x.

Krasnova, Larissa, and Chi-Huey Wong. 2016. "Understanding the Chemistry and Biology of Glycosylation with Glycan Synthesis." *Annual Review of Biochemistry* 85 (June): 599–630.

Kristiansen, Troels Zakarias, Jakob Bunkenborg, Mads Gronborg, Henrik Molina, Paul J. Thuluvath, Pedram Argani, Michael G. Goggins, Anirban Maitra, and Akhilesh Pandey. 2004. "A Proteomic Analysis of Human Bile." *Molecular and Cellular Proteomics*. https://doi.org/10.1074/mcp.m400015-mcp200.

Krivanek, O. L., S. L. Friedman, A. J. Gubbens, and B. Kraus. 1995. "An Imaging Filter for Biological Applications." *Ultramicroscopy*. https://doi.org/10.1016/0304-3991(95)00034-x.

Kroos, M. J., J. S. Starreveld, C. E. Verrijt, H. G. van Eijk, and J. P. van Dijk. 1996. "Regulation of Transferrin Receptor Synthesis by Human Cytotrophoblast Cells in Culture." *European Journal of Obstetrics, Gynecology, and Reproductive Biology* 65 (2): 231–34.

Kuo, Chi-Lin, Eline van Meel, Kassiani Kytidou, Wouter Willem Kallemeijn, Martin Witte, Herman Stephen Overkleeft, Marta Elena Artola, and Johannes Maria Aerts. 2018. "Activity-Based Probes for Glycosidases: Profiling and Other Applications." *Chemical*

Glycobiology Part B. Monitoring Glycans and Their Interactions.
https://doi.org/10.1016/bs.mie.2017.06.039.

Laemmli, U. K. 1970. "Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4." *Nature* 227 (5259): 680–85.

Lambert, Lisa A. 2012. "Molecular Evolution of the Transferrin Family and Associated Receptors." *Biochimica et Biophysica Acta* 1820 (3): 244–55.

Lammens, Willem, Katrien Le Roy, Lindsey Schroeven, André Van Laere, Anja Rabijns, and Wim Van den Ende. 2009. "Structural Insights into Glycoside Hydrolase Family 32 and 68 Enzymes: Functional Implications." *Journal of Experimental Botany* 60 (3): 727–40.

Lawrence.1999. "Ray S. Babyonyshev M. Galluser R. Borhani DW Harrison SC Crystal Structure of the Ectodomain of Human Transferrin Receptor." *Science*.

Lee, Byonghoon. 2017. "Extremophilic Carbohydrate Active Enzymes (CAZymes)." *Journal of Nutritional Health and Food Engineering*.
https://doi.org/10.15406/jnhfe.2017.07.00230.

Legler, Günter. 1990. "Glycoside Hydrolases: Mechanistic Information from Studies with Reversible and Irreversible Inhibitors." *Advances in Carbohydrate Chemistry and Biochemistry*. https://doi.org/10.1016/s0065-2318(08)60034-7.

Levina, Aviva, and Peter A. Lay. 2019a. "Transferrin Cycle and Clinical Roles of Citrate and Ascorbate in Improved Iron Metabolism." *ACS Chemical Biology* 14 (5): 893–900.

Liebschner, Dorothee, Pavel V. Afonine, Matthew L. Baker, Gábor Bunkóczi, Vincent B. Chen, Tristan I. Croll, Bradley Hintze, et al. 2019. "Macromolecular Structure Determination Using X-Rays, Neutrons and Electrons: Recent Developments in Phenix." *Acta Crystallographica. Section D, Structural Biology* 75 (Pt 10): 861–77.

Li, H., Sadler, P.J. and Sun, H. 1996. 'Rationalization of the strength of metal binding to human serum transferrin', *European journal of biochemistry / FEBS*, 242(2), pp. 387–393.

Li, Kah-Yee, Jianbing Jiang, Martin D. Witte, Wouter W. Kallemeijn, Wilma E. Donker-Koopman, Rolf G. Boot, Johannes M. F. G. Aerts, Jeroen D. C. Codée, Gijsbert A. van der Marel, and Herman S. Overkleeft. 2014. "Exploring Functional Cyclophellitol Analogues as Human Retaining Beta-Glucosidase Inhibitors." *Organic and Biomolecular Chemistry* 12 (39): 7786–91.

Lombard, Vincent, Hemalatha Golaconda Ramulu, Elodie Drula, Pedro M. Coutinho, and Bernard Henrissat. 2014. "The Carbohydrate-Active Enzymes Database (CAZy) in 2013." *Nucleic Acids Research* 42 (Database issue): D490–95.

Luck, Ashley N., and Anne B. Mason. 2012. "Transferrin-Mediated Cellular Iron Delivery." *Metal Transporters*. https://doi.org/10.1016/b978-0-12-394390-3.00001-x.

Lunin, V. Yu, V. Yu. Lunin, and M. M. Woolfson. 1993. "Mean Phase Error and the Map-Correlation Coefficient." *Acta Crystallographica Section D Biological Crystallography*.
https://doi.org/10.1107/s0907444993005852.

MacGillivray, R. T., E. Mendez, J. G. Shewale, S. K. Sinha, J. Lineback-Zins, and K. Brew. 1983. "The Primary Structure of Human Serum Transferrin. The Structures of Seven Cyanogen Bromide Fragments and the Assembly of the Complete Structure." *Journal of Biological Chemistry*. https://doi.org/10.1016/s0021-9258(18)32696-6.

MacGillivray, R. T., E. Mendez, S. K. Sinha, M. R. Sutton, J. Lineback-Zins, and K. Brew. 1982. "The Complete Amino Acid Sequence of Human Serum Transferrin." *Proceedings of the National Academy of Sciences of the United States of America* 79 (8): 2504–8.

MacGillivray, R. T., S. A. Moore, J. Chen, B. F. Anderson, H. Baker, Y. Luo, M. Bewley, et al. 1998. "Two High-Resolution Crystal Structures of the Recombinant N-Lobe of Human

Transferrin Reveal a Structural Change Implicated in Iron Release." *Biochemistry* 37 (22): 7919–28.

Malac, Marek, Simon Hettler, Misa Hayashida, Emi Kano, Ray F. Egerton, and Marco Beleggia. 2021. "Phase Plates in the Transmission Electron Microscope: Operating Principles and Applications." *Microscopy*. https://doi.org/10.1093/jmicro/dfaa070.

Matthews, Brian W. 2005. "The Structure of *E. coli* Beta-Galactosidase." *Comptes Rendus Biologies* 328 (6): 549–56.

McCarter, J. D., and S. G. Withers. 1994. "Mechanisms of Enzymatic Glycoside Hydrolysis." *Current Opinion in Structural Biology* 4 (6): 885–92.

McGregor, Nicholas G. S., Chi-Lin Kuo, Thomas J. M. Beenakker, Chun-Sing Wong, Wendy A. Offen, Zachary Armstrong, Bogdan I. Florea, et al. 2022. "Synthesis of Broad-Specificity Activity-Based Probes for -β-Mannosidases." *Organic and Biomolecular Chemistry* 20 (4): 877–86.

McKie, Andrew T. 2005. "A Ferrireductase Fills the Gap in the Transferrin Cycle." *Nature Genetics*.

McMullan, G., S. Chen, R. Henderson, and A. R. Faruqi. 2009. "Detective Quantum Efficiency of Electron Area Detectors in Electron Microscopy." *Ultramicroscopy* 109 (9): 1126–43.

McNicholas, S., E. Potterton, K. S. Wilson, and M. E. M. Noble. 2011. "Presenting Your Structures: The CCP4mg Molecular-Graphics Software." *Acta Crystallographica. Section D, Biological Crystallography* 67 (Pt 4): 386–94.

McNicholas, Stuart, and Jon Agirre. 2017. "Glycoblocks: A Schematic Three-Dimensional Representation for Glycans and Their Interactions." *Acta Crystallographica. Section D, Structural Biology* 73 (Pt 2): 187–94.

Mendez, Joshua, and Scott Stagg. 2018. "Pros and Cons of Integrating Versus Counting with a Direct Electron Detecting Camera." *Microscopy and Microanalysis*. https://doi.org/10.1017/s1431927618004932.

Merk, Alan, Takuma Fukumura, Xing Zhu, Joseph E. Darling, Reinhard Grisshammer, Jana Ognjenovic, and Sriram Subramaniam. 2020. "1.8 Å Resolution Structure of β-Galactosidase with a 200 keV CRYO ARM Electron Microscope." *IUCrJ* 7 (Pt 4): 639–43.

Mesters, Jeroen R., and Rolf Hilgenfeld. 2007. "Protein Glycosylation, Sweet to Crystal Growth?" *Crystal Growth and Design* 7 (11): 2251–53.

Meyerson, Joel R., Prashant Rao, Janesh Kumar, Sagar Chittori, Soojay Banerjee, Jason Pierson, Mark L. Mayer, and Sriram Subramaniam. 2014. "Self-Assembled Monolayers Improve Protein Distribution on Holey Carbon Cryo-EM Supports." *Scientific Reports* 4 (November): 7084.

Miao, S., J. D. McCarter, M. E. Grace, G. A. Grabowski, R. Aebersold, and S. G. Withers. 1994. "Identification of Glu340 as the Active-Site Nucleophile in Human Glucocerebrosidase by Use of Electrospray Tandem Mass Spectrometry." *The Journal of Biological Chemistry* 269 (15): 10975–78.

Mills, Deryck J. 2021. "Setting up and Operating a Cryo-EM Laboratory." *Quarterly Reviews of Biophysics* 54 (January): e2.

Milne, Jacqueline L. S., Mario J. Borgnia, Alberto Bartesaghi, Erin E. H. Tran, Lesley A. Earl, David M. Schauder, Jeffrey Lengyel, Jason Pierson, Ardan Patwardhan, and Sriram Subramaniam. 2013. "Cryo-Electron Microscopy - a Primer for the Non-Microscopist." *FEBS Journal*. https://doi.org/10.1111/febs.12078.

Moore, S. A., B. F. Anderson, C. R. Groom, M. Haridas, and E. N. Baker. 1997. "Three-Dimensional Structure of Diferric Bovine Lactoferrin at 2.8 A Resolution." *Journal of Molecular Biology* 274 (2): 222–36.

Moremen, Kelley W., Michael Tiemeyer, and Alison V. Nairn. 2012. "Vertebrate Protein Glycosylation: Diversity, Synthesis and Function." *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm3383.

Murakami, Yuta, Kiyoshi Saito, Hiromi Ito, and Yasuhiro Hashimoto. 2019. "Transferrin Isoforms in Cerebrospinal Fluid and Their Relation to Neurological Diseases." *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences* 95 (5): 198–210.

Murata, Kazuyoshi, Xiangan Liu, Radostin Danev, Joanita Jakana, Michael F. Schmid, Jonathan King, Kuniaki Nagayama, and Wah Chiu. 2010. "Zernike Phase Contrast Cryo-Electron Microscopy and Tomography for Structure Determination at Nanometer and Subnanometer Resolutions." *Structure* 18 (8): 903–12.

Murata, Kazuyoshi, and Matthias Wolf. 2018a. "Cryo-Electron Microscopy for Structural Analysis of Dynamic Biological Macromolecules." *Biochimica et Biophysica Acta (BBA) - General Subjects*. https://doi.org/10.1016/j.bbagen.2017.07.020.

Nakane, Takanori, Abhay Kotecha, Andrija Sente, Greg McMullan, Simonas Masiulis, Patricia M. G. Brown, Ioana T. Grigoras, et al. 2020. "Single-Particle Cryo-EM at Atomic Resolution." *Nature*, October. https://doi.org/10.1101/2020.05.22.110189.

National Center for Biotechnology Information. 2024a. PubChem Compound Summary for CID 109374151, Ferric Pyrophosphate Citrate. Retrieved March 21, 2024 from https://pubchem.ncbi.nlm.nih.gov/compound/Ferric-Pyrophosphate-Citrate.

National Center for Biotechnology Information.2024b. PubChem Compound Summary for CID 118984355, Ferric Ammonium Citrate. Retrieved March 21, 2024 from https://pubchem.ncbi.nlm.nih.gov/compound/118984355.

Naydenova, Katerina, and Christopher J. Russo. 2017. "Measuring the Effects of Particle Orientation to Improve the Efficiency of Electron Cryomicroscopy." *Nature Communications*. https://doi.org/10.1038/s41467-017-00782-3.

Nguyen, Nguyen Phuoc, Ilker Ersoy, Jacob Gotberg, Filiz Bunyak, and Tommi A. White. 2021. "DRPnet: Automated Particle Picking in Cryo-Electron Micrographs Using Deep Regression." *BMC Bioinformatics* 22 (1): 55.

Nicholls, Robert A., Michal Tykac, Oleg Kovalevskiy, and Garib N. Murshudov. 2018. "Current Approaches for the Fitting and Refinement of Atomic Models into Cryo-EM Maps Using CCP-EM." *Acta Crystallographica. Section D, Structural Biology* 74 (Pt 6): 492–505.

Noble, Alex J., Hui Wei, Venkata P. Dandey, Zhening Zhang, Yong Zi Tan, Clinton S. Potter, and Bridget Carragher. 2018. "Reducing Effects of Particle Adsorption to the Air-Water Interface in Cryo-EM." *Nature Methods* 15 (10): 793–95.

Nogales, Eva. 2018. "Profile of Joachim Frank, Richard Henderson, and Jacques Dubochet, 2017 Nobel Laureates in Chemistry." *Proceedings of the National Academy of Sciences of the United States of America* 115 (3): 441–44.

Nogales, Eva, and Sjors H. W. Scheres. 2015. "Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity." *Molecular Cell* 58 (4): 677–89.

Noinaj, Nicholas, Nicole C. Easley, Muse Oke, Naoko Mizuno, James Gumbart, Evzen Boura, Ashley N. Steere, et al. 2012. "Structural Basis for Iron Piracy by Pathogenic Neisseria." *Nature* 483 (7387): 53–58.

Notredame, C., D. G. Higgins, and J. Heringa. 2000. "T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment." *Journal of Molecular Biology* 302 (1): 205–17.

Nute, Michael, Ehsan Saleh, and Tandy Warnow. 2019. "Evaluating Statistical Multiple Sequence Alignment in Comparison to Other Alignment Methods on Protein Data Sets." Systematic Biology 68 (3): 396–411.

Odziomek, Katarzyna, Daniela Ushizima, Przemyslaw Oberbek, Krzysztof Jan Kurzydłowski, Tomasz Puzyn, and Maciej Haranczyk. 2017. "Scanning Electron Microscopy Image Representativeness: Morphological Data on Nanoparticles." *Journal of Microscopy*. https://doi.org/10.1111/jmi.12461.

Ohgami, Robert S., Dean R. Campagna, Eric L. Greer, Brendan Antiochos, Alice McDonald, Jing Chen, John J. Sharp, Yuko Fujiwara, Jane E. Barker, and Mark D. Fleming. 2005. "Identification of a Ferrireductase Required for Efficient Transferrin-Dependent Iron Uptake in Erythroid Cells." *Nature Genetics* 37 (11): 1264–69.

Ohtsubo, Kazuaki, and Jamey D. Marth. 2006. "Glycosylation in Cellular Mechanisms of Health and Disease." *Cell* 126 (5): 855–67.

Park, I., E. Schaeffer, A. Sidoli, F. E. Baralle, G. N. Cohen, and M. M. Zakin. 1985. "Organization of the Human Transferrin Gene: Direct Evidence That It Originated by Gene Duplication." *Proceedings of the National Academy of Sciences of the United States of America* 82 (10): 3149–53.

Pascreau, Tiffany, Claire Auditeau, and Delphine Borgel. 2023. "Hemostatic Defects in Congenital Disorders of Glycosylation." *Research and Practice in Thrombosis and Haemostasis* 7 (3): 100142.

Passmore, L. A., and C. J. Russo. 2016. "Specimen Preparation for High-Resolution Cryo-EM." Methods in Enzymology 579 (June): 51–86.

Pelley, John W. 2011. Elsevier's Integrated Review Biochemistry E-Book: With STUDENT CONSULT Online Access. Elsevier Health Sciences.

Perello, Edward. 2018. "CRISPR Genome Editing: A Technical & Policy Primer." jbox.gmu.edu. http://jbox.gmu.edu/bitstream/handle/1920/11339/Perello_final.pdf?sequence=1.

Pettersen, Eric F., Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. 2004. "UCSF Chimera--a Visualization System for Exploratory Research and Analysis." *Journal of Computational Chemistry* 25 (13): 1605–12.

Pinho, Salomé S., and Celso A. Reis. 2015. "Glycosylation in Cancer: Mechanisms and Clinical Implications." *Nature Reviews. Cancer* 15 (9): 540–55.

Pratt, Raymond, Garry J. Handelman, Thomas E. Edwards, and Ajay Gupta. 2018. "Ferric Pyrophosphate Citrate: Interactions with Transferrin." *Biometals: An International Journal on the Role of Metal Ions in Biology, Biochemistry, and Medicine* 31 (6): 1081–89.

Price, Morgan N., Adam P. Arkin, and Eric J. Alm. 2006. "The Life-Cycle of Operons." *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.0020096.

Princiotto, J. V., and E. J. Zapolski. 1975. "Difference between the Two Iron-Binding Sites of Transferrin." *Nature* 255 (5503): 87–88.

Ptashne, Mark. 2004. *A Genetic Switch: Phage Lambda Revisited*. CSHL Press.

Ptashne, Mark, and Alexander Gann. 2002. *Genes and Signals*. CSHL Press.

Punjani, Ali, Ali Punjani, John Rubinstein, David Fleet, and Marcus Brubaker. 2017. "Protocol for Rapid Unsupervised Cryo-EM Structure Determination Using cryoSPARC Software." *Protocol Exchange*. https://doi.org/10.1038/protex.2017.009.

Quaroni, A., E. Gershon, and G. Semenza. 1974. "Affinity Labeling of the Active Sites in the Sucrase-Isomaltase Complex from Small Intestine." *The Journal of Biological Chemistry* 249 (20): 6424–33.

Ravelli, Raimond B. G., Frank J. T. Nijpels, Rene J. M. Henderikx, Giulia Weissenberger, Sanne Thewessem, Abril Gijsbers, Bart W. A. M. M. Beulen, Carmen López-Iglesias, and Peter

J. Peters. 2020. "Cryo-EM Structures from Sub-Nl Volumes Using Pin-Printing and Jet Vitrification." *Nature Communications* 11 (1): 2563.

Rayleigh, John William Strutt. 1879. *Investigations in Optics, with Special Reference to the Spectroscope*.

Rempel, Brian P., and Stephen G. Withers. 2008. "Covalent Inhibitors of Glycosidases and Their Applications in Biochemistry and Biology." *Glycobiology* 18 (8): 570–86.

Renaud, Jean-Paul, Ashwin Chari, Claudio Ciferri, Wen-Ti Liu, Hervé-William Rémigy, Holger Stark, and Christian Wiesmann. 2018. "Cryo-EM in Drug Discovery: Achievements, Limitations and Prospects." *Nature Reviews. Drug Discovery* 17 (7): 471–92.

Rohou, Alexis, and Nikolaus Grigorieff. 2015. "CTFFIND4: Fast and Accurate Defocus Estimation from Electron Micrographs." *Journal of Structural Biology* 192 (2): 216–21.

Romero-Campero, Francisco José, and Mario J. Pérez-Jiménez. 2008. "Modelling Gene Expression Control Using P Systems: The Lac Operon, a Case Study." *Bio Systems* 91 (3): 438–57.

Rooijen, J. J. van, U. Jeschke, J. P. Kamerling, and J. F. Vliegenthart. 1998. "Expression of N-Linked Sialyl Le(x) Determinants and O-Glycans in the Carbohydrate Moiety of Human Amniotic Fluid Transferrin during Pregnancy." *Glycobiology* 8 (11): 1053–64.

Rosenthal, Peter B., and Richard Henderson. 2003. "Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-Particle Electron Cryomicroscopy." *Journal of Molecular Biology* 333 (4): 721–45.

Rouwenhorst, R. J., J. T. Pronk, and J. P. van Dijken. 1989. "The Discovery of β-Galactosidase." *Trends in Biochemical Sciences* 14 (10): 416–18.

Sadaghiani, Amir M., Steven Hl Verhelst, and Matthew Bogyo. 2007. "Tagging and Detection Strategies for Activity-Based Proteomics." *Current Opinion in Chemical Biology* 11 (1): 20–28.

Safarian, S., A. Hahn, D. J. Mills, M. Radloff, M. L. Eisinger, A. Nikolaev, J. Meier-Credo, et al. 2019. "Active Site Rearrangement and Structural Divergence in Prokaryotic Respiratory Oxidases." *Science* 366 (6461): 100–104.

Sass-Kuhn, S. P., R. Moqbel, J. A. Mackay, O. Cromwell, and A. B. Kay. 1984. "Human Granulocyte/pollen-Binding Protein. Recognition and Identification as Transferrin." *The Journal of Clinical Investigation* 73 (1): 202–10.

Satomi, Yoshinori, Yasutsugu Shimonishi, Toshiharu Hase, and Toshifumi Takao. 2004. "Site-Specific Carbohydrate Profiling of Human Transferrin by Nano-Flow Liquid Chromatography/electrospray Ionization Mass Spectrometry." *Rapid Communications in Mass Spectrometry: RCM* 18 (24): 2983–88.

Satomi, Yoshinori, Yasutsugu Shimonishi, and Toshifumi Takao. 2004. "N-Glycosylation at Asn(491) in the Asn-Xaa-Cys Motif of Human Transferrin." *FEBS Letters* 576 (1-2): 51–56.

Saur, Michael, Michael J. Hartshorn, Jing Dong, Judith Reeks, Gabor Bunkoczi, Harren Jhoti, and Pamela A. Williams. 2020. "Fragment-Based Drug Discovery Using Cryo-EM." *Drug Discovery Today* 25 (3): 485–90.

Scheres, Sjors H. W., and Shaoxia Chen. 2012. "Prevention of Overfitting in Cryo-EM Structure Determination." *Nature Methods* 9 (9): 853–54.

Scheres, Sjors H. W., Haixiao Gao, Mikel Valle, Gabor T. Herman, Paul P. B. Eggermont, Joachim Frank, and Jose-Maria Carazo. 2007. "Disentangling Conformational States of Macromolecules in 3D-EM through Likelihood Optimization." *Nature Methods*. https://doi.org/10.1038/nmeth992.

Schmidt, A., M. Teeter, E. Weckert, and V. S. Lamzin. 2011. "Crystal Structure of Small

Protein Crambin at 0.48 A Resolution." https://doi.org/10.2210/pdb3nir/pdb.

Schröder, Sybrin P., Wendy A. Offen, Alexandra Males, Yi Jin, Casper de Boer, Jacopo Enotarpi, Laura Marino, et al. 2021. "Development of Non-Hydrolysable Oligosaccharide Activity-Based Inactivators for Endoglycanases: A Case Study on α-1,6 Mannanases." *Chemistry* 27 (37): 9519–23.

Shamsi Kazem Abadi, Saeideh, Michael Tran, Anuj K. Yadav, Pal John Pal Adabala, Saswati Chakladar, and Andrew J. Bennet. 2017. "New Class of Glycoside Hydrolase Mechanism-Based Covalent Inhibitors: Glycosylation Transition State Conformations." *Journal of the American Chemical Society* 139 (31): 10625–28.

Shen, Peter S. 2018. "The 2017 Nobel Prize in Chemistry: Cryo-EM Comes of Age." *Analytical and Bioanalytical Chemistry* 410 (8): 2053–57.

Shimizu, Kazuyuki. 2013. "Regulation Systems of Bacteria such as Escherichia Coli in Response to Nutrient Limitation and Environmental Stresses." *Metabolites* 4 (1): 1–35.

Sieber, Stephan A. 2012. *Activity-Based Protein Profiling*. Springer Science and Business Media.

Sigworth, Fred J. 2016. "Principles of Cryo-EM Single-Particle Image Processing." *Microscopy* 65 (1): 57–67.

Silva, André M. N., Tânia Moniz, Baltazar de Castro, and Maria Rangel. 2021. "Human Transferrin: An Inorganic Biochemistry Perspective." *Coordination Chemistry Reviews*. https://doi.org/10.1016/j.ccr.2021.214186.

Sinnott, M. L., and I. J. Souchard. 1973. "The Mechanism of Action of Beta-Galactosidase. Effect of Aglycone Nature and -Deuterium Substitution on the Hydrolysis of Aryl Galactosides." *Biochemical Journal* 133 (1): 89–98.

Snijder, Joost, Andrew J. Borst, Annie Dosey, Alexandra C. Walls, Anika Burrell, Vijay S. Reddy, Justin M. Kollman, and David Veesler. 2017. "Vitrification after Multiple Rounds of Sample Application and Blotting Improves Particle Density on Cryo-Electron Microscopy Grids." *Journal of Structural Biology* 198 (1): 38–42.

Song, Boyuan, Julian Lenhart, Vanessa Judith Flegler, Cihan Makbul, Tim Rasmussen, and Bettina Böttcher. 2019. "Capabilities of the Falcon III Detector for Single-Particle Structure Determination." *Ultramicroscopy*. https://doi.org/10.1016/j.ultramic.2019.01.002.

Sorzano, C. O. S., J. Vargas, J. Otón, V. Abrishami, J. M. de la Rosa-Trevín, J. Gómez-Blanco, J. L. Vilas, R. Marabini, and J. M. Carazo. 2017. "A Review of Resolution Measures and Related Aspects in 3D Electron Microscopy." *Progress in Biophysics and Molecular Biology* 124 (March): 1–30.

Spik, G., B. Bayard, B. Fournet, G. Strecker, S. Bouquelet, and J. Montreuil. 1975. "Studies on Glycoconjugates. LXIV. Complete Structure of Two Carbohydrate Units of Human Serotransferrin." *FEBS Letters* 50 (3): 296–99.

Stubbs, Keith A. 2014. "Activity-Based Proteomics Probes for Carbohydrate-Processing Enzymes: Current Trends and Future Outlook." *Carbohydrate Research* 390 (May): 9–19.

Stubbs, Keith A., Adrian Scaffidi, Aleksandra W. Debowski, Brian L. Mark, Robert V. Stick, and David J. Vocadlo. 2008. "Synthesis and Use of Mechanism-Based Protein-Profiling Probes for Retaining Beta-D-Glucosaminidases Facilitate Identification of Pseudomonas Aeruginosa NagZ." *Journal of the American Chemical Society* 130 (1): 327–35.

Sturiale, Luisa, Rita Barone, Agata Fiumara, Marta Perez, Marco Zaffanello, Giovanni Sorge, Lorenzo Pavone, et al. 2005. "Hypoglycosylation with Increased Fucosylation and Branching of Serum Transferrin N-Glycans in Untreated Galactosemia." *Glycobiology* 15

(12): 1268–76.

Sun, H., M. C. Cox, H. Li, A. B. Mason, R. C. Woodworth, and P. J. Sadler. 1998. "[1H,13C] NMR Determination of the Order of Lobe Loading of Human Transferrin with Iron: Comparison with Other Metal Ions." *FEBS Letters* 422 (3): 315–20.

Sun, Hongzhe, Hongyan Li, and Peter J. Sadler. 1999. "Transferrin as a Metal Ion Mediator." *Chemical Reviews*. https://doi.org/10.1021/cr980430w.

Sun, X. L., H. M. Baker, S. C. Shewry, G. B. Jameson, and E. N. Baker. 1999. "Structure of Recombinant Human Lactoferrin Expressed in Aspergillus Awamori." *Acta Crystallographica. Section D, Biological Crystallography* 55 (Pt 2): 403–7.

Sutanto, Fandi, Markella Konstantinidou, and Alexander Dömling. 2020. "Covalent Inhibitors: A Rational Approach to Drug Discovery." *RSC Medicinal Chemistry* 11 (8): 876–84.

Tailford, Louise E., Victoria A. Money, Nicola L. Smith, Claire Dumon, Gideon J. Davies, and Harry J. Gilbert. 2007. "Mannose Foraging by Bacteroides Thetaiotaomicron: Structure and Specificity of the Beta-Mannosidase, BtMan2A." *The Journal of Biological Chemistry* 282 (15): 11291–99.

Tanaka, TANAKA, TSUDA, TERAUCHI, TSUNO, KANEYAMA, HONDA, and ISHIDA. 1999. "A New 200 keV Ω-Filter Electron Microscope." *Journal of Microscopy*. https://doi.org/10.1046/j.1365-2818.1999.00446.x.

Tan, Jianbo, Dongdong Liu, Yuhao Bai, Chundong Huang, Xueliang Li, Jun He, Qin Xu, and Li Zhang. 2017. "Enzyme-Assisted Photoinitiated Polymerization-Induced Self-Assembly: An Oxygen-Tolerant Method for Preparing Block Copolymer Nano-Objects in Open Vessels and Multiwell Plates." *Macromolecules*. https://doi.org/10.1021/acs.macromol.7b01219.

Templeton, Douglas M. 2002. *Molecular and Cellular Iron Transport*.

Thomas, Sabu, Miran Mozetic, Uros Cvelbar, Petr Spatenka, and K. M. Praveen. 2018. *Non-Thermal Plasma Technology for Polymeric Materials: Applications in Composites, Nanostructured Materials, and Biomedical Fields*. Elsevier.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. "CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice." *Nucleic Acids Research* 22 (22): 4673–80.

Thompson, Rebecca F., Matthew G. Iadanza, Emma L. Hesketh, Shaun Rawson, and Neil A. Ranson. 2019. "Collection, Pre-Processing and on-the-Fly Analysis of Data for High-Resolution, Single-Particle Cryo-Electron Microscopy." *Nature Protocols* 14 (1): 100–118.

Thomson, N.M., Shirai, T., Chiapello, M., Kondo, A., Mukherjee, K.J., Sivaniah, E., Summers, D.K. and Numata, K., 2016. Metabolomic and proteomic analyses of a quiescent Escherichia coli cell factory reveal the mechanisms behind its production efficiency. *bioRxiv*, p.082305.

Tsai, Charng-Sheng, Yaw-Kuen Li, and Lee-Chiang Lo. 2002. "Design and Synthesis of Activity Probes for Glycosidases." *Organic Letters*. https://doi.org/10.1021/ol0265315.

Tsai, Charng-Sheng, Hsin-Yung Yen, Meng-I Lin, Tsung-I Tsai, Shi-Yun Wang, Wen-I Huang, Tsui-Ling Hsu, Yih-Shyun E. Cheng, Jim-Min Fang, and Chi-Huey Wong. 2013. "Cell-Permeable Probe for Identification and Imaging of Sialidases." *Proceedings of the National Academy of Sciences of the United States of America* 110 (7): 2466–71.

Tsai, Yu-Xi, Ning-En Chang, Klaus Reuter, Hao-Ting Chang, Tzu-Jing Yang, Sören von Bülow, Vidhi Sehrawat, et al. 2024. "Rapid Simulation of Glycoprotein Structures by Grafting

and Steric Exclusion of Glycan Conformer Libraries." *Cell* 187 (5): 1296–1311.e26.

Tung, Ching-Hsuan, Qing Zeng, Khalid Shah, Dong-Eog Kim, Dawid Schellingerhout, and Ralph Weissleder. 2004. "In Vivo Imaging of Beta-Galactosidase Activity Using Far Red Fluorescent Switch." *Cancer Research* 64 (5): 1579–83.

Valle, Mikel, Jayati Sengupta, Neil K. Swami, Robert A. Grassucci, Nils Burkhardt, Knud H. Nierhaus, Rajendra K. Agrawal, and Joachim Frank. 2002. "Cryo-EM Reveals an Active Role for Aminoacyl-tRNA in the Accommodation Process." *The EMBO Journal* 21 (13): 3557–67.

Veen, Harrie A. van, Marlieke E. J. Geerts, Patrick H. C. van Berkel, and Jan H. Nuijens. 2004. "The Role of N-Linked Glycosylation in the Protection of Human and Bovine Lactoferrin against Tryptic Proteolysis." *European Journal of Biochemistry / FEBS* 271 (4): 678–84.

Vinothkumar, Kutti R., Greg McMullan, and Richard Henderson. 2014. "Molecular Mechanism of Antibody-Mediated Activation of β-Galactosidase." *Structure* 22 (4): 621–27.

Viratelle, O. M., and J. M. Yon. 1973. "Nucleophilic Competition in Some -Galactosidase-Catalyzed Reactions." *European Journal of Biochemistry / FEBS* 33 (1): 110–16.

Vocadlo, David J., and Carolyn R. Bertozzi. 2004. "A Strategy for Functional Proteomic Analysis of Glycosidase Activity from Cell Lysates." *Angewandte Chemie* 43 (40): 5338–42.

Vogel, Hans J. 2012. "Lactoferrin, a Bird's Eye View." *Biochemistry and Cell Biology = Biochimie et Biologie Cellulaire* 90 (3): 233–44.

Wally, Jeremy, Peter J. Halbrooks, Clemens Vonrhein, Mark A. Rould, Stephen J. Everse, Anne B. Mason, and Susan K. Buchanan. 2006a. "The Crystal Structure of Iron-Free Human Serum Transferrin Provides Insight into Inter-Lobe Communication and Receptor Binding." *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.m604592200.

Wang, Hongbin, Zheng Li, Junling Niu, Yongfen Xu, Li Ma, Ailing Lu, Xun Wang, et al. 2018. "Antiviral Effects of Ferric Ammonium Citrate." *Cell Discovery*. https://doi.org/10.1038/s41421-018-0013-6.

Warnecke, Athanasia, Thomas Averbeck, Ulrich Wurster, Meike Harmening, Thomas Lenarz, and Timo Stöver. 2004. "Diagnostic Relevance of beta2-Transferrin for the Detection of Cerebrospinal Fluid Fistulas." *Archives of Otolaryngology--Head and Neck Surgery* 130 (10): 1178–84.

Wehrman, Thomas S., Georges von Degenfeld, Peter O. Krutzik, Garry P. Nolan, and Helen M. Blau. 2006. "Luminescent Imaging of Beta-Galactosidase Activity in Living Subjects Using Sequential Reporter-Enzyme Luminescence." *Nature Methods* 3 (4): 295–301.

Wheatley, Robert W., Summie Lo, Larisa J. Jancewicz, Megan L. Dugdale, and Reuben E. Huber. 2013. "Structural Explanation for Allolactose (lac Operon Inducer) Synthesis by lacZ β-Galactosidase and the Evolutionary Relationship between Allolactose Synthesis and the Lac Repressor." *The Journal of Biological Chemistry* 288 (18): 12993–5.

Willems, Lianne I., Thomas J. M. Beenakker, Benjamin Murray, Berend Gagestein, Hans van den Elst, Erwin R. van Rijssel, Jeroen D. C. Codée, et al. 2014. "Synthesis of α- and β-Galactopyranose-Configured Isomers of Cyclophellitol and Cyclophellitol Aziridine." *European Journal of Organic Chemistry*. https://doi.org/10.1002/ejoc.201402589.

Williams, Spencer J., Omid Hekmat, and Stephen G. Withers. 2006. "Synthesis and Testing of Mechanism-Based Protein-Profiling Probes for Retaining Endo-Glycosidases." *Chembiochem: A European Journal of Chemical Biology* 7 (1): 116–24.

Withers, S. G., and R. Aebersold. 1995. "Approaches to Labeling and Identification of Active

Site Residues in Glycosidases." *Protein Science: A Publication of the Protein Society* 4 (3): 361–72.

Witte, Martin D., Wouter W. Kallemeijn, Jan Aten, Kah-Yee Li, Anneke Strijland, Wilma E. Donker-Koopman, Adrianus M. C. H. van den Nieuwendijk, et al. 2010. "Ultrasensitive in Situ Visualization of Active Glucocerebrosidase Molecules." *Nature Chemical Biology* 6 (12): 907–13.

Wu, Liang, Zachary Armstrong, Sybrin P. Schröder, Casper de Boer, Marta Artola, Johannes Mfg Aerts, Herman S. Overkleeft, and Gideon J. Davies. 2019. "An Overview of Activity-Based Probes for Glycosidases." *Current Opinion in Chemical Biology*. https://doi.org/10.1016/j.cbpa.2019.05.030.

Wu, Mengyu, and Gabriel C. Lander. 2020. "How Low Can We Go? Structure Determination of Small Biological Complexes Using Single-Particle Cryo-EM." *Current Opinion in Structural Biology*. https://doi.org/10.1016/j.sbi.2020.05.007.

Yamashita, Keitaro, Colin M. Palmer, Tom Burnley, and Garib N. Murshudov. 2021. "Cryo-EM Single-Particle Structure Refinement and Map Calculation Using Servalcat." *Acta Crystallographica. Section D, Structural Biology* 77 (Pt 10): 1282–91.

Yanatori, Izumi, and Fumio Kishi. 2019. "DMT1 and Iron Transport." *Free Radical Biology and Medicine* 133 (March): 55–63.

Yang, H-W, R. T. A. MacGillivray, J. Chen, Y. Luo, Y. Wang, G. D. Brayer, A. Mason, R. C. Woodworth, and M. E. P. Murphy. 2000. "HUMAN SERUM TRANSFERRIN." https://doi.org/10.2210/pdb1d3k/pdb.

Yang, Nan, Hongmin Zhang, Minji Wang, Quan Hao, and Hongzhe Sun. 2012a. "Iron and Bismuth Bound Human Serum Transferrin Reveals a Partially-Opened Conformation in the N-Lobe." *Scientific Reports* 2 (December): 999.

Yang, Shuang, Yingwei Hu, Lori Sokoll, and Hui Zhang. 2017. "Simultaneous Quantification of N- and O-Glycans Using a Solid-Phase Method." *Nature Protocols* 12 (6): 1229–44.

Yoshioka, Craig, Bridget Carragher, and Clinton S. Potter. 2010. "Cryomesh™: A New Substrate for Cryo-Electron Microscopy." *Microscopy and Microanalysis*. https://doi.org/10.1017/s1431927609991310.

Yu, Guimei, Kunpeng Li, and Wen Jiang. 2016. "Antibody-Based Affinity Cryo-EM Grid." *Methods*. https://doi.org/10.1016/j.ymeth.2016.01.010.

Zak, O., B. Tam, R. T. MacGillivray, and P. Aisen. 1997. "A Kinetically Active Site in the C-Lobe of Human Transferrin." *Biochemistry* 36 (36): 11036–43.

Zhang, Kai. 2016. "Gctf: Real-Time CTF Determination and Correction." *Journal of Structural Biology* 193 (1): 1–12.

Zhang, Kaiming, Shanshan Li, Kalli Kappel, Grigore Pintilie, Zhaoming Su, Tung-Chung Mou, Michael F. Schmid, Rhiju Das, and Wah Chiu. 2019. "Cryo-EM Structure of a 40 kDa SAM-IV Riboswitch RNA at 3.7 Å Resolution." *Nature Communications* 10 (1): 5511.

Zhao, Ningning, and Caroline A. Enns. 2013. "N-Linked Glycosylation Is Required for Transferrin-Induced Stabilization of Transferrin Receptor 2, but Not for Transferrin Binding or Trafficking to the Cell Surface." *Biochemistry*. https://doi.org/10.1021/bi4000063.

Zhao, Xinyu, Guoshun Li, and Shufang Liang. 2013. "Several Affinity Tags Commonly Used in Chromatographic Purification." Journal of Analytical Methods in Chemistry 2013 (December): 581093.

Zheng, Shawn Q., Eugene Palovcak, Jean-Paul Armache, Kliment A. Verba, Yifan Cheng, and David A. Agard. 2017. "MotionCor2: Anisotropic Correction of Beam-Induced Motion for Improved Cryo-Electron Microscopy." *Nature Methods* 14 (4): 331–32.

Zivanov, Jasenko, Takanori Nakane, Björn O. Forsberg, Dari Kimanius, Wim J. H. Hagen, Erik Lindahl, and Sjors H. W. Scheres. 2018. "New Tools for Automated High-Resolution Cryo-EM Structure Determination in *RELION*-3." *eLife*. https://doi.org/10.7554/elife.42166.

Zivanov, Jasenko, Takanori Nakane, and Sjors H. W. Scheres. 2020. "Estimation of High-Order Aberrations and Anisotropic Magnification from Cryo-EM Data Sets in *RELION*-3.1." *IUCrJ*. https://doi.org/10.1107/s2052252520000081.

Zuo, Chao, Jiasong Sun, Jiaji Li, Jialin Zhang, Anand Asundi, and Qian Chen. 2017. "High-Resolution Transport-of-Intensity Quantitative Phase Microscopy with Annular Illumination." *Scientific Reports* 7 (1): 7654.

Zworykin, Vladimir K. 1942. "The Scanning Electron Microscope." *Scientific American*. https://doi.org/10.1038/scientificamerican0942-111.

Annex

WORLDWIDE wwPDB
PROTEIN DATA BANK

Full wwPDB EM Validation Report ⓘ

Sep 27, 2023 – 04:23 pm BST

PDB ID	:	8QLZ
EMDB ID	:	EMD-18488
Title	:	E. coli beta-galactosidase with covalently attached 4-epi cyclophellitol
Deposited on	:	2023-09-20
Resolution	:	2.70 Å(reported)

This wwPDB validation report is for manuscript review

This is a Full wwPDB EM Validation Report.

This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 2.70 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Percentile Ranks	Value
Clashscore		5
Ramachandran outliers		0
Sidechain outliers		2.1%

Worse *Better*

▮ Percentile relative to all structures

◻ Percentile relative to all EM structures
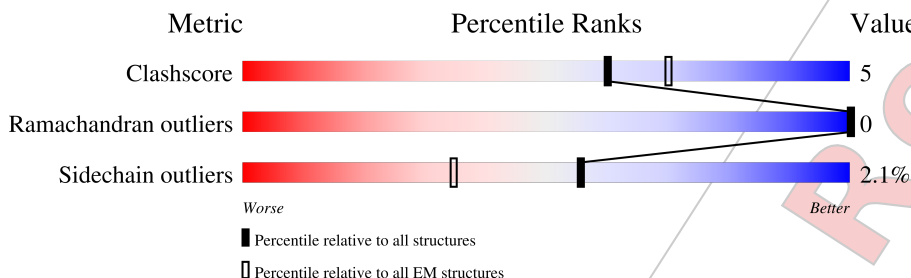
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the EM map (all-atom inclusion < 40%). The numeric value is given above the bar.

| Mol	Chain	Length	Quality of chain
1	A	1024	11% ▬▬ 88% ▬▬▬▬▬▬▬▬ 11% •

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 8222 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Beta-galactosidase.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	1022	Total 8206	C 5190	N 1452	O 1525	S 39	0	0

- Molecule 2 is 4-epi-Cyclophellitol (three-letter code: H9Q) (formula: $C_7H_{12}O_5$) (labeled as "Ligand of Interest" by depositor).
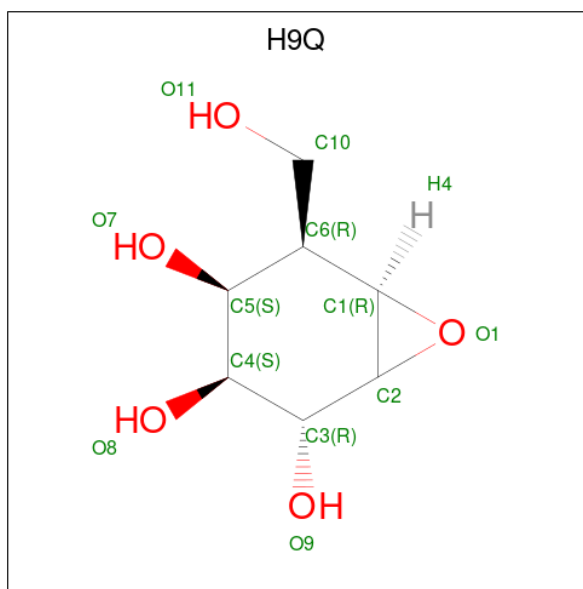
Mol	Chain	Residues	Atoms			AltConf
2	A	1	Total 12	C 7	O 5	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
3	A	1	Total 1	Mg 1	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		AltConf
4	A	3	Total O		0
			3 3		

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion in map density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red diamond above a residue indicates a poor fit to the EM map for this residue (all-atom inclusion < 40%). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: Beta-galactosidase
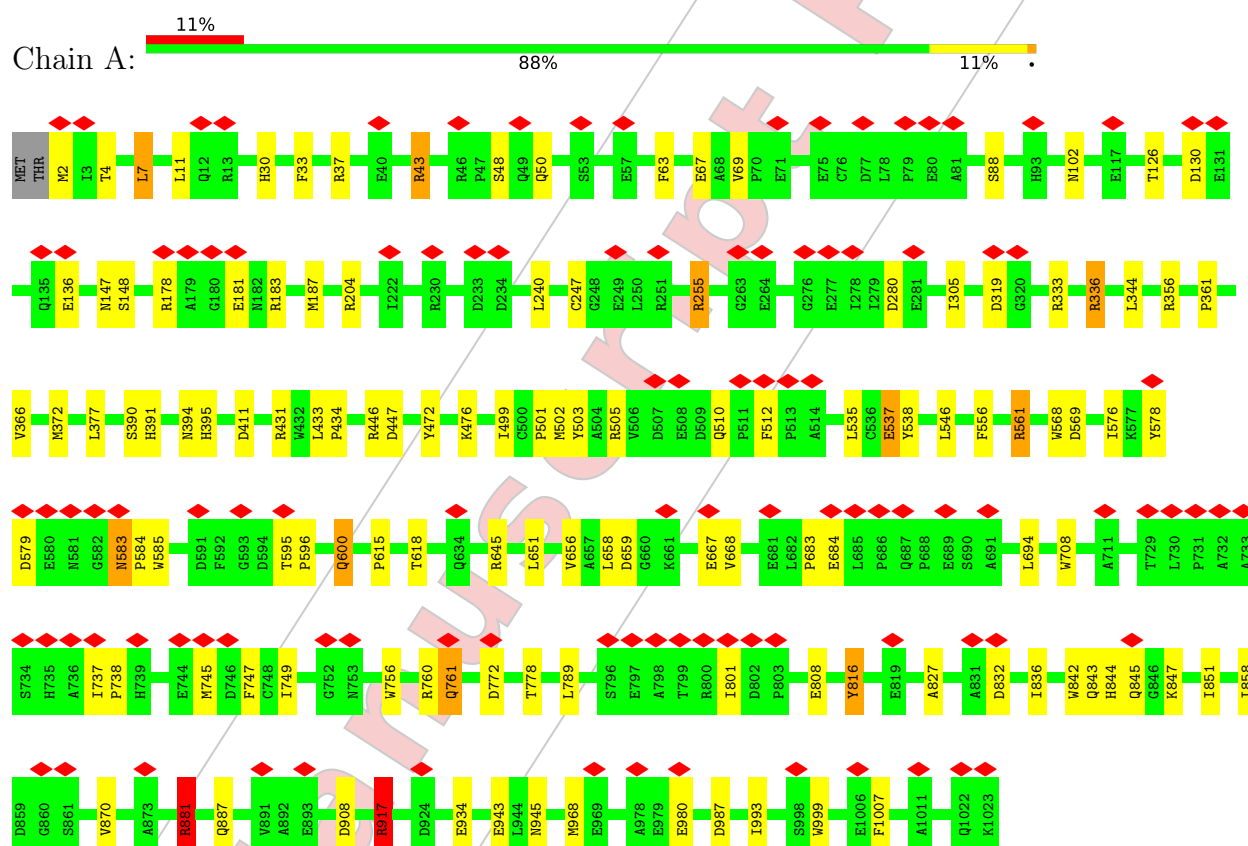
4 Experimental information ⓘ

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, D2	Depositor
Number of particles used	204567	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	TFS GLACIOS	Depositor
Voltage (kV)	200	Depositor
Electron dose $(e^-/\text{Å}^2)$	50	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	2000	Depositor
Magnification	Not provided	
Image detector	FEI FALCON IV (4k x 4k)	Depositor
Maximum map value	0.102	Depositor
Minimum map value	-0.044	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.003	Depositor
Recommended contour level	0.02	Depositor
Map size (Å)	399.6, 399.6, 399.6	wwPDB
Map dimensions	450, 450, 450	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	0.888, 0.888, 0.888	Depositor

5 Model quality ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: MG, H9Q

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$|Z| > 5$	RMSZ	#$|Z| > 5$
1	A	0.46	1/8448 (0.0%)	0.86	15/11526 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	6

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	537	GLU	CD-OE2	9.29	1.35	1.25

All (15) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	561	ARG	NE-CZ-NH1	-12.86	113.87	120.30
1	A	561	ARG	NE-CZ-NH2	12.68	126.64	120.30
1	A	917	ARG	NE-CZ-NH1	-12.54	114.03	120.30
1	A	37	ARG	NE-CZ-NH2	10.69	125.65	120.30
1	A	336	ARG	NE-CZ-NH1	-7.60	116.50	120.30
1	A	917	ARG	NE-CZ-NH2	7.46	124.03	120.30
1	A	183	ARG	NE-CZ-NH2	6.99	123.79	120.30
1	A	446	ARG	NE-CZ-NH2	6.74	123.67	120.30
1	A	842	TRP	CB-CG-CD2	-6.29	118.42	126.60
1	A	816	TYR	CB-CG-CD2	6.23	124.74	121.00
1	A	881	ARG	NE-CZ-NH2	6.01	123.31	120.30
1	A	336	ARG	NE-CZ-NH2	5.60	123.10	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	505	ARG	NE-CZ-NH1	-5.32	117.64	120.30
1	A	280	ASP	CB-CA-C	-5.23	99.94	110.40
1	A	556	PHE	CB-CG-CD1	5.03	124.32	120.80

There are no chirality outliers.

All (6) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	204	ARG	Sidechain
1	A	255	ARG	Sidechain
1	A	356	ARG	Sidechain
1	A	43	ARG	Sidechain
1	A	561	ARG	Sidechain
1	A	917	ARG	Sidechain

5.2 Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	8206	0	7805	76	0
2	A	12	0	0	0	0
3	A	1	0	0	0	0
4	A	3	0	0	0	0
All	All	8222	0	7805	76	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 5.

All (76) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:881:ARG:NH1	1:A:987:ASP:OD2	2.04	0.90
1:A:908:ASP:HB3	1:A:1007:PHE:CD2	2.17	0.80
1:A:255:ARG:HG2	1:A:255:ARG:HH21	1.48	0.79
1:A:600:GLN:HE21	1:A:600:GLN:H	1.33	0.75

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:943:GLU:OE1	1:A:945:ASN:ND2	2.16	0.73
1:A:503:TYR:CE2	1:A:537:GLU:OE1	2.45	0.69
1:A:43:ARG:HG2	1:A:43:ARG:HH21	1.59	0.67
1:A:887:GLN:NE2	1:A:980:GLU:O	2.28	0.66
1:A:737:ILE:HD13	1:A:832:ASP:HA	1.78	0.66
1:A:431:ARG:HG2	1:A:431:ARG:HH11	1.61	0.65
1:A:503:TYR:CE2	1:A:999:TRP:HH2	2.16	0.63
1:A:881:ARG:HH11	1:A:987:ASP:CG	2.02	0.63
1:A:749:ILE:HD12	1:A:749:ILE:N	2.16	0.61
1:A:737:ILE:HD12	1:A:738:PRO:HD2	1.83	0.60
1:A:816:TYR:HE1	1:A:968:MET:SD	2.28	0.57
1:A:908:ASP:HB3	1:A:1007:PHE:CE2	2.39	0.57
1:A:502:MET:CB	1:A:537:GLU:HB2	2.36	0.56
1:A:510:GLN:HB3	1:A:512:PHE:CE1	2.42	0.55
1:A:843:GLN:HA	1:A:847:LYS:O	2.07	0.55
1:A:2:MET:O	1:A:4:THR:N	2.40	0.54
1:A:30:HIS:CE1	1:A:33:PHE:CD2	2.96	0.54
1:A:377:LEU:HD22	1:A:708:TRP:HA	1.89	0.53
1:A:816:TYR:CE1	1:A:968:MET:SD	3.01	0.53
1:A:881:ARG:NH2	1:A:934:GLU:OE1	2.41	0.53
1:A:503:TYR:CE2	1:A:999:TRP:CH2	2.96	0.53
1:A:600:GLN:HE21	1:A:600:GLN:N	2.05	0.53
1:A:578:TYR:HE1	1:A:584:PRO:HB3	1.74	0.52
1:A:844:HIS:CE1	1:A:845:GLN:HG3	2.44	0.52
1:A:656:VAL:CG1	1:A:694:LEU:HD22	2.40	0.52
1:A:130:ASP:OD1	1:A:130:ASP:N	2.43	0.52
1:A:63:PHE:CD2	1:A:69:VAL:HG22	2.45	0.52
1:A:747:PHE:CE1	1:A:827:ALA:HB3	2.45	0.51
1:A:48:SER:OG	1:A:50:GLN:OE1	2.28	0.50
1:A:255:ARG:HG2	1:A:255:ARG:NH2	2.19	0.50
1:A:583:ASN:OD1	1:A:583:ASN:N	2.42	0.50
1:A:816:TYR:CE1	1:A:968:MET:CE	2.95	0.50
1:A:917:ARG:NH1	1:A:943:GLU:OE1	2.44	0.50
1:A:7:LEU:HD12	1:A:11:LEU:HD11	1.93	0.49
1:A:772:ASP:O	1:A:772:ASP:CG	2.51	0.49
1:A:595:THR:HA	1:A:596:PRO:C	2.32	0.49
1:A:651:LEU:HD12	1:A:668:VAL:C	2.34	0.48
1:A:881:ARG:NH1	1:A:987:ASP:CG	2.64	0.48
1:A:789:LEU:HD11	1:A:993:ILE:HG22	1.97	0.47
1:A:411:ASP:OD2	1:A:447:ASP:OD2	2.32	0.47
1:A:147:ASN:HA	1:A:148:SER:HA	1.62	0.46

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:372:MET:HE1	1:A:395:HIS:HB3	1.98	0.46
1:A:756:TRP:CD2	1:A:858:ILE:HD13	2.52	0.45
1:A:361:PRO:HB2	1:A:576:ILE:HG12	1.99	0.45
1:A:568:TRP:CD2	1:A:569:ASP:HB3	2.52	0.44
1:A:801:ILE:HD12	1:A:808:GLU:CD	2.37	0.44
1:A:305:ILE:HD11	1:A:645:ARG:HB3	1.99	0.44
1:A:431:ARG:HG2	1:A:431:ARG:NH1	2.30	0.44
1:A:651:LEU:HD11	1:A:667:GLU:HB3	1.99	0.44
1:A:319:ASP:OD1	1:A:319:ASP:C	2.56	0.44
1:A:433:LEU:N	1:A:434:PRO:CD	2.81	0.44
1:A:240:LEU:C	1:A:240:LEU:HD23	2.38	0.43
1:A:683:PRO:O	1:A:684:GLU:C	2.56	0.43
1:A:747:PHE:CD1	1:A:827:ALA:CB	3.02	0.43
1:A:615:PRO:O	1:A:618:THR:HG22	2.18	0.43
1:A:658:LEU:O	1:A:659:ASP:C	2.56	0.43
1:A:745:MET:CE	1:A:761:GLN:HG3	2.49	0.43
1:A:88:SER:HA	1:A:366:VAL:HG21	2.01	0.43
1:A:7:LEU:HD11	1:A:187:MET:HE3	2.01	0.42
1:A:499:ILE:HG22	1:A:501:PRO:HD3	2.01	0.42
1:A:472:TYR:O	1:A:476:LYS:HG2	2.20	0.42
1:A:502:MET:HB3	1:A:537:GLU:HB2	2.00	0.42
1:A:102:ASN:C	1:A:102:ASN:OD1	2.58	0.42
1:A:651:LEU:HD12	1:A:668:VAL:O	2.19	0.41
1:A:801:ILE:HD12	1:A:808:GLU:OE2	2.19	0.41
1:A:851:ILE:O	1:A:870:VAL:HA	2.20	0.41
1:A:390:SER:HA	1:A:391:HIS:HA	1.91	0.41
1:A:579:ASP:OD1	1:A:585:TRP:CD1	2.73	0.41
1:A:747:PHE:CD2	1:A:760:ARG:HG2	2.56	0.41
1:A:30:HIS:ND1	1:A:33:PHE:CE2	2.89	0.41
1:A:7:LEU:HD12	1:A:11:LEU:CD1	2.51	0.40
1:A:126:THR:HG23	1:A:181:GLU:HG2	2.04	0.40

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	1020/1024 (100%)	985 (97%)	35 (3%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	874/876 (100%)	856 (98%)	18 (2%)	53	80

All (18) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	7	LEU
1	A	67	GLU
1	A	136	GLU
1	A	178	ARG
1	A	247	CYS
1	A	333	ARG
1	A	336	ARG
1	A	344	LEU
1	A	394	ASN
1	A	535	LEU
1	A	538	TYR
1	A	546	LEU
1	A	583	ASN
1	A	600	GLN
1	A	761	GLN
1	A	778	THR
1	A	836	ILE
1	A	881	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	245	GLN
1	A	262	GLN
1	A	394	ASN
1	A	485	GLN
1	A	600	GLN
1	A	624	GLN
1	A	718	GLN
1	A	1022	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	H9Q	A	1101	1	13,13,13	10.93	2 (15%)	16,20,20	12.29	4 (25%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	H9Q	A	1101	1	-	0/2/27/27	0/2/2/2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1101	H9Q	O1-C2	39.21	2.11	1.44
2	A	1101	H9Q	C2-C1	3.38	1.53	1.46

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	1101	H9Q	O1-C1-C2	40.29	91.63	59.61
2	A	1101	H9Q	O1-C2-C1	-22.19	41.97	59.61
2	A	1101	H9Q	C2-O1-C1	-15.56	46.40	60.78
2	A	1101	H9Q	O1-C2-C3	-6.85	106.31	115.73

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand H9Q A 1101

Bond lengths

Bond angles

Torsions

Rings
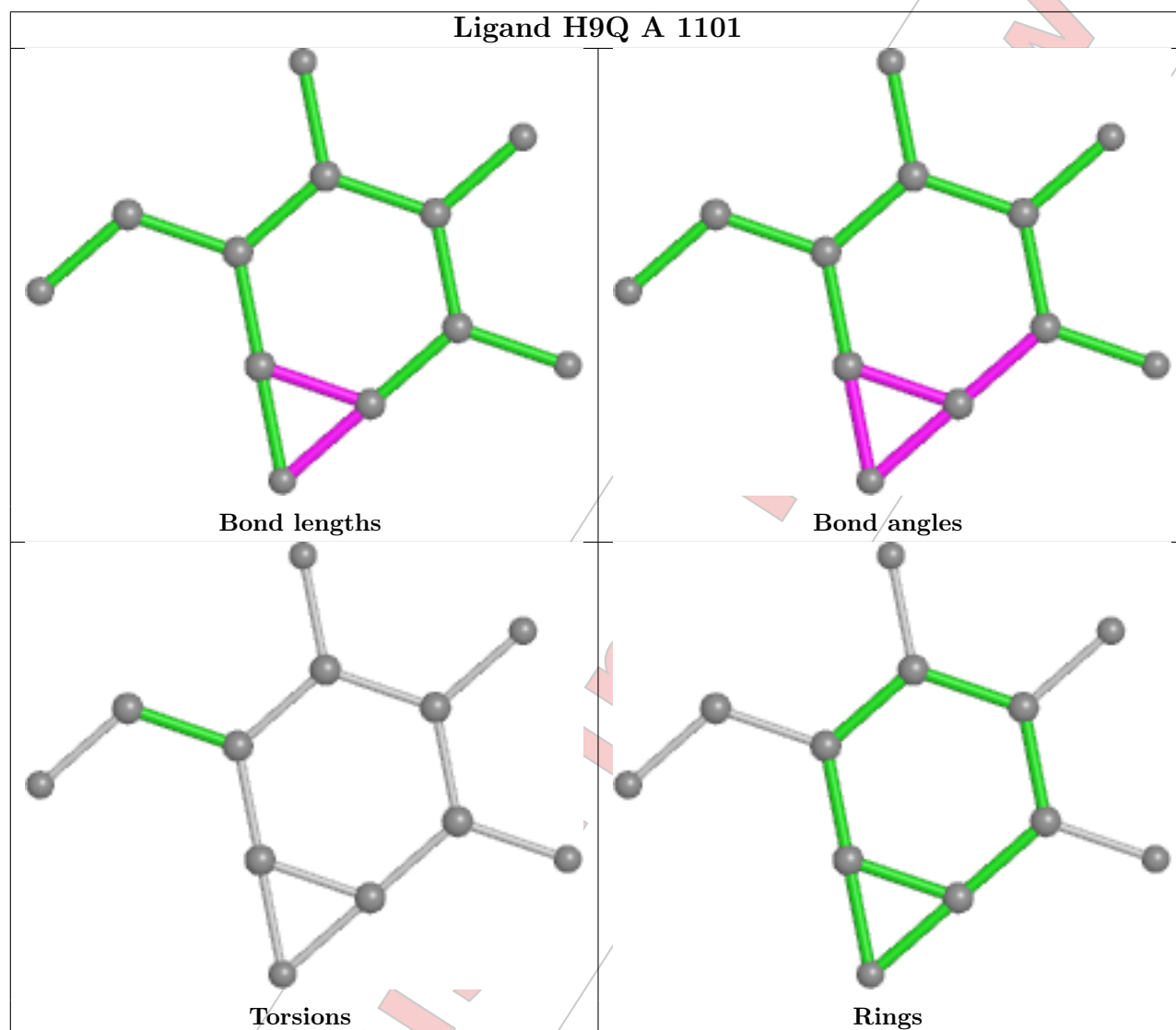
5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6 Map visualisation ⓘ

This section contains visualisations of the EMDB entry EMD-18488. These allow visual inspection of the internal detail of the map and identification of artifacts.

Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections ⓘ
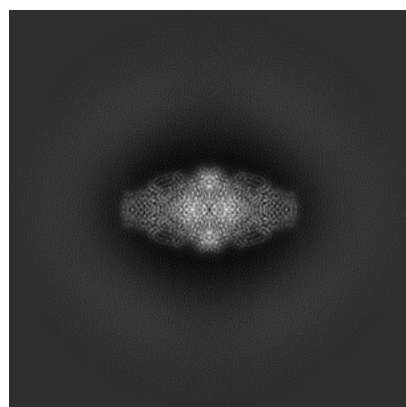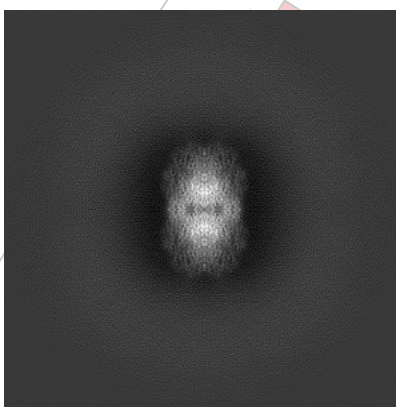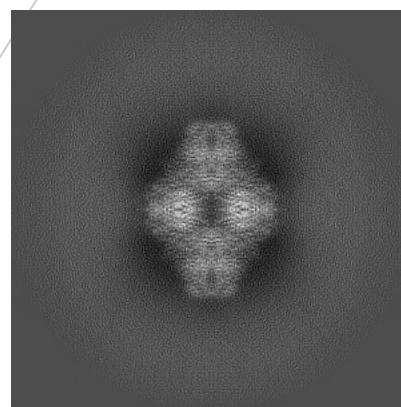
6.1.1 Primary map

| X | Y | Z |

6.1.2 Raw map

| X | Y | Z |
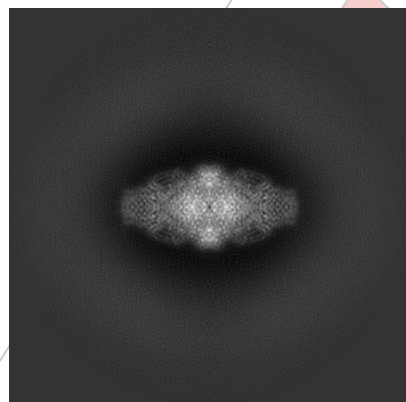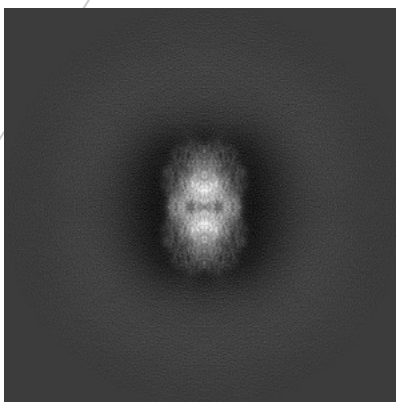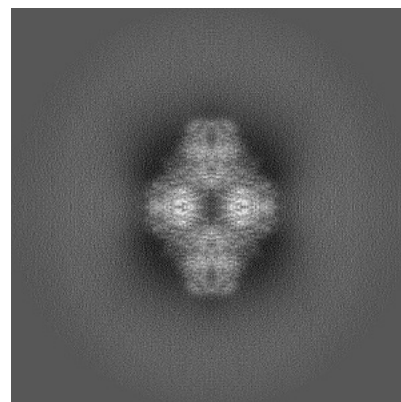
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

6.2.1 Primary map
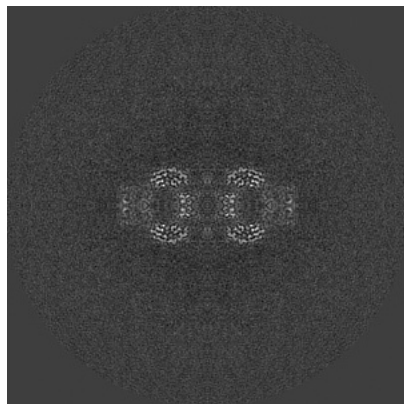
X Index: 225
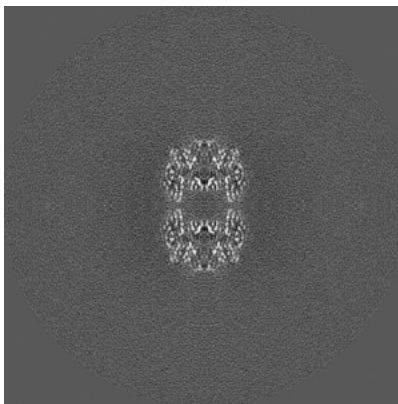
Y Index: 225
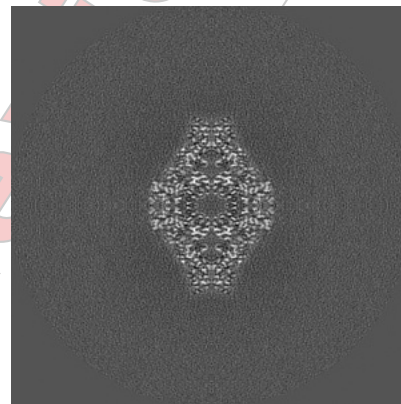
Z Index: 225

6.2.2 Raw map
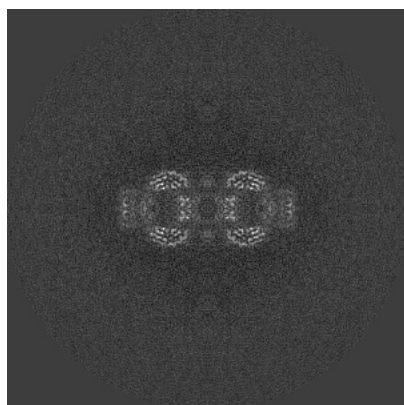
X Index: 225
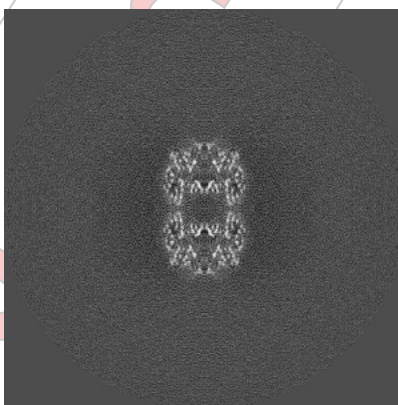
Y Index: 225
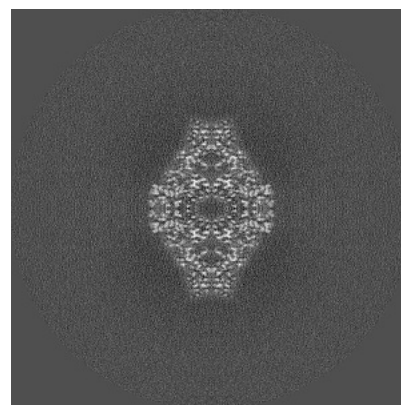
Z Index: 225

The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices ⓘ

6.3.1 Primary map
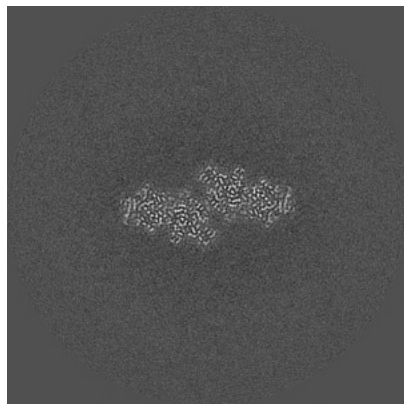
X Index: 244
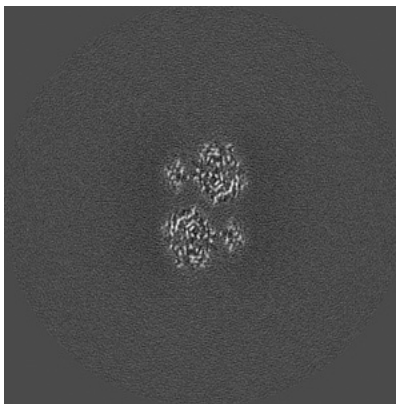
Y Index: 234
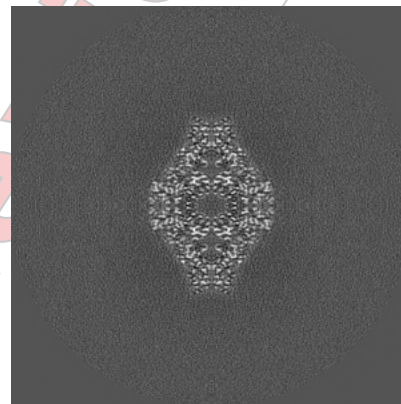
Z Index: 225

6.3.2 Raw map
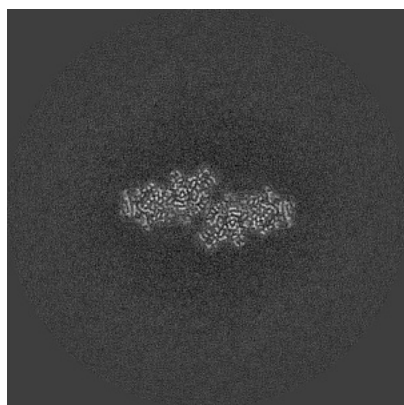
X Index: 206
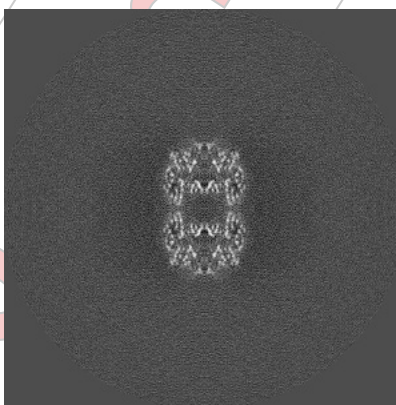
Y Index: 225
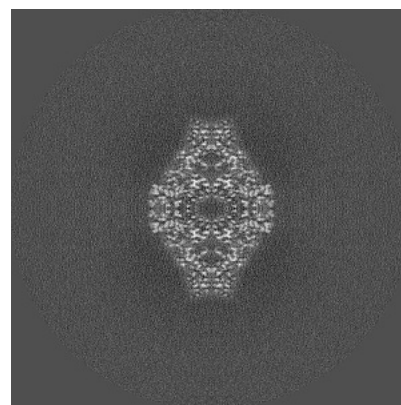
Z Index: 225

The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal standard-deviation projections (False-color) ⓘ

6.4.1 Primary map

X

Y

Z

6.4.2 Raw map

X

Y

Z
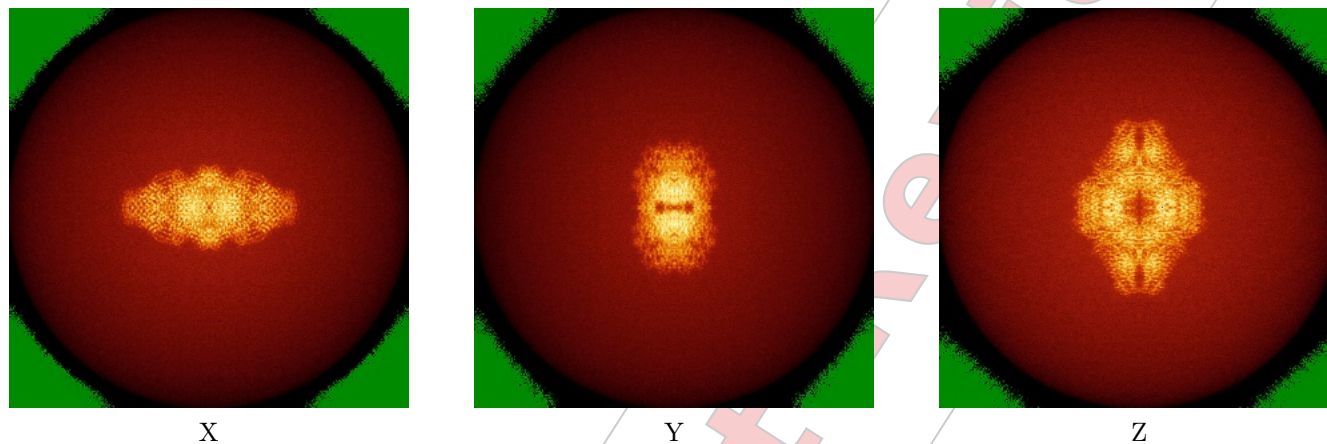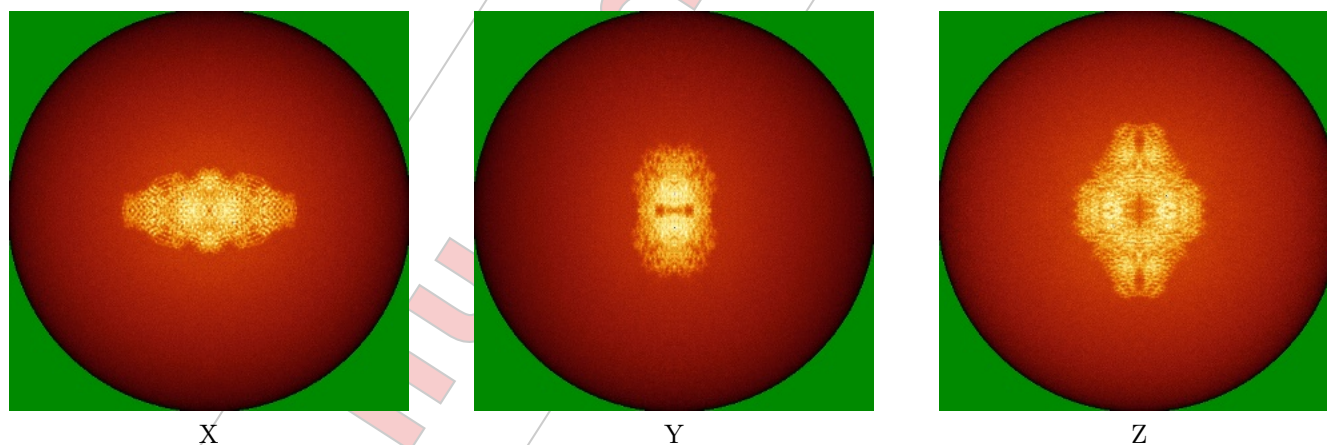
The images above show the map standard deviation projections with false color in three orthogonal directions. Minimum values are shown in green, max in blue, and dark to light orange shades represent small to large values respectively.

6.5 Orthogonal surface views ⓘ

6.5.1 Primary map

X Y Z

The images above show the 3D surface view of the map at the recommended contour level 0.02. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

6.5.2 Raw map

X Y Z
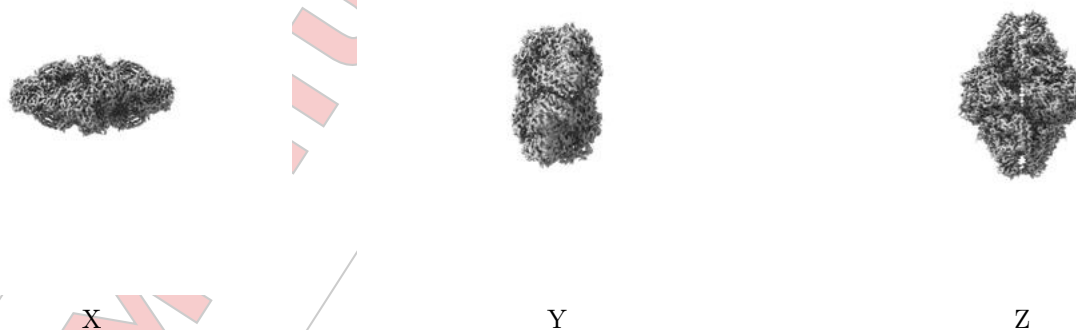
These images show the 3D surface of the raw map. The raw map's contour level was selected so that its surface encloses the same volume as the primary map does at its recommended contour level.

6.6 Mask visualisation ⓘ

This section was not generated. No masks/segmentation were deposited.

7 Map analysis ⓘ

This section contains the results of statistical analysis of the map.

7.1 Map-value distribution ⓘ
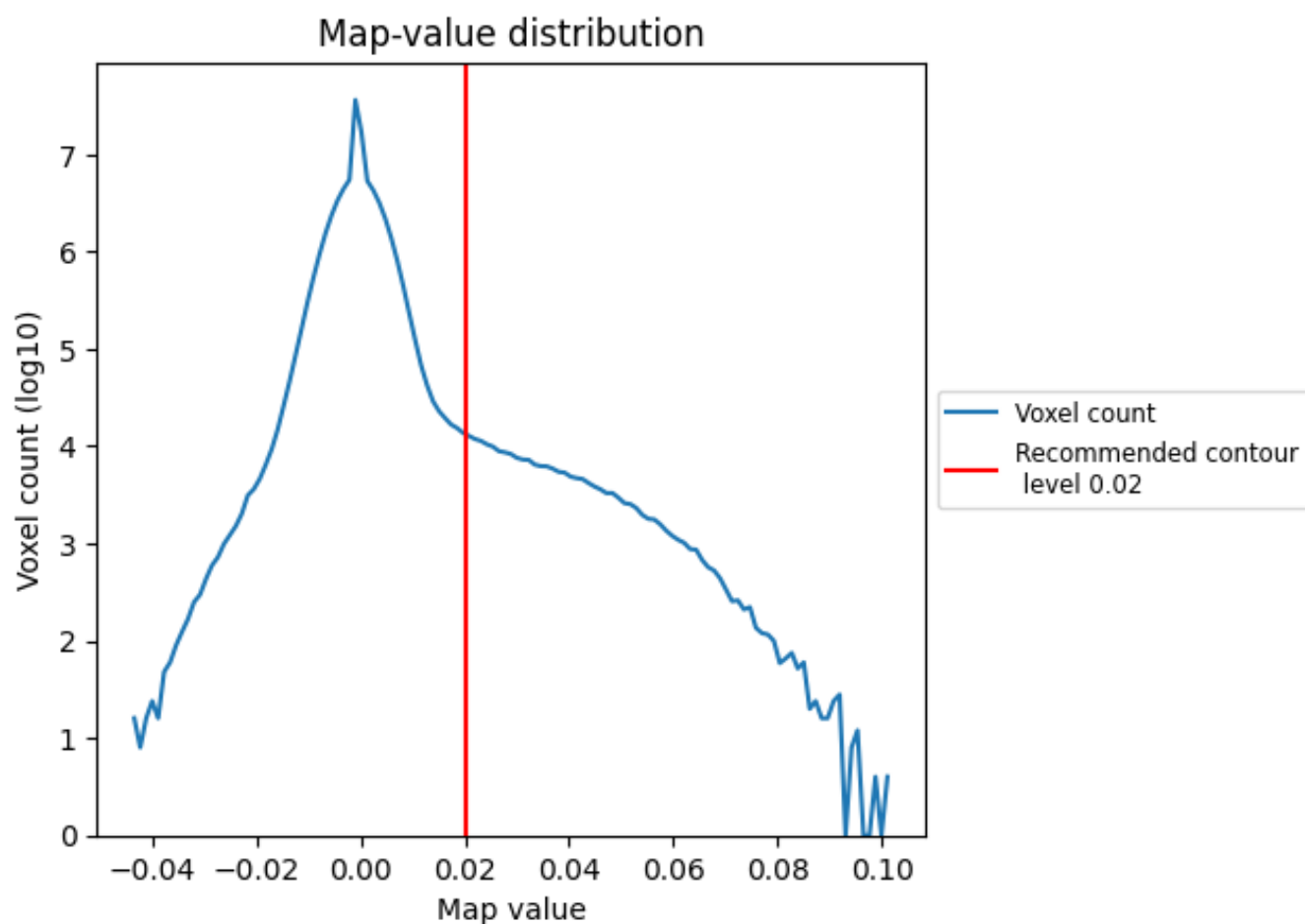
The map-value distribution is plotted in 128 intervals along the x-axis. The y-axis is logarithmic. A spike in this graph at zero usually indicates that the volume has been masked.

7.2 Volume estimate ⓘ
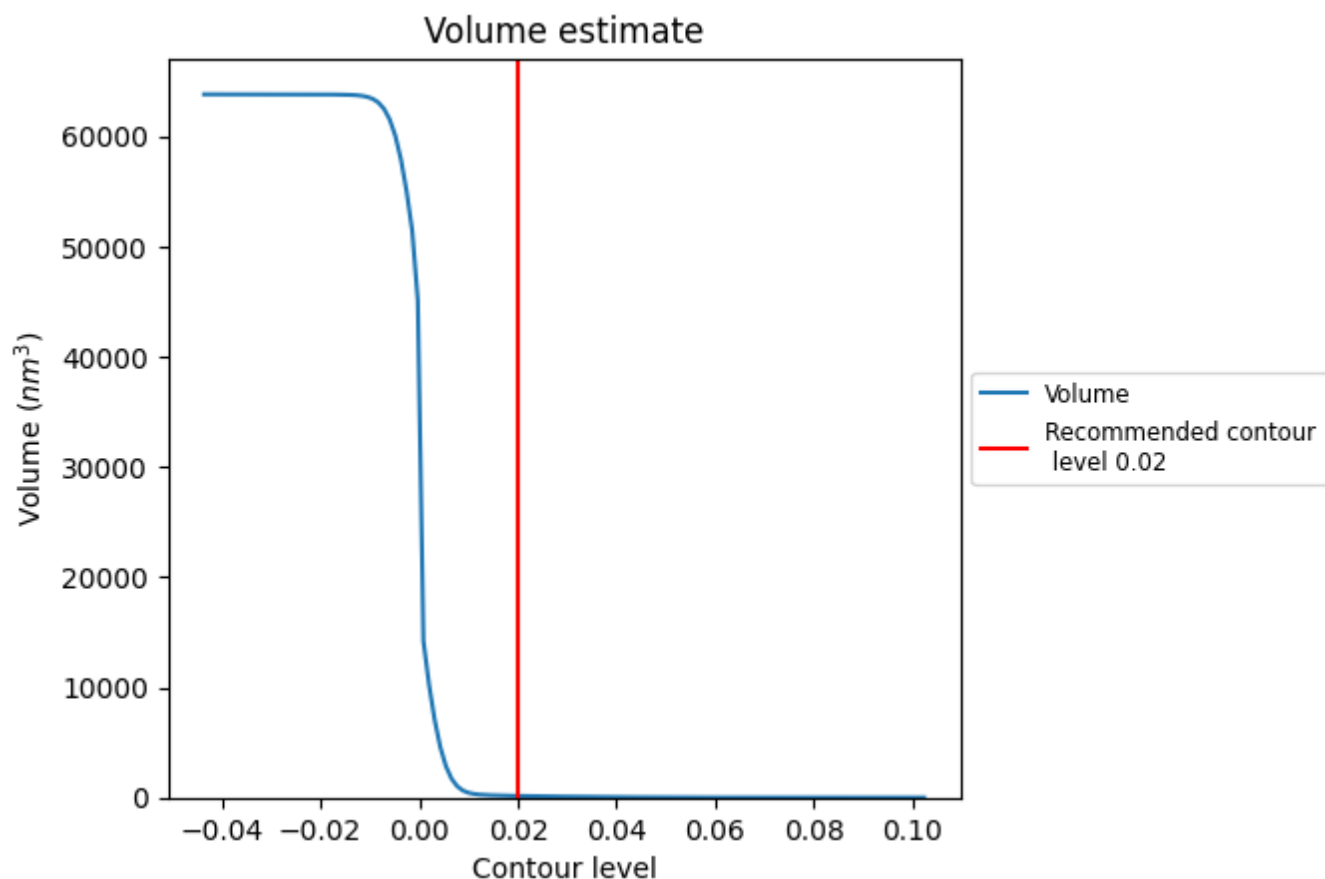
Volume estimate

The volume at the recommended contour level is 147 nm^3; this corresponds to an approximate mass of 133 kDa.

The volume estimate graph shows how the enclosed volume varies with the contour level. The recommended contour level is shown as a vertical line and the intersection between the line and the curve gives the volume of the enclosed surface at the given level.

7.3 Rotationally averaged power spectrum ⓘ
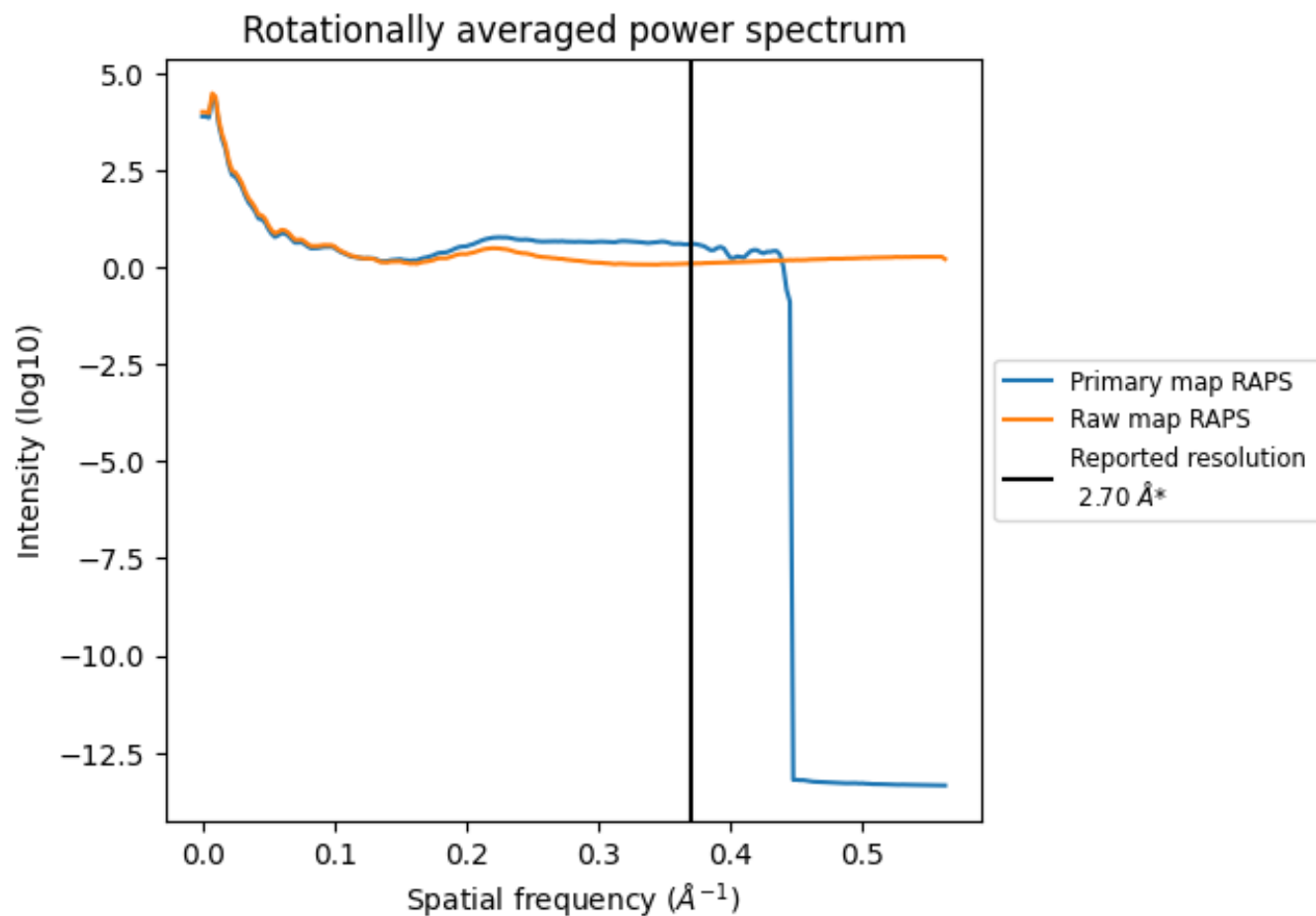
*Reported resolution corresponds to spatial frequency of 0.370 Å^{-1}

8 Fourier-Shell correlation ⓘ

Fourier-Shell Correlation (FSC) is the most commonly used method to estimate the resolution of single-particle and subtomogram-averaged maps. The shape of the curve depends on the imposed symmetry, mask and whether or not the two 3D reconstructions used were processed from a common reference. The reported resolution is shown as a black line. A curve is displayed for the half-bit criterion in addition to lines showing the 0.143 gold standard cut-off and 0.5 cut-off.

8.1 FSC ⓘ
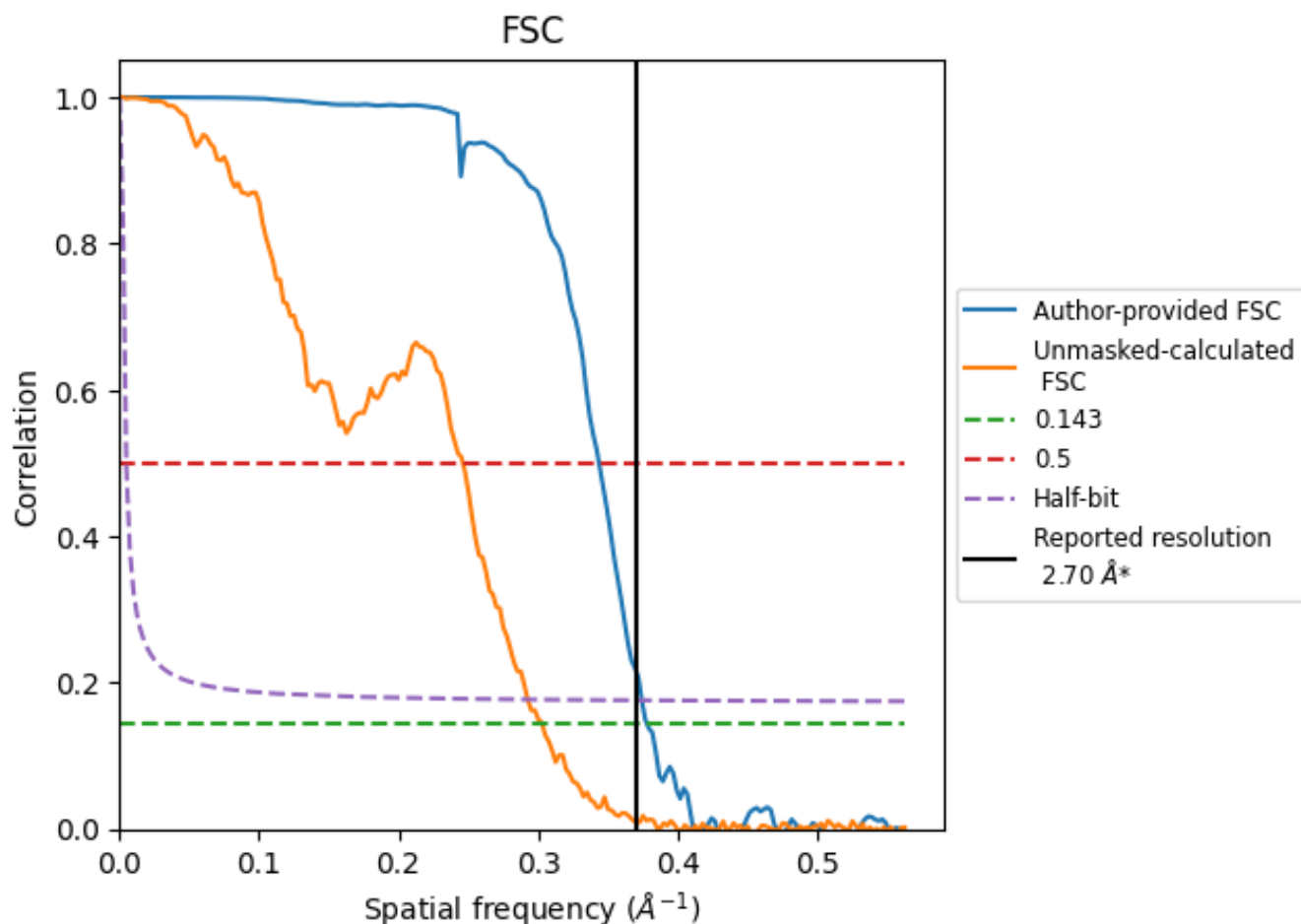
*Reported resolution corresponds to spatial frequency of 0.370 Å^{-1}

8.2 Resolution estimates (i)

| Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	2.70	-	-
Author-provided FSC curve	2.64	2.91	2.67
Unmasked-calculated*	3.30	4.06	3.42

*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps. The value from deposited half-maps intersecting FSC 0.143 CUT-OFF 3.30 differs from the reported value 2.7 by more than 10 %
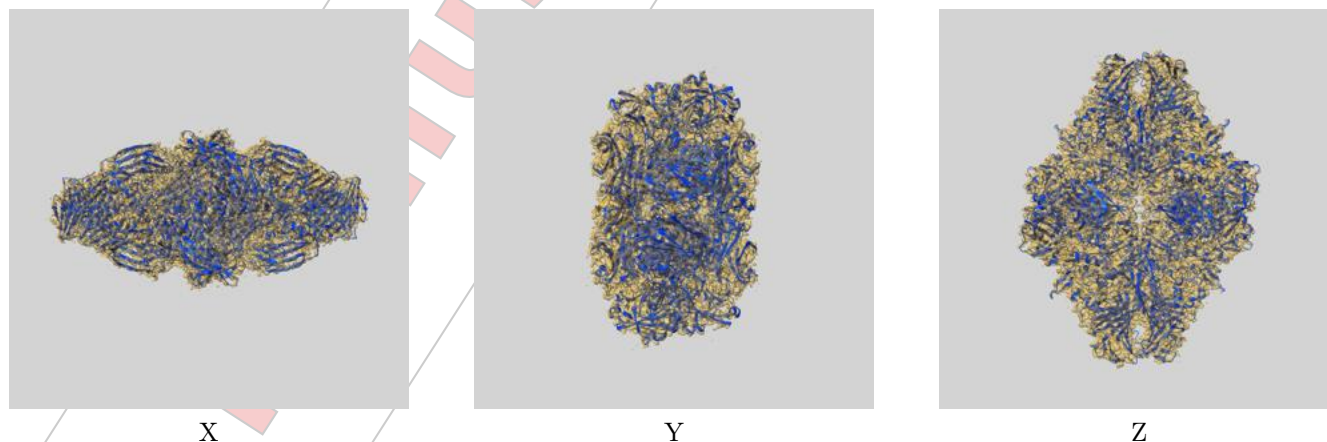
9 Map-model fit ⓘ

This section contains information regarding the fit between EMDB map EMD-18488 and PDB model 8QLZ. Per-residue inclusion information can be found in section 3 on page 5.

9.1 Map-model overlays

9.1.1 Map-model overlay ⓘ

X

Y


Z
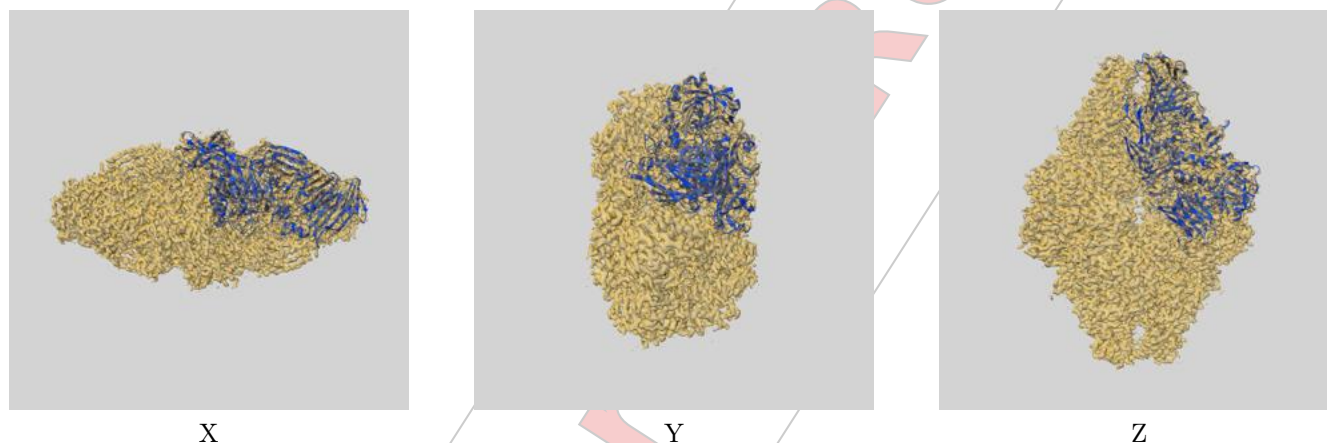
9.1.2 Map-model assembly overlay ⓘ

X

Y

Z

The images above show the 3D surface view of the map at the recommended contour level 0.02 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

## 9.2 Q-score mapped to coordinate model ⓘ

X　　　　　　　　Y　　　　　　　　Z

The images above show the model with each residue coloured according its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions at which atoms can be resolved. Low Q-score values may therefore be expected for many entries.

9.3 Atom inclusion mapped to coordinate model ⓘ

X　　　　　　　　Y　　　　　　　　Z
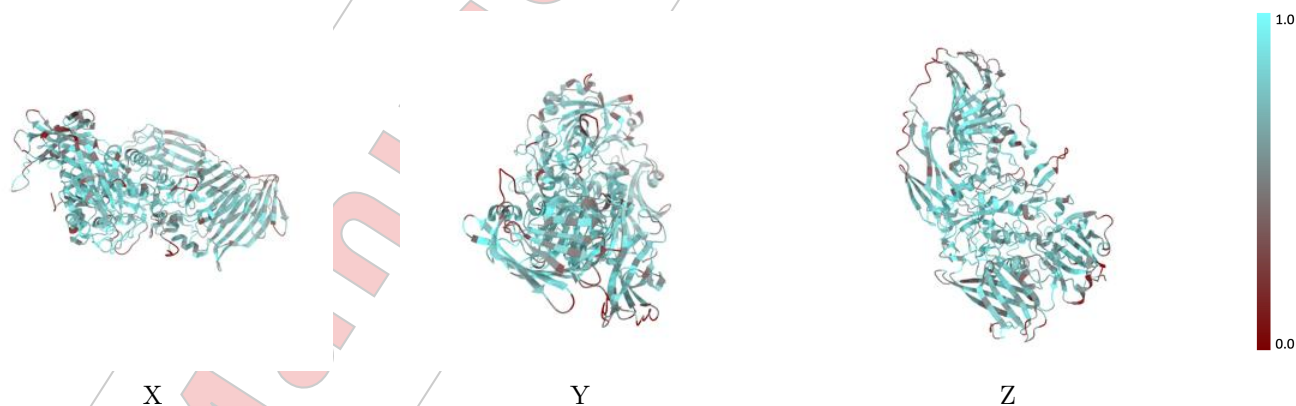
The images above show the model with each residue coloured according to its atom inclusion. This shows to what extent they are inside the map at the recommended contour level (0.02).

9.4 Atom inclusion ⓘ
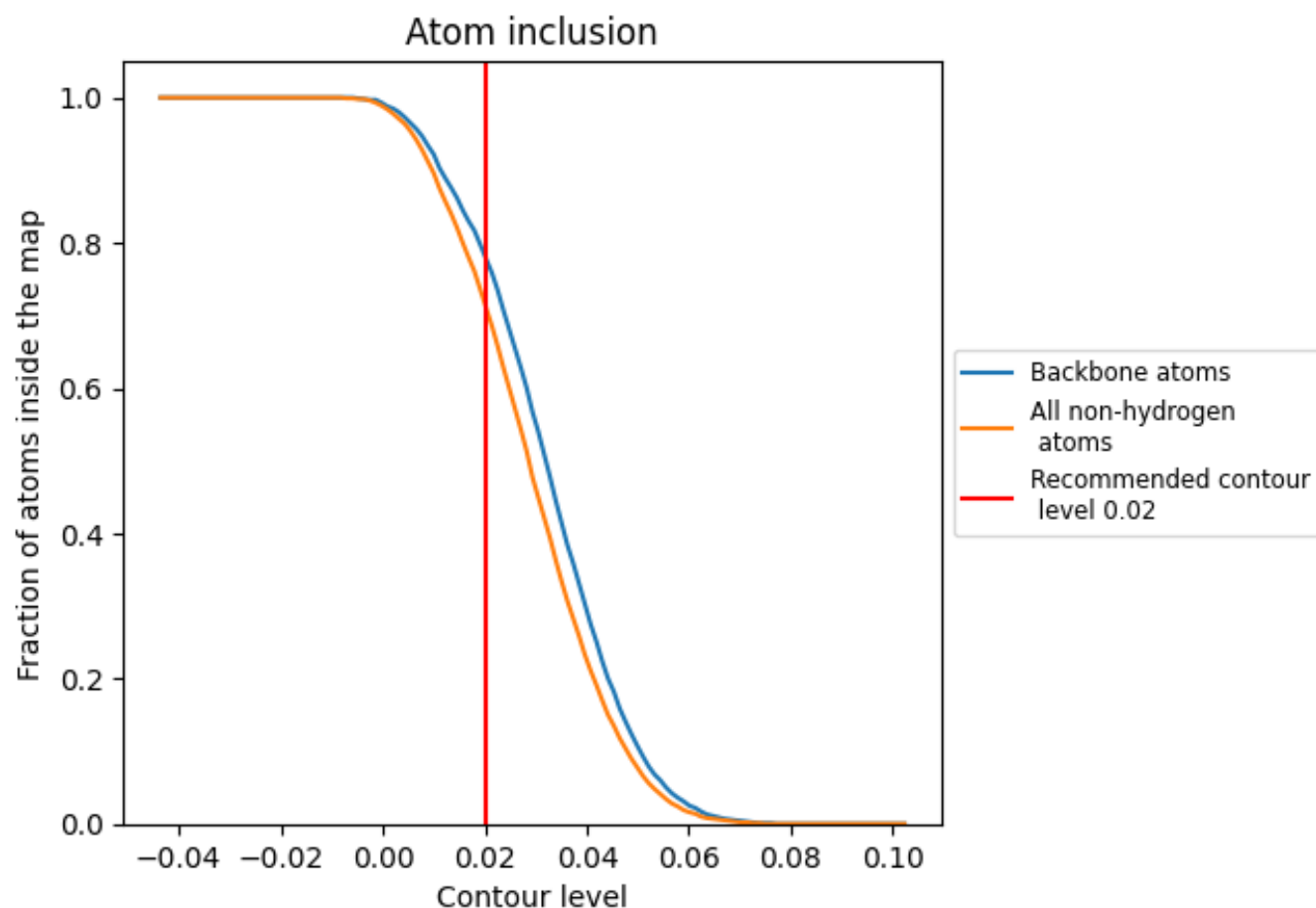
At the recommended contour level, 78% of all backbone atoms, 72% of all non-hydrogen atoms, are inside the map.

9.5 Map-model fit summary ⓘ

The table lists the average atom inclusion at the recommended contour level (0.02) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.7170	0.4750
A	0.7170	0.4750