

Towards time-resolved structural studies of membrane transporters

Dissertation for a doctorate at

Universität Hamburg

Chemistry Department

Diogo Melo, Hamburg 2024

Thesis Reviewers

Prof. Dr. Arwen Pearson

Universität Hamburg, Department of Physics, Nanostructure and Solid-State Physics

Prof. Dr. Daniel N. Wilson

Universität Hamburg, Department of Chemistry, Biochemistry and Molecular Biology

Examination committee

Prof. Dr. Tobias Beck

Universität Hamburg, Department of Chemistry, Physical Chemistry

Prof. Dr. Arwen Pearson

Universität Hamburg, Department of Physics, Nanostructure and Solid-State Physics

Dr. Thomas Hackl

Universität Hamburg, Department of Chemistry, Analytical and Scientific Services - NMR

Date of disputation: 01 March 2024



This project has received funding from the European Union's Horizon 2020 research and Innovation programme under the Marie Skłodowska-Curie grant agreement No 722687



This work was done under the supervision of Prof. Dr. Arwen Pearson at Universität Hamburg, in the DESY campus, firstly in the Centre for Free Electron Laser Science (CFEL) and later Hamburg Advanced Research Centre for Bioorganic Chemistry (HARBOR).

Abstract

This work describes progress towards carrying out time-resolved crystallographic studies of membrane transporters, using the bacterial sodium symporters LeuT and Mhp1 as model systems. Specifically, a new approach to reaction initiation was explored, where a disulfide crosslink was used to lock the protein into an inactive state. The disulfide could then be released by UV-laser excitation or by rapid mixing with a reductant solution. The degree of crosslinking was quantified, and a protocol developed to deliver a near 100% crosslinked sample for crystallization. Several spectroscopic methods were then tested to identify a protocol which could be used to monitor the on-set of ligand binding and transporter function upon crosslink cleavage. Finally, first steps towards the optimization of microcrystal growth were taken.

In parallel a number of other studies were made, some of which are also presented in this thesis.

The optimization of large-scale Mhp1 production, in order to allow SAXS studies of Mhp1 conformational changes and oligomerization under different buffer conditions are presented. In addition, Mhp1 datasets from a range of ligand complexes were reanalyzed and the utility of a novel torsion angle-based model and ensemble refinement were compared to standard atom-based modelling and refinement.

Preliminary work is presented on enabling time-resolved studies of a second class of membrane protein target, diacylglycerol kinase (DgKa), where reaction initiation will be via decaging of the substrate ATP. This includes first SSX diffraction data collections as well as an exploration of the compatibility of LCP-crystallized samples with laser excitation. Here the effects of light refraction in solid targets as well as laser induced phase changes in the LCP were explored.

As part of a wider programme to develop improved methods for membrane protein crystallization, a study of detergent diffusion rates across dialysis membranes, commonly used for diffusion-based crystallization, was made. This showed that not only were the diffusion rates of different detergents quite distinct, but also that the dialysis membranes strongly absorbed detergents, resulting in a lower actual detergent concentration than expected. This potentially explains some of the challenges in using diffusion-based approaches for membrane protein crystallization.

Finally, the large-scale production of HRV-3C protease and concentrative nucleoside transporter (CNT) proteins are described.

Zusammenfassung

Diese Arbeit beschreibt den Fortschritt bei der Durchführung zeitabhängiger kristallographischer Studien von Membrantransportern unter Verwendung der bakteriellen Natriumsymporter LeuT und Mhp1 als Modellsysteme. Insbesondere wurde ein neuer Ansatz zur Reaktionsinitiierung untersucht, bei dem eine Disulfidvernetzung verwendet wurde, um das Protein in einen inaktiven Zustand zu bringen. Das Disulfid konnte dann durch UV-Laseranregung oder schnelle Mischung mit einer Reduktionslösung freigesetzt werden. Der Grad der Vernetzung wurde quantifiziert und ein Protokoll wurde entwickelt, um eine nahezu 100%ig vernetzte Probe für die Kristallisation bereitzustellen. Anschließend wurden mehrere spektroskopische Methoden getestet, um ein Protokoll zu identifizieren, mit dem der Beginn der Ligandenbindung und Transporterfunktion nach Vernetzungsspaltung überwacht werden kann. Schließlich wurden erste Schritte zur Optimierung des Wachstums von Mikrokristallen unternommen.

Parallel dazu wurden eine Reihe von weiteren Studien durchgeführt, von denen einige ebenfalls in dieser Arbeit präsentiert werden. Es werden die Optimierung der groß angelegten Mhp1-Produktion zur Durchführung von SAXS-Studien zu konformationellen Veränderungen und Oligomerisierung von Mhp1 unter verschiedenen Pufferbedingungen präsentiert. Darüber hinaus wurden Mhp1-Datensätze von einer Reihe von Ligandenkomplexen erneut analysiert, und der Nutzen eines neuartigen Modells auf der Basis von Torsionswinkeln und Ensembleverfeinerung wurde mit der Standardmodellierung und -verfeinerung auf Atomniveau verglichen.

Vorläufige Arbeiten zur Ermöglichung von zeitabhängigen Studien an einem zweiten Typ von Membranprotein, der Diacylglycerin-Kinase (DgKa), werden vorgestellt. Dabei wird die Reaktionsinitiierung durch Freisetzung des Substrats ATP erfolgen. Dies umfasst erste SSX-Diffraktionsdatensammlungen sowie eine Untersuchung der Kompatibilität von LCP-gekristallisierten Proben mit Laseranregung. Hier wurden die Effekte von Lichtbrechung in festen Zielen sowie laserinduzierte Phasenänderungen in der LCP untersucht.

Im Rahmen eines umfassenderen Programms zur Entwicklung verbesserter Methoden für die Kristallisation von Membranproteinen wurde eine Studie über die Diffusionsraten von Detergenzien durch Dialysemembranen, die häufig für diffusionsbasierte Kristallisation verwendet werden, durchgeführt. Dabei wurde gezeigt, dass die Diffusionsraten unterschiedlicher Detergenzien deutlich voneinander abweichen und dass die Dialysemembranen Detergenzien stark absorbieren, was zu einer niedrigeren tatsächlichen Detergenzkonzentration als erwartet führt. Dies erklärt möglicherweise einige der Herausforderungen bei der Verwendung von diffusionsbasierten Ansätzen für die Kristallisation von Membranproteinen.

Abschließend wird die großtechnische Produktion von HRV-3C-Protease und konzentrativen Nukleosidtransporterproteinen (CNT) beschrieben.

Acknowledgements

Firstly, I want to thank Prof. Dr. Arwen for the exceptional opportunity I was presented with. Besides the multiple productive scientific discussions and tutorship, I think an extra dimension must be acknowledged here. Even though multiple hard times came through the years, both from my perspective as a student and hers as a supervisor, despite all difficulties, I know that she had at all times the best of intents, and efforts, looking out not only for the best of the research, but most importantly, for the best of those working under her or around her. Things might become hard, and not go as intended, but I think in the end, all anyone can ask for is genuine good intentions and a true attempt at practicing them, of which in contrast to other tutorships, this one did not have a shortage of.

I also want to acknowledge all the good colleagues I worked with, that provided either or both of positive learning experiences, and general support or good times, those are: Yunyun Gao, Henry Gieseler, Brandon Seychell, Sam Horrel, Marta Sans Valls, Jo Doyle, and Susanne Meier.

Lastly, I want to acknowledge everyone I worked with throughout my doctorate, many of whom are mentioned throughout this thesis, in this case a list of names would be too long to present in here, and so a general thank you will have to do, with a special remark for all the RAMP members I interacted with, the network was indeed something remarkable, to be cherished over the years in my memory – and occasionally, hopefully, in future in person meetings.

Contents

1.	Introduction.....	1
1.1.	Membrane Proteins	1
1.1.1.	Importance of membrane proteins and transporters.....	1
1.1.2.	Peripheral membrane proteins.....	2
1.1.3.	Membrane embedded proteins.....	2
1.1.4.	Membrane Protein Transport types	4
1.1.5.	Protein production and purification considerations.....	4
1.1.6.	Protein targets of this work.....	5
1.2.	Protein Structure.....	8
1.2.1.	Structure determination methods.....	8
1.2.2.	Macromolecular X-ray Crystallography.....	10
1.3.	Detergents and LCP	19
1.3.1.	Detergent general characteristics and characterizations.....	19
1.3.2.	Critical micellar concentration.....	20
1.3.3.	Membrane and protein solubilization by detergents	20
1.3.4.	Considerations in structural biology	21
1.3.5.	Detergent diffusion rates	22
1.3.6.	Lipidic cubic phase	23
2.	Methods.....	23
2.1.	LeuT.....	23
2.1.1.	Production and purification.....	23
2.1.2.	Sulfhydryl quantification	25
2.1.3.	Cysteine crosslinking	25
2.1.4.	Crystallization and optimization	25

2.1.5.	Circular dichroism	26
2.1.6.	Fluorescence ligand titration	26
2.1.7.	Crystallography data collection and processing	26
2.2.	Mhp1	28
2.2.1.	Large scale production and purification	28
2.2.2.	SAXS.....	30
2.2.3.	Structure refinement.....	30
2.3.	DgKa	31
2.4.	LCP-refraction	32
2.5.	Detergent studies.....	32
2.6.	HRV-3C.....	33
2.7.	CNT.....	34
3.	Results.....	35
3.1.	LeuT	35
3.1.1.	Purification problems/successes	36
3.1.2.	Cross-linking.....	39
3.1.3.	Binding assays.....	40
3.1.4.	LeuT Crystals	50
3.1.5.	LeuT X-ray data	52
3.2.	Mhp1	55
3.2.1.	Large scale protein production	56
3.2.2.	SAXS of Mhp1	56
3.2.3.	Structure refinement.....	60
3.3.	Detergent studies.....	66
3.3.1.	Diffusion rate	66
3.4.	Solid target and LCP compatibility studies for time-resolved crystallography.....	70

3.4.1.	LCP in solid targets	72
3.4.2.	DgKa LCP crystal slurry on solid targets	72
4.	Discussion	74
4.1.	LeuT	74
4.2.	Mhp1	76
4.3.	Detergent studies and LCP	78
5.	References and index.....	79
6.	List of hazardous substances	88
7.	Declaration on Oath.....	90

Figure 1-1: LeuT crystal structures compiled in outward, occluded, and inward conformations.	6
Figure 1-2: Types of membrane protein crystals	22
Figure 3-1: Lack of LeuT induction.....	36
Figure 3-2: LeuT protein expression SDS-PAGE and Western blot.....	37
Figure 3-3 Final regained LeuT expression	38
Figure 3-4: Fluorescence potential amino acids and corresponding results.	41
Figure 3-5: Partial helix distortion and unwinding between different LeuT conformations.	43
Figure 3-6: Average error distribution and leucine introduced background.	44
Figure 3-7: CD titration spectra of the WT and 3M LeuT.....	45
Figure 3-8: Ellipticity at 221 nm and their variations in percentage.	46
Figure 3-9 CD: spectra prediction of various LeuT conformational states.	48
Figure 3-10: Reproduction of the best crystal hit from commercial crystallization screen HR2-110.	51
Figure 3-11: Example pictures of the PEG screen made around the best crystallization hit from the HR2-110 screen which had 20% PEG 8K.....	52
Figure 3-12: Pictures of example data frames with different options.	53
Figure 3-13: Comparison of spot finding results with default options and refined.	54
Figure 3-14: Fermenter growth profile with various metrics.	56
Figure 3-15: SEC-SAXS elution profiles of the wild-type Mhp1 in complex with NM.....	57
Figure 3-16: SEC-SAXS elution profiles of Mhp1 in DDM.....	58
Figure 3-17: Coot models and density maps of IMH original and refined structures comparing ligand positions.....	63
Figure 3-18: Superposition of the Vagabond refined IMH model, the multiple REFMAC refined model, and the original model.	64
Figure 3-19: Cavities of original and refined Mhp1.....	65
Figure 3-20: First experiments for detergent diffusion rate measurement.	67
Figure 3-21: Experiment showing detergent absorption by dialysis membrane.	68
Figure 3-22: Additional studies made with non-detergent absorbing membranes.....	69
Figure 3-23: LCP mylar sandwich soaked with fluorophore and targeted with a laser.....	71
Figure 3-24: Frames of a video separated in seconds showing crystal movement on solid targets.....	73

Table 2-1: Buffers used in LeuT protein production.....	24
Table 2-2: Buffers used in large scale protein production.....	29
Table 3-1: Cross-linking efficiency assessment.	39
Table 3-2: α -helix contents of crystal structures, their predicted CD spectrums, and measured CD spectrums.	49
Table 3-3: SEC-SAXS elution peaks details and oligomer states.	59
Table 3-4: Statistics for each dataset.	60

1. Introduction

1.1. Membrane Proteins

Membrane bilayers made of phospholipids with embedded proteins are the boundary between living cells and their exterior. These bilayers / membranes limit molecular exchange to the slow passive diffusion of small molecules (*i.e.* O₂, CO₂, NH₃). Embedded proteins enable selective enhanced transport via various energy-free (channels) or energy-consuming (active transport) methods. The following sections elaborate on some of these characteristics of cell membranes and their proteins.

1.1.1. Importance of membrane proteins and transporters

Proteins, regulated by genes and other proteins, perform a vast, not yet fully catalogued or scientifically explored array of functions, each intertwined with multiple other functions via a complex mass of metabolic pathways.

Considering the significant and diverse role proteins play in all life forms, it is not an exaggeration to say that research into proteins has potential applications for a wide range of activities. Examples of applications in the health industry include understanding proteins' activity as neuro-regulators, antibodies and cell recognition, rapid tests for pregnancy and diseases such as COVID-19. Making use of the properties and functionality of proteins is all around us from mundane impacts such as creating chocolates with a molten core achieved by enzymes three days post production, to major life changing impacts such as neuro-regulators have on (partially) understood human G protein-coupled receptors, or insulin related diseases [1]–[3].

Membrane proteins are more difficult to handle than soluble proteins. This is due to several reasons, such as the need to use an added variable of solubilization agent due to their partially hydrophobic surfaces, their increased flexibility, often lack of stability in *in vitro* conditions, and lower yields of protein production [4], [5].

The importance of membrane proteins is, however, not at all diminished by the difficulty in studying them. By the end of 2022, there were 81018 total unique protein structures deposited at protein data bank of which only 1516 were membrane protein structures, making membrane proteins less than 2% of all unique known protein structures today, whereas over 80% of drug targets are membrane proteins [4], [6], [7]. This shows a stark contrast between our knowledge and the need for more understanding of the two protein types. It is therefore imperative that research into membrane proteins be improved.

1.1.2. Peripheral membrane proteins

Membrane proteins can be categorized by considering their location and function with respect to the membranes. Extrinsic or peripheral membrane proteins are a group of membrane proteins that interact with proteins or lipids at the membrane, usually through electrostatic interactions.

Peripheral membrane proteins attach or anchor themselves to the membranes via a flexible part of the protein. This mobile part of the protein interacts with the membrane in a number of possible ways: for example via electrostatic interactions with multiple phospholipid head groups, with a specific phospholipid headgroup only, with an amphipathic helix at the interfacial volume of the membrane, or an anchored phospholipid in the membrane. In all these cases, the mobile portion that is responsible for interacting with the membrane is covalently bound to the protein.

A second type of peripheral membrane proteins are proteins, or even peptides, that partially insert themselves in the membrane bilayer. Colicins are an example of such a protein, used by bacteria as bacteriocins. These are capable of interacting with bacteria receptors by inserting themselves into the membrane and then using the transmembrane translocation system to reach the cytoplasmic side of the bilayer, at which point they act as high-rate ion leaking channels, killing the host bacteria.

1.1.3. Membrane embedded proteins

Transmembrane proteins fully transverse the membrane bilayer and can be further divided depending on how often they cross the entire membrane bilayer. Monotopic membrane proteins do not span completely from one side to the other of the membrane bilayer, but they are still considered to be embedded in the membrane. They differentiate from the peripheral membrane proteins by their high propensity to be at the bilayer, and need to be solubilized when purifying from membranes. This in contrast to the peripheral membrane proteins that do not require solubilization, as a larger portion of these are water soluble.

Bitopic membrane proteins cross the membrane once, before being further subdivided as to whether they have their N-terminus towards the inner or outer side of the cell. Polytopic proteins cross the membrane multiple times, connecting trans-membrane sections with loops. Lastly, multiple bitopic proteins can come together to interact to form trans-membrane oligomers, making up their own group of membrane embedded proteins.

The structures of trans-membrane proteins, most commonly solved by X-ray crystallographic methods, show that the secondary structure of membrane proteins is distinct from their water-soluble

counterparts. In membrane proteins the nonpolar amino-acids in α -helices commonly point outwards, favouring nonpolar interactions with the fatty acid carbon chains. Acidic or basic amino acids are in their uncharged state, form ion pairs to neutralize their charge, or play specific roles involving proton and electron binding or transport. Hydrogen-bonds are commonly found stabilizing intra α -helix backbones and oligomeric α -helices, or forming helix caps. In water soluble proteins, glycine and proline usually terminate α -helices, whereas in trans-membrane proteins they are more often conserved amino acids within α -helices. Glycines allow for a tighter α -helix oligomerization, giving rise to a tighter packing, and proline introduces a tilt in α -helices due to its restricted backbone angles. Amino acids capable of being both donor and acceptors of H-bonding are found promoting oligomerization.

There are also common functional alterations to the secondary structure of trans-membrane proteins, one of them being the tilt induced by prolines in α -helices. There are π -bulges caused by missing H-bonding of a backbone carbonyl amino acid, helix unwinding (as in the case of LeuT), two half helices joining to span the membrane bilayer, and portions of 3_{10} helices.

As might be expected, lipids themselves also play an important role with regards to membrane proteins. Folded proteins are significantly more "rigid" than lipids and their insertion into the membrane bilayer has a strong impact on the immediate surrounding lipids, constraining their movement and forming nonpolar interactions. These lipids are called boundary lipids, whereas the more distant lipids that are not affected by this effect are bulk lipids. It is also often the case that membrane proteins require specific lipids for their function, even when these lipids are not the substrates themselves, giving them the name of cofactors.

It is also worth noting the hydrophobic mismatch phenomenon arising from possible differences of the nonpolar region height of a protein and its surrounding membrane bilayer thickness. Depending on the situation, the mismatch can either force a change in the bilayer thickness, or a conformation/orientation change of the protein. Single α -helices are not able to deform the surrounding lipid bilayer, instead they tilt within the bilayer to match the membrane local thickness. Larger protein structures tend to deform the surrounding lipids, extending or compressing its local height to accommodate the proteins' nonpolar area, satisfying the mismatch. Hydrophobic mismatch is thought to play a part in membrane protein sorting to different cell locations, where proteins with adequate hydrophobic region heights stay for example at the Golgi complex vesicles rather than being transferred to the cells' membranes due to differences in lipid bilayer thickness.

1.1.4. Membrane Protein Transport types

Passive diffusion is the free slow diffusion of small molecules across the membrane bilayer towards the lower concentration side of the membrane, assisted by the second law of thermodynamics and lowering the overall system entropy. This transport type is slow compared to all other types of transport across the membrane, but it not only is energy-free to the cell, but also requires no protein production as opposed to facilitated diffusion.

Larger and/or hydrophilic molecules that do not significantly permeate the membrane bilayer require accelerated transport, are often transported via facilitated diffusion. In this case protein channels or diffusion transporters are used. Membrane protein channels can open a pathway from one side to the other of the membrane bilayer, enabling diffusion of molecules not possible by passive diffusion only. These channels are molecular specific and often regulated between open or closed states, enabling quick responses to substrate concentration changes [8].

Primary active transport is when a protein transports molecules against their gradient, requiring active energy consumption in the form of metabolic molecules such as ATP or light absorption (*e.g.* light absorption enabling proton pumping in photosynthesis) [9], [10].

Secondary active transport happens when proteins transport molecules against a gradient as in the case of the primary transporters. These require energy, but instead of consuming metabolites such as ATP, they couple the transport to a gradient favourable molecule, transporting different molecules towards the same (symport) or to opposite sides (antiport).

An example of secondary active transporter function is in neurons. After an action potential with opposed sodium and potassium ion concentrations is established across the membrane bilayer, sodium channels open leading to quick sodium diffusion. Depolarization then happens via potassium channels re-establishing ion equilibrium; and finally repolarization occurs when sodium potassium pumps actively (secondary) transport sodium and potassium ions outwards and inwards respectively, re-establishing the action potential [10]–[13].

1.1.5. Protein production and purification considerations

Membrane protein purification requires a few more considerations and steps than water soluble proteins. One of the major and typical bottlenecks is expression levels and yields. High protein concentration at the cellular membranes can constrain functions in the bilayers, eventually disrupting significant cellular processes. Although this is not always the case (*i.e.*, inclusion bodies), overexpressed

proteins at the membrane bilayer become toxic to the cells and therefore limit potential yields. Post-translation modifications and additional methods might also be required to have the membrane proteins correctly folded, further pressuring successful protein production and purification protocols[14].

Typical additional steps include centrifugation to isolate the overexpressed proteins from their membranes, sometimes requiring multiple cycles of centrifugation and washing. Membrane proteins often aggregate requiring different expression systems and strategies besides *E.coli* [15], [16]. Protein solubilization is required after cell disruption and various detergents and conditions must be tested and optimized [17], [18]. These conditions then need to be satisfying enough for the protein to remain stable for reasonable periods, throughout all the purification procedures and following experiments of interest.

For protein crystallography, potential low protein yields and highly disordered detergents negatively contrasts with the methods to produce high protein concentration and crystal ordering. The development of methods such as LCP and HiLiDe (high lipid detergent) now provide additional options for the crystallization of membrane proteins. More details regarding membrane protein crystallization and detergent and lipid usage are presented in the corresponding later sections of the introduction.

1.1.6. Protein targets of this work

In this thesis I worked on various projects, of which 3 different membrane proteins were involved: LeuT, Mhp1, and DgkA. Various methods were used, including X-ray crystallography, circular dichroism, fluorescence ligand titration, small angle X-ray scattering and large-scale protein production.

1.1.6.1 *LeuT*

Major Facilitator Superfamily (MFS) are membrane proteins that transport various drugs, metabolites, sugars, amino-acids, and ions. Within this superfamily the Neurotransmitter Transporters are responsible for the re-uptake of neurotransmitters at the synapses of the neurons via secondary transport dependent of sodium and chloride (NSS, neurotransmitter sodium symporter) or sodium and potassium (EAAT, excitatory amino-acid transporter). This function is required after nerve impulses to avoid neuronal over-stimulation and returning to a ready state for subsequent signaling. Neurotransmitter transporters are special medicinal targets, they are involved in various medical conditions including schizophrenia, depression, strokes, Parkinson's disease, autism and epilepsy, and they are targets of antidepressants fluoxetine, desipramine, and imipramine as well as chemicals like cocaine [4], [19].

LeuT (leucine transporter) from *Aquifex aeolicus* was the first ortholog within the neurotransmitter sodium symporter family of the serotonin, GABA, and biogenic amine transporters which a crystal structure was determined (1.7 Å resolution) significantly increasing our understanding of this family's structure[4].

LeuT shows 12 transmembrane helices (TM 1-12), with a structural internal repeat of TM 1-5 and 6-10. The helices 3-5 and 8-10 form a scaffold, in which TM helices 1,2,6, and 7 are the core. TM 1/6 and TM 2/7 form the central translocation pathway where the substrate and sodium ions bind. Transport is proposed to be done by movement of the protein core (TM 1-2, 6-7). The outer parts of TM 1/6 (TM1a/6a), with help from the external loop 4 (EL4), block the outwards substrate path in the inward conformations and the interior TM1b/6b protrude outwards on the open conformation; alternating to the outward open conformation when TM1b/6b tightly close and TM1a/6a and EL4 open[4], [20]–[22].

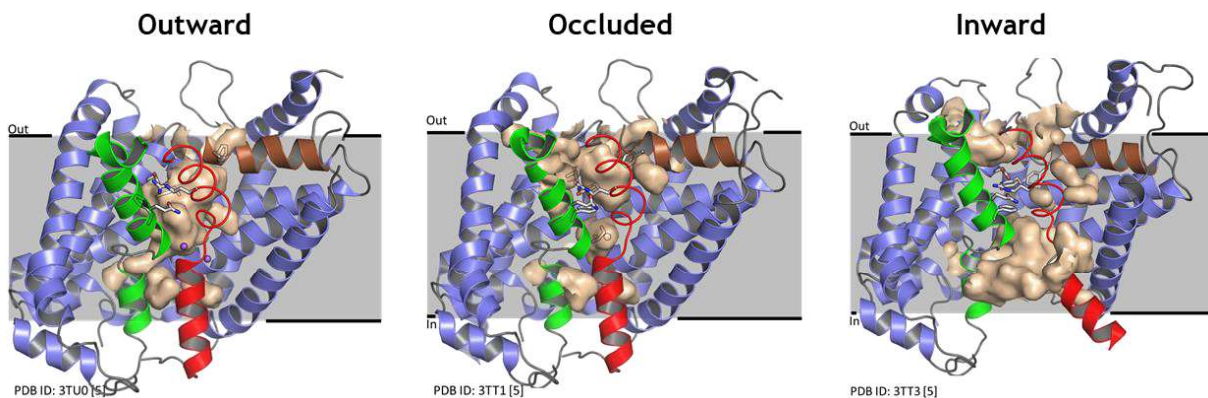


Figure 1-1: LeuT crystal structures compiled in outward, occluded, and inward conformations.

The various crystal structures of LeuT in 3 conformations. In red and green TM 1 and 6, and representing protein cavities. Image made with the PyMol software.

A combination of three crystal structures and corresponding PDB IDs are presented (Figure 1-1) showing the outward open, outward occluded, and inward open conformations.

Some key differences between this and the human ortholog is the lack of N and C terminal extended portions that interact with modular proteins at the synapse, as well as the lack of phosphorylation sites or terminal domains, and intracellular loops involved in function regulation. Tricyclic antidepressants are noncompetitive inhibitors of LeuT, however, mutations can make some of these act as competitive inhibitors similar to the human orthologs [4].

1.1.6.2 Mhp1

Mhp1 belongs to the LeuT superfamily. As observed in other family members it also has the same 12 transmembrane helices (TM1-12), a structural repeat from TM1-5 and TM6-10, and TM 1,2,6,7 at its core. *Microbacterium liquefaciens* Mhp1 is from the NCS-1 family (nucleobase-cation-symporter) which has over 1000 proteins in archaea, bacteria, plants, and fungi all with the function of importing organic compounds such as cytosine, uracil, thiamine, and various hydantoin compounds [23], [24].

Mhp1 was the first protein belonging to the LeuT family to be structurally determined in three different crucial conformations (inwards, outwards, and closed) by protein crystallography [25], [26]. These structures served to further advance our understanding of the mechanism of action of these proteins. Mhp1 couples sodium ion transport towards their gradient to hydantoin compounds into bacteria cytoplasm [27]. A few datasets of Mhp1 with mutations and ligands were collected years before this thesis and were passed onto me to do final refinement cycles and later deposit in the PDB.

1.1.6.3 DgkA

Diacylglycerol kinase (DgkA) catalyzes the conversion of diacylglycerol to phosphatidic acid. In *E. coli* it is involved in osmotic response, and it has served as a membrane protein model for many fields such as enzymology [28]–[30], folding and assembly [30]–[35], and stability [36], [37]. The protein DgkA is a trimer with positive heteroallostery and some form of lipidic substrate promiscuity [28]. Crystal structures were obtained of both WT and mutants in LCP, and although at first structural results seem to conflict with NMR data [38], it was later shown the protein was active within the crystals [39].

DgkA is an interesting protein for crystallographic time-resolved studies. It has been extensively studied, with crystallization conditions found using LCP, and crystal structures showing the protein binding to a substrate and an inert version of DgkA, suggesting it is ready to be activated, needing only an activating ATP [40]. On top of this, NMR studies have shown enzyme activation can be controlled via decaging NPE-caged ATP [41]. Therefore, we began a collaboration with Martin Caffrey and Boland Coillín where they provided one syringe with DgkA crystal slurry.

Experimental work involving DgkA was limited to a few experiments on sample availability. Here the aim was to perform a few preliminary tests in order to support subsequent synchrotron time-resolved experiments. The tests revolved around stress-testing sample delivery via solid target, using chips which were previously mainly used in non-LCP protein samples [42], [43]. The main question here was if we could physically, practically, and efficiently deploy the LCP crystal slurry into the chip wells. Knowing that

LCP is quite viscous it had to be shown that sufficient crystals could indeed be positioned in the wells, to provide a satisfying hit-rate during crystallographic experiments.

Additionally, I posed the question if we could instead use the usually limiting LCP viscosity factor to our own advantage, using chip holding mylar sandwiches, usually used to hold the solid targets, but without any chip itself. In this case the mylar sandwich would hold the crystals still enough due to the LCP high viscosity. Furthermore, we could also possibly use the grid screen software to locate crystals in the whole LCP area. For this to be a time-resolved experiment we would have to validate that the triggering laser light would not uncontrollably refract to surrounding regions. To do this we set out to experiment with fluorescent doped LCP in a mylar sandwich, laser triggering the fluorescence and a camera with a triggering laser filter that would let the fluorescence emission light pass, thereby showing which regions in the LCP would be triggered after laser absorbance.

1.2. Protein Structure

1.2.1. Structure determination methods

Historically, scientists first discovered the function of individual proteins, before studying and understanding their structure. However, proteins are able to perform their function as a result of the protein and its surroundings providing adequate chemical and physical conditions. There is a chemical, physical, and space arrangement that the protein structure needs to satisfy for its function to be performed. If we agree with this statement, it becomes clear that the structure information can explain how the protein performs the function, which in turn is also invaluable information to understand how we may generate ligands for a desired effect for example in drug design, or to modify its function as in bioengineering. The importance of determining a protein structure, even though we might know its function, is the same as knowing how something works instead of just what it does.

Indeed, structural information is valuable but also difficult to obtain. Over the years a number of new techniques have been developed that try to address issues involved in e.g., sample preparation or radiation damage. Each method comes with its own sets of advantages and disadvantages, so no one approach is *the* best approach for solving the structure of all types of proteins. Instead, each method should be viewed as a tool with specific requirements and advantages in a growing toolbox for protein structure solving.

One of the oldest techniques used for solving the structure of proteins is X-ray Crystallography. Here X-Rays are used to produce diffraction patterns from crystallized proteins. To date this technique has

resulted in the most structures and those of the highest resolution. Its advantages are the high-resolution potential and vast history, knowledge, and the improvement of the method. The disadvantages of the method are that it requires crystallization of the protein samples, which not only demands a high amount of material making low protein production yields a struggle, but also having to the need to identify ways to crystallize the protein itself, which may just not be possible. The crystallization aspect of X-ray Crystallography has been studied extensively, but it still shows a degree of undesired empiricism on how to obtain crystallization conditions. A lot of work has been done to understand and work with the empiric nature of crystallization, such as forming numerous crystallization screens that use many different, previously successful crystallization solutions, together with the lowest possible amount of protein per condition to see if any crystallize. An example of this is the combination of these screens with robots to pipette the tiniest amount of liquid per experimental well, maximizing the efficiency of the “empirical” method towards time and sample consumption.

Nuclear Magnetic Resonance (NMR) was developed sometime later than X-ray crystallography. This method makes use of the nuclear spins to ascertain the distances between individual atoms, gradually adding information of such distances until the whole (macro)molecule can be assembled in 3D space. The advantage of this technique is that it works in solution, allowing for a sample environment more *in vivo* than if crystallized. The disadvantages, however, are that as the number of atoms increases, so does the complexity of measuring atomic distances, requiring not only higher computational power, but most importantly, stronger magnetic fields. Due to the technical limitations of stronger magnetic fields, the technique is limited to smaller, less complex proteins.

The newest structural technique is single particle Cryo-Electron Microscopy. This technique makes use of a fast-cooling system. Thin layers of protein samples are required to take images of the sample in an electron microscope, recording images of individual proteins in hopefully random orientations. These 2D projections of the proteins are then pooled together to increase signal to noise ratio and reconstruct a 3D model of the protein. This technique has the advantage of not requiring crystals. A huge advantage of this technique is that it can work for large proteins and complexes. However, smaller proteins can be difficult to image due to a lack of contrast, presenting a size limitation quite in contrast to that of NMR. Furthermore, the protein being studied may prefer certain orientations on the grids used, making it difficult to have enough 2D projections for a full 3D structure.

1.2.2. Macromolecular X-ray Crystallography

Macromolecular X-ray Crystallography is a method used to obtain atomic resolution 3D protein structures by using X-rays and protein crystals. The first protein atomic resolution structures were obtained via this technique, which today still leads the way in experimentally determined protein structures. At the time of writing (August 2022), the protein data bank (<https://www.rcsb.org/stats>) holds 164777 protein structures of which 145544 (88 %) were provided by X-ray crystallography.

1.2.2.1 Protein Crystallography

1.2.2.1.1. Crystals

The crystallization of proteins is a procedure which, by various means, gradually increases protein concentration to eventually obtain highly pure and organized protein crystals. This process must be slow and gentle enough so that proteins precipitate in a steady and organized fashion to form crystals. This contrasts with disorganized protein precipitation. Crystals are highly organized and can be represented by a unit cell – the smallest group of molecules that can reconstruct the crystal by translational symmetry operations. Unit cells are characterized by the lengths “a”, “b”, “c” and angles “ α ”, “ β ” and “ γ ” [44], [45].

Unit cells contain at least one asymmetric unit, which, for proteins, reconstructs the unit cell by rotational and translational symmetry operations and belong to one of four types: Primitive lattice (P), base-centered lattice (A, B or C), internal-centered lattice (I) or face-centered lattice (F). When combined with the seven crystalline systems, 28 crystalline lattices are available. Since proteins are chiral, the only symmetry operations available are translation and rotation, excluding any mirror planes or inversion centers. Because of this, the possible space groups for protein crystals are reduced from 230 to 65, all of them available for consultation at “The International Tables for Crystallography”. An important distinction from i.e., salt crystals is that protein crystals contain solvent channels representing 30-70% of the crystal as first described by Matthews, in part justifying why protein crystals are less stress resistant, and allow for small molecule diffusion through them even after crystallization [44]–[46].

1.2.2.1.2. Crystallization process

The first step in the growth of protein crystals is spontaneous nuclei formation, these are a primary structure to which further protein molecules can be added, hence growing into a crystal. Nuclei are not the same as small crystals, and fundamental studies on crystal/nuclei formation are still under way. Alternatively, nuclei or “seeds” can be added to the solution to stimulate or serve as nuclei to which protein can be added, growing into crystals without themselves having to form nuclei of their own. A typical and popular way of doing this is by streaking the seeds using cat’s whiskers [44], [45], [47], [48].

Phase diagrams represent the physical state of the crystallization mixture and are useful tools to track and explain the crystallization process. Crystallographic phase diagrams are composed of various areas or phases; commonly precipitation, nucleation and crystal growth. More generally, undersaturation, metastable, and supersaturation regions can also be depicted. Within the supersaturation region, proteins will start to precipitate out of solution, and if the conditions are right, within this region a nucleation zone may be found, where nuclei are formed. The protein in solution will gradually precipitate either by addition to already formed nuclei and/or crystals, or in a disorganized manner which will produce amorphous precipitate.

As protein comes out of solution, its concentration is reduced, and we move towards the metastable or crystal growth region. Here proteins will continue to be added to favorable lattices if nuclei or crystals are formed, but no spontaneous nucleation will occur. In the case that nuclei or crystals are present, the protein solution concentration will continue to decrease until it reaches the undersaturated region of the phase space, at which point crystals will stop growing. At this stage crystal size will remain constant. If, for some reason, protein crystals are placed in an undersaturated region, they will dissolve, until either the metastable interface is reached (as protein in solution will increase as the crystals dissolve), or they are completely dissolved.

Various methods are employed to attempt to find or navigate the phase diagram in favorable ways. As an example, in vapor diffusion the droplet can start in the undersaturated region. As water is lost towards the reservoir solution via air diffusion, both precipitant agent and protein concentrations will increase (since there is less water dissolving them), so that the solution eventually reaches the metastable region. If no nuclei or crystals are present, it will carry on into the precipitation zone, and if successful the nucleation zone. Nuclei are formed and protein concentration starts to lower as more proteins are added to the formed nuclei and grow into crystals. This continues until the metastable-undersaturation interface is reached. If more protein is added to the droplet, precipitant concentration increases, or temperature decreases, the crystals can grow even further[44], [49].

1.2.2.1.3. Crystallization methods

There are several crystallization methods and most use a closed system “well” which is covered by a coverslip to stop or slow down air exchange with the surroundings. The most common method is vapor diffusion, which can be performed by a so-called hanging or sitting drop. In hanging drop, the crystallization solution and protein mixture are applied on the top of a siliconized coverslip and the protein

is added. Then follows the sealing of the well with the cover slip. As for the sitting drop, the droplet is prepared on top of a bridge which sits onto the mother liquor at the bottom of the well.

Dialysis buttons consist of a small button enveloped by a dialysis membrane and immersed in the crystallization condition. The dialysis button has a depression where the protein is added prior to immersion. Free interface diffusion usually makes use of capillaries to which protein and crystallization solutions are added on opposite ends. These then meet somewhere along the capillary. A gradient is achieved between 100 % protein and 0 % (or near values) which will create larger crystals near the protein end and smaller crystals towards the precipitant end. This technique is useful to roughly screen various precipitant concentrations quickly and easily with lower protein sample usage.

Micro batch is a technique that makes use of the protein and water immiscibility in oil. The protein and crystallization solutions are added to oil at the bottom of the well as a single drop. Water will migrate from the drop through the oil at an even slower rate than standard vapor diffusion.

The use of crystallization robots for automated crystallization is a popular and appealing option. Crystallization robots are used for initial screening assays. Using specific crystallization plates, the robots automatically add protein and crystallization solutions to the wells. The machines complete the screening assays much faster, but also use significantly less protein sample (down to the nanoliter-scale) when compared to manual screens.

1.2.2.2 X-ray crystallographic experiment and data

1.2.2.2.1. X-ray sources

The first generation of X-ray sources were X-ray tubes, in which electrons collided with an anode under vacuum conditions. In these conditions, inner orbital electrons are ejected when the electrons hit the metal. An outer orbital electron then fills the empty orbital, releasing energy in the form of X-rays. Each metal emits photons of specific wavelengths. However, this procedure heavily degrades the metal in the region in which the electrons collide. To solve this problem, new X-ray sources were developed – rotating anodes. In this case the anode is rotated and cooled, and the area of impact of electrons is distributed along the anode. By doing this, the anode degradation was dispersed across a larger area. However, the temperature rises under vacuum conditions remain a problem, limiting the achievable flux despite water cooling of the anode [44].

Today, synchrotron radiation is widely used. Electrons are accelerated and then maintained in a stable orbit in storage rings and emit X-ray photons when forced to change direction by a series of

magnets. The photons released then pass through monochromators that make X-ray wavelength selection possible. Synchrotron radiation revolutionized X-ray crystallography by the significant time reduction required for data collection, increased data quality and reduced crystal size requirement.

More recently, X-ray free electron lasers (XFEL) sources have been built. Here electrons are generated by a laser hitting a metal target, before being accelerated in superconducting niobium resonators. The electron bunches then go through undulators thereby producing X-rays. During this process, electrons and photons are grouped in bunches, concentrating them into shorter and more intense pulses, ultimately producing increased brilliance on the order of thousands compared to synchrotron sources, and extremely short pulse lengths of several femtoseconds. The short pulses can interact with protein crystals before significant radiation damage has been caused, and hence allow recording of (near) radiation damage free data. A high repetition rate also means that data can be collected at a much faster pace, provided fresh sample can be delivered quickly enough to the X-ray interaction region.

1.2.2.2.2. X-ray experiment and diffraction

During a diffraction experiment, X-rays pass through a protein crystal and are diffracted by the electrons of the protein. The diffracted beams hit the detector and their intensities and relative positions are recorded, although the phase information is not measured and therefore lost. The experiment results in a diffraction pattern, a plane with various spots varying in size and intensity. The crystal orientation is changed over the course of the diffraction experiment, and so many diffraction patterns are recorded from the different lattice planes of the crystal. The diffraction patterns are indexed according to the symmetry of the crystal and the result of the experiment is a list of intensities across the reciprocal space sphere and an estimate of the uncertainty of each measured intensity.

Since the crystal is a periodic organization of molecules, for any orientation of the crystal with respect to the incident X-ray beam, some atoms will diffract X-rays in the same direction and in phase across all illuminated unit cells. These atoms are said to diffract X-rays in a constructive manner. In contrast atoms not structurally repetitive inside the crystal (*i.e.*, solvent channel waters) cannot scatter the incoming X-rays constructively into individual reflections and contribute only to the background scatter. W.H. Bragg and his son W. L. Bragg discovered and mathematically described this phenomenon, now known as Bragg's Law.

The planes of atoms within a crystal that produce constructive interference can be represented by three numbers, called the Miller indices (or lattice indices) – h , k and l . These indices represent sets of parallel planes with constant interplanar spaces of d_{hkl} . If the scattered X-rays from these planes are in phase, they interfere constructively, resulting in multiple diffracted/reflected X-rays exiting the crystal in the same direction and in phase –producing a spot or reflection in the diffraction pattern. Each crystal orientation has a different sets of planes that meet Bragg’s law condition. The crystal is either rotated in order to capture the diffraction from all planes of atoms, or alternatively multiple smaller crystals are fed to the beam in random orientations, in which case it is called a serial crystallographic experiment.

1.2.2.2.3. [From diffraction images to merged data](#)

After crystallographic data collection we are left with multiple frames or diffraction patterns. The process to transform these images into usable data is summarized here, with a more pronounced focus on any information required to understand later methodology and results. An emphasis is made on which tasks are done and required for the processing, as opposed to which programs are used for each of them.

The first step is always to manually inspect a few images. This can save a lot of downtime by checking for any big defects or difficult characteristics in the data that will later cause problems. Some examples are checking for detector inconsistencies such as large dead pixel areas, identifying whether the spots look good, well defined, or streaked, or even if spots are missing at the beginning or end of the frames. Other characteristics to look out for are water rings or other unexpected powder rings and multiple lattices. Noticing these early makes it possible to be proactive about data processing later on. For example, this could be by masking a questionable detector region, selecting only a subset of frames to be processed, or masking specific resolution rings (water or other) etc. It is also worthwhile doing this for the multiple datasets collected, and choosing the ones that look better overall, while filtering away datasets that are not really worth looking at.

The next step is spot finding. This is where the software will go over every selected frame, looking for pixels with a significantly higher signal as compared to the background. Sometimes the general default parameters are not adequate enough for specific datasets. This is especially the case for difficult or weak datasets; these must be tweaked. Typical values to adjust are shoebox sizing (area accepted by the software around the maximum intensity peak), minimal signal to noise accepted ratio (lowering will improve spot finding but increase false positives), signal to local mean for the pixel to be considered as peak/strong pixel (related to shoebox settings) and minimal and maximum spot size. Various iterations are often done in difficult cases attempting to optimize these settings.

The software can then determine the lattice of the dataset after running the spot finding program with established settings. Spot prediction (from the chosen lattice) is carried out and the solution quality for the lattice choice is assessed. After defining masked areas (shadows or ice-rings) cell refinement is done. This is followed by integration. During the integration step, spot intensities are counted, and their errors estimated. This is usually a computationally heavy step. Analysis of the Laue group and space group are then carried out. If we are happy with the current outcome, scale and merging comes next, where the data are combined into one set of structure factors, normalized according to symmetry, and an estimate of mosaicity is made. Finally, intensities are converted to amplitudes. At this point the images have been converted into merged datasets of $h\ k\ l$ values, intensities, amplitudes, with associated error estimates. The next step is solving the phase problem, which is explained in the following section with emphasis on molecular replacement as this is the most used method and the only one used for this work.

1.2.2.2.4. Phase problem

Fourier showed that any complex periodic function can be described as the sum of simpler periodic functions. Diffracted X-rays can be described as three-dimensional waves. They are therefore a complex periodic function and can be described as the sum of various simpler functions.

The 3D electronic distribution of the atoms in the protein can be calculated using the Fourier Transform (FT). The FT transforms the values of a periodic function into their reciprocal values, in the case of X-ray Crystallography it transforms the reflection data (reciprocal space) into electron density data (real space). FT is reversible, therefore applying it to electron density data gives back the reflection data.

The FT requires the amplitudes and phase angles to transform diffraction data into real space information. Amplitudes are calculated directly from the indexed list of reflection intensities obtained experimentally. The phase angles are, however, missing, resulting in the “phase problem” where the crystallographer needs to obtain phase information by other means. Molecular replacement is one way of obtaining phase information. It is popular and widely used because it does not require any additional experiments, however, does require an already existing homology model of the protein being studied. The molecular replacement method consists of placing the already obtained structure into the new unit cell using a series of rotation and translation operations. Once this is complete, the predicted phases from the model are calculated. These phases are taken as initial estimations of the real phases [44], [45], [49].

1.2.2.3 Model improvement

With the estimated phases and the measured intensities, the FT is applied, and a 3D map of the protein electron density is obtained. This map is however significantly influenced by the model introduced by molecular replacement, and in fact the phase information dominates the amplitude information [22].

The crystallographer interprets the electronic density and makes adjustments that, in principle, will approximate the model to the real structure based on model expectations from *e.g.*, biochemistry. After significant adjustments are made, another FT is calculated updated reciprocal-space information about the structure, comprising the measured amplitudes and updated phase estimates. Refinement software uses this information to further optimize phases, approximating the model's calculated structure factors to the data obtained experimentally. Cycles of FT and real space model improvement are made, converging the 3D model to the structure that experimentally diffracted the X-rays [44], [45], [49].

To evaluate whether this process is over-fitting or diverging from the original data, as well as to assess overall data quality, various parameters are used in combination. Some of the most used parameters and their expected values are: 1) Ratio of signal to noise ($\langle I/\sigma \rangle$) (≥ 2); 2) completeness, which represents the percentage of measured expected reflections ($\geq 90\%$ at the highest resolution shell); 3) correlation coefficient (CC1/2) representing the correlation between two halves of a randomly split data set (≥ 0.5); and 4) Rmerge which is the averaging of multiple symmetry-related reflections (depending on the study case ≤ 0.25). Comparisons of these values across different structures are best made when using similar resolution datasets.

1.2.2.4 Time-resolved X-ray crystallography

1.2.2.4.1. Photosystem II

Photosystem II catalyzes light-induced water splitting producing oxygen, hydrogen, and electrons. An experiment reported in 2014 showed 5 Å resolution structures of the dark S_1 and putative S_3 states [50]. The experimental trigger method was two laser flashes, and a third flash was used to reach S_4 . Because each pulse could not guarantee full state transition, later states become more heterogeneous in successful transitions, and so S_4 was used to corroborate successful S_3 data collection, which had different unit cell dimensions than S_1 . S_4 preliminary data evaluation showed unit cell recovery, therefore suggesting that the S_3 unit cell expansion was indeed due to conformational changes over the reaction mechanism, as opposed to merely crystal degradation after laser pumps. Crystal optimization and production was monitored by DLS and SONICC and crystals averaged 1 μm in size. The experiment was performed at the Coherent X-ray Imaging instrument in the Linac Coherent Light Source (LCLS). Gas

focused liquid jets were used and crystals sat on an anti-settler temperature-controlled device. The jet diameter was 4 μm at the X-ray intersection region, with a focal area of 2 μm^2 and X-ray photon energy of 6.0 keV. The laser pump was 527 nm, and lengths for succeeding pulses were 90 ns and 150 ns, 400 μm in diameter, 100 μm above the X-ray interaction region, with the jet running 12 m per second, and with an energy per pulse of 3 μJ . Time delays were 570 μs and 780 μs .

In a separate experiment done in 2020, the authors experimented with flow rates. Here they wanted to capture the boundary jet speed at which sample is consumption is minimized, but the sample refresh rate is sufficient that the experiment does not suffer from pump laser contamination due to incomplete sample replenishment. They noted that the contaminated jet region goes significantly beyond the pump laser interaction region – possibly due to stream scattering. For this experiment, the authors used a high viscosity jet (flow rates of 2.5 – 9.8 $\mu\text{l}/\text{min}$) at the SACLA “diverse application platform for hard X-ray diffraction” (DAPHNIS). The pump laser was 532 nm and contacted the jet from 2 directions, 160° apart each with 42 mJ/cm^2 (focal diameter of 240 μm). The pump-probe delay was 10 ms. The microcrystals had a maximum size of 100 μm . They were pre-flashed with a laser at 532 nm and 52 mJ/cm^2 to oxidize the tyrosine D and decrease contamination with the S_0 state. The crystals were then dehydrated and mixed with silicone grease. Every step from purification to grease introduction was carried out in dark or dim-green light [51], [52].

1.2.2.4.2. Photosystem I

The first time-resolved SFX experiment was done in 2012 at LCLS SLAC [53]. The authors used the pre-existing SFX delivery system consisting of a GDVN (Gas dynamic virtual nozzle) and added a pump laser consisting of approximately 8 mJ, delivered in 10 ns length pulses and focused to 380 μm (jet diameter was 4 μm). The 532 nm pump pulses corresponded to 10 times the calculated required fluence to excite every protein within a crystal of an average 2 μm size. The jet flow rates were maintained to guarantee speeds of over 13 m/s and sample replenishment before the next pump-probe sequence. The experiment measured timepoints of 5 and 10 μs delays. This was a groundbreaking experiment illustrating how to carry out these types of experiments and produced the first ever time-resolved SFX results.

1.2.2.4.3. Purple bacteria photosynthetic reaction center

The photosynthetic reaction center of purple bacteria *B. viridis* is responsible for transporting 2 protons across the membrane per photon absorbed, building the proton gradient that is later used for energy production by ATP synthase. A time-resolved crystallographic experiment reported in 2017 [54] observed multiple electron changes over the 1 ps, 4 ps, 20 ps, 300 ps, and 8 μs timepoints. At LCLS,

microcrystals were delivered onto the X-ray interaction point via gas dynamic virtual nozzles, producing 10 μm diameter jets. The authors triggered the reaction with 960 nm 40 fs length laser pulses, focused to a spot size of 190 μm , with an energy per pulse of 25 mJ/cm².

1.2.2.4.4. Cytochrome *c* oxidase

Cytochrome *c* oxidase is an important part of the respiratory chain. It reduces oxygen molecules (O₂) consuming four electrons and two hydrogen molecules, thereby producing water and releasing energy that is used to transport a total of four electrons from the mitochondrial matrix into the mitochondrial inner membrane, contributing to the proton motive force. In 2017 a time-resolved crystallography experiment carried out at SACLA was reported with insight into the closure and opening mechanism of the proton channel [55]. Infrared and visible light absorption time-resolved spectroscopies were used before the experiment in preparation for the time-resolved crystallographic experiment. In the experiment itself, the authors employed serial femtosecond rotational crystallography using a total of 76 larger crystals (>100 μm) instead of the typical SFX approach where small crystals, usually under 10-20 μm , are used. This choice was justified by the requirement of higher resolution datasets (1.9 Å) to enable proton movement observation via side-chain changes; not provided by smaller crystals. Hence, the experiment used larger crystals mounted in loops, rotated in 0.1° steps. They were then moved 50 μm away from the previous shot area of the crystal, so that each X-ray shot hit a fresh and reoriented area of the crystals. Hereby the authors confirmed the suitability of the translation steps between X-ray pulses by both visual damage of shot crystals, and analysis of the total reflected peaks variance. The pump laser used in the experiment was 532 nm with a pulse length of 4 ns and 0.5 ns jitter. The laser was split and focused onto the crystals from the back and front sides of the crystals for trigger homogeneity; datasets were collected for 100 μs and 20 ns time delays. Data collection was done >1 Hz to avoid photoproducts accumulation, and of note is that such an experiment is possible only due to the reaction reversibility.

1.2.2.4.5. Rhodopsin and bacteriorhodopsin

Opsins are light responsive proteins, of which bacteriorhodopsin and halorhodopsin are well studied light-driven outward and inward ion pumps, respectively.

Data from early time-resolved absorption spectroscopy experiments [56], [57], together with structural data showed the mechanism of *Nonlabens marinus* rhodopsin-3 ranged from 10 μs to 200 ms, with potentially five different intermediary conformations. Time-resolved SFX was later performed at SACLA, using high-viscosity jets to deliver the LCP crystals to the X-ray interaction point, and triggering the reaction with a pump laser at 540 nm [58], [59]. Static datasets were previously obtained with up to 1.5 Å

resolution, and the successful time-resolved structure at 1 ms was 2.1 Å. Structural changes were observed and together with previous research advanced the current understanding of the reaction mechanism.

The synchrotron experiment at Spring-8 in 2015 was done at the BL38B1 beamline. In this experiment the authors employed soaking solutions in combination with red-light exposure and flash cooling to lock the protein in different reaction states. Finally, crystals were exposed to 2×10^{12} photons/mm²/s and datasets were reported showing the ground state and an intermediate state structure (PDB ID: 4QRY; [58]).

More recently a TR-SFX experiment was performed at BL3 in the XFEL facility SACLA. This experiment used 540 nm light pulses as triggering events, the pump synchronization with the XFEL pulses had a delay time jitter of ± 0.5 ns, and the laser was split into two paths, to interact with the sample from opposite directions. The pump energy was 6 μ J (3 μ J from each split beam) and because the pump repetition rate was one third that of the probe, data was collected with a sequence of laser-on, 2x laser-off pattern. The X-ray pulses were 7.0 KeV and 10 fs long, with a repetition rate of 30 Hz. The sample was delivered in LCP and injected at 2.5 μ L/min and over 30 k images were collected for the 10- μ s and 1 ms timepoints, (PDB IDs: 7VGT, 7VGU, 7VGV – including the dark state) [59].

1.3. Detergents and LCP

The study of membrane proteins is a complex version of the process of studying typical water-soluble proteins. This is not only due to the usually lower expression yields of membrane proteins, but also due to the requirement of detergent usage for protein extraction and solubilization, aiming to minimize protein aggregation as much as possible.

Detergents are amphipathic molecules which can form protein-detergent complexes, thereby solubilizing them. They have various important characteristics that are discussed in the following sections.

1.3.1. Detergent general characteristics and characterizations

Detergents are water-soluble surfactants and amphiphiles, which are typically used for membrane protein purification and membrane protein studies. Examples are natural detergents such as bile salts and saponins, and a vast collection of synthesized detergents. They can be ionic (*e.g.*, SDS), zwitterionic (*e.g.* LDAO lauryldimethylamine oxide), or non-ionic (*e.g.* Triton X-100). Detergents are usually a combination of headgroups which are polar, and non-polar hydrocarbon tails, or even bile salts which have a tetracyclic

structure similar to cholesterol. Some of the commonly used nonionic detergents have sugars for headgroups such as β -Octylglucoside (β -OG) and dodecylmaltoside (DDM).

Variety in detergent availability is an asset as these can be chosen and used for different projects, purification stages, and experiments depending on each detergents' characteristic. However, some detergents can have limiting prices as opposed to cheaper popular options such as DDM. Examples of where detergent variety is a benefit is the use of Triton X-100 as a nonionic mild detergent which can stabilize proteins while still retaining minimal function. However, this detergent gives UV absorbance due to aromatic ring, possibly invalidating related experiments. SDS strongly binds to proteins and can be efficient at extracting them but often is too harsh and denatures membrane proteins.

1.3.2. Critical micellar concentration

Most detergents form micelles at a point known as the critical micellar concentration (CMC). Micelles are rough spheres or elliptical assemblies of detergent molecules or monomers. The hydrophilic portions of detergents are oriented outwards, interacting with water and other polar or ionic solution components *e.g.*, salts. The hydrophobic portion of the detergents are oriented inwards, minimizing solvent contacts and maximizing apolar-apolar interactions. As opposed to initial expectations, the hydrophobic tails at the interior of micelles are highly disorganized and not fully extended as demonstrated by NMR studies[60]. The micelle shape and size vary with the monomer characteristics, and the micellar radius is usually 10-30 % smaller than a fully extended monomer.

Because the micellar formation is cooperative, a critical micellar concentration exists at which point micelles start forming while the dissolved monomer concentration usually remains constant. Below the CMC only monomers exist in solution. There are also minimal temperatures below which detergents to form micelles called critical micellar temperature (CMT).

1.3.3. Membrane and protein solubilization by detergents

As detergent is added to a suspension of mechanically disrupted membranes, detergent monomers start being introduced into the bilayers. As more detergent is added and the detergent to lipid ratio increases, the bilayer itself is eventually disrupted and protein lipid detergent complexes are formed. If the ratio of detergent to lipid is high enough, minimal quantities of lipids remain interacting with the proteins and the proteins become mainly complexed by detergent molecules. At this point the membrane protein extraction from the membranes is considered successful. Naturally, the detergent to protein ratio,

and detergent to lipid ratio are major factors to be considered, in combination with the CMC of the specific detergent used, for a successful membrane protein extraction.

After membrane protein extraction, subsequent steps will either reduce or swap the detergent solubilizing the membrane proteins, typically achieved via chromatography techniques. Additionally, some experiments can require detergent replacement by lipids and liposomes. In these cases, methods for detergent removal include techniques such as gel filtration, absorption to polystyrene beads, pH detergent precipitation, and dialysis. The methods used will vary depending on the specific detergent and membrane protein characteristics. For instance, where there are large differences between detergent-protein complexes versus empty detergent micelles (due to large protein sizes or small non-polar detergent domains), the micelles can be efficiently separated by gel filtration. In the case of SDS, it can easily be precipitated out of solution by sodium exchange to potassium as potassium dodecylsulfate is insoluble at room temperature.

1.3.4. Considerations in structural biology

Crystallization of membrane proteins is significantly more complex and difficult than for their water-soluble counterparts, due to the requirement that they are solubilized in detergents (or stabilized using the LCP method). These detergents are often at least partially disordered and do not provide specific inter-particle interactions, directly counteracting the requirement for an orderly molecular arrangement across the crystal lattice. This factor makes crystallization more difficult and potentially lowers the final structure resolution due to lower quality of the obtained crystals.

Crystallized membrane proteins can be found exhibiting more or less favorable packing. The most desirable protein crystal packing is type 1 (Figure 1-2), where the proteins interact in a head-to-head, or head to tail manner with a large surface area of protein-protein interaction. This results in a low number of disorganized molecules (lipids, detergents) between each protein unit, making a more stable and stronger interaction between units than is possible in type 2 packing, where the protein-protein interaction surface is usually lower, and the micelles are relatively spatially unconstrained. Due to the tighter packing in type 1 crystals, a higher diffraction power is expected.

It is also important to understand that different detergents and/or lipids can enable or disable protein crystallization, by impacting the interaction possibilities and stability of the crystal packing. Hence crystallization screens varying detergent and/or lipid can also be carried out to find optimal conditions, unfortunately thereby adding an extra step to the already time-consuming and difficult crystallization

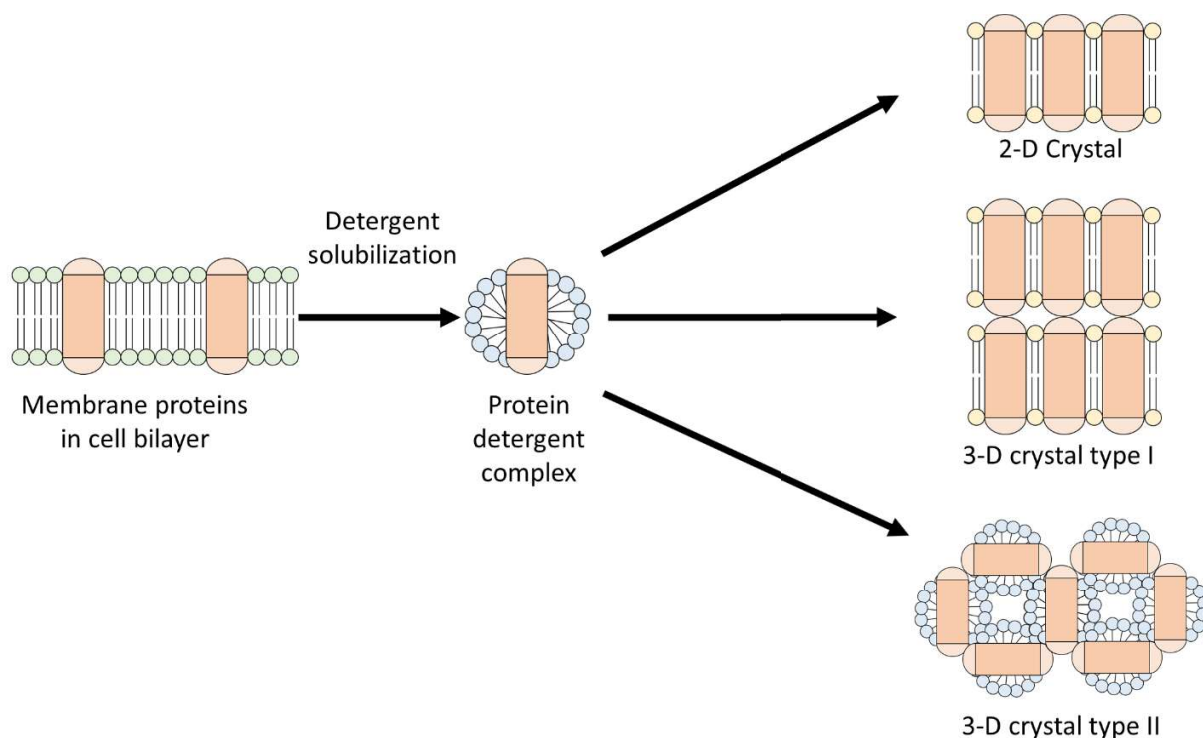


Figure 1-2: Types of membrane protein crystals

Illustration of various crystal types for membrane proteins, where they interact only horizontally (2D), horizontally and vertically (3D I), and only between the protein heads (3D II) [90].

optimization process. Commercial screens that focus on this exact problem have been produced; commercial or in-house screens are typically used when required.

1.3.5. Detergent diffusion rates

Diffusion techniques can be used to achieve remarkable effects in membrane protein crystallization, for example, using dialysis membranes with crystallization dialysis buttons. One improved dialysis crystallization method is the controlled manipulation of the crystallization buffer components and temperature coupled to automated real-time microscopy monitorization [61]. In this case a fluidics system can constantly manipulate the outside-button buffer components, typically changing salt concentrations to achieve the results desired for the crystallization process (*e.g.*, large crystals optimization). The system can, however, also manipulate the pH and concentration of other components such as detergent.

Previous studies have been done using this setup, measuring the speed at which the typical crystallization component PEG and salt diffuse into the crystallization button. This has not been previously done with detergents, however, and for this work I tackled this question in collaboration with Prof. Monika Spano and Dr. Elham Vahdat from the university of Grenoble, at ESRF in IBS.

1.3.6. Lipidic cubic phase

A successful way of handling membrane proteins and crystallizing them is by using LCP (lipidic cubic phase). This phase is obtained by mixing specific lipid to water ratios (usually 3 to 2) while considering temperature and lipid specific phase diagram. This and other mesophases are more ordered than liquids but less than solids. The LCP can be described as an organization of lipids in bilayers, forming two interweaved but non-contacting water channels [62], [63].

Due to the nature of the lipid bilayer, this phase is a better approximation of physiological conditions (as opposed to protein-detergent complexes) and is more ordered than solubilized membrane protein solutions. Some membrane proteins can only be stabilized at higher concentrations by LCP (often required for crystallization), and the crystals are expected to be type 1 – the most ordered and desirable crystal packing.

The LCP shortcomings are often related to its high viscosity characteristic, and sensitivity to humidity and temperature changes. It can be difficult to handle LCP samples without proper tools or training and delivering it to the beamlines also requires additional considerations. Even though this technique is efficient, it can be considered daunting to begin working with, as well as introducing an additional set of variables to the sample handling – different lipids (and even mixtures) give rise to different LCP characteristics. For example, larger hydrophobic tails and/or insaturation enlarges the bilayer cross-section and can shorten the water channel diameter.

2. Methods

2.1. LeuT

2.1.1. Production and purification

Plasmids were kindly provided by Thomas Stockner and Jawad Khan from the Medical University of Vienna. Plasmids were introduced into electro-competent C41 *E. coli* cells via electroporation which were then incubated for one hour before inoculation in 100 µg/ml ampicillin agar plates overnight.

Single colonies were used to inoculate pre-cultures (5-10% of the growth media volume) or the actual growth media consisting of 100 µg/ml ampicillin LB. Growth media flasks had a 1:5 ratio of media to air volume and were incubated at 37 °C until OD₆₀₀ was 0.6-0.8 (2-4 hours), at which point the temperature was lowered to 18-20 °C and 0.1 mM IPTG induction overnight started.

Twenty hours after induction, cell harvesting was done by centrifugation at 9000 x g for 10 minutes at 4 °C. Cells were either frozen at -80 °C for later use or resuspended in lysis buffer at 4 °C, 3:1 ratio of buffer (ml) to cell paste (g), homogenized with a glass homogenizer, and mixed for 1 hour. Cell lysis was done using a cell disrupter machine at 15 – 20 k PSI, for 2-4 cycles depending on achieved pressure.

Lysed sample was centrifuged at 4300 x g for 15 min and 4 °C, the cell debris pellet discarded and the supernatant ultracentrifuged at 120 k x g for 2 hours at 4 °C. The resulting membrane pellet was either frozen at -80 °C for later use or homogenized using a glass homogenizer and suspended in solubilization buffer with gentle mixing for 1.5 hours at 4 °C.

After the solubilization step, the sample was again ultracentrifuged at 120 k x g for 30 min and 4 °C. The supernatant consisting of solubilized material was then incubated with HisPur™ cobalt resin pre-equilibrated with equilibration buffer for 1.5 hours or overnight. Wash steps using wash buffer consisted of 10 CV (column volumes) 5 mM, 10 CV 10 mM, and 20 CV 20 mM imidazole concentrations. Protein

Table 2-1: Buffers used in LeuT protein production.

A list of various buffers used in purification.

Buffer name	Components
Lysis	20 mM HEPES 7.5, 200 mM NaCl, 5 mM MgCl ₂ , 0.2 mg/ml Lysozyme, 0.01 mg/ml DNase, 0.5 mM PMSF, 10% v/v glycerol
Solubilization	20 mM HEPES 7.5, 200 mM NaCl, 1 % DDM, 5 mM MgCl ₂ , 0.5 mM PMSF, 10% v/v glycerol
Wash/Elution	20 mM HEPES 7.5, 20 mM salt (NaCl or KCl), 0.05 % DDM, various imidazole concentrations
Dialysis	20 mM Tris 8.0 or HEPES 7.5, 20 mM salt (NaCl or KCl), detergent near CMC (<i>i.e.</i> ,
SEC	0.05 % w/v DDM or 40 mM β-OG)

elution was done with 5 CV 350 mM imidazole buffer. The protein sample was then concentrated with 30 kDa molecular weight cutoff concentrators, pre-washed with water to remove sodium azide.

His-tag cleavage was carried out using a Thrombin CleanCleave™ Kit resin and simultaneously dialyzed to the desired detergent buffer overnight at 4 °C. The dialyzed sample was separated from the thrombin kit with two gentle (~500 x g) 5 min centrifugation cycles.

The protein sample was then concentrated to 0.5 ml, centrifugated at max speed in a bench centrifuge for 5 min and injected in a SEC s200 (or s200 Increase) column pre-equilibrated with the desired buffer. The collected sample was again concentrated.

2.1.2. Sulfhydryl quantification

Sulfhydryl quantification was done using Ellman's reagent (DTNB) as described in [64]. Working solution was obtained by dissolving the 10 mM DTNB DMSO stock solution 100-fold with 100 mM Tris-HCl pH 7.5. Sample volumes of 50 μ l were added to tubes containing DTNB working solution and briefly mixed by vortexing, followed by 2 min of incubation at room temperature, and absorbance measurements with a spectrophotometer at 412 nm. A calibration curve with BSA was made to confirm values given in [64] and the free sulfhydryl concentration was calculated.

2.1.3. Cysteine crosslinking

Cysteines were cross-linked using a Cu(II)-(1,10 phenanthroline)₃ oxidant catalyst (CuPh) as in [65] with a few modifications. Previous unpublished results showed protein precipitation when the CuPh reagent was added to isolate membranes containing expressed 3M LeuT. Instead, the reaction took place after the desalting and immobilized metal affinity chromatography (IMAC) columns. The CuPh stock solution was 0.5 M 1,10-phenanthroline (in 1:1 water to ethanol) and 0.15 M CuSO₄. The reaction was done using a working solution of diluted stock CuPh in phosphate buffered saline (PBS) down to 0.05-1.00 μ M, at room temperature for 15 min. The reaction was stopped with 100mM EDTA and a size exclusion chromatography (SEC) was carried out to separate the protein from the reagents.

2.1.4. Crystallization and optimization

Purified protein concentrated to 10 mg/ml was used to screen for crystallization conditions around both those previously discovered [66]–[68] and with a Hampton Research crystallization screen [69]. Experiments were done at 20 °C, all necessary solutions and equipment were left in the same room to mitigate any uncontrolled temperature change. To each well of a 24 well crystallization plate the corresponding crystallization solutions were added. This was immediately followed by 0.5 μ l of crystallization solution on a siliconized cover-slip (hanging drop) and 0.5 μ l of protein solution. Once a row of cover-slips were readied they were set on each well, and the wells sealed with vacuum grease. This exact sequence of events was always repeated to avoid variation due to uncontrolled evaporation of crystallization drops or well solutions. Plates were then placed in 20 °C incubators.

2.1.5. Circular dichroism

Protein samples were diluted to 0.4 mg/ml in 20 mM HEPES pH 7.5, 0.05% DDM, 10 mM NaCl. Considering that sodium chloride was needed for binding activation but simultaneously caused a sharp increase in background noise, its concentration was tested to find the best compromise between binding activation and background. CD spectra were collected between 190-260 nm, 1 nm steps, using 0.5 s/step. Leucine titration consisted of 0.5 mM, 1.5 mM, and 3.0 mM using a 100 mM leucine stock solution in CD buffer with 5 minutes incubation time (although measurements would take up to 25 minutes). After the final leucine titration measurement, the sample was incubated with 0.5 mM TCEP for 5 minutes at 4 °C using a 0.5 M TCEP stock. Each sample was prepared in triplicate, and each triplicate CD measurement was done 15 times.

Four LeuT published crystal structures (ID: 3tt1, 3tu0, 6xwm, and 3tt3) had their secondary structure content read with YASARA (<http://www.yasara.org/>) software after removing any non-LeuT protein molecules in each file (*e.g.* antibodies used in crystallization). CD spectra prediction was done for each of these pdb files (<https://pdbmd2cd.cryst.bbk.ac.uk/>), and then corresponding secondary structure analysis was done with the Bestsel tool (<https://bestsel.elte.hu/>).

2.1.6. Fluorescence ligand titration

Fluorescence was measured in dark bottomed 96-well plates containing 0.25 mg/ml LeuT in 20 mM HEPES 7.5 pH, 10 mM NaCl, and 0.05% DDM. Leucine was added and incubated for 5-10 minutes at 1.5 mM, and 1.5 mM plus 0.5 mM TCEP (added from a 0.5 M stock solution) at room temperature ~25 °C. A total of 14 wells per data point were used, from 3 different plate and protein sample preparations. Buffer titrations were also done and subtracted to the corresponding protein titration samples.

2.1.7. Crystallography data collection and processing

Crystallographic data were collected on the P14-1 beamline at PETRA III, DESY at the end of 2019. Several different software programs were used in the early stages of the crystallography structural biology pipeline, from spot finding up to scaling and molecular replacement solutions. For a number of reasons (discussed later), the data proved difficult to work with so that several programs were used for the same stages of data processing. By using a range of software solutions, we could attempt to trouble shoot and find the appropriate tool for processing sections of the data.

Software used included DIALS [70], XDS [71], iMOSFLM [72], and Xia2 [73], of which the first two mentioned were used the most. In each case the general protocol used was the one provided by the

developers, with various tweaks, *e.g.*, in the case of spot finding where the initial default protocol did not work. Many attempts were made, too many to mention here, so for this work I will present the general tweaks made in an effort to improve the data processing. Specific examples and/or details will be shown only in the case of DIALS. Some of these tweaks could have also been used with the other mentioned software. In most cases the command line version of the software was used as opposed to the GUI versions.

Diffraction images were imported with each software, general manual overview of a few frames was done and noted, for later comparison with each change in spot-finding settings. Standard spot-finding software settings were used and then frames were inspected to assess how successful the process was. A few water (and possibly detergent) rings were observed that sometimes conflicted with the default spot-finding settings. For these cases, settings related to ice rings were tested (*i.e.*, `ice_rings=True` on DIALS). Streaked spots were also frequently observed but did not dominate the various datasets. This probably had a negative impact on the success rate of the spot-finding programs. We tried a few settings. These included shoe box, (area accepted by the software around the maximum intensity peak), and max/min size settings. As expected in the case of the latter if the maximum shoe box was increased, a few more streaked spots would be accepted, but a few portions of the ice ring would then become false positives. Therefore, both settings were used together. It was also frequently observed that various lower intensity spots were being missed out and not picked up by the spot-finding programs, in which case we tweaked the minimal signal to noise accepted ratio (`sigma_background`) and signal to local mean for the pixel to be considered as peak/strong pixel (`sigma_strong`). The size of the local area around a spot considered to calculate the mean and variance was also tried at higher numbers, because if this number is too small, the peak values will end up being deflated, and more false negatives will occur.

In DIALS the option for a reciprocal 3D lattice viewer is an option, and it was used in parallel with the 2D image viewer tools. In the case of the 3D lattice viewer, it was useful to notice areas of the detector to be manually masked out, and also to check the quality of the data (or at least, the peaks found in it, as false negatives will not be present here). It was also useful to check twinning possibilities. When aligning the reciprocal spots along a lattice when there are multiple lattices similar to each other, we can observe the second lattice slightly unaligned to the first lattice.

In some of the datasets where multiple lattices existed I attempted to process data accordingly. The same standard method was used in DIALS, until the indexing step, at which point the option for multiple lattices would be added to the indexing software. From then on, I selected one of the two lattices, refined

the Bravais settings, and re-indexed it with the refined settings. Then the re-indexed data could then be used as a normal single lattice data for downstream data processing.

Then followed the indexing refinement of the unit cell settings (or Bravais settings), in which the best options were usually between P1 and C2. Subsequent steps were refinement followed by integration.

Integration was often the most difficult step to solve. Various programs crashed or had errors triggered at this stage. With persistence and changes in spot-finding settings, some of the datasets could be carried over this step. Symmetry analysis, scaling, and merging were the last steps of X-ray image integration.

2.2. Mhp1

2.2.1. Large scale production and purification

Mhp1 cell growth and membrane preparations were done at the University of Leeds with kind help from David Sharples and Adrian Goldman. There we used a 30-liter fermenter for Mhp1 growth and 100 liter fermenter for HRV-3C protease.

The growth media was prepared in isolated concentrated solutions and autoclaved at 120 °C before adding to the fermenter; from previous experience it was expected that if autoclaved together some of the ingredients would react affecting, the growth profile. The concentrated solutions were prepared as follows: 14.79 g of magnesium sulfate heptahydrate in 50 ml water; 1.7664 g of calcium chloride dihydrate in 50 ml water; 60 g of casamino acids in 250 ml water; 30 g of ammonium chloride in 250 ml water; 55.2 g of glycerol in 250 ml. Before the fermenter inoculation 29 L of distilled water were sterilized inside the fermenter. Once this cooled down to 37 °C, all the previously prepared media plus 300 ml of the antibiotic carbenicillin at a concentration of 30 mg/ml was added to the fermenter.

Carbenicillin working concentration in all cases was 100 µg/ml. One Mhp1 transformed cell colony (kindly provided by Adrian Goldman and Peter Henderson) was used to inoculate 5 ml media at 37 °C for 16 hours. The 5 ml media was used to inoculate 300 ml of media at 37 °C for 6 hours, which was finally added into the fermenter, completing the 30 L volume threshold resulting in a 1% final volume inoculation.

The fermenter initial OD_{600nm} was 0.06, temperature was maintained at 37 °C, agitation at 100 rpm, airflow 0.5 VVM (volume of liquid per minute) and pH was only monitored, not maintained. After 3 hours and 26 minutes the OD_{600nm} was 0.49, at which point the temperature was decreased to 27 °C and

induction was done by adding a filtered IPTG solution obtaining a final concentration of 0.2 mM. Addition of 5 ml of autoclaved antifoam was required before 5 hours had passed and the cell harvesting procedure started.

The harvesting was carried out after the temperature was lowered to 15 °C and the final measured OD_{600nm} was 1.54. A total of 95.4 g of cell paste was collected and later frozen at - 80 °C after centrifuging the total volume of the fermenter over 20 min using a continuous flow centrifuge. Cells were then resuspended with lysis buffer at 4 °C, 7:1 ratio of buffer (ml) to cell paste (g), homogenized with a glass homogenizer, and mixed for 1 hour. Cell lysis was made using a cell disrupter machine at 30 k PSI, ensuring the cells passed through it twice.

The lysed sample was centrifuged at 14 k x g for 45 min and 4 °C, the cell debris pellet discarded, and the supernatant proceeded to ultracentrifugation at 100 k x g for 2 hours at 4 °C. The membrane pellet was then frozen at -80 °C for transport to Hamburg.

Membrane preparations were thawed, and total protein concentration calculated via Bradford or Schaffner and Weissman (SW) assays [74]. Membrane pellets were resuspended and diluted in solubilizing buffer in a 1:3 mg of total protein to ml of final volume. The sample was left to incubate at 4 °C for 4 hours with mixing. The solubilizing sample was then ultracentrifuged at 100 k x g and 4 °C for 1 hour before discarding the pellet. The supernatant was introduced into a pre-packed nickel column and left recirculating over the column overnight. The column was pre-equilibrated with a 5ml equilibration buffer. Wash steps with the wash buffer were 10 CV DDM 5 mM imidazole, 20 CV DDM/NM 5 mM imidazole, with elution in 4 CV DDM/NM 300 mM imidazole, with the later 2 steps introducing NM instead of DDM for protein production batches which required detergent exchange (*i.e.* NM for crystallization).

The eluate was concentrated down to 1.5 ml and injected into a 5 ml desalting column. The protein containing elution fractions were re-concentrated to approximately 10 mg/ml and frozen at 80°C.

Table 2-2: Buffers used in large scale protein production.

Buffer name	Components
Lysis	20 mM Tris 8.0, 150 mM NaCl, 5% glycerol
Solubilization	20 mM Tris 8.0, 300 mM NaCl, 1 % DDM, 20 mM imidazole, 20% glycerol
Wash/Elution	20 mM Tris 8.0, 0.05 % DDM (or NM), various imidazole concentrations

2.2.2. SAXS

Size exclusion chromatography – small angle X-ray scattering (SEC-SAXS) experiments were performed at the BM29 beamline at the European Synchrotron Radiation Facility (ESRF), and P12 beamline at the PETRA III Synchrotron (DESY).

Protein samples were shipped at -20 °C on dry ice to synchrotron facilities. Concentrated 25 µl Mhp1 samples were diluted in the buffer of interest for each experiment, and in some cases re-concentrated on site at the ESRF facilities to 5 mg/ml or 10 mg/ml. Protein quantification was assessed via SW assays. All samples were centrifuged at bench centrifuge max speed for 10 minutes prior to size-exclusion chromatography (SEC) column injection.

Samples were injected into SEC columns connected to a High-Performance Liquid Chromatography (HPLC) system and the SAXS beam line flow-through exposure capillary. SEC columns used were Superdex200-increase (GE) which allowed for chromatographic separation of detergent-protein complexes and “empty” detergent micelles. Columns were pre-equilibrated with the respective run buffers with 0.050 – 0.075 ml/min flows for 3 CV at room temperature and all buffers were filtered and degassed. Buffer baseline subtraction was done by subtracting the signal immediately before sample elution from the rest of the chromatogram.

2.2.3. Structure refinement

We had access to Mhp1 crystallography data collected from 2013-2015 by Dr. Anna Polyakova. This dataset was already merged and had already been through refinements at the time of collecting with Refmac. We slightly improved the refined structures statistics with Refmac [75], and also compared it to refinement using Vagabond [76].

The Refmac refinement cycles were the default settings with the addition of jelly-body restraints and map sharpening (this weights heavier contributions from the highest resolution shells). Various cycles of coot ([77]) model improvement and refinement cycles were also carried out until a satisfactory compromise between model detail addition and restraint was achieved, so as to avoid over fitting with respect to data resolution. Vagabond refinement was done following instructions on the author’s webpage; no significant detours from the default settings were made.

Following refinement, structure validation tools were used to assess the model’s quality and overfitting metrics. For this purpose, the online tool MolProbity [78] was used in conjunction with “metric

model geometry validation” task at CCP4 [79] which gives insight into important model statistics in comparison to other X-ray structures and of similar resolution.

In some cases, we also attempted to redo the molecular replacement but the result was redundant with the already obtained models. This as an attempt to restart the data processing as far back in the crystallography structure solving pipeline as possible, however the original data would have to be obtained for this to be further pushed back.

2.3. DgKa

Preliminary tests were made regarding a possible time-resolved synchrotron serial crystallography experiment. These experiments consisted of loading test LCP DgKa crystal sample (kindly provided by Professor Martin Caffrey and Dr. Coilin Boland) with low crystal density, on mylar-chip-mylar sandwiches. These tests attempted to understand if we could manage to get the crystal slurry deposited in the chips, as opposed to the usual less viscous crystal samples.

Two mylar sheets were cut to the chip holder pieces sizes. These sheets were then attached to the holders using screws, and in a way to be slightly stretched along the chip holders. These were then put to the side and used in the later step of this procedure.

To load sample onto the chip, the crystal slurry from the crystallization syringes typically used with LCP samples was smeared as much as possible over a wide range of wells. In this step it is important to maximize the spread and minimize the LCP height that sits on the chip, as this will impact the following steps.

Vacuum was applied underneath the chip (between 0.3 and 0.5 bar), as an attempt to drive the slurry and crystals through and into the chip wells. This step needs to be carefully done, as it will easily dry the LCP and hence destabilize the crystals. The vacuum had to be applied just until enough LCP is settled in the wells; more vacuum will not necessarily place crystals in the wells, and will definitely dry the LCP, compromising the experiment.

After the previous steps were done, the chip with LCP was placed on top of the chip holder, with the correct combination of X-ray facing the sides of both chip and chip holder. The next piece of the chip holder was put on top of the chip, and the chip holder pieces were attached with screws.

A last important step that I discovered, as stated later in this work, is that the mylar sheet (side where the LCP is sitting) needs to be gently and homogeneously pressed against the chip. This final step will

settle the crystals in the wells and stop the slow movement of the LCP and crystals (showed later in results and discussion). With this step, a few crystals are smashed against the in-between wells chip areas, and significant LCP volumes will be sitting inside the wells, contributing to background noise. If however, this step is not carried out, less crystals are settled and significant crystal movement happens, making the experiment impossible.

It is also worth noting that all previous steps must be done as fast as possible, as soon as the LCP samples exit the syringes. Just like other non-viscous samples, the LCP will dry over time when exposed to air, however, the average chip preparation time will be longer than that of non-viscous samples due mainly to the vacuum step. This increased air exposure time combined with LCP fine water-lipid ratio requirements (and the crystals in the slurry) make these samples especially sensitive to air exposure. It is, therefore, crucial to have all the tools, and importantly the mylar sheets, already prepared and in place for the procedure. Although not a requirement, this process can also be done within a small humidity-controlled box, which will attenuate the sample to air exposure risks.

2.4. LCP-refraction

Monoolein (MAG) 9.9 was mixed with water in gas tight syringes in a 3:2 ratio. Two syringes with water and lipid were connected with a metal coupler and ferrule in each connection. The syringe contents were then pushed back and forth into each syringe, until a difference in plunger resistance was felt and the syringe content turned clear, signaling that LCP formation has been achieved. In the cases where the fluorophore was used, it was added to the water syringe, subtracting the volume of 6-aminoquinoline from the water volume to maintain the lipid to water fraction ratio.

The LCP was then added to a chipless chip holder and mylar sandwich. This was then mounted in front of a fast camera. A bandpass filter (no transmission below 400 nm) was added to the camera, so that the only light detected would be the fluorescence. The excitation laser was at 355 nm wavelength and the fluorophore had an absorption maximum of 355 nm and emission peak at 550 nm.

A series of pictures and movies were then collected by shining the laser onto the LCP, with and without fluorophore, and capturing the light emission from the fluorophore. The distance from the laser target area, and other parameters tested are described further in the results section.

2.5. Detergent studies

Diffusion rates were determined by closing a crystallization button filled with water and a dialysis membrane inside a falcon tube with 5 ml detergent solution. Multiple buttons and falcons were prepared

this way and opened at different times. The dialysis membrane was either ruptured with a pipette tip, a needle, or by dislodging it completely from the button. Then two to three 50 μ l samples were added to individual 2 ml plastic tubes where the detergent quantification assay would take place.

Dialysis membrane absorbance of detergent was assessed by placing dialysis membranes (in comparable sizes to typical crystallization in dialysis buttons) in falcon tubes containing 5 ml 0.05 % DDM and left at room temperature for three days. Tubes lacking membranes were placed in the same conditions. The detergent quantification assay was then made to assess and compare how much detergent was present. A few detergents, detergent concentrations, membrane cutoff size, and membrane types were tested using both above methods.

The colorimetric assay was carried out as described by [80], under a fume hood and with necessary protective equipment. In brief, 200 μ l of 5 % phenol was added to the sample followed by 650 μ l of concentrated (>97%) sulfuric acid, the tube was closed and vortexed at maximum speed for a few seconds. The sample drastically heats which might cause the tube to burst open, hence the vortex machine was placed inside the fume hood for this step.

Samples were left for two minutes to cool down, at which point they were transferred to plastic UV-vis cells. Due to the significant acrid vapors released by the samples, the UV-vis cells used had lids to limit exposure to humans. Finally, each sample absorbance at 490_{nm} was read at a spectrometer.

2.6. HRV-3C

HRV-3C is a protease used in the purification protocols of concentrative nucleoside transporter (CNT) and other protein constructs that colleagues and/or collaborators have used. I purified it on several occasions and although its impact on the results of this work were not noticeable, its purification protocol is stated here for completeness.

HRV-3C cell growth were done in Leeds with kind help from David Sharples and Adrian Goldman using the 100 liter fermenter.

The growth media was LB and was prepared in isolated concentrated solutions and autoclaved at 120 °C before adding to the fermenter. The used antibiotic was Kanamycin at a working concentration of 50 μ l/ml. A pre-culture of 1 L was grown overnight to OD_{600nm} 1.5 and provided by the lab staff.

The initial temperature of the fermenter was maintained at 37 °C, with agitation at 100 rpm, and airflow 0.5 VVM (volume of liquid per minute). pH was only monitored, not maintained. When the OD_{600nm} was 1.2, filtered IPTG solution was added, obtaining a final concentration of 0.5 mM.

After 24 hours we proceeded with harvesting; the final measured OD_{600nm} was 1.6. A total of 225 g of cell paste was collected and later frozen at - 80 °C after centrifuging the total volume of the fermenter over 20 min using a continuous flow centrifuge.

Stock 10x PBS buffer was prepared by dissolving 80 g NaCl, 2 g KCL, 14.4 g Na₂HPO₄, and 2.4 g KH₂PO₄ per 1 L of water. The final pH was 8. The 1x stock buffer was a 10 times dilution to which 1 mM DTT was added. The equilibration, wash, and elution buffers were the working PBS buffer, with 7.5 mM, 30 mM, and 250 mM imidazole.

Working PBS buffer with 0.5 EDTA was used to resuspend the transformed cells and they were disrupted in a cell disruptor at 30 kPSI over 2 cycles. Following centrifugation at 27000 x g over 15 min at 4 °C, the supernatant was collected, and Triton X100 and imidazole added to 0,1 % and 7.5 mM final concentrations, respectively. The protein was then loaded into a nickel column, the resin (pre-equilibrated with equilibration buffer) was washed with 20 CV equilibration buffer, then 5 CV wash buffer, and finally eluted with 5 CV elution buffer. After the IMAC column, the sample was gel filtered with a Superdex 75 column, the sample was eluted after 1,1 CV and later concentrated to 5 mg/ml.

2.7. CNT

Cell transformation was done with 50 µl BL21 Star™ competent cells, and the antibiotic Kanamycin was used at 50 µg/ml. Heat shock at 42 °C over 30 seconds was done after plasmid addition and 30 seconds incubation. After two minutes LB medium was added to a final volume of 1 ml and incubated with shaking for 1 hour at 37 °C. This medium was then added onto agar plates with Kanamycin and incubated overnight at 37 °C.

Protein expression was done by inoculating 300-600 ml of LB with Kanamycin overnight. This pre-culture was then distributed over 3-6 L TB in 6-12 flasks. These were incubated at 37 °C until the OD_{600nm} was 0.6, at which point induction was done by IPTG at 0.5 mM concentration. After six hours the cells were harvested and either frozen or we proceeded with cell disruption.

The cell disruption was done by resuspending the cells in PBS buffer pH 7.4, and then 2 cycles through a cell disruptor at 30 kPSI. Then centrifugation at 14000x g was done for over 45 minutes and the resulting

supernatant was ultracentrifuged at 100000x g for two hours. Two washes were done which consisted of resuspending the pellet in PBS and again ultracentrifuging at the same g force but for one hour. Finally, the pellet was resuspended and homogenized with a glass homogenizer.

The solubilization was done with 1 % DDM, PBS buffer, 5 % glycerol, and 150 mM NaCl over two hours at 4 °C. Ultracentrifugation followed again at 100000x g force for one hour and the pellet discarded.

The purification was done with 4 ml of HisPur™ cobalt resin equilibrated with PBS buffer and 5 mM imidazole and 0.05 % DDM, the protein sample was added to the resin and left to incubate with gentle shaking for one hour. A wash was done with 10 CV of 15 mM imidazole. The column was then unpacked and HRV-3C was added in a 5:1 ratio to CNT and incubated for two hours. The flow-through was then collected and incubated with a Ni-NTA nickel resin for 15 minutes. The flow-through was then collected and concentrated to 10 mg/ml with a 30 kDa cut-off centrifuge concentrator.

3. Results

3.1. LeuT

This project was done in collaboration with Prof. Dr. Thomas Stockner and Dr. Jawad Khan from the Institute of Pharmacology, Medical University of Vienna, with the aim of carrying a time-resolved structural study of LeuT. The experiment was intended to be done via serial crystallography, following the conformation changes between the various LeuT states previously captured in static structures.

We planned to trigger the reaction by breaking a cysteine crosslink which was introduced into a mutant variation of LeuT in order to lock it into an inward occluded conformation. The cleavage can be done by either laser, X-rays, or reducing agents. For this experiment, high quantities of protein need to be purified, crystallized, optimized to microcrystals, and the reaction activation validated. As a starting point we tested the cross-linking efficiency, and ligand binding activity of crosslinked LeuT. This was to assess whether the mutant variant was indeed cross-linked, and whether cleavage could trigger activity in a way suitable for a time-resolved crystallography experiment.

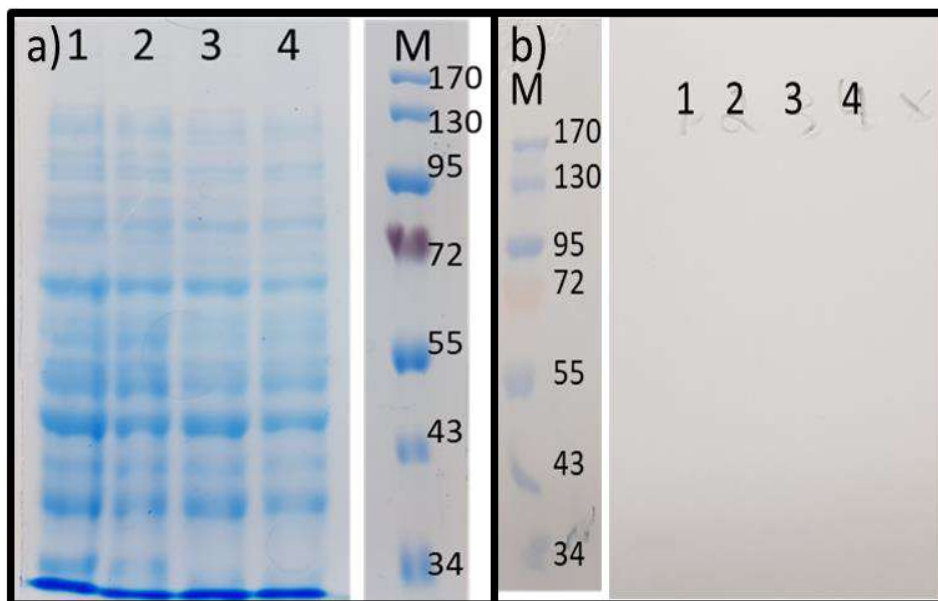


Figure 3-1: Lack of LeuT induction.

Comparison of WT and 3M LeuT induction results, a) SDS-PAGE and b) western-blot of a twin gel as a). Samples 1 and 2 were WT and 3M without IPTG induction, and 3 and 4 were WT and 3M with IPTG induction respectively. Notice no significant difference at the expected LeuT region just under the 43 kDa band, both in the SDS-PAGE and western-bot.

Over the following sections the results for protein production, cross-linking, binding tests with fluorescence and CD, crystallization, and data processing are detailed. The steps that would follow in the future of this project are crystal optimization towards microcrystals, and reaction triggering assays, finally enabling controlled time-resolved X-ray crystallography experiments.

3.1.1. Purification problems/successes

During a secondment to Vienna, I worked with the Stockner group to produce a first batch of purified LeuT, resulting in just under 3 mg of pure protein. I also prepared several batches of membrane preparations which were frozen and taken to Hamburg where I purified the LeuT and carried out initial cross-linking characterization, and crystallization tests. In total, five batches of 6 L growth media were grown, of which four were frozen at -80°C as membrane preparations (after cell lysis and centrifugation), and one of the batches was purified in Vienna. On average each batch would produce ~ 2 mg of purified LeuT. Once this protein was consumed, I began to express LeuT in Hamburg using an expression plasmid kindly supplied by the Stockner group. This proved problematic and I began to troubleshoot to identify the problem. At first glance the issue seemed to be with the competent cells. However, after making new competent cells, we were still not getting significant LeuT induction (see Figure 3-1). We therefore requested a new plasmid preparation from the Stockner group which resulted in successful

transformation and expression. We could now again induce LeuT production, but at a lower yield than before, and so we proceeded to troubleshoot this new issue.

It was at this time that the SARS-COV-2 pandemic started and my project was suspended as my research group pivoted towards supporting a repurposing library ligand screen against SARS-CoV-2 main protease. In this project I contributed to the intensive protein production campaign. After months of lockdown I was allowed to restart the LeuT project, however the new expression tests failed to yield LeuT. I restarted trouble shooting, although this was complicated with the still limited lab access and competition with the ongoing SARS-CoV-2 project and the move of our labs from CFEL to HARBOR. Nevertheless, with support from Prof. Dr. Henning Tidow and Dr. Inokentij Josts we regained satisfactory protein yields (see Figure 3-2).

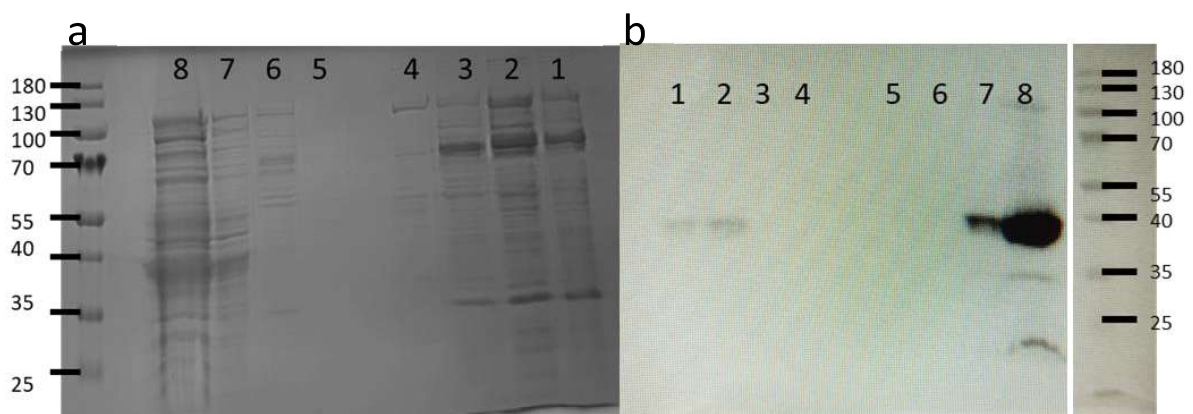


Figure 3-2: LeuT protein expression SDS-PAGE and Western blot

Confirmation of LeuT protein expression. In a), SDS-PAGE and b) western-blot of the following samples: 1- membrane preparation, 2- solubilized fraction, 3- flow-through metal affinity column, 4-6 metal affinity column washes of 5, 10, and 20 mM imidazole, 7- pooled metal affinity column fractions before the desalting column, 8- after desalting column. Notice 8 was over concentrated in this PAGE intensifying contaminants in relation to sample number 7. The western-blot shows protein bands containing his-tags, confirming the stronger protein band after the metal affinity column was indeed his-tagged LeuT.

Factors I found were contributing to lower protein yield were usage of aerated (filter caps) culture flasks with 1:1.5 ratio of growth media to air. Protein expression was increased after changing this ratio to 1:5, both in aerated and non-aerated flasks. This demonstrated that for this case, the aeration provided by the filters was clearly insufficient, and having a higher initial media to air volume important, as air filter flasks with 1:5 ratio showed the same result as non-aerated caps with the same 1:5 media:air ratio. Secondly, significant protein precipitation was happening upon each concentration step, which was resolved by a more intense pre-washing of the concentrators. It is possible that residual sodium azide was inducing some precipitation, and concentrator filter hydration was not enough. Finally, after many months of lockdown and troubleshooting, LeuT protein production was again operational (Figure 3-3).

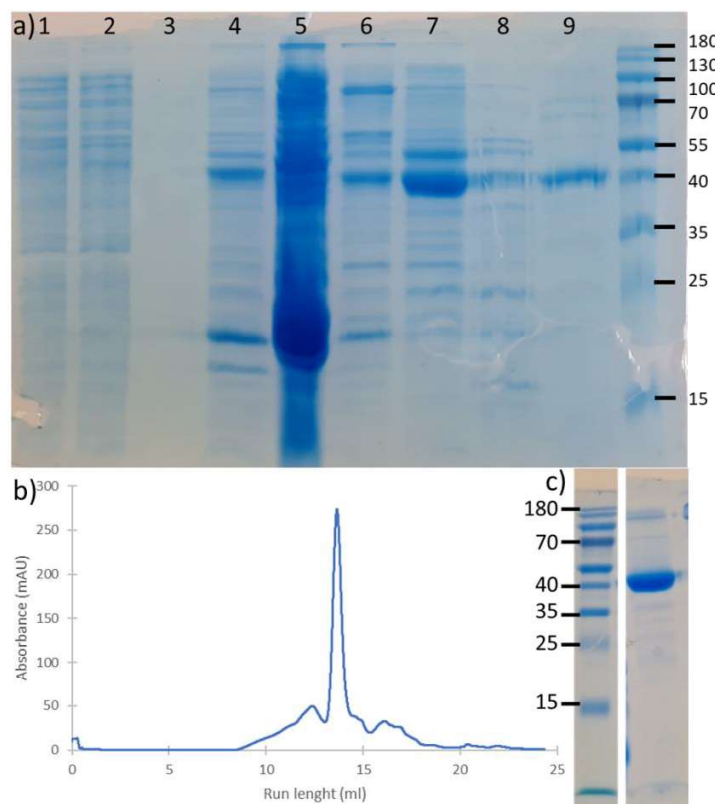


Figure 3-3 Final regained LeuT expression

Example of LeuT protein expression after SEC columns. a) WT LeuT SDS-PAGE from metal affinity column to after SEC column, b) WT LeuT SEC chromatogram, and c) example final protein sample of a 3M LeuT. Samples in a) are: 1- flow-through metal affinity column, 2- 5 mM imidazole wash, 3- 10 mM imidazole wash, 4- LeuT elution (metal affinity column), 5- precipitated protein before SEC, 6- first SEC peak, 7- second SEC peak, 8- third SEC peak, 9- 3M comparison sample.

The final adjustment made to the purification protocol was related to the use of copper phenanthroline complex (CuPh) to induce full crosslinking and locking of LeuT in the inward occluded form. As I was able to show that the purified protein is fully crosslinked (see §3.1.2) I removed these steps,

lowering the protocol complexity, protein usage, and reducing the time needed for the entire purification procedure. As a final note, when using CuPh, it had to be used on already purified protein instead of membrane preparations as using the latter resulted in precipitation[81].

3.1.2. Cross-linking

The triggering strategy for our LeuT time-resolved crystallography experiment was planned to be the cleavage of a cysteine disulfide bridge that locked the protein into an inward-occluded state. Cleavage of the disulfide would mean the protein conformation would no longer be restrained, in principle allowing for ligand binding and further conformation changes. Therefore, I assessed the degree of mutant LeuT crosslinking and determined whether additional crosslinking steps were needed.

To test the degree of crosslinking I used Ellman’s reagent to quantify the free sulfhydryl groups in LeuT after different treatments. Whereas WT LeuT has no cysteines, the 3M mutation introduces two cysteines into the LeuT sequence. To calculate the free sulfhydryl concentration following the Ellman’s reagent protocol the following formula was used:

$$\text{mM free sulfhydryls} = \text{Absorbance} / (\text{path length} \times 14.15) \times 20 \times \text{sample dilution factor}$$

Table 3-1: Cross-linking efficiency assessment.

Free sulfhydryl groups concentration in percentage of total cysteines in the sample calculated from protein concentration (3M). Protein buffer and WT samples had zero cysteines and results were negligible for each case (absorbance equal to the background baseline). In the “after purification” column samples were tested just after the normal purification procedure as reported on the methodology chapter. In the “TCEP incubation” column, a desalting column step was done before Ellman’s reaction.

Sample	% Sulfhydryl groups	
	After purification	TCEP incubation
Protein Buffer	0	0
3M	3	98
WT	0	0

A 3M sample, as purified with no CuPh treatment, at 0.6 mM (1.2 mM sulfhydryl groups) after DTNB reaction gave an absorbance at 412 nm of 0.036 (after background subtraction), equivalent to 0.04 mM free sulfhydryl groups (total dilution factor was 100). 0.04/1.2 indicates only 3% of the sulfhydryl groups in 3M are accessible to the DTNB.

If reacted with TCEP for 20 minutes and then passed through a desalting column before DTNB addition, the absorption measured was 0.2695, equivalent to 0.38 mM sulfhydryl groups (total dilution factor was 20). The sample after the desalting column was re-concentrated to 0.19 mM. With a theoretical

maximum free thiol signal corresponding to 0.39 mM (twice the protein concentration), the final free thiol content was now 98%.

Control experiments without protein sample and with WT LeuT gave negligible absorbance values, showing the increased signal in Ellman's reagent protocol was not due to vestiges of TCEP after the desalting column.

These results showed that contrary to what was initially thought, no pre-treatment with CuPh for disulfide bonding was required, as the protein sample was by already mostly oxidized after purification (3% free sulfhydryl groups). A desalting column step was necessary to separate the protein fraction from other reagents (*i.e.* TCEP) and allow for adequate Ellman's reagent reaction for free sulfhydryl quantification. After TCEP protein reduction, and a desalting column step, the free sulfhydryl numbers increased from 3% to 98% of calculated total cysteines in the sample, showing that indeed TCEP was reducing and breaking the disulfide bonds of the 3M, and that it took 20 minutes or less for it to happen. Controls were also made for WT and protein buffer only, but in every case they showed negligible results.

3.1.3. Binding assays

In order to determine whether a) the disulfide crosslink was able to lock the protein in an inward-occluded form and b) breakage of the disulfide could restore function, I carried out leucine binding assays. For this purpose, we saw potential in both fluorescence and CD assays as tryptophan fluorescence quenching and mechanism related secondary structure change would not require new protein modifications such as mutations and labeling as in previous reported experiments where the authors used single-molecule fluorescence resonance energy transfer (smFRET) or site-directed fluorescence quenching spectroscopy (SDFQS) [82]–[84] to study LeuT function. The fluorescence assays aimed to test whether ligand binding would induce fluorescence quenching due to aromatic residues being involved in the ligand interaction. The circular dichroism assays looked to confirm if the signal would change due to secondary structure alterations induced by ligand binding, confirming both that the protein was active and that there were indeed significant secondary structure alterations along the mechanism.

3.1.3.1 Binding (fluorescence)

Substrate binding can be detected using tryptophan fluorescence if ligand binding alters the environment of a tryptophan residue [85]. If the tryptophan interacts with the ligand, a decrease in fluorescence is expected, as part of the absorbed energy is transferred to the ligand instead of being

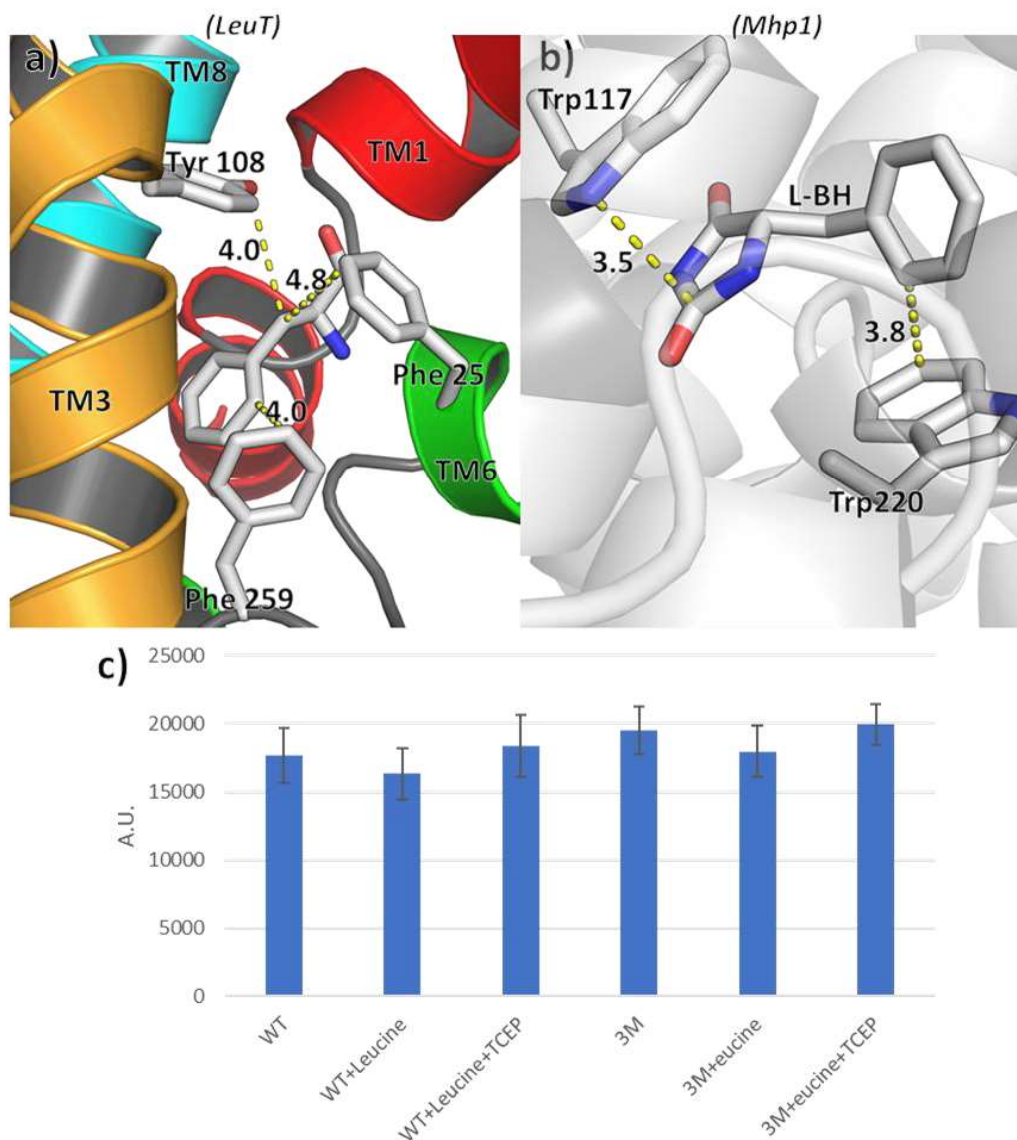


Figure 3-4: Fluorescence potential amino acids and corresponding results.

a) Distances to a phenylalanine ligand from surrounding potential fluorescence quench responsive amino acids, PDB ID: 3TU0 (LeuT outward occluded). b) amino acids responsible for native fluorescence quenching of mhp1 upon ligand binding [add reference], PDB ID: 4D1B. c) fluorescence results for WT and 3M LeuT.

released as fluorescence. This is the case for LeuT, as crystal structures of the outward occluded state show that tryptophan interacts with the leucine substrate (Figure 3-4). Hence, we proceeded with fluorescence assays to determine whether we could use this approach to monitor ligand binding to LeuT. If so, we could use this method to determine how much the 3M variant affects ligand binding. TCEP was added at the end of the leucine titration for both WT and 3M variant. If the disulfide bridge locks the

transporter in an inactive state, an increase in leucine binding and hence reduction in fluorescence would be expected upon TCEP addition.

Although there is a slight decrease in fluorescence when leucine is added to both WT and 3M LeuT, the change is less than the measurement error, making it statistically irrelevant.

Why is no significant change seen? Possible causes for this are 1) a lack of experiment setup optimization, for example leucine and sodium chloride concentrations, or the usage of dark bottom plates, which in the case of detergent using samples is ill-advised since detergent bubbles can disrupt adequate excitation and emission measurement from the top of the wells (even though this was minimized while setting up the plates); 2) fluorescence itself is not an adequate method for measuring binding of leucine in this specific study case, even when considering leucine should be interacting with a tryptophan in the binding pocket; or 3) the protein is not leucine-binding competent in the protein detergent complex. In this case, we do not expect that the DDM causes problems as this detergent has been used in previous structural studies of LeuT, and related Mhp1 experiments, showing that the larger fatty acid chain (as compared to shorter fatty acid chains like NM) correlated with increased protein flexibility and activity [82].

3.1.3.2 Binding (CD)

As the fluorescence experiments appeared not to be sensitive to LeuT binding of leucine, I explored circular dichroism as an alternative approach. Based on crystal structures, we saw that as LeuT binds to leucine and sodium and moves to an inward open state, trans-membrane helix 10 (TM10)

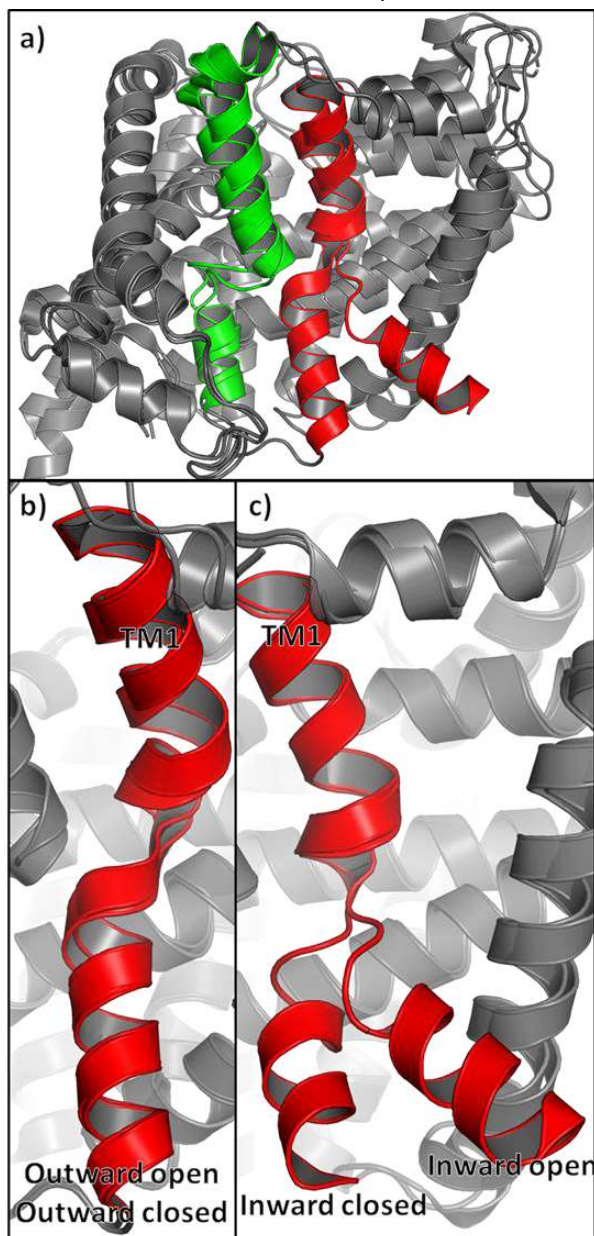


Figure 3-5: Partial helix distortion and unwinding between different LeuT conformations.

a) general overview of LeuT, with the TM1 colored in red and TM6 in green, superimposed structures of outward open and inward open (3tt1 and 3tt3); b) superposition of outward open and outward closed structures (pdb ID: 3tt1 and 3tu0 respectively). c) inward closed and inward open structures (pdb ID: 6xwm and 3tt3). Notice a partial distortion of the helix in the first 2 states, and then on the third and fourth states there is a pronounced helix unwinding, which is the feature that brought up the question if we could measure activity via CD spectrums.

protrudes outward, unwinding a few amino acids of the helix as it does so (Figure 3-5). As CD spectroscopy is very sensitive to secondary structure, it should be possible to detect leucine binding to LeuT using CD. If the LeuT is competent for binding, we would expect a gradual decrease in secondary structure signal as leucine is titrated in, until a plateau corresponding to the protein maximum binding capacity is reached. For the 3M variant, if this is unable to bind leucine, we would expect either a lower or no response to ligand titration until a reducing agent is added, breaking the disulfide link and allowing binding of leucine.

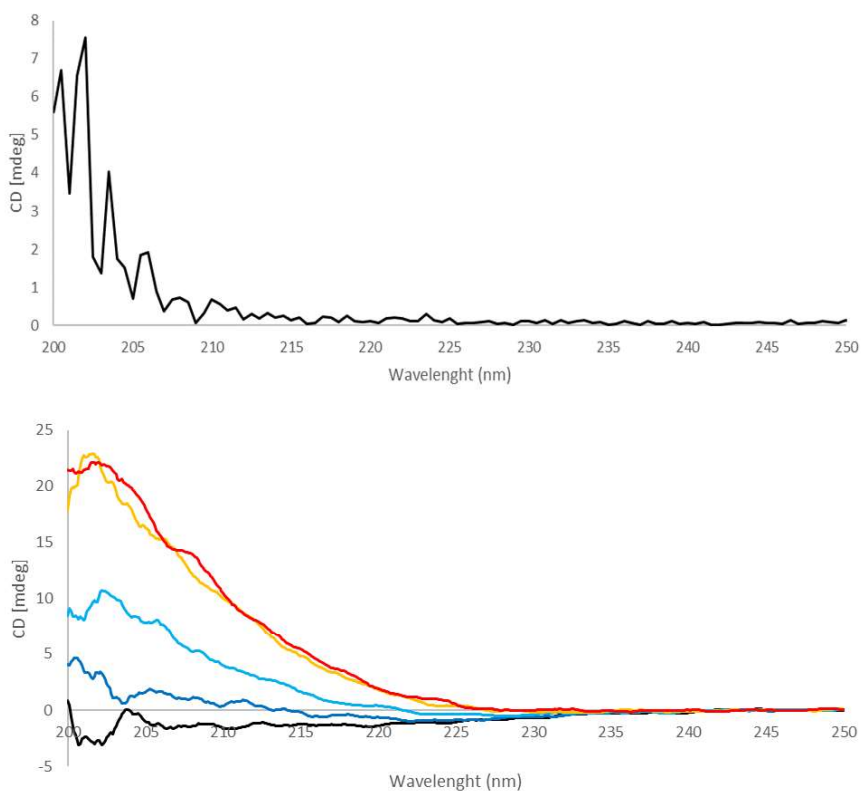


Figure 3-6: Average error distribution and leucine introduced background.

On top, average error from a 3 mM leucine and buffer contribution calculated from repeated measurements. On the bottom measured background for the final buffer composition and various leucine concentrations: black 0.0 mM, dark blue 0.5 mM, light blue 1.5 mM, orange 3.0 mM, and red 3.0 mM leucine with 0.5 mM TCEP. Buffer conditions were 20 mM HEPES pH 7.5, 0.05% DDM, 10 mM NaCl.

CD is a membrane protein “friendly” method as it works well even with various detergents. The major obstacle for this experiment, however, is salt concentration. LeuT’s interaction with ligands is salt dependent, especially on sodium [anna]. However, CD is highly sensitive to salt concentration, with noise increasing proportionally to salt concentration at the lower wavelengths, and at high concentrations it can overwhelm the signal even up to 230 nm. A second problem in this method is that the ligand used in

this study is leucine, which also contributes to the measured signal due its chiral nature, meaning both ligands used in this experiment (sodium chloride and leucine) contribute heavily to the measured signal.

I tested several buffer conditions with varying sodium chloride concentrations and settled on 10 mM NaCl as a satisfying compromise between background and ligand binding efficiency. Figure 3-6 shows the experimental noise (measured as standard deviation between 3 buffer baselines) using a 10 mM sodium chloride buffer. Below 200 nm the signal is too unreliable (jumping from 7 to 20-70 mdeg) for measurements to be reliable, hence the subsequent spectra shown are all plotted from 200 nm. Of these, the 200 – 203 nm zone was also noisy but still usable; 203-210 nm was satisfactory; and the range >210 nm was the most reliable (Figure 3-6).

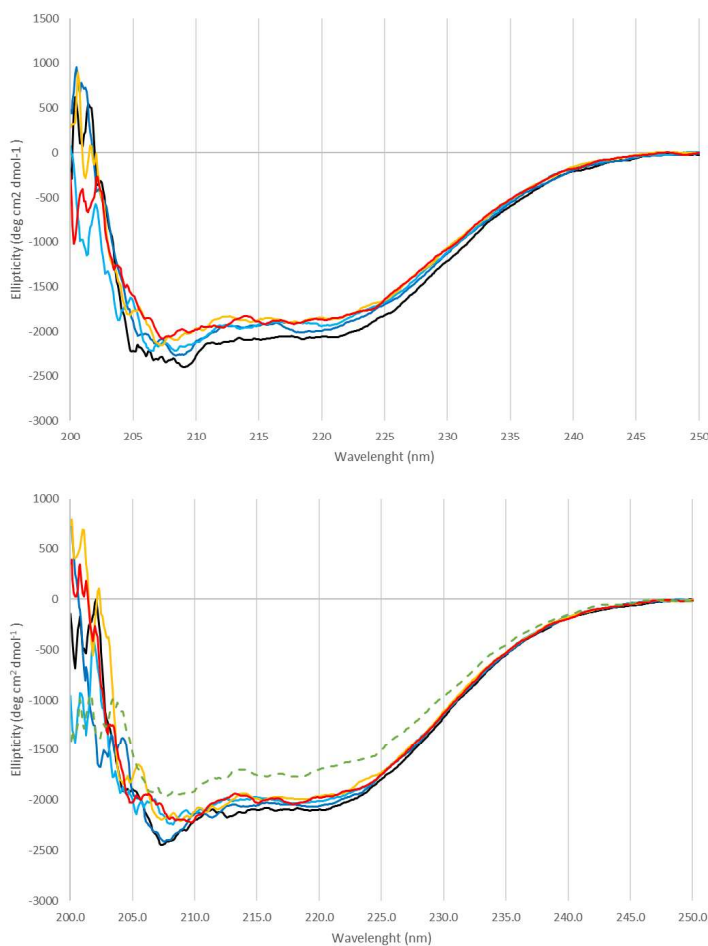


Figure 3-7: CD titration spectra of the WT and 3M LeuT.

Wild type (top) and mutant (bottom) LeuT CD spectra with black 0.0 mM, dark blue 0.5 mM, light blue 1.5 mM, orange 3.0 mM, and red 3.0 mM leucine with 0.5 mM TCEP. In dashed green is the same 3.0 mM leucine with 0.5 mM TCEP but with incubation time over night.

Alpha-helices appear in CD spectra as three peaks near 190, 208, and 221 nm. Due to the presence of 10 mM of sodium chloride, for direct peak change observation I disregarded the 190 nm and 208 nm peaks, relying mainly on the 221 nm peak. In the leucine titration spectra, a gradual decrease in the 208 and 221 nm peaks is observed in both WT and 3M cases (Figure 3-7).

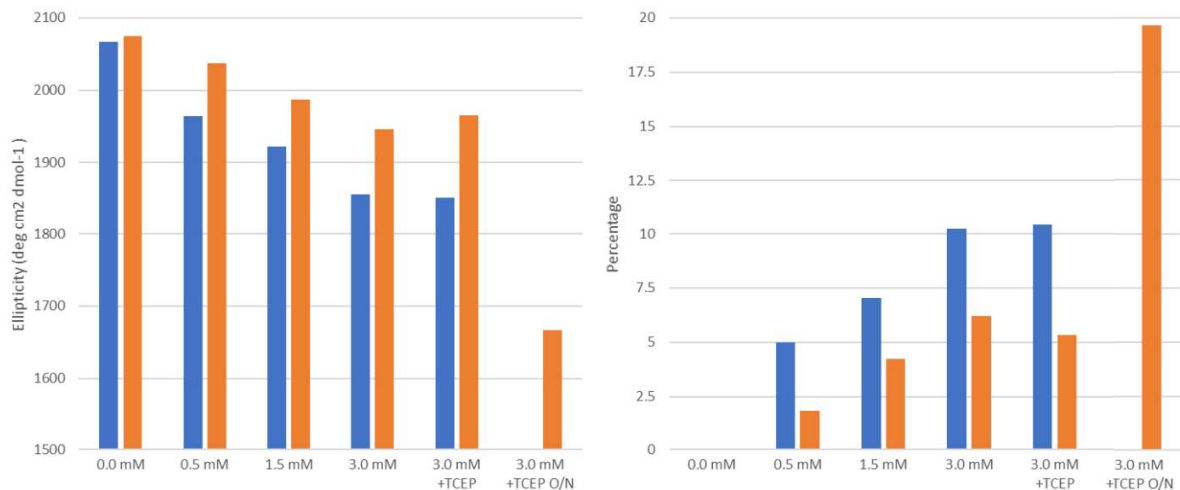


Figure 3-8: Ellipticity at 221 nm and their variations in percentage.

In the left ellipticity values at 221 nm and to the right variation of WT (blue) and 3M (orange) in percentage of each initial peak intensity.

Using the 221 nm peak as the most reliable activity indicator due to its lower noise from salt contribution, a steady decrease in signal for the WT is observed even after ligand signal subtraction (Figure 3-8). This was expected if binding is occurring, as a portion of the TM10 α -helix is unwinding. The peak intensity decreased in 5 % to 2.5 % steps of the original signal from 0 mM to 3 mM Leucine. After TCEP addition no further signal change for WT was observed, as expected since the WT lacks the disulfide crosslink.

Looking at the same peak for the 3M variant, we still observe a gradual decrease in signal but to a lesser extent than for WT (1 - 2% steps, as opposed to 2.5 - 5 % steps), this seems to suggest the mutant has indeed been locked in a conformation, where it cannot bind leucine as efficiently, but where binding is not fully ablated. At 3 mM leucine, there is nearly 50 % less signal variation than for the wild type protein (6 % versus WT 10 %). Based on the WT data and initial expectations, I expected that upon TCEP treatment the 221 nm peak for the 3M variant would decrease to a similar level as for the WT. On the contrary, a slight increase was observed, suggesting TCEP treatment caused a slight increase in the alpha helical content of 3M. Possibly the incubation time was not long enough (5 minutes) for the TCEP to effectively

break the disulfide bond, releasing the protein from a locked conformation and allowing it to change conformation. Of note is that the incubation time before the desalting column for the sulfhydryl quantification lasted 20 minutes, resulting in full disulfide bond reduction.

To determine whether there were further structural changes occurring over long time-scales, I incubated the mutant LeuT sample treated with 3 mM leucine and TCEP at 4 °C overnight. This sample showed a significantly lowered secondary structure the next day, even after correcting for precipitated protein (at least 96 % of the LeuT remained in solution overnight). This loss of secondary structure could indicate that more extensive structural rearrangements continued over a long period after TCEP treatment of the mutant. Unfortunately, this measurement was not made for the wild type which means we cannot currently distinguish these possibilities.

To better explain these observations, I calculated secondary structure predictions of the collected spectra, and compared these to theoretical values based on the reported crystal structures of LeuT in different conformational states. For theoretical values, I used four pdb structures each with LeuT in a different state: outward open (pdb ID 3tt1), outward occluded (pdb ID 3tu0), inward occluded (pdb ID 6xwm) and inward open (pdb ID 3tt3). For each file, I excluded anything that was not part of the LeuT protein (e.g. antibodies used in crystallization).

For each of these files, I obtained a secondary structure content in percentage from their structure using the YASARA software (<http://www.yasara.org/>). Additionally, for each file I ran "PDBMD2CD" (<https://pdbmd2cd.cryst.bbk.ac.uk/>) which outputs a predicted CD spectrum of the input protein structure. With the predicted spectrum of each LeuT mechanistic state, I then did a secondary structure analysis using the Bestsel tool (<https://bestsel.elte.hu/>) which outputs secondary structure in percentages. The Bestsel tool was also used for the measured LeuT WT and 3M data to obtain and compare with the theoretical results of each LeuT state. The result was: 1) secondary structure reading from crystal structures (YASARA), 2) secondary structure analysis of predicted CD spectra (PDBMD2CD + Bestsel tools), and 3) secondary structure analysis of measured CD spectra (Bestsel).

The CD spectra calculation produced by PDBMD2CD (from 3tt1, 3tu0, 6xwm, 3tt3) showed, as expected, similar overall profiles (Figure 3-9). However, a few differences were noticeable without any other analytical tool (like Bestsel), namely the peaks near 195 nm and 221 nm varied across LeuT states. The larger variation was in the 195 nm peak as α -helices heavily contribute here, but unfortunately, due to my experimental salt concentration requirements, this could not be directly compared with my data.

Nevertheless, the 221 nm peak also showed a significant difference between LeuT states. Even more interestingly, the signal change here is superimposed in alternate states; that is, the pairs outward occluded and inward open superimpose, as do the outward open and inward occluded. This means, that as the protein moves from outward open – outward occluded – inward occluded – inward open, the intensity of this peak would alternate between lower – higher – lower – higher intensities respectively. This possibly would explain why we saw an increase in peak intensity of 3 mM Leucine plus TCEP for the 3M, and its abrupt decrease overnight, but further experiments to test this are necessary.

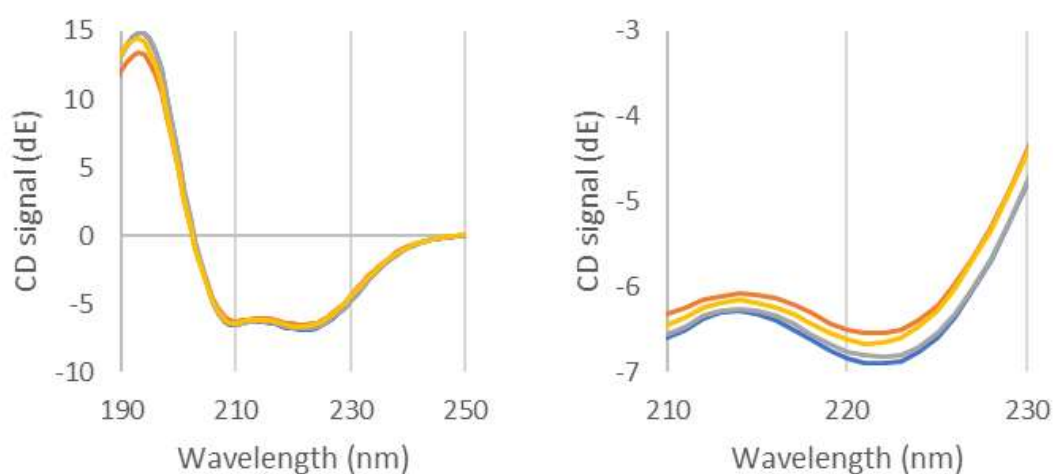


Figure 3-9 CD: spectra prediction of various LeuT conformational states.

Spectra prediction of various LeuT conformations, to the right zoom in near the 221 nm peak. The color scheme is: outward open orange (3tt1), outward occluded blue (3tu0), inward occluded yellow (6xwm), inward open gray (3tt3).

As seen in

Table 3-2, there are small differences in secondary structure content between each LeuT state, which also show an alternation with positive and negative variations. This includes some smaller variations and others that are larger, for example, outward open to outward closed (+0.1) and inward closed to inward open (+3.3). In the same table, the CD secondary structure prediction shows the same overall behavior. There is a lower overall total α -helix content, but this can be explained by both error accumulation between CD spectra prediction and secondary structure analysis of the prediction itself, and CD prediction spectra truncation (200 nm – 260 nm data, even though the CD spectra prediction outputs are as low as 180 nm). This was done in order to maintain a fair comparison to the actual measured data, which started only from 200 nm.

The secondary structure analysis of measured LeuT CD spectra again shows a pattern of alternation between lower and higher total α -helix content, in this case even more pronounced than

Table 3-2: α -helix contents of crystal structures, their predicted CD spectrums, and measured CD spectrums.

Table with the α -helix content in percentage of the crystal structures (YASARA), secondary structure analysis of their predicted CD spectrums, and below for the measured CD spectra of each leucine titration point. In parenthesis, value change from the previous data point, organized from outward open, outward occluded, inward occluded, and inward open. To the right, a few possibilities for which conformation each data point represents. *crystal structure secondary structure content was read with YASARA software after removing any non-LeuT protein molecules in their pdb file. **firstly a CD spectra prediction was done from each previously prepared pdb file (<https://pdbmd2cd.cryst.bbk.ac.uk/>), and then secondary structure analysis of these were made with Bestsel tool (<https://bestsel.elte.hu/>).

<u>THEORETICAL DATA</u>	Structure (YASARA)*		CD prediction**		
	α -helix content (%)	Difference from previous (%)	α -helix content (%)	Difference from previous (%)	
Outward open (3tt1)	74.9	(-0.7)	59.5	(-0.5)	
Outward closed (3tu0)	75.0	(+0.1)	60.7	(+1.2)	
Inward closed (6xwm)	72.3	(-2.7)	59.3	(-1.4)	
Inward open (3tt3)	75.6	(+3.3)	60.0	(+0.7)	
<u>MEASURED DATA</u>	WT**		3M**		<i>Expected conformations</i>
0.0 mM Leucine	63.41	(+13.21)	59.72	(6.00)	<i>Out closed / In open</i>
0.5 mM Leucine	59.73	(-3.68)	48.43	(-11.29)	<i>In closed / Out open</i>
1.5 mM Leucine	50.47	(-9.26)	48.29	(-0.14)	
3.0 mM Leucine	56.98	(+6.51)	58.86	(+10.57)	<i>In open / Out closed</i>
3.0 mM Leucine TCEP	50.2	(-6.78)	51.72	(-7.14)	
3.0 mM Leucine TCEP O/N	-	-	29.39	(-22.33)	<i>Out open / In closed</i>

predictions, with value changes as large as 10 % (discounting TCEP over-night with 22 %). Indeed, this seems to suggest the protein passed through various states during the leucine titration. For both WT and 3M, each starting at state “a”, moving through state “b” between 0.5 and 1.5 mM leucine titration, reaching state “c” at 3.0 mM Leucine, and possibly moving into state “d” with TCEP incubation and possible overnight incubation too. Comparing the WT and 3M results, and summing the 0.5 and 1.5 mM variations, values change with similar intensities between the potential states: -12.9/11.43, +6.51/10.57, -6.78/7.14 for WT/3M and states “b”, “c”, and “d”.

The theoretical data shows that the biggest negative α -helical content change is from outward closed to inward closed. If we attribute this relation to the data, it suggests the starting conformation is outward closed. However, the TCEP over-night spectra raises more questions. When considered with the other data points, it is the largest signal variation, and we know this is not due to lesser protein concentration as the protein precipitation overnight was less than 4 %. The α -helical content variation of -22.33 now becomes the largest variation and would now suggest this is the change from outward closed to inward closed state, making the initial state inward open, as would be expected for the 3M LeuT variant with its crosslinking cysteine.

Overall, these results show that we can indeed follow a change in secondary structure with ligand titration via CD, and potentially attribute these changes to changes in LeuT conformations. In total we have: 1) 221 nm peak intensities, 2) pdb total α -helix content (YASARA), 3) secondary structure prediction (PDBMD2CD – Bestsel), and 4) secondary structure analysis from CD measured spectra (Bestsel) all showing alternate positive and negative variations between LeuT conformations, strongly corroborating each other and suggesting we are indeed observing conformational states over the leucine titration.

Comparing WT and 3M results, the 3M unexpectedly shows a stronger reacting to 0.5 mM leucine, suggesting enhanced binding which goes against expectations, even though when summing the 0.5 and 1.5 mM leucine variations we have a higher change for the WT (-12.9 vs -11.43). It might be that total α -helical variation intensities means less than we are accounting for here, and should be taken more seriously as qualitative variations for the current data quality. Finer titration points, and comparison with higher and lower incubation times would be valuable to confirm any quantitative conclusions.

Ultimately, the aim of this experiment was to confirm that LeuT 3M was active, and that TCEP would break the cysteine crosslink from a LeuT locked conformation. The results show that we can investigate this with CD spectra experiments, but in order to validate that the 3M is in a locked conformation, further experiments will be needed. Unfortunately, current results cannot guarantee that a time-resolved experiment using the 3M would be feasible, but the results so far are promising.

3.1.4. LeuT Crystals

With the protein available after purification from transported membrane preparations, a few crystallization trays were carried out. Two initiatives were done in parallel, 1) a screen surrounding known crystallization conditions from Gouaux et al.[66], and 2) a commercial crystallization screen HR2-110 [69] to possibly find new or alternative crystallization conditions.

We obtained a few crystal hits from the Gouaux conditions tray. However, more promising hits came from the commercial screen. A total of four possible crystal hits were obtained out of the 99 conditions tested, one of which seemed particularly good at yielding plenty of small, fine, mostly singular needle like crystals (Figure 3-10). Since this was a new condition for us I decided to explore and screen around this novel condition.

It is of note that all crystal hits from the commercial kit contained acetate (3/4 being ammonium acetate) and sodium containing salt. The best hit was number 18, magnesium acetate, sodium cacodylate pH 6.5, PEG 8k. The Gouaux condition also used sodium for its salt component: sodium chloride, HEPES

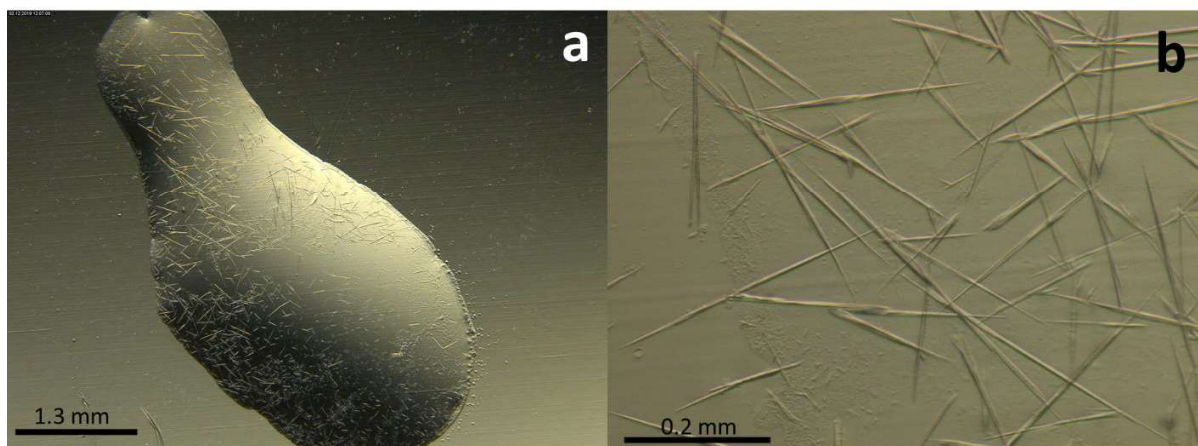


Figure 3-10: Reproduction of the best crystal hit from commercial crystallization screen HR2-110.

a) Picture of the drop which was the same conditions as the vial number 18 from the screen HR2-110, containing 20% PEG 8K, 0.2 M magnesium acetate, 0.1 M sodium cacodylate pH 6.5, protein concentration of 5.5 mg/ml. b) zoom in of the same crystallization drop.

pH 7.0, PEG 550. It is also interesting that these successful conditions spanned a wide pH range, namely from 4.6 - 7.0. This is useful information since it can easily be used to manipulate protein activity rates across those crystal compatible pH values, both for cases of each crystal tolerating pH changes, or just crystallization at different pH values to obtain the desired effect.

Further screening around hit 18 showed that these conditions were reproducible. The largest single crystals were obtained at PEG concentrations of 16 % (Figure 3-11), being large, long rod-like crystals with lengths up to 1 mm. 20-24 % PEG yielded multiple small or micro crystals and should be the starting point for optimizing crystallization towards micro crystals. It was at this point that the last protein samples were used from the transported membrane preparations, and I did not have another chance to work and

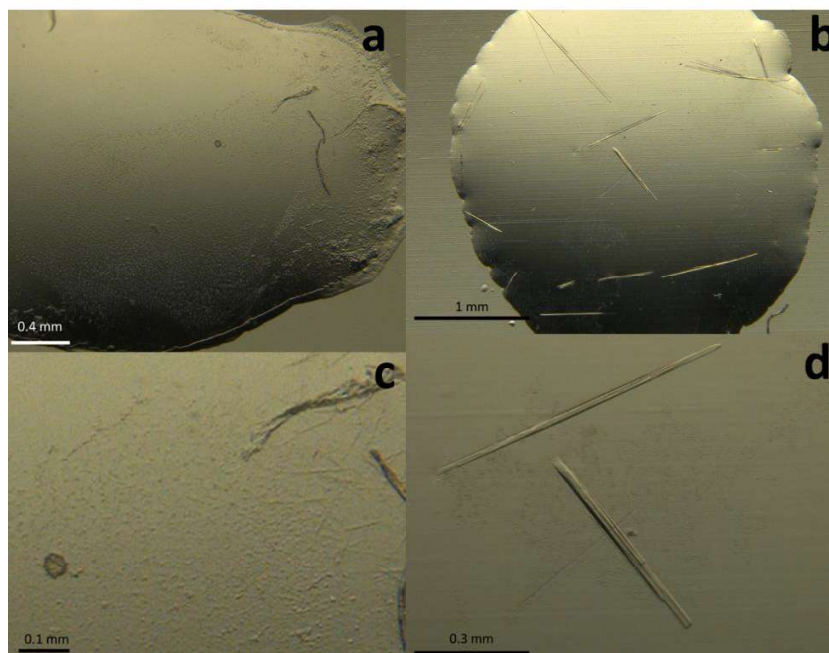


Figure 3-11: Example pictures of the PEG screen made around the best crystallization hit from the HR2-110 screen which had 20% PEG 8K.

All pictures are from the same two drops, which were 0.2 M magnesium acetate, 0.1 M sodium cacodylate pH 6.5, protein concentration of 5.5 mg/ml and a) 22 % PEG 8K, b) 16 % PEG 8K, c) zoom in of “a”, and d) zoom in of “b”.

improve these conditions towards micro crystals and a time-resolved experiment in the time remaining for my thesis.

Most crystals observed not only tended to be needles or rods, but were also often 2D. Indeed, various rod-like crystals after manipulation with crystal fishing loops were observed to be more like thin long plates than rods. Many of the crystals were unfortunately damaged when fished. This particular type of morphology is especially fragile when using inadequate needle loop sizes, and non-needle/rod crystal loops. My inexperience at fishing such crystal types, urgency to be ready in time for the beamtime, and a momentary lack of needle/rod loops led to increased numbers of damaged crystals and may also have been a major contributing factor for compromised data collection described in the following chapter.

3.1.5. LeuT X-ray data

Data collected at PETRA III beamline P14 were processed with various software in an attempt to overcome difficulties in data processing. XDS, mosflm, and DIALS were all tried for the first stages, and multiple options tried to improve spot finding. With all the various software the data processing was difficult even with multiple attempts at optimizing the settings for the specific data. Multiple datasets had

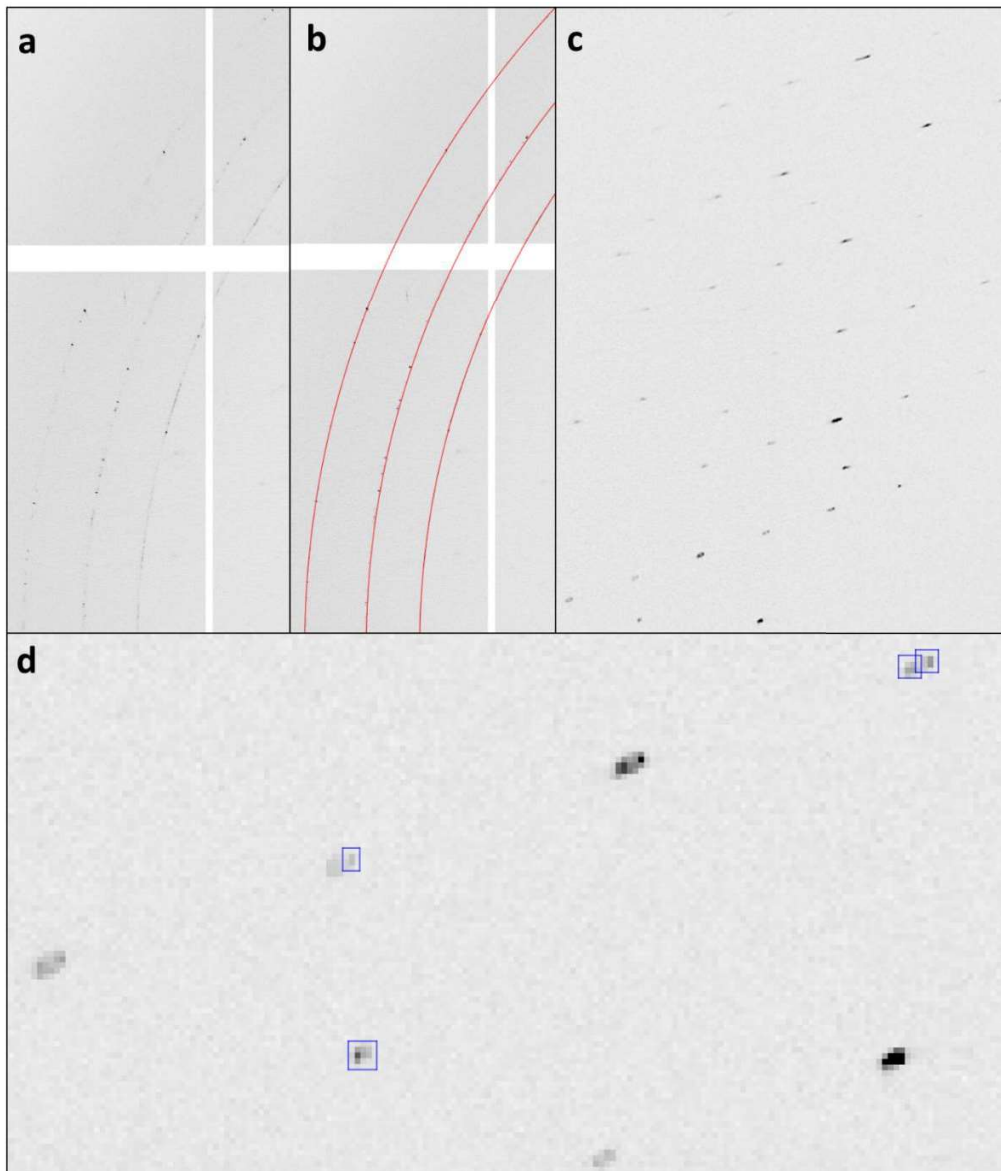


Figure 3-12: Pictures of example data frames with different options.

All the pictures here are from frames 1 (a and b) or 2000 (c and d) with a stacking of 10 frames each. a) ice rings, b) same as a but with ice rings option loaded, c) zoom in frame 2000 to show a few spots with streaking and d) same frame as c but further zoom showing a few multiple strong pixel spots, and with shoebox option enabled, identifying which spots the software accounted for and their area.

significant ice rings as shown in the Figure 3-12. More importantly, multiple frames showed strong streaking, which is the elongation of the spot along one direction. On top of this problem, spots were also found with multiple strong pixels, indicating that perhaps there were multiple lattices present. Finally, there was immense variation on the number of spots present per frame, across all datasets, due to the

anisotropy of the diffraction. All these combined factors proved too much for both the data processing software and the users.

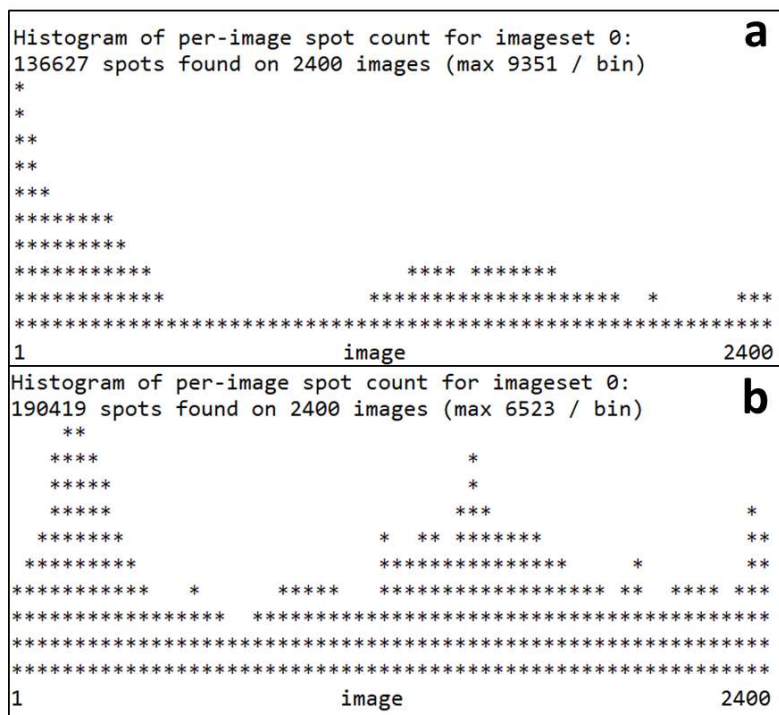


Figure 3-13: Comparison of spot finding results with default options and refined.

a) default options and b) optimized options. Of note is the heavy variance in spot finding across frames.

With various tweaking as mentioned in the methods, we observed an improvement in true positives in terms of spot finding, but still a significant number of spots were not being picked up by the software. Also, possible false positives were being captured as spots – even if these did not outnumber the positive change due to different settings.

In the example shown in Figure 3-13, the difference in spots found with the default options (136 k) versus the optimized options (190 k) can be observed (even though spot heterogeneity over frames persists). This could be due to simple crystal morphology, whereby one of the axes of the crystal was much slimmer (thin long plate-like crystals), or perhaps that the alignment was not properly done so that the beam was not hitting the crystal in a particular direction. This last possibility, however, seems less likely because all the crystals presented heterogeneity in spots per frame the same way, and at least two different people did alignments at various times. Moreover, even parts of the data that show lower spots

per frame, protein Bragg peaks were found, meaning the crystals must have been hit by the beam for the whole period.

The indexing and integration showed many problems, most likely due to the already poor initial data quality and spot finding capabilities. No matter the possible options selected regarding unit cell selection, P1, C2, multiple lattice options enabled or otherwise, the indexing showed very poor indexing rate. The rate was on average 20% for default parameters, and raised up to 50% with multi-lattices options enabled. Even so, the cell differences captured by the software were still minimal. Next the integration often failed, reported requiring unsustainable amounts of RAM and finding surprisingly high mosaicity. Even the cases where the data was forcefully pushed through these and later stages of data processing, it was impossible to solve the phase problem later with molecular replacement. There is still hope in further optimizing these data processing steps, and meetings with experts of the software developers' teams are being arranged.

3.2. Mhp1

This work was done in collaboration with Dr. Yunyun Gao and Dr. Maria Kokkinidou. The large-scale protein production was done to provide sample for their experiments. The experiments were aimed at studying the crystallization behavior of Mhp1, and hence also gaining some general insight into membrane protein crystallization. Through SAXS and various buffer conditions, we could observe different oligomerization protein behavior, and compare it to crystallization results, asserting if there were oligomer-crystallization dependencies.

Additionally, interesting Mhp1 datasets collected up to 2014 by Dr. Anna Polyakova that were not deposited in the PDB were re-refined and interpreted. The datasets were collected from samples with various ligands, although not all of them were visible on the density maps.

3.2.1. Large scale protein production

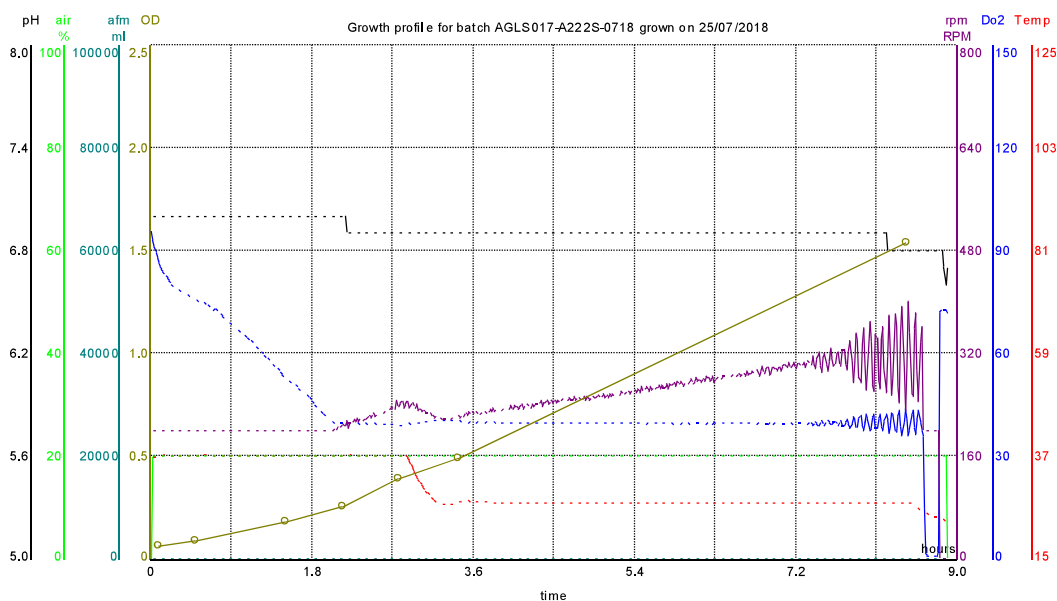


Figure 3-14: Fermenter growth profile with various metrics.

Growth profile of the 30 L fermenter growth of Mhp1 for later usage in Drs. Maria and Yunyun research projects. Oxygen (blue) was fixed and maintained by stirring (rpm, standing for rotation per minute of the fermenter mixing blades). Temperature was maintained and changed near 0.5 OD, while pH was followed and recorded but no set value was enforced. Antifoam (afm) was added manually at the start of the growth and not changed again. CD measurements were taken by sampling out media and using a bench spectrometer at various time intervals.

Each batch of transformed cells was inoculated into 30 L of media (Figure 3-14). Each such growth produced 95 g of cell paste on average, making it approximately 3.2 g per liter of growth media. In flasks, usually 10 L growths, the cell paste obtained was 14-24 g, meaning only 1.4-2.4 g per litre of cell paste per growth media. When comparing the cell paste obtained per liter (3.2 vs ~2.0 g) it becomes clear how efficient large-scale protein production can be, up to 56 % higher cell yields and no less than 25 % in the best-case scenario for flask growths.

The products of this large-scale protein production, and other growths, were purified by the methods described earlier and used for various studies done by Maria Kokkinidou including SAXS beamtimes. The results are described in her dissertation [86]. In this dissertation, the SAXS results from this collaboration are summarized and discussed below.

3.2.2. SAXS of Mhp1

This project was part of a collaboration between Dr. Yunyun Gao and Dr. Maria Kokkinidou throughout their PhDs, where firstly Dr. Kokkinidou prepared protein samples and Dr. Gao processed the SAXS data.

Nearing the second half of the project I started producing the necessary protein samples using large-scale protein production and helping at data collection experiments, until firstly Dr. Kokkinidou finished her thesis, and later Dr. Gao his. These results have, therefore, been presented in previous dissertations, but since these were products of collaboration and the time investment from my part was very significant, they are also presented in this dissertation.

The aim of the project was to follow and understand the relationship between buffer components (detergent and ligands), oligomerization states, and crystallization propensity. HPLC systems coupled to SAXS beamlines were used due to their capacity for oligomer separation by HPLC, and the ability of SAXS to provide overall shapes and sizes of the same oligomers, including the detergent corona, which in some cases can be useful. With this method it becomes easier to study the various oligomers present in solution since the resulting readouts are chromatograms with adequate separation between peaks. These peaks are caused by different existing oligomers that can then be individually studied via SAXS.

It was also initially hoped that we could discern between outward and inward conformations of Mhp1. Previous data showed that WT-MHp1 in DDM adopts predominantly an inward conformation [85], and that this gradually shifts towards an outward conformation with increasing amounts of sodium and hydantoin. WT Mhp1 in DDM has so far not been crystallizable, detergent exchanges are required for crystallization, *e.g.* NM. WT-Mhp1 in NM, however, shows decreased activity [87]. Based on the reported relationships, the following experiments studied the effects of NM, DDM, sodium, and benzylhydantoin on Mhp1.

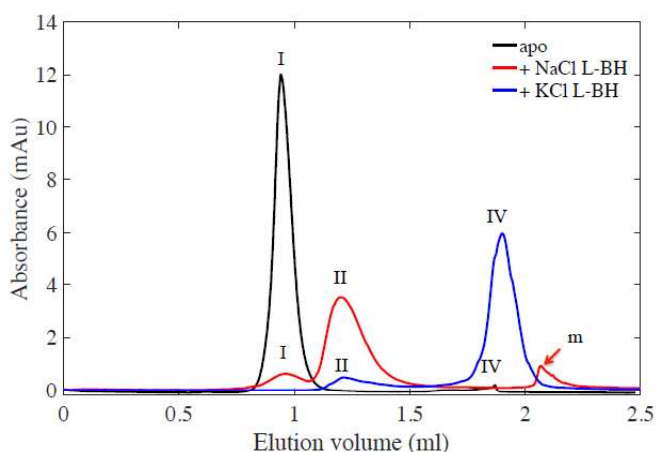


Figure 3-15: SEC-SAXS elution profiles of the wild-type Mhp1 in complex with NM.

WT Mhp1 in NM micelles without salt or L-BH (black), in presence of 1M NaCl and 2 mM L-BH (red), 1M KCl and 2mM L-BH (blue). Peaks are named in order of elution.

A single peak was observed for WT Mhp1 in NM at 0.9 ml (peak I, Figure 3-15), whereas the major peak shifted to 1.25 ml (peak II) and 1.9 ml (peak IV) for in samples incubated with sodium chloride and benzyl hydantoin, and potassium chloride plus benzyl hydantoin, respectively. The R_g and D_{max} for each peak varied between 91-110 Å and 300-345 Å for peak I, 57-118 Å and 221-322 Å for peak II, and 35 Å and 100 Å for peak IV, suggesting that peaks I, II, and IV are trimers, dimers, and monomers respectively (peak

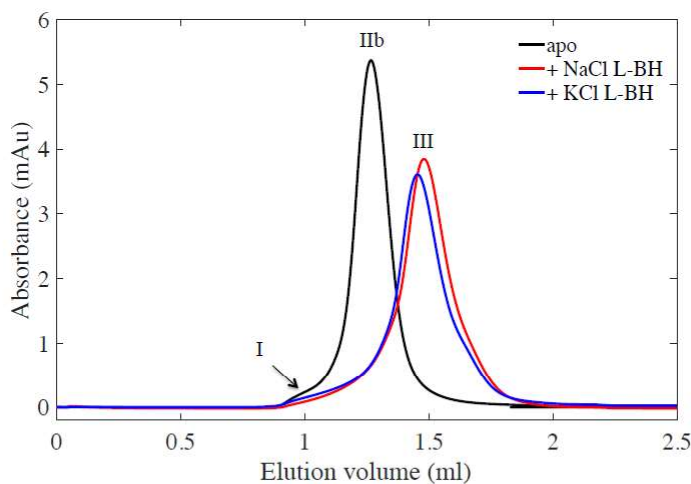


Figure 3-16: SEC-SAXS elution profiles of Mhp1 in DDM.

Mhp1 in DDM micelles without salt or L-BH (black), with 1 M NaCl and 2 mM L-BH (red), and 1 M KCl and 2 mM L-BH (blue).

II had larger R_g and D_{max} values for the singular case of samples incubated with potassium chloride plus benzyl hydantoin run, possibly having signal contributions from trimers also). It therefore seems that WT-Mhp1 in NM is predominantly composed of trimers, whereas addition of sodium chloride and benzyl hydantoin shifts it towards dimers, and monomers for the case of potassium chloride and benzyl hydantoin. Furthermore, modelling of the SAXS data suggested that the trimers were head-to-tail oligomers, resembling reported the crystal packing of Mhp1 structures.

For the WT-Mhp1 in DDM (Figure 3-16), peaks at 1.27 ml (peak IIb), and 1.45 ml (peak III) were observed for WT Mhp1 in DDM without and with potassium/sodium chloride and benzyl hydantoin, respectively. The calculated R_g and D_{max} were 43-59 Å and 200 Å, and 40-58 Å and 161-178 Å for peaks IIb and III respectively.

Additional experiments regarding WT-Mhp1 in DDM were done, showing that with a sodium chloride titration the peak IIb remained until both sodium chloride and benzyl hydantoin were present, shifting the peak to III. This showed that both salt and benzyl hydantoin are required for this shift to happen in DDM. Also, addition of only benzyl hydantoin, or choline chloride, or both benzyl hydantoin and choline

chloride did not shift the IIb peak towards III, showing that not only are sodium and benzyl hydantoin both required, but that there is an ion selectivity between potassium and sodium.

Table 3-3: SEC-SAXS elution peaks details and oligomer states.

SEC-SAXS various elution peak details and corresponding attributed oligomeric states by modelling performed in Dr. Yunyun Gao's PhD work. For each peak, the buffer, gyration radius (Rg), maximum distance (Dmax), and modelling suggested oligomeric state are showed.

Detergent	Buffer	Elution volume (ml)	Rg (Å)	Dmax (Å)	Suggested Oligomeric state
NM	apo	Peak I: 0.94	91-100	300-345	Trimer
	1M Na, 2mM L-BH	Peak I: 0.90	110	338	Trimer
		Peak II: 1.25	57	221	Dimer
	1M K, 2mM L-BH	Peak II: 1.25	118	322	Dimer
Peak IV: 1.90		35	100	Monomer	
DDM	apo	Peak IIb: 1.27	43-59	200	Dimer
	1M Na, 2mM L-BH	Peak IIb: 1.27	40-54	161	Dimer
	1M K, 2mM L-BH	Peak IIb: 1.27	44-58	178	Dimer

The project carried on profiling the HPLC-SAXS data of Mhp1 mutants that caused preferred outward (Met39Cys and Ala222Ser) and inward (Ala309Asn and Thr313Ala) conformations. Those profiles further corroborated the here formed observations. The inward conformation promoting mutations shifted the oligomeric state of Mhp1 in NM from trimers to dimers and monomers, which in turn gave poor crystallization outcomes. Mutants that stabilized outward conformations showed a predominance of trimers in NM (similarly to WT-Mhp1 in NM) and were in turn associated with improved crystallization outcomes. This suggests that trimer oligomers are related to outward conformations since WT-Mhp1 in NM (outward in solution) and outward locked mutations provided predominantly trimer oligomers and crystallized readily, whereas inward locked mutations lacked trimers and did not crystallize well.

In total, this project 1) showed a relationship between oligomer state and crystallization propensity in which the formation of stable trimers was associated with high crystallization propensity; 2) showed that detergent and ligands influence oligomer states; and 3) demonstrated that outward conformations were more likely to assemble into trimers while inward conformations were found predominantly as dimers and monomers. Hence an observation can be made that buffer conditions/mutants can provoke (out/inward) conformations that in turn are compatible with trimer and ultimately crystal formation.

3.2.3. Structure refinement

The structures refined here were obtained from experiments carried out over eight years ago, the data were already merged, and phases solved with molecular replacement, and even a few cycles of refinement had been done. I resumed and continued the refinement until no further reasonable advancements could be made. In total we had five data sets, each with different ligands in the sample, although not all of them were visible on the density maps. The datasets were: apo (no ligand), IMH (L-5-indolylhydantoin), BH (L-5-benzylhydantoin), BH-NaCl (L-5-benzylhydantoin, lower salt), and BVH (Bromovinylhydantoin). The difference between BH and BH-NaCl was that, up until crystallization lower sodium chloride concentrations were used, as an attempt to diminish protein ligand binding which is dependent on the salt concentration. However, both structures are from crystals grown at the same sodium chloride concentration of 100 mM, which the author of the work at the time (Dr. Anna Polyakova) suspected would override the possible changes in samples ligand binding and/or conformations caused by the different salt concentrations.

Table 3-4: Statistics for each dataset.

The five datasets are organized in columns with corresponding maximum resolution indicated after their names. Cells present the values of R and Rfree on top of clash score and molprobtity score, each separated by a forward slash. The rows are: original values taken from the pdb obtained at the start of this study; first cycle of refinement without changing any parameters except as instructed by the author of the prior refinement cycles; after multiple refinements using map sharpening and jelly body restrictions as instructed by the author; same as previous row except without jelly body restrictions; and finally for comparison Vagabond refinement.

	Apo 3.8	BH-salt 3.8	BH 3.7	BVH 3.5	IMH 4.0
Original	28/31	28/32	28/31	24/27	28/33
	15/2.9	5/2.0	10/2.5	6/2.2	12/2.4
First cycle reftmac	29/33	30/33	31/33	24/27	31/35
	15/2.9	3/1.5	10/2.5	6/2.2	14/3.1
First cycle no jelly body	29/33	30/33	31/33	24/27	31/35
	18/3.2	13/2.7	23/3.3	21/3.2	15/2.9
Multiple refinements	31/32	33/33	32/36	25/27	28/36
	7/2.5	5/2.7	14/2.8	10/2.6	12/2.4
Multiple refinements no jelly body	29/32	30/33	27/33	22/27	28/36
	10/2.8	13/3.1	22/3.2	24/3.3	14/3.0
Vagabond refinement	54/34	58/30	58/38	44/29	57/34
	15/2.8	13/3.2	23/8.2	17/3.0	16/3.2

Table 3-4 presents the resulting R and R_{free} values, clash and molprobity scores organized per dataset and processing stage from which the values were taken. Multiple observations can be made from these results.

Firstly, comparing the first and second rows, we notice the R and R_{free} values are inflated except for the best resolution dataset BVH. The second row corresponds to doing one REFMAC refinement cycle on top of the originally provided data (pdb and mtz files) without any deviance from the instructions given by the author at [87], *i.e.* no coot assisted model modification was done before the REFMAC cycle. The difference in the inflated R values should be solely due to different REFMAC/CCP4 software versions and is, therefore, interesting to note that these increase rather than decrease as would be expected. This might be due to over-fitting of the model, where too much detail is introduced that the data can not verify. This seems to be the case even if the added detail better corresponds to the reality of the crystallized molecule (because of the improved software). It is interesting that this difference in R values does not occur for the BVH dataset which is the highest resolution dataset, possibly due to this very reason, *i.e.* the refinement software did not inflate the R values as it had access to better quality data that could justify the automated modifications.

The clashscore and molprobity score (combination of clashscore, Ramachandran not favored, and bad sidechain rotamers) can be used as indicators of model quality. For three out of five datasets these scores remained the same for the original data and first cycle of REFMAC (apo, BH, BVH). For the BH-salt dataset, one cycle of REFMAC resulted in higher R values but simultaneously improved the model geometry regarding clashes, Ramachandran non-favored, and bad rotamers.

The “multiple refinements” row presents values given by datasets that were processed over a few Coot assisted model modifications and REFMAC refinement cycles, all with the jelly body restrictions and map sharpening features turned on. Over the various iterations, no major modifications were made or felt to be necessary. Sidechains were added even in cases where no electron density was present from lacking enough resolution to solve their positions (although the backbone had electron density). This was a conscious decision and it was done with the original aim of depositing the structures. Future possible users of the data may benefit from those added sidechains, *e.g.*, for projects exploring molecular dynamics. Adding these sidechains along the peptide chain is most probably an important factor for the increase of R values when comparing the multiple refinements and original values rows. An attempt to limit over-fitting was otherwise done, *e.g.* for the IMH dataset there was not enough electron density to

justify the ligand, which was removed and later added by the end of the refinement cycles for tests and comparisons.

For comparison with the “multiple refinements” row, one cycle of refinement was done without jelly body restrictions, instead this was done on top of the previous refinements containing the jelly body restrictions option on. It is expected that jelly body restrictions add a layer of refinement information that can help compensate for the datasets’ low resolution. Interestingly, and perhaps as expected, when comparing these two rows (multiple refinements with vs without jelly body restrictions) it is observable that the clashscore and molprobity score improve but at the cost of inflating the R values even further. We can expect this difference to be due to the jelly body restrictions helping the refinement software find conformations that better accommodate inter backbone and side-chain interactions and packing. Therefore this better satisfies clashes, Ramachandran outliers and non-favored conformations, and expected sidechain rotamers, even though there is not enough resolution to validate such detailed model coordinates in at least some areas of the protein.

To further elucidate this situation, I did a first refinement cycle on the original data, the same as the row number 2 in the previous table, but without the jelly body restrictions option. Interestingly, the R values for all the datasets remained the same, even with the clash and molprobity scores worsening significantly, and more intensely in most cases than when comparing to the “multiple refinements” with and without jelly body refinements. It is possible this difference is somehow related to the increased amount of refinement cycles done with the “jelly body” option on, where slight model differences add on each other over multiple cycles. When finally, the option is toggled off, the slight model differences over multiple cycles make for a much more significant impact on statistics than taking the option on or off over fewer refining cycles as in the case of the “first cycle” data rows, and original.

Overall, the multiple refinement structures were similar to the original ones, with less than 0.4 RMSD for aligned structures. The major features of the protein remained, and only in the IMH dataset did one specific feature differ from the refined structure. The original model included an IMH molecule at the active site, interacting with tryptophan 117. The electron density supporting this ligand placement was minor, covering only a small fraction of the ligand. The author herself pointed out that although it was suspected to be the ligand, the data did not fully support the affirmation. The now refined structure

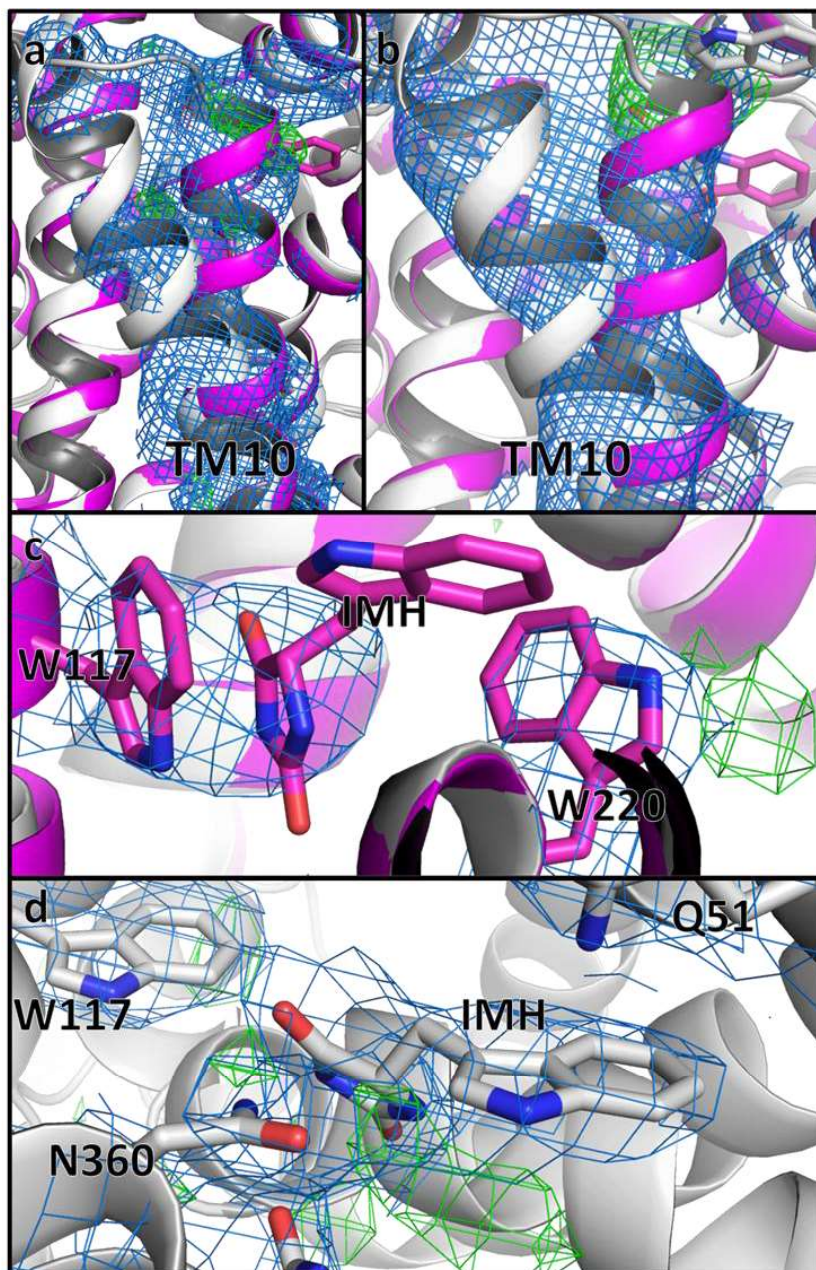


Figure 3-17: Coot models and density maps of IMH original and refined structures comparing ligand positions.

For all images, in white is refined models, and magenta the original model. a) comparison of TM10, with electron density of the original model b) now electron density from the more recent refinements, but without ligand addition (note the positive density where the ligand is added at d). c) original ligand placement, with the tryptophans 117 and 220 and corresponding electron densities; d) ligand placement in recent refinement cycles, notice the tryptophan 117 is now rotated occupying the previous ligand position.

showed that instead of the ligand, the tryptophan 117 side chain was now shifted by 90 degrees (Figure 3-17, c and d) and occupying the previous ligand position. No positive density appeared at the previous

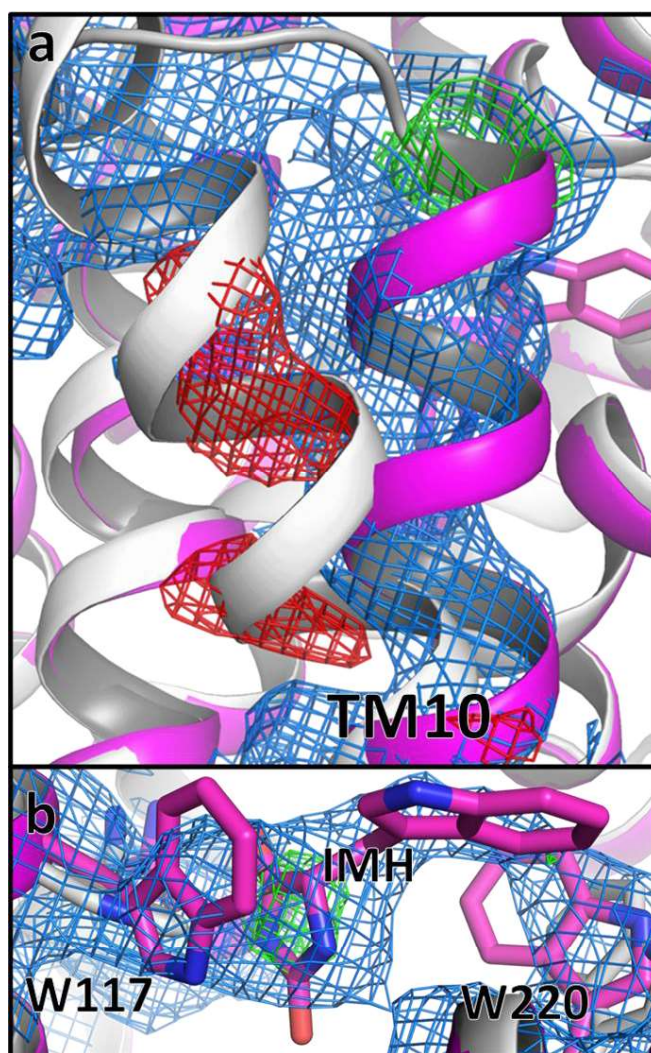


Figure 3-18: Superposition of the Vagabond refined IMH model, the multiple REFMAC refined model, and the original model.

a) main chain representation of the Vagabond refined structure superimposed with the REFMAC multiple refinement model (filling the negative density volume) b) vagabond model focusing on the tryptophan 117.

tryptophan position (before rotation), and no negative density was present at its new position, whether using sharpened or unsharpened electron density maps. Additionally, new positive electron density appeared a few angstroms away, closer to asparagine 360. As a test the ligand was placed there and a few cycles of coot and REFMAC refinement were done, still some positive electron density remained. However, when comparing both original and refined IMH structures, a new potential justification for the electron density became clear. The refined structure had a different main chain folding at that position (TM10), and major change in sidechains positioning due to this. The original structure was indeed filling out this new electron density with its main and sidechains (Figure 3-17 a and b). The question then was

which of those was the correct model. The original (without any first REFMAC cycles) showed a better Rfree value, the new refined structure seemed to better satisfy the electron density regarding the ligand.

The original structure, just like its author mentioned, did not have enough electron density to validate the ligand position. For the newly refined data, one could argue that if the backbone can be guaranteed to be in the position which accommodates the ligands' new position, it could be the most likely correct option; but this could not be assured. The backbone portion of TM10 facing the outward surface of the protein lacks electron satisfying electron density to justify it (Figure 3-17). Finally, the new ligand position did not have enough meaningful interactions with the surrounding amino acids to support its location. It indeed seemed plausible that this shift did not occur at all, and the backbone was just badly refined on the specific position, leaving then no remaining electron density to justify any ligand placement.

The new software Vagabond was tested to compare its performance with the final refined and original models [76]. Vagabond uses a bond-based model, instead of atoms which results in a lower number of refinable parameters. This was expected to be particularly helpful for low resolution data (where the danger of overfitting is high).

The Vagabond refined structures still largely resembled the original and refined structures. Interestingly, the major difference was found again for the IMH dataset, in which the helix portion that

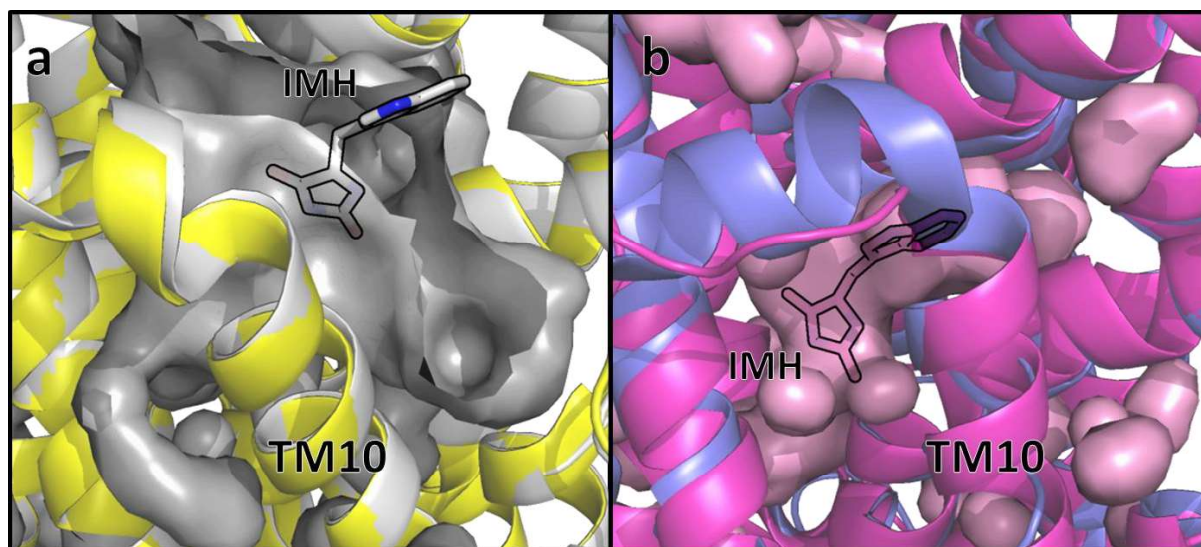


Figure 3-19: Cavities of original and refined Mhp1.

a) in white refined Mhp1 with ligand placement, superimposed with 2jln, showing an outward open conformation with the binding cavity surface shown in grey. b) in magenta the original model, its occluded cavity showing in light-magenta, and superimposed with 2x79 in blue (inward conformation).

was refined in a new position here showed clear negative electron density, suggesting TM10 was indeed in the occluded conformation. This further supported the assumption that this new conformation was just an artifact (Figure 3-18).

This conformation was most probably induced due to using 2jln for molecular replacement (Figure 3-19), in which TM10 stays in the outward open conformation. Although REFMAC and manual refinement iterations never showed negative density in this region, Vagabond did pick this up, and clearly showed negative density here, making the correct position of TM10 clearer – indeed showing the software’s promise for low resolution datasets. Positive density was also found on the previously expected ligand position, near to tryptophan 117. However, inspection showed the sidechain of tryptophan 117 had collapsed upon itself; this bug is most probably the reason for the positive density.

3.3. Detergent studies

One of the aims of this work was to optimize membrane protein crystallization towards micro-crystals, and provide a general approach for such procedures. For this, we used the existing “crystallization bench” at IBS, in collaboration with Prof. Monika Spano and Dr. Elham Vahdat from the Université Grenoble Alpes. The “crystallization bench” consists of a temperature-controlled fluidics system connected to a dialysis button and constantly monitored by a microscope and camera.

Previously salt and PEG diffusion rates across typically used dialysis membranes were measured [61]. We were now interested in knowing these values for typically used membrane protein detergents, therefore I designed an experimental setup to measure the detergent diffusion rates.

3.3.1. Diffusion rate

At the time of my experiments, there were no reports of measuring diffusion rates through the dialysis membranes commonly used in dialysis crystallization experiments. Therefore, I designed and planned an experiment that could measure the diffusion of detergent across dialysis membranes used both in typical crystallographic dialysis button experiments as well as in the “crystallization bench”.

Firstly, we needed a way to detect detergent. One possibility was the refractive index for which we tried an old wood refractometer at Université Grenoble Alpes, but this did not have adequate resolution for the purpose. There are modern refractometers that can measure detergent at even higher resolutions than the method we employed here, but I did not have access to such machines at the start of the project, so I sought an alternate solution.

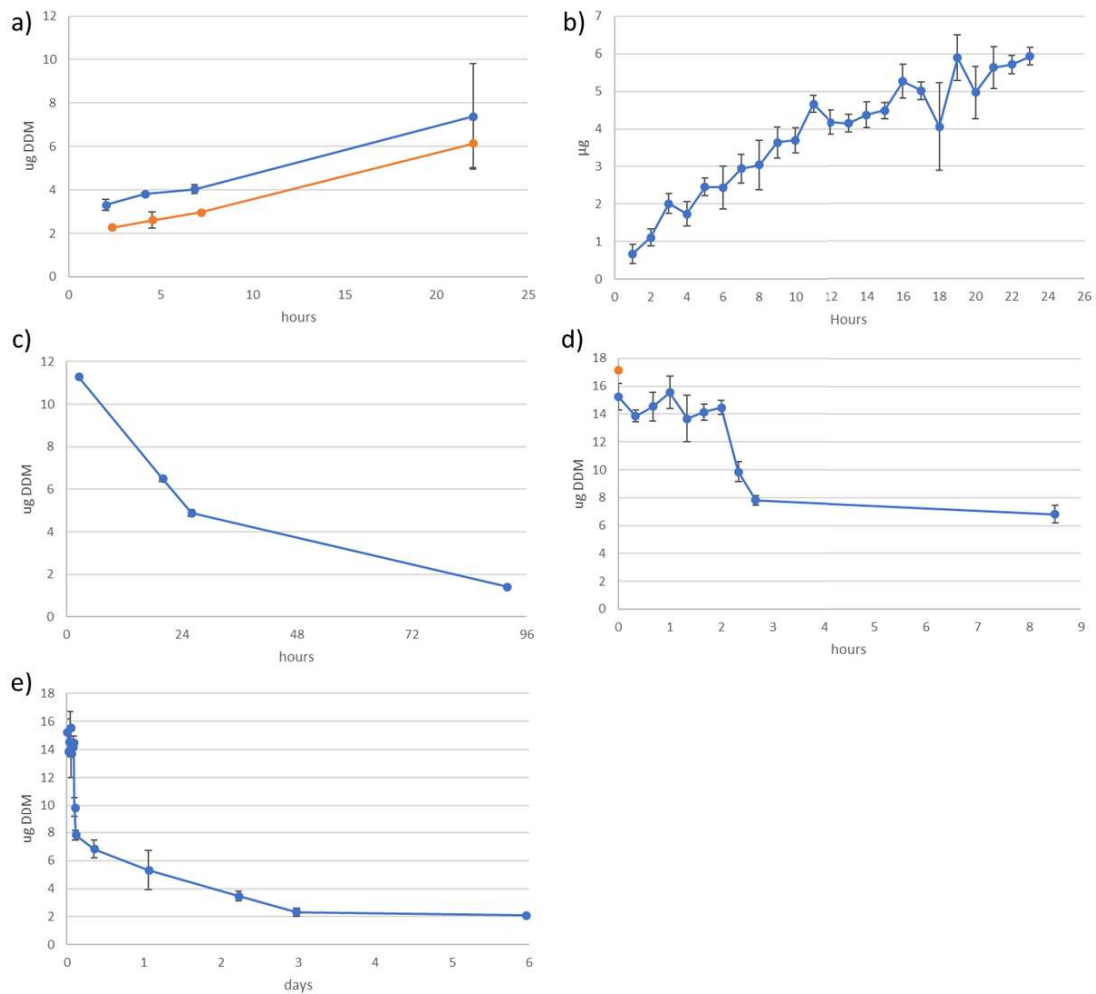


Figure 3-20: First experiments for detergent diffusion rate measurement.

First experiments that were done to measure detergent diffusion rate across dialysis membranes. a) blue and orange are 100 and 20 kDa cutoff membranes respectively, the last point of the 100 kDa had increased error due to lower repetitions at this first experimental test (1 button, 2 measurements). detergent was measured outside the buttons. b) and following experiments were done with only the 100 kDa dialysis membranes. c) and following experiments detergent measurement were done inside the dialysis buttons. d) and e) are the same experiment but the first in the hours scale and following in the days scale. Each data point is now 2 buttons, 4 measurements. In all cases 0.05 % DDM was the initial concentration, corresponding to 23 μg of DDM. Notice the b-c first data points are all below the expected 23 μg DDM; which later we found out was due to dialysis membrane detergent absorbance.

I selected the method described by Urbani et al. [80] which uses sulfuric acid to dehydrate the samples' carbohydrates and then phenol to react specifically with the sugar molecules present after the detergent dehydration, producing an absorbance at 495 nm that correlates to the sugar concentration. This method was sensitive down to 3-5 μg of DDM.

The experiment itself involved taking 0.2 ml dialysis buttons covered with dialysis membranes of varying cut-off sizes and inserting these into falcon tubes containing 5 ml solutions, with detergent either in the button or in the surrounding solution. The detergent concentration inside the button over time was followed with the colorimetric assay [80].

I assumed that dialysis membrane with cut-off sizes larger than the detergent micelles (72 kDa) would allow for both detergent monomers and micelle diffusion, whereas smaller cut-off sizes would allow only for monomers to diffuse, therefore lower diffusion rates.

As a preliminary assay, I did a test experiment (Figure 3-20) with four time point measurements from 2.5 hours to 22 hours, for 100 and 20 kDa cut-off sizes with 0.05% detergent solution outside the buttons. This preliminary test showed that both membranes gave a similar detergent diffusion rate, although with a high error at the 100 kDa membrane's last time point and with a slight difference in starting DDM concentration that seemed to be maintained for the 22 hours.

I then wanted to confirm the time required for the detergent concentration to reach equilibrium and changed the DDM solution from being outside the buttons to inside, to use less detergent overall. This experiment used only the 100 kDa cut-off membrane and four time points were measured to a maximum of 92 hours, after which the detergent outside the button was near the final expected concentration of 1 μg .

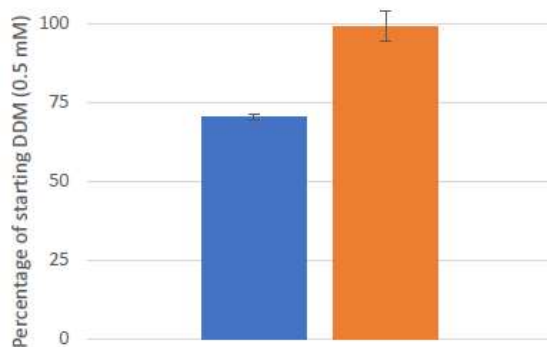


Figure 3-21: Experiment showing detergent absorption by dialysis membrane.

Experiment where tubes containing a DDM detergent solution with (blue) or without (orange) 100 kDa dialysis membranes were left for three days, and then measured the detergent remaining in solution. It becomes clear more than 25 % of detergent was absorbed by the dialysis membranes, explaining why the zero points were not equal to the expected in previous experiments.

With this information, I wanted to more finely sample the initial diffusion over the first three hours, as well as continue to monitor the system for up to six days. In this experiment I recorded 15 time point measurements, using 30 buttons, and making 60 measurements over six days. An increased

diffusion rate was indeed observed in the first 3 hours of the experiment, especially pronounced between hours two and three. From the third day to the sixth day no difference was detected between measurements, although the 2 μg detergent detected inside the button was still higher than the 0.7 μg expected for the equilibrium point.

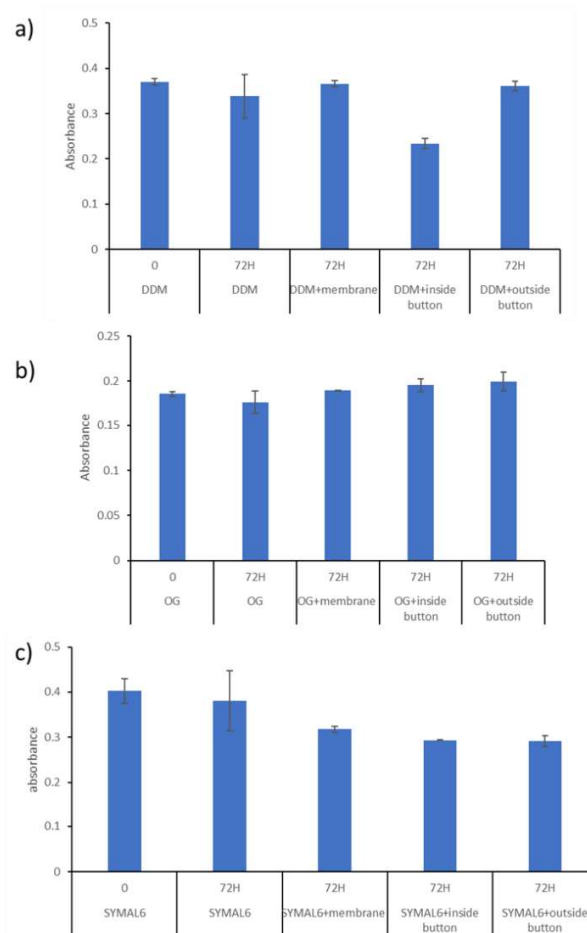


Figure 3-22: Additional studies made with non-detergent absorbing membranes.

Study executed by Elham Vahdatahar and Monika Spano after realizing the detergent absorption phenomenon, and testing with the cellulose membranes, preventing the majority of detergent absorption. Results for a) DDM, b) OG, c) SYMAL6 are shown for time zero, 72 hours without membranes, with membranes, with membranes inside and outside the dialysis button.

After this experiment, we wondered whether the membranes might be absorbing detergent, and if so, how this would impact the results. For this a simple experiment was designed, in which a comparison was made between falcons with detergent solution, and falcons with the same solution plus a dialysis membrane added (Figure 3-21). Measurements of free detergent in solution were made after 3 days. This showed that for the 100 kDa dialysis membranes, 25% of the detergent was absorbed after three days. It

is interesting to note that in the previous time-series (figure 3.21) the $t=$ "zero" time points had consistently around 25 % less detergent than expected, suggesting perhaps this absorption onto the dialysis membrane does not need a long time to occur, and instead occurs under 5 to 10 minutes. In the previous experiments the $t=$ zero time points were in practice a few minutes after the detergent and membrane first came into contact as it took several minutes to seal the button and then insert the button in the falcon tube containing detergent free solution.

This result adds a layer of complexity to the question of measuring detergent diffusion over time, as absorption is taking place at the same time as diffusion. We then delved into the literature to find more about this effect and discovered that of the two main dialysis membranes used in our laboratories, one of them (cellulose acetate, which was our 100 kDa membrane) is hydrophobic, and does absorb detergent, whereas the 12 -14 kDa membranes (cellulose) are hydrophilic and do not absorb detergent.

With this new information, we set out to confirm this, and did a series of tests with three different detergents, as well as detergent solution outside the buttons, taking measurements on the third day. This showed that indeed the 100 kDa cellulose acetate membrane absorbed detergent, more so in the case of DDM and SYMAL6, but not OG. Also, in the 12-14 kDa membranes (cellulose), the equilibrium was reached after three days for OG, but not DDM, which was at about 60% of the DDM stock solution.

In order to complete the diffusion rate study, a repeat of the experiments should be done with only cellulose membranes, therefore, preventing most of the detergent absorption. In any case, time taken to reach detergent concentration equilibrium is now known; for OG under three days, for DDM using a 100 kDa membrane MWCO three days, and for DDM using a 12-14 kDa membrane MWCO more than three days (Figure 3-22). Importantly, the fact that dialysis membranes absorb detergent, potentially dropping the concentration below the CMC has been highlighted and may explain challenges in maintaining the stability of and using diffusion-based approaches to crystallize membrane proteins. I do not believe this issue is well known in the crystallization and membrane protein research communities.

3.4. Solid target and LCP compatibility studies for time-resolved crystallography

The objective of the following experiments was to assess if it would be possible to do time-resolved crystallographic experiments using LCP grown protein crystals and light-triggering using synchrotron radiation sources. A major challenge in such experiments is how to deliver the crystals to the beam.

Serial crystallographic experiments have been done, for example at XFEL facilities, using an injector to deliver a string of crystal slurry LCP into the beamline. The question arose of how much the triggering-

light would travel to adjacent areas of the LCP slurry through, for instance, refraction. This would create a scenario of uncontrolled triggering of areas depending not only on the distance from the laser impact area, but also in relation to when those areas would cross the X-ray beam, and even depending on the depth of the crystal in the slurry compared to where the laser hit.

To access this potential problem, we devised an experiment for the purpose. We used a laser with emission at 355 nm, a fast camera and filter with a wavelength cutoff of 400 nm or lower, LCP slurry loaded with a fluorophore (6-aminoquinoline), for which the maximum absorption was at 350 nm and emission at 550 nm. The experiment consisted of illuminating the the fluorescence doped LCP slurry with

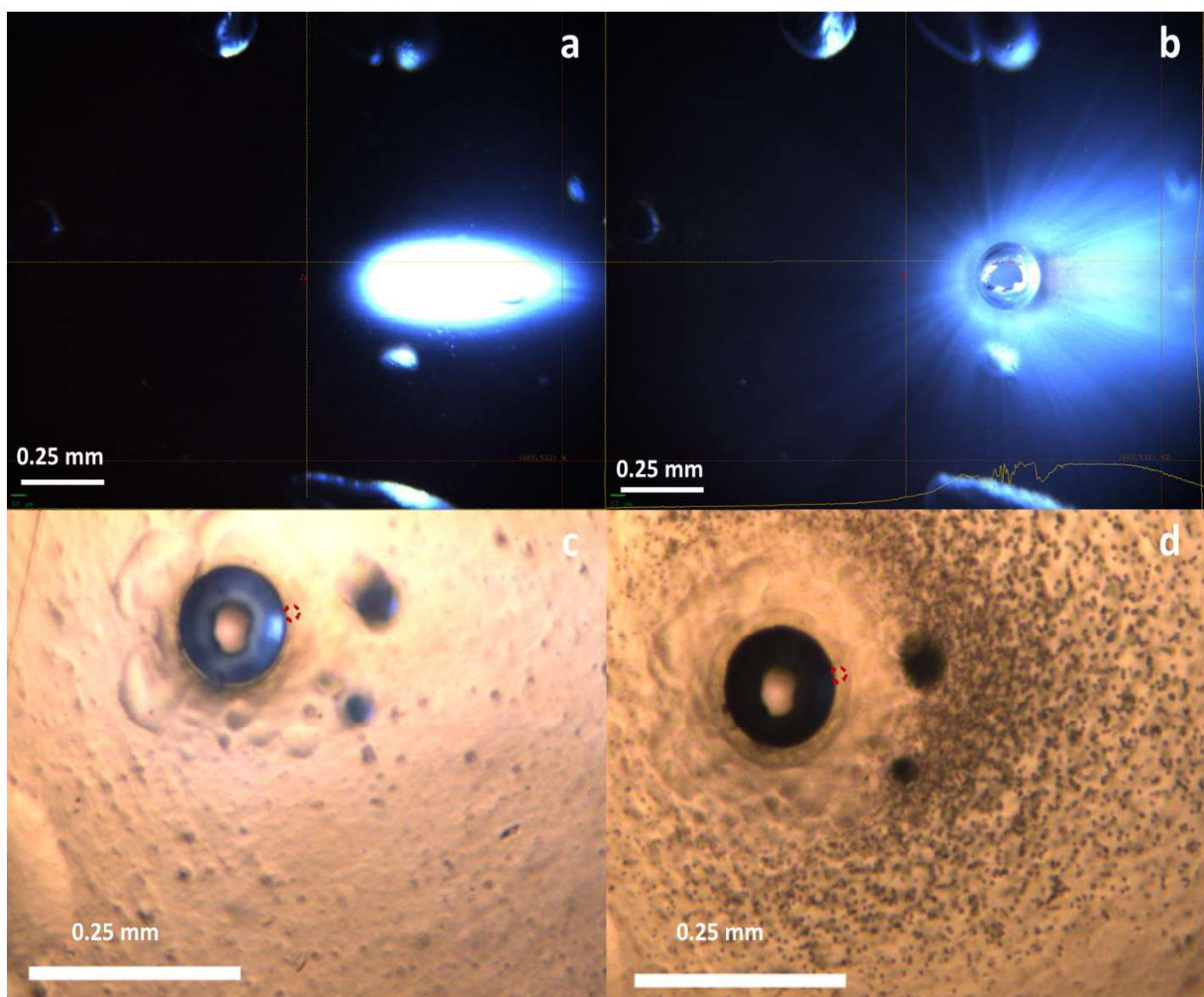


Figure 3-23: LCP mylar sandwich soaked with fluorophore and targeted with a laser.

a) first image with laser on, showing already fluorescence away from the laser target area of up to 1 mm b) after a few seconds exposure, a hole appears caused by the laser, and the fluorescence further spreads along the LCP. c) last frame with laser on d) moments after laser off, showing phase separation spherulites.

the laser and then quantifying the distance away from the laser spot that fluorescence was detected. Because of the filter and fluorescence, the experiment could be done in complete darkness and all light detected was the product of fluorescence.

3.4.1. LCP in solid targets

In these tests I mounted the LCP in a mylar sandwich using the chip holder. Initially, when the laser was hitting the LCP soaked in the fluorophore, fluorescence was observed as far as 1 mm away from the laser spot. This value later more than doubled, when after a few moments with the laser on, a hole in the mylar started to appear and grow. With the hole formation the fluorescence glow further spread. Although the hole in the mylar is probably not representative of a typical pump-probe experiment (the laser pump was triggered on seconds timescale), this simple experiment clearly showed that the triggering laser light spreads significantly around the target spot, which would make laser triggered experiments difficult without extra measures to prevent this effect.

Interestingly, in one of the control experiments (*i.e.* no fluorophore, and hutch light on to see the chip), we not only observed the mylar hole formations, but also a strange phenomenon. After the laser was turned off, hundreds of small darker spheres would appear surrounding the spot. If the laser was switched back on these would disappear, reforming when the laser was off. This is most probably the result of heat and humidity changes. The laser heats the surrounding area; and the hole that forms in the mylar will certainly result in changing humidity conditions immediately around it. Each time the laser hits the sample, heating the LCP causing a change in lipid phase. This, in combination with humidity changes, would likely result in different lipid phases surrounding the laser illumination spot, *i.e.*, phase separation. Although curious, this phenomenon might not have a significant impact on time resolved experiments, except as a reminder that when dealing with delicate phases (LCP) or crystals, it is important to have in mind that the laser might significantly change them depending on the setup, laser power and pulse length.

3.4.2. DgKa LCP crystal slurry on solid targets

We used a crystal slurry with DgKa crystals provided by our collaborators Coillin Boland and Martin Caffrey. This protein is a potential target for time-resolved crystallographic experiments due to previously acquired structures. These structures [40] show DgKa in complex with two monoacylglycerol molecules and an ATP like inactive molecule (ACP), all of these close to, and at the active site. By using a photo-caged ATP molecule, as long as this same crystallization solution and conformation can be maintained, it should be a straight-forward time-resolved experiment. It was because of this that we collaborated to work towards a DgKa time-resolved experiment.

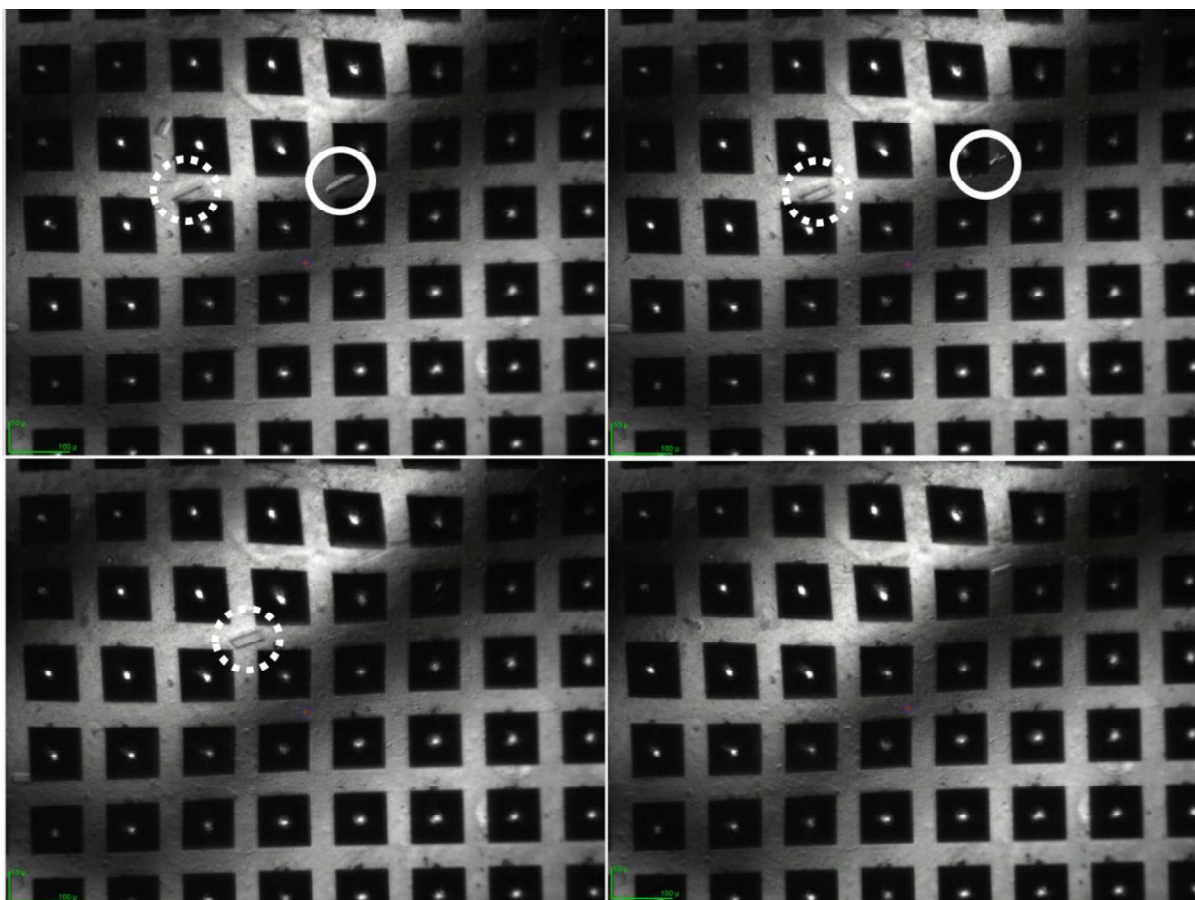


Figure 3-24: Frames of a video separated in seconds showing crystal movement on solid targets.

The frames show significant movement of crystals over the solid target chip. Some crystals are marked for assisting in tracking them over frames. Crystals were DgkA in LCP, kindly provided by Martin Caffrey and Coilin Boland.

After creating the typical mylar sandwich containing the LCP slurry smeared onto a silicon chip and mounting it on the beamline, we realized the crystals were moving across the chip. The image below shows successive pictures at 10 second intervals and the last one after one minute. It was clear this movement was very significant and data collection for such experiments would not be possible.

We tried applying more vacuum for longer to the LCP during chip loading, which did not change the results. After a few attempts we pressed on the closed mylar sandwich, pressing the crystals and LCP down into the chip features. Despite the mechanical stress caused to the crystals this seemed to be the best approach to get crystals to be stably positioned on the chips.

4. Discussion

4.1. LeuT

The aim of this work was to move towards a time resolved X-ray crystallography experiment imaging the conformational changes associated with LeuT transport. The trigger strategy was to use a mutant variant (3M) which consisted of two carefully chosen cysteine mutation positions (see Stockner's work [88]). After crosslinking these cysteines, the disulfide bridge would lock the protein in the inward conformation, and after its cleavage via reducing agents, X-ray or laser pulses, structural relaxation and ligand binding would be allowed.

For this project, high protein production was required. If we assume that 2 mg of crystallized protein are roughly needed for one dataset, and that flask production yields 1-2 mg of protein per 10 L, we can see that numerous batches would be needed for the crystallographic experiment. Large scale protein production would ideally be used, increasing the protein yield and time efficiency by 67-80% (see Mhp1 large scale protein production section for details). Sadly, various difficult events and protocol troubleshooting problems coincided, greatly inhibiting protein production for a very long time. Nonetheless, finally the protein production was solved, and it showed how seemingly small details (like flask air to medium ratio) can have a big impact on final total protein yields.

To assess whether the mutant LeuT 3M was indeed cross-linked so to enable the triggering strategy, I used CuPh to cross-link cysteines [65]. To verify the total free thiol groups in solution, I used the Ellman's reaction [64]. After a few tests, I realized that directly after purification, only 3% of the cysteines of the 3M LeuT were reacting with DTNB. To validate the hypothesis that this was due to a major fraction of the sample being already cross-linked after purification, I reduced the protein and disulfide bridges with TCEP. Since TCEP itself reduces the Ellman's reagent, producing very high false positives, I injected the reduced protein onto a desalting column, separating the TCEP and protein fractions. After this procedure, 98 % of the cysteines were now free to react with the Ellman's reagent. After doing the same procedure with wild type LeuT and a blank sample (only buffer), I could confirm that the protein was indeed crosslinked after purification.

I next needed to see if the mutant LeuT was binding to leucine. From Mhp1 studies [87] I knew that binding could be followed by fluorescence decay after ligands were bound, interacting with a tryptophan at the binding pocket. After observing the LeuT structure, I questioned whether this same experiment would also be possible for the surrounding aromatic amino acids (Figure 3-4), and therefore carried out a

few fluorescence assays. The tests showed no reaction to ligand, and so I explored CD as an alternative. It is known that since LeuT transports leucine, helices 1 and 6 change conformation, possibly unwinding portions of themselves (Figure 3-5). Because of this, I hypothesized that the resulting change in secondary structure would be enough to show a signal difference by CD. Using Leucine titration assays I could verify that indeed binding was detected. The 3M LeuT variant showed a lower binding activity when following the 221 nm CD spectrum peak. On the other hand, it showed a more pronounced change in the total α -helical content than the WT counterpart. Additionally, we found that both WT and 3M α -helix content bounced in between adjacent conformations, which was then corroborated by secondary structure analysis of predicted CD spectrums of published crystal structures and also modeled conformations.

Interestingly, after TCEP treatment, both 3M and WT changed their secondary structure. This was contrary to what I expected to observe; it was expected only the 3M would change as TCEP would be reducing the crosslink of LeuT. Additionally, I observed a major change in secondary structure with a 24 hour incubation time at 4 °C, suggesting that the incubation time for TCEP to reduce disulfide bonds and protein conformation change could be longer than five minutes. An additional study is necessary to pinpoint how long TCEP should be incubated with LeuT, and compare with the WT to assess if indeed the protein was being released from a cross-link induced conformation lock. Useful following experiments would also consider testing finer ligand titration steps, tuning salt concentration (with attention to sodium chloride to ligand binding activity dependence), other reducing agents, or testing laser induced disulfide breakage followed by CD measurements, allowing us to assess disulfide bond breaking efficiency for the various methods.

Finally, I obtained LeuT crystals of the mutant crosslinked form. The crystals were thin long 2D plates, and data were collected at PETRA III P14. Unfortunately, this data was low in quality and a structure could not be solved. Besides other problems, the major issues seemed to be too high mosaicity and major diffraction anisotropy. This was coupled with the difficulty of crystal harvesting, where the thin plate crystals bend at the loops due to surface tension. This bending of the crystals is expected to have pressured the unit cells at different crystal length positions, with various tension strength along the crystal (due to the curvature nature), highly contributing to mosaicity of the dataset. Sadly, after protein production, a world-wide pandemic, and other factors (described at the corresponding results section), meant that no more protein was available for crystallization before the end of this project. Naturally, thoughts on the project from here-on are highly speculative, but it would be interesting to observe if the crystal structure of the cross-linked mutant LeuT 3M was indeed in a closed conformation, whether the

probing X-rays were breaking the disulfide bridge, laser pulses, or reducing agents, and if the protein still was capable of binding leucine in the same site – further clarifying the binding CD results.

4.2. Mhp1

Mhp1 protein was produced on a large scale as an effort to assist crystallography, SAXS, and other experiments by Drs. Kokkinidou and Gao. This was done with large scale fermenters capable of producing 30 L and 100 L worth of transformed cells for protein production with each batch.

Multiple variables are at play when comparing small scale versus large scale protein production. However, the assured airflow control associated with the fermenters could be having a major impact on the amount of protein produced when compared to the flasks. It is known that airflow, oxygen levels, and air to growth media ratios play an important role in cell growth. Factors such as airflow become even more important for proteins that inhibit cell function or become toxic when produced in large numbers (which is commonly the case for membrane proteins since they can impair membrane functions). In this set-up, the cells are already experiencing high levels of stress, which can indeed be a limiting factor for protein production. If airflow requirements are then also being neglected, the yield becomes even more compromised.

When multiple batches of flask growths must be made (i.e. final volume of 30 L), it is arguably best to use large scale protein production when available. The variation between fermenter growths is much smaller than between each flask. This is because the fermenter machines are built with parameters controls in mind, whereas flasks are not. In a fermenter, only one transformation, inoculation, and induction event occur, whereas in flasks there is one transformation per batch, and one inoculation and induction per flask. Furthermore, considering the amount of waste, material used, objects requiring sterilization, and overall manual work, the production of 30 L in flasks (30 x 1 L flasks, or even 60 x 0.5 L flasks) is much less demanding when using fermenters. In addition, more time is needed to prepare the flasks and media for the growths. Most crucially, however, the production per week using fermenters is significantly higher. Using the equipment available in the labs I worked at (which I do not consider a bottleneck), it would take on average 4 days to grow 10 L, so 12 days for a total of 30 L. The same 30 L growth using fermenters took us 4-5 days, this considering there were two workers where one was tutoring another how to operate the fermenters. I estimate that with two experienced workers, a 30 L fermenter could be done every 3-4 days with assurance, and possibly 2.5-3 days with more optimization of the in-house procedure. This means that whereas it would take one worker 12 days to produce 30 L, 2 workers could produce the same 30 L every 2.5-4 days with regularity, making an 80-67 % improvement

in cell yield output. This improvement can also be multiplied by the improvement in yield of cells per liter of growth, giving a final value of 2.1 (1.67x1.25) to 2.8 (1.80x1.56) (or 210-280%) times higher yields for large scale protein production.

There are many reasons why this practice is not more widely used for academic projects, but some of them are also due to the disadvantages of the approach itself. This method requires a great deal of equipment and resources including the fermenters themselves, the associated software, a continuous flow centrifuge, steam generator, considerable space and potentially a dedicated room with specific safety requirements, regular maintenance of all of the mentioned equipment, and trained workers. On top of these requirements, there is a significant initial financial investment requirement to get this started and running in a lab. All of these costs combined must be weighed against the needs the lab has for cell growth which usually does not make it feasible.

As a final analysis, for cases where constant large amount of cell growths is required, and where protein production becomes the bottleneck, large scale protein production is an extremely attractive option. For labs where this is the case, starting large scale protein production can greatly improve work flow and output. For other labs where the necessary financial resources are not available, or production requirements do not justify the investment, usage large-scale protein production companies and facilities can still be a very significant improvement.

SAXS and crystallography studies done by Dr. Yunyun Gao and Dr. Maria Kokkinidou were done on Mhp1 produced and/or purified from membrane preparations by me. Joining as a collaborator, I attended the later stages of the projects, and summary conclusions and discussion regarding the relation between Mhp1's oligomer state and crystallizability are presented here. The project showed clearly that the oligomer state of the protein, induced by buffer differences, indeed plays a role in crystallization. Even though it is debatable whether the observation that the protein buffer impacts crystallization is obvious, it is interesting to pinpoint this consequence at least in part, to the protein oligomer state. SEC protein peaks, SAXS data and modeling, and crystallization trials showed that specific mutations, buffers and detergents (NM) induced outward conformation preference and trimerization, which enabled crystallization. On the other hand, conditions that induced the inward conformation also shifted the oligomerization towards dimers and monomers, which led to poor crystallization results. The complete scenario is not yet fully understood, and sometimes SAXS and SEC data appear to disagree, although discrepancies could also be accounted for by micelle size differences in protein-detergent complexes.

4.3. Detergent studies and LCP

Detergent studies in collaboration with Elham Vahdatahar and Monika Spano were able to get rough detergent diffusion rates across dialysis membranes, and these results and studies could be further exploited in their group. I also found that a very significant detergent absorption was occurring in some experiments, leading me to the conclusion that any detergent sensitive experiment or methodology is well advised to use cellulose dialysis membranes as these show the least absorption.

I also tested ways to deliver LCP in solid targets to synchrotron X-ray sources, finding that the normal non-LCP sample implication using a mild vacuum was not enough. One way of solving the issue was smearing the LCP layer gently against the solid target. By the time of writing, I have been happily informed this approach has already been successfully employed for other samples at the P14-2 beamline. I also saw that the typical laser power used for pump-probe experiments will quickly damage the mylar sandwich, and destabilize surrounding LCP. Whether this phenomenon is significant enough to alter crystallographic time-resolved studies using solid targets and LCP remains to be confirmed.

5. References and index

- [1] S. Hasnain, "Impact and influence of crystallography across the sciences," *IUCrJ*, vol. 3. International Union of Crystallography, pp. 389–390, 2016. doi: 10.1107/S2052252516017012.
- [2] S. Kulshrestha, P. Tyagi, V. Sindhi, and K. S. Yadavilli, "Invertase and its applications – A brief review," *J Pharm Res*, vol. 7, no. 9, pp. 792–797, Sep. 2013, doi: 10.1016/j.jopr.2013.07.014.
- [3] J.-M. Beaulieu, S. Espinoza, and R. R. Gainetdinov, "Dopamine receptors – IUPHAR Review 13," 2014, doi: 10.1111/bph.12445/abstract.
- [4] M. Luckey, *Membrane Structural Biology*. Cambridge: Cambridge University Press, 2008. doi: 10.1017/CBO9780511811098.
- [5] E. P. Carpenter, K. Beis, A. D. Cameron, and S. Iwata, "Overcoming the challenges of membrane protein crystallography," *Curr Opin Struct Biol*, vol. 18, no. 5, Oct. 2008, doi: 10.1016/j.sbi.2008.07.001.
- [6] Stephen White, "mpstruc - Membrane Proteins of Known 3D Structure," <https://blanco.biomol.uci.edu/mpstruc/>.
- [7] "PDB Statistics - Growth in Number of Protein Sequences in Released PDB Structures (Cumulative) at identity 95%," <https://www.rcsb.org/stats/growth/sequence/cluster-ids-95>.
- [8] Z. L. Johnson *et al.*, "Structure and mechanism of the bacterial transporter LeuT," *Nature*, vol. 49, no. 6, pp. 1655–1661, 2012, doi: 10.1021/bi901738k.
- [9] N. Chaffey, "Raven biology of plants, 8th edn," *Ann Bot*, vol. 113, no. 7, pp. vii–vii, Jun. 2014, doi: 10.1093/aob/mcu090.
- [10] D. A. Bryant and N.-U. Frigaard, "Prokaryotic photosynthesis and phototrophy illuminated," *Trends Microbiol*, vol. 14, no. 11, pp. 488–496, Nov. 2006, doi: 10.1016/j.tim.2006.09.001.
- [11] H. Tamagawa, M. Funatani, and K. Ikeda, "Ling's Adsorption Theory as a Mechanism of Membrane Potential Generation Observed in Both Living and Nonliving Systems," *Membranes (Basel)*, vol. 6, no. 1, p. 11, Jan. 2016, doi: 10.3390/membranes6010011.

- [12] D. C. Gadsby, F. Bezanilla, R. F. Rakowski, P. De Weer, and M. Holmgren, "The dynamic relationships between the three events that release individual Na⁺ ions from the Na⁺/K⁺-ATPase," *Nat Commun*, vol. 3, no. 1, p. 669, Jan. 2012, doi: 10.1038/ncomms1673.
- [13] K. Skene, "Life's a Gas: A Thermodynamic Theory of Biological Evolution," *Entropy*, vol. 17, no. 12, pp. 5522–5548, Jul. 2015, doi: 10.3390/e17085522.
- [14] F. Gubellini *et al.*, "Physiological Response to Membrane Protein Overexpression in *E. coli*," *Molecular & Cellular Proteomics*, vol. 10, no. 10, p. M111.007930, Oct. 2011, doi: 10.1074/mcp.M111.007930.
- [15] D. Drew, L. Fröderberg, L. Baars, and J.-W. L. de Gier, "Assembly and overexpression of membrane proteins in *Escherichia coli*," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1610, no. 1, pp. 3–10, Feb. 2003, doi: 10.1016/S0005-2736(02)00707-1.
- [16] S. Wagner, M. L. Bader, D. Drew, and J.-W. de Gier, "Rationalizing membrane protein overexpression," *Trends Biotechnol*, vol. 24, no. 8, pp. 364–371, Aug. 2006, doi: 10.1016/j.tibtech.2006.06.008.
- [17] A. M. Seddon, P. Curnow, and P. J. Booth, "Membrane proteins, lipids and detergents: not just a soap opera," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1666, no. 1–2, pp. 105–117, Nov. 2004, doi: 10.1016/j.bbamem.2004.04.011.
- [18] G. G. Privé, "Detergents for the stabilization and crystallization of membrane proteins," *Methods*, vol. 41, no. 4, pp. 388–397, Apr. 2007, doi: 10.1016/j.ymeth.2007.01.007.
- [19] M. D. Marger and M. H. Saier, "A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport," *Trends Biochem Sci*, vol. 18, no. 1, pp. 13–20, Jan. 1993, doi: 10.1016/0968-0004(93)90081-W.
- [20] Z. Zhou, J. Zhen, N. K. Karpowich, C. J. Law, M. E. A. Reith, and D. N. Wang, "Antidepressant specificity of serotonin transporter suggested by three LeuT-SSRI structures," *Nat Struct Mol Biol*, vol. 16, no. 6, pp. 652–657, 2009, doi: 10.1038/nsmb.1602.
- [21] H. Wang and E. Gouaux, "Substrate binds in the S1 site of the F253A mutant of LeuT, a neurotransmitter sodium symporter homologue," *EMBO Rep*, vol. 13, no. 9, pp. 861–866, 2012, doi: 10.1038/embor.2012.110.

- [22] C. L. Piscitelli and E. Gouaux, "Insights into transport mechanism from LeuT engineered to transport tryptophan," *EMBO Journal*, vol. 31, no. 1, pp. 228–235, 2012, doi: 10.1038/emboj.2011.353.
- [23] S. Weyand *et al.*, "The alternating access mechanism of transport as observed in the sodium-hydantoin transporter Mhp1," *J Synchrotron Radiat*, vol. 18, no. 1, pp. 20–23, Jan. 2011, doi: 10.1107/S0909049510032449.
- [24] G. Diallinas, "An Almost-Complete Movie," *Science (1979)*, vol. 322, no. 5908, pp. 1644–1645, Dec. 2008, doi: 10.1126/science.1168107.
- [25] T. Shimamura *et al.*, "Molecular Basis of Alternating Access Membrane Transport by the Sodium-Hydantoin Transporter Mhp1," *Science (1979)*, vol. 328, no. 5977, pp. 470–473, Apr. 2010, doi: 10.1126/science.1186303.
- [26] S. Weyand *et al.*, "Structure and Molecular Mechanism of a Nucleobase–Cation–Symport-1 Family Transporter," *Science (1979)*, vol. 322, no. 5902, pp. 709–713, Oct. 2008, doi: 10.1126/science.1164440.
- [27] S. Suzuki and P. J. F. Henderson, "The Hydantoin Transport Protein from *Microbacterium liquefaciens*," *J Bacteriol*, vol. 188, no. 9, pp. 3329–3336, May 2006, doi: 10.1128/JB.188.9.3329-3336.2006.
- [28] W. D. Van Horn and C. R. Sanders, "Prokaryotic diacylglycerol kinase and undecaprenol kinase," *Annual Review of Biophysics*, vol. 41, no. 1, pp. 81–101, Jun. 09, 2012. doi: 10.1146/annurev-biophys-050511-102330.
- [29] F. W. Lau, X. Chen, and J. U. Bowie, "Active sites of diacylglycerol kinase from *Escherichia coli* are shared between subunits," *Biochemistry*, vol. 38, no. 17, pp. 5521–5527, Apr. 1999, doi: 10.1021/bi982763t.
- [30] P. Badola and C. R. Sanders, "*Escherichia coli* diacylglycerol kinase is an evolutionarily optimized membrane enzyme and catalyzes direct phosphoryl transfer," *Journal of Biological Chemistry*, vol. 272, no. 39, pp. 24176–24182, Sep. 1997, doi: 10.1074/jbc.272.39.24176.
- [31] D. Mi, H. J. Kim, A. Hadziselimovic, and C. R. Sanders, "Irreversible misfolding of diacylglycerol kinase is independent of aggregation and occurs prior to trimerization and membrane association," *Biochemistry*, vol. 45, no. 33, pp. 10072–10084, Aug. 2006, doi: 10.1021/bi060887x.

- [32] D. Li and M. Caffrey, "Renaturing membrane proteins in the lipid cubic phase, a nanoporous membrane mimetic," *Sci Rep*, vol. 4, Jul. 2014, doi: 10.1038/srep05806.
- [33] J. K. Nagy, F. W. Lau, J. U. Bowie, and C. R. Sanders, "Mapping the oligomeric interface of diacylglycerol kinase by engineered thiol cross-linking: Homologous sites in the transmembrane domain," *Biochemistry*, vol. 39, no. 14, pp. 4154–4164, Apr. 2000, doi: 10.1021/bi991781n.
- [34] R. L. Smith, J. F. O'toole, M. E. Maguire, and C. R. Sanders, "Membrane Topology of Escherichia coli Diacylglycerol Kinase," 1994. [Online]. Available: <https://journals.asm.org/journal/jb>
- [35] "Exploring the allowed sequence space of a membrane protein".
- [36] Y. Zhou and J. U. Bowie, "Building a thermostable membrane protein," *Journal of Biological Chemistry*, vol. 275, no. 10, pp. 6975–6979, Mar. 2000, doi: 10.1074/jbc.275.10.6975.
- [37] F. W. Lau, S. Nauli, Y. Zhou, and J. U. Bowie, "Changing Single Side-chains can Greatly Enhance the Resistance of a Membrane Protein to Irreversible Inactivation." [Online]. Available: <http://www.idealibrary.com>
- [38] W. D. Van Horn *et al.*, "Solution nuclear magnetic resonance structure of membrane-integral diacylglycerol kinase," *Science (1979)*, vol. 324, no. 5935, pp. 1726–1729, Jun. 2009, doi: 10.1126/science.1171716.
- [39] D. Li *et al.*, "Crystal structure of the integral membrane diacylglycerol kinase," *Nature*, vol. 497, no. 7450, pp. 521–524, 2013, doi: 10.1038/nature12179.
- [40] D. Li *et al.*, "Ternary structure reveals mechanism of a membrane diacylglycerol kinase," *Nat Commun*, vol. 6, 2015, doi: 10.1038/ncomms10140.
- [41] J. de Mos, A. Jakob, J. Becker-Baldus, A. Heckel, and C. Glaubitz, "Light-Induced Uncaging for Time-Resolved Observations of Biochemical Reactions by MAS NMR Spectroscopy," *Chemistry - A European Journal*, vol. 26, no. 30, pp. 6789–6792, May 2020, doi: 10.1002/chem.202000770.
- [42] E. C. Schulz, B. A. Yorke, A. R. Pearson, and P. Mehrabi, "Best practices for time-resolved serial synchrotron crystallography," *Acta Crystallogr D Struct Biol*, vol. 78, pp. 14–29, Jan. 2022, doi: 10.1107/S2059798321011621.
- [43] P. Mehrabi *et al.*, "The HARE chip for efficient time-resolved serial synchrotron crystallography," *J Synchrotron Radiat*, vol. 27, pp. 360–370, Mar. 2020, doi: 10.1107/S1600577520000685.

- [44] B. Rupp, *Biomolecular Crystallography*. Garland Science, 2009. doi: 10.1201/9780429258756.
- [45] D. G. Rhodes, "Crystallography made crystal clear: A guide for users of macromolecular models (3rd Ed.)," *Biochemistry and Molecular Biology Education*, vol. 35, no. 5, pp. 387–388, 2007, doi: 10.1002/bmb.89.
- [46] Matthews, "Solvent content of protein crystals.," *J. Mol. Biol*, 1968.
- [47] M. C. Wiener, "A pedestrian guide to membrane protein crystallization," *Methods*, vol. 34, no. 3, pp. 364–372, 2004, doi: 10.1016/j.ymeth.2004.03.025.
- [48] A. Meyer *et al.*, "Single-drop optimization of protein crystallization," *Acta Crystallogr Sect F Struct Biol Cryst Commun*, vol. 68, no. 8, pp. 994–998, 2012, doi: 10.1107/S1744309112024074.
- [49] A. Wlodawer, W. Minor, Z. Dauter, and M. Jaskolski, "Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination," *FEBS Journal*, vol. 280, no. 22, pp. 5705–5736, 2013, doi: 10.1111/febs.12495.
- [50] C. Kupitz *et al.*, "Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser," *Nature*, vol. 513, no. 7517. Nature Publishing Group, pp. 261–265, 2014. doi: 10.1038/nature13453.
- [51] A. Aquila *et al.*, "Ultrafast technology; (170.7440) X-ray imaging; (140.3450) Laser-induced chemistry; (140.7090) Ultrafast lasers; (170.0170) Medical optics and biotechnology," 2012. [Online]. Available: <http://arxiv.org/abs/1105.2104>
- [52] H. Li *et al.*, "Capturing structural changes of the S1to S2transition of photosystem II using time-resolved serial femtosecond crystallography," *IUCrJ*, vol. 8, pp. 431–443, May 2021, doi: 10.1107/S2052252521002177.
- [53] A. Aquila *et al.*, "Ultrafast technology; (170.7440) X-ray imaging; (140.3450) Laser-induced chemistry; (140.7090) Ultrafast lasers; (170.0170) Medical optics and biotechnology," 2012. [Online]. Available: <http://arxiv.org/abs/1105.2104>
- [54] R. Dods *et al.*, "Ultrafast structural changes within a photosynthetic reaction centre," *Nature*, vol. 589, no. 7841, pp. 310–314, Jan. 2021, doi: 10.1038/s41586-020-3000-7.
- [55] A. Shimada *et al.*, "A nanosecond time-resolved XFEL analysis of structural changes associated with CO release from cytochrome c oxidase," 2017. [Online]. Available: <https://www.science.org>








- [56] T. Lazarides *et al.*, "Ligand-field excited states of hexacyanochromate and hexacyanocobaltate as sensitizers for near-infrared luminescence from Nd(III) and Yb(III) in cyanide-bridged d-f assemblies," *Photochemical and Photobiological Sciences*, vol. 6, no. 11, pp. 1152–1157, 2007, doi: 10.1039/b708683k.
- [57] T. Tsukamoto, S. Yoshizawa, T. Kikukawa, M. Demura, and Y. Sudo, "Implications for the Light-Driven Chloride Ion Transport Mechanism of Nonlabens marinus Rhodopsin 3 by Its Photochemical Characteristics," *Journal of Physical Chemistry B*, vol. 121, no. 9, pp. 2027–2038, Mar. 2017, doi: 10.1021/acs.jpcc.6b11101.
- [58] T. Kouyama, H. Kawaguchi, T. Nakanishi, H. Kubo, and M. Murakami, "Crystal Structures of the L1, L2, N, and O States of pharaonis Halorhodopsin," *Biophys J*, vol. 108, no. 11, pp. 2680–2690, Jun. 2015, doi: 10.1016/j.bpj.2015.04.027.
- [59] T. Hosaka *et al.*, "Conformational alterations in unidirectional ion transport of a light-driven chloride pump revealed using X-ray free electron lasers," 2022, doi: 10.1073/pnas.2117433119/-/DCSupplemental.
- [60] J. N. Atherton, "Dynamics of Solutions and Fluid Mixtures by NMR Edited by Jean-Jacques Delpuech (Université Henri Poincaré). Wiley: New York. 1995. xi + 587 pp. \$79.95. ISBN 0-471-95411-X.," *J Am Chem Soc*, vol. 119, no. 30, pp. 7172–7172, Jul. 1997, doi: 10.1021/ja955294g.
- [61] N. Junius, E. Oksanen, M. Terrien, C. Berzin, J. L. Ferrer, and M. Budayova-Spano, "A crystallization apparatus for temperaturecontrolled flow-cell dialysis with real-time visualization," *J Appl Crystallogr*, vol. 49, pp. 806–813, 2016, doi: 10.1107/S1600576716004635.
- [62] M. Caffrey and C. Porter, "Crystallizing membrane proteins for structure determination using lipidic mesophases.," *J Vis Exp*, vol. 4, no. 45, pp. 706–731, 2010, doi: 10.1038/nprot.2009.31.
- [63] M. Caffrey, "A lipid's eye view of membrane protein crystallization in mesophases," *Curr Opin Struct Biol*, vol. 10, no. 4, pp. 486–497, 2000, doi: 10.1016/S0959-440X(00)00119-6.
- [64] A. G. Technology, "DTNB (Ellman's Reagent) (5,5-dithio-bis-(2-nitrobenzoic acid)," *Thermo Fisher Scientific*, pp. 2–5, 2011, [Online]. Available: <https://www.thermofisher.com/order/catalog/product/22582>



- [65] R. Valdés, U. Shinde, and S. M. Landfear, "Cysteine cross-linking defines the extracellular gate for the *Leishmania donovani* nucleoside transporter 1.1 (LdNT1.1)," *Journal of Biological Chemistry*, vol. 287, no. 53, pp. 44036–44045, 2012, doi: 10.1074/jbc.M112.414433.
- [66] A. Yamashita, S. K. Singh, T. Kawate, Y. Jin, and E. Gouaux, "Crystal structure of a bacterial homologue of Na⁺/Cl⁻-dependent neurotransmitter transporters," *Nature*, vol. 437, no. 7056, pp. 215–223, 2005, doi: 10.1038/nature03978.
- [67] H. Krishnamurthy and E. Gouaux, "X-ray structures of LeuT in substrate-free outward-open and apo inward-open states," *Nature*, vol. 481, no. 7382, pp. 469–474, 2012, doi: 10.1038/nature10737.
- [68] M. Quick, A.-M. L. Winther, L. Shi, P. Nissen, H. Weinstein, and J. A. Javitch, "Binding of an octylglucoside detergent molecule in the second substrate (S2) site of LeuT establishes an inhibitor-bound conformation," *Proceedings of the National Academy of Sciences*, vol. 106, no. 14, pp. 5563–5568, 2009, doi: 10.1073/pnas.0811322106.
- [69] Hampton Research, "Crystallization Screen HR2-110," https://hamptonresearch.com/uploads/support_materials/HR2-110_Binder.pdf.
- [70] G. Winter *et al.*, "DIALS as a toolkit," *Protein Science*, vol. 31, no. 1, pp. 232–250, Jan. 2022, doi: 10.1002/pro.4224.
- [71] W. Kabsch, "XDS," *Acta Crystallogr D Biol Crystallogr*, vol. 66, no. 2, pp. 125–132, Feb. 2010, doi: 10.1107/S0907444909047337.
- [72] T. G. G. Battye, L. Kontogiannis, O. Johnson, H. R. Powell, and A. G. W. Leslie, "iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM," *Acta Crystallogr D Biol Crystallogr*, vol. 67, no. 4, pp. 271–281, Apr. 2011, doi: 10.1107/S0907444910048675.
- [73] R. J. Gildea *et al.*, "Xia2.multiplex: A multi-crystal data-Analysis pipeline," *Acta Crystallogr D Struct Biol*, vol. 78, pp. 752–769, Jun. 2022, doi: 10.1107/S2059798322004399.
- [74] W. Schaffner and C. Weissmann, "A rapid, sensitive, and specific method for the determination of protein in dilute solution," *Anal Biochem*, vol. 56, no. 2, pp. 502–514, Dec. 1973, doi: 10.1016/0003-2697(73)90217-0.

- [75] G. N. Murshudov *et al.*, "REFMAC5 for the refinement of macromolecular crystal structures," *Acta Crystallogr D Biol Crystallogr*, vol. 67, no. 4, pp. 355–367, Apr. 2011, doi: 10.1107/S0907444911001314.
- [76] H. M. Ginn, "Vagabond: Bond-based parametrization reduces overfitting for refinement of proteins," *Acta Crystallogr D Struct Biol*, vol. 77, pp. 424–437, Apr. 2021, doi: 10.1107/S2059798321000826.
- [77] B. L. P. Emsley, W. G. S. And, and K. Cowtan, "Features and development of Coot research papers," *Acta Crystallogr*, pp. 486–501, 2010, doi: 10.1107/S0907444910007493.
- [78] C. J. Williams *et al.*, "MolProbity: More and better reference data for improved all-atom structure validation," *Protein Science*, vol. 27, no. 1, pp. 293–315, Jan. 2018, doi: 10.1002/pro.3330.
- [79] M. D. Winn *et al.*, "Overview of the CCP 4 suite and current developments," *Acta Crystallogr D Biol Crystallogr*, vol. 67, no. 4, pp. 235–242, Apr. 2011, doi: 10.1107/S0907444910045749.
- [80] A. Urbani and T. Warne, "A colorimetric determination for glycosidic and bile salt-based detergents: Applications in membrane protein research," *Anal Biochem*, vol. 336, no. 1, pp. 117–124, 2005, doi: 10.1016/j.ab.2004.09.040.
- [81] M. Radchenko, R. Nie, and M. Lu, "Disulfide cross-linking of a Multidrug and toxic compound extrusion transporter impacts Multidrug efflux," *Journal of Biological Chemistry*, vol. 291, no. 18, pp. 9818–9826, 2016, doi: 10.1074/jbc.M116.715227.
- [82] D. S. Terry *et al.*, "A partially-open inward-facing intermediate conformation of LeuT is associated with Na⁺ release and substrate transport," *Nat Commun*, vol. 9, no. 1, Dec. 2018, doi: 10.1038/s41467-017-02202-y.
- [83] M. V. LeVine *et al.*, "The allosteric mechanism of substrate-specific transport in SLC6 is mediated by a volumetric sensor," *Proc Natl Acad Sci U S A*, vol. 116, no. 32, pp. 15947–15956, Aug. 2019, doi: 10.1073/pnas.1903020116.
- [84] C. B. Billesbølle *et al.*, "Substrate-induced unlocking of the inner gate determines the catalytic efficiency of a neurotransmitter: Sodium symporter," *Journal of Biological Chemistry*, vol. 290, no. 44, pp. 26725–26738, Oct. 2015, doi: 10.1074/jbc.M115.677658.

- [85] A. N. Calabrese *et al.*, "Topological Dissection of the Membrane Transport Protein Mhp1 Derived from Cysteine Accessibility and Mass Spectrometry," *Anal Chem*, vol. 89, no. 17, pp. 8844–8852, Sep. 2017, doi: 10.1021/acs.analchem.7b01310.
- [86] Maria C. Kokkinidou, "Structural and functional studies of the alternating access mechanism of secondary transporters," 2019.
- [87] A. Polyakova, "Applying complementary structural techniques to elucidate structure-function relationships of the bacterial Na⁺-hydantoin transporter Mhp1," 2015.
- [88] A. Sohail *et al.*, "The Environment Shapes the Inner Vestibule of LeuT," *PLoS Comput Biol*, vol. 12, no. 11, pp. 1–24, 2016, doi: 10.1371/journal.pcbi.1005197.

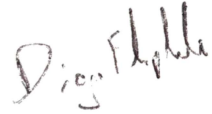
6. List of hazardous substances

Chemical	Hazard pictograms	Hazard statements	Precautionary statements
Imidazole	  	<p>Harmful if swallowed.</p> <p>Causes severe skin burns and eye damage.</p> <p>May damage the unborn child.</p>	<p>Do not breathe dust/fume/gas/mist/vapours/spray.</p> <p>Wear protective gloves/protective clothing/eye protection/face protection.</p> <p>IF SWALLOWED: rinse mouth. Do NOT induce vomiting.</p> <p>IF ON SKIN (or hair): Take off immediately all contaminated clothing. Rinse skin with water or shower.</p> <p>IF IN EYES: Rinse cautiously with water for several minutes. Remove contact lenses, if present and easy to do. Continue rinsing. Immediately call a POISON CENTER/doctor.</p> <p>Dispose of contents/container to industrial combustion plant.</p>
Phenol	   	<p>Highly flammable liquid and vapour.</p> <p>Toxic if swallowed, in contact with skin or if inhaled.</p> <p>Causes severe skin burns and eye damage.</p> <p>Suspected of causing genetic defects.</p> <p>Causes damage to organs.</p> <p>May cause damage to organs through prolonged or repeated exposure.</p>	<p>Keep away from heat, hot surfaces, sparks, open flames and other ignition sources. No smoking.</p> <p>Do not breathe dust/fume/gas/mist/vapours/spray.</p> <p>Wear protective gloves/protective clothing/eye protection/face protection.</p> <p>IF SWALLOWED: Immediately call a POISON CENTER/doctor.</p> <p>IF ON SKIN (or hair): Take off immediately all contaminated clothing. Rinse skin with water or shower.</p> <p>IF IN EYES: Rinse cautiously with water for several minutes. Remove contact lenses, if present and easy to do. Continue rinsing.</p> <p>In case of fire: Use sand, carbon dioxide or powder extinguisher to extinguish.</p> <p>Store in a well-ventilated place. Keep container tightly closed.</p> <p>Store in a well-ventilated place. Keep cool.</p>

<p>Nickel (contained in HisPur™ nickel resin columns)</p>		<p>Flammable liquid and vapor. Causes serious eye irritation. Causes skin irritation. May cause allergy or asthma symptoms or breathing difficulties if inhaled. May cause an allergic skin reaction. May cause cancer. May damage fertility or the unborn child. May cause damage to organs through prolonged or repeated exposure.</p>	<p><i>Nothing to note</i></p>
<p>Cobalt (contained in HisPur™ cobalt resin columns)</p>		<p>Flammable liquid and vapor. Causes serious eye irritation. Causes skin irritation. May cause allergy or asthma symptoms or breathing difficulties if inhaled. May cause an allergic skin reaction. May cause cancer. May damage fertility or the unborn child.</p>	<p><i>Nothing to note.</i></p>

7. Declaration on Oath

I hereby declare on oath that this doctoral dissertation is written independently and solely by my own based on the original work of my PhD and has not been used other than the acknowledged resources and aids.

A handwritten signature in black ink, appearing to read "Diogo Melo". The signature is written in a cursive, slightly slanted style.

Diogo Melo