Structural and biochemical characterisation of human integral membrane proteins regarded as potential therapeutic targets

Tomás da Costa Cabral Gomes Páscoa

Centre for Medicines Discovery

Nuffield Department of Medicine

and

Somerville College

University of Oxford



A thesis submitted for the partial fulfilment of the degree of Doctor of Philosophy

May 2022

Structural and biochemical characterisation of human integral membrane proteins regarded as potential therapeutic targets

Tomás da Costa Cabral Gomes Páscoa

Nuffield Department of Medicine and Somerville College, University of Oxford

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Abstract

Integral membrane proteins are responsible for a wide range of cellular functions and are important targets for the development of novel therapeutics. Therefore, obtaining structural and mechanistic insights into their function is of central importance for rational drug design. In this thesis, I focused on characterising three human integral membrane proteins using X-ray crystallography, cryo-EM, mass spectrometry and other biochemical methods.

MFSD10 is a member of the drug:H^+ antiporter family 1 (DHA1) of the Major Facilitator Superfamily of solute carriers and has been proposed to act as an efflux pump for organic anions, including several non-steroidal anti-inflammatory drugs (NSAIDs). Here, I present a crystal structure of human MFSD10 in an outward-facing partially occluded conformation, with a bound citrate ion from the crystallisation solution. The structure provides insights into the recognition of organic anions by MFSD10 and suggests a potential mechanism of proton coupling. An *in vitro* assay was employed to confirm the NSAID:H^+ antiporter function of MFSD10 and will constitute the starting point for additional functional assays to dissect the mechanism of this transporter.

The second focus of this thesis falls on elucidating the mechanisms of two enzyme families responsible for the acyl chain specificity in fatty acid elongation and sphingolipid biosynthesis, respectively. ELOVL elongases catalyse the first rate-limiting condensation step of the fatty acid elongation cycle in the endoplasmic reticulum (ER). The recently solved crystal structure of human ELOVL7 provided the first structural insights into this enzyme family. Here, I followed up on the structural data by characterising the mechanism of this enzyme using mass spectrometry. I demonstrated that catalysis by human ELOVL7, and likely other ELOVLs, proceeds via a ping-pong type mechanism involving an unusual covalent acyl-imidazole intermediate. Lastly, I focused on obtaining structural insights into the function of human ceramide synthases, which catalyse the N-acylation of sphingoid bases in the ER and determine the acyl chain composition of sphingolipids. I conducted a systematic screening of expression and purification conditions for two members of this family, identified suitable truncations for structural studies, obtained and validated nanobody binders, and optimised cryo-EM conditions. Together, these provide an advanced position for future structural studies to obtain valuable three-dimensional information on these important drug targets.

Acknowledgements

Firstly, I would like to thank my supervisors, Prof Liz Carpenter, Prof Christian Siebold, Prof Wyatt Yue and Dr Ashley Pike, for the guidance and support. Specifically, I would like to thank Liz for the fantastic opportunities to pursue several exciting projects in her group and Ashley for all the help, training and discussions over the years. I am very grateful to Wyatt and Christian, who stepped in as my supervisors after Liz left the University and ensured I had all the support I could ask for during the remainder of my DPhil. In particular, I am incredibly grateful to Wyatt for his help during the last few months.

Furthermore, I would like to thank every present and past member of the Integral Membrane Proteins group, especially Dr Ashley Pike, Dr Gamma Chi, Dr Agnese Baronina, Dr Chady Nasrallah, Dr Ali Flayhan, Dr Simon Bushell, Dr Didi He, Dr Karin Rodstrom, Dr Yin Dong, Dr David Sauer, David Speedman and Adam Evans. It was a pleasure to work with and learn from all of you. Likewise, I am grateful to the Biotechnology Group for maintaining the pipeline of the SGC/CMD, which has made the work described in this thesis possible. In particular, I would like to thank Shubhashish Mukhopadhyay for coordinating cell culture work, Leela Shrestha and Dr Cassandra Dix for cloning work, Tiago Moreira and Dr Rod Chalk for maintaining an excellent mass spectrometry facility and for sharing their advice and Tina Bohstedt for sharing her immense knowledge with me. I would also like to thank everyone within the Oxford Cellular Structural Biology DPhil programme for their continued support and the Wellcome Trust for funding this outstanding programme.

Finally, my deepest gratitude goes to my family and Lucrezia for always encouraging me to pursue my objectives and supporting me whenever I needed it. I could not have done it without you.

Authorship declaration

The work presented in this thesis was completed in the Structural Genomics Consortium / Centre for Medicines Discovery, Nuffield Department of Medicine, University of Oxford, between April 2018 and April 2022, under the supervision of Prof Liz Carpenter, Prof Christian Siebold, Prof Wyatt Yue and Dr Ashley Pike.

The initial MFSD10 crystals were obtained under the supervision of Dr Simon Bushell. The initial MFSD10 dataset was collected by Dr Ashley Pike. Soaking studies of MFSD10 were conducted with assistance from Dr Ashley Pike. The CerS6-c025 Titan Krios cryo-EM dataset was collected with assistance from Dr Gamma Chi. Other cryo-EM datasets were collected with assistance from Dr Gamma Chi and/or Dr Ashley Pike. Tryptic digestion of SDS-PAGE gel bands and LC-MS/MS analysis of the resulting peptides were conducted by Mr Tiago Moreira and Dr Rod Chalk of the Biotechnology group of the SGC/CMD. All cloning of CerS constructs and initial 3 mL purification tests were also conducted by the Biotechnology group of the SGC/CMD. Other contributions, including optimisation work carried out prior to the start of my DPhil project, are clearly stated in the text.

Publications resulting from this work

Nie L*, Pascoa TC*, Pike ACW*, Bushell SR, Quigley A, Ruda GF, Chu A, Cole V, Speedman D, Moreira T, Shrestha L, Mukhopadhyay SMM, Burgess-Brown NA, Love JD, Brennan PE, Carpenter EP. The structural basis of fatty acid elongation by the ELOVL elongases. Nature Structural & Molecular Biology. 2021;28(6):512-20. http://doi.org/10.1038/s41594-021-00605-6

* These authors contributed equally.

Some material from this publication is reproduced in modified form in Chapter 4 of this thesis. Contributions from other authors are clearly described in the text.

Pascoa TC, Pike ACW, Bushell SR, Quigley A, Mukhopadhyay SMM, Burgess-Brown NA, Carpenter EP. Major facilitator superfamily domain-containing protein 10 (MFSD10) A Target Enabling Package (TEP) (Version 1) [Data set]. 2021. http://doi.org/10.5281/zenodo.5126783

Some material from this publication is reproduced in Chapter 3 of this thesis. Contributions from other authors are clearly described in the text.

List of abbreviations

ABC	ATP-binding cassette
ACP	acyl carrier protein
ADA	N-(2-Acetamido)iminodiacetic acid
AMTF1	atypical MFS transporter family 1
ATP	adenosine triphosphate
BLI	bio-layer interferometry
C12E8	3,6,9,12,15,18,21,24-octaoxahexatriacontan-1-ol
C12E10	3,6,9,12,15,18,21,24,27,30-decaoxadotetracontan-1-ol
$CC_{1/2}$	correlation coefficient between random half-datasets
CDR	complementarity-determining region
CerS	ceramide synthase
CHS	cholesteryl hemisuccinate
CK2	casein kinase II
CMC	critical micelle concentration
CMD	centre for medicines discovery
CoA	coenzyme A
COX-1/2	cyclooxygenase 1/2
cryo-EM	electron cryomicroscopy
CTF	contrast transfer function
CYMAL-5	5-cyclohexyl-1-pentyl-β-D-maltoside
DDM	n-dodecyl-β-D-maltoside
DES	dihydroceramide Δ4-desaturase
DHA1	drug:H^+ antiporter family 1
DM	n-decyl-β-D-maltoside
DMNG	2,2-dioctylpropane-1,3-bis-β-D-maltopyranoside
DMSO	dimethyl sulfoxide
DSF	differential scanning fluorimetry
DTT	dithiothreitol
ECH	extracellular helix
EDTA	ethylenediaminetetraacetic acid

EGTA	ethylenebis(oxyethylenenitrilo)tetraacetic acid
ELISA	enzyme-linked immunosorbent assay
ELO	elongation of fatty acids protein
ELOVL	elongation of very long chain fatty acids protein
EndoH	endoglycosidase H
ER	endoplasmic reticulum
ESI	electrospray ionization
FAE1	*A. thaliana* fatty acid elongase 1
FB1	fumonisin B1
FC-12	n-dodecylphosphocholine
FSEC	fluorescence-detection size exclusion chromatography
Fuc	fucose
G3P	filamentous phage gene-3-protein
GalNac	N-acetylgalactosamine
GDN	glyco-diosgenin
GFP	green fluorescent protein
GlcNAc	N-acetylglucosamine
GnTI	N-acetylglucosaminyltransferase I
HACD	3-hydroxyacyl-CoA dehydratase
HEK293	human embryonic kidney 293 cells
HEPES	N-(2-hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid)
HexNAc	N-acetylhexosamine
HFD	high fat diet
HPLC	high performance liquid chromatography
IAM	iodoacetamide
ICH	intracellular helix
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
KAR	3-keto acyl-CoA reductase
KCS	3-keto acyl-CoA synthase
KDSR	3-ketodihydrosphingosine reductase
LB	Luria-Bertani broth
LC	liquid chromatography

LCP	lipidic cubic phase
LDAO	N,N-dimethyl-1-dodecanamine-N-oxide
LMNG	2,2-didecylpropane-1,3-bis-β-D-maltopyranoside
LMV	large multilamellar vesicles
LUVs	large unilamellar vesicles
MAG7.7	1-(7Z-tetradecenoyl)-rac-glycerol
MAG7.8	1-(7Z-pentadecenoyl)-rac-glycerol
MAG7.9	1-(7Z-hexadecenoyl)-rac-glycerol
MAG9.9	1-(9Z-octadecenoyl)-rac-glycerol
Man	mannose
MBP	maltose binding protein
MDR	multidrug resistance
Mff	mitochondrial fission factor
MFS	major facilitator superfamily
MFSD10	major facilitator superfamily domain-containing protein 10
MS	mass spectrometry
MW	molecular weight
NAFLD	nonalcoholic fatty liver disease
NASH	non-alcoholic steatohepatitis
NG	n-nonyl-β-D-glucopyranoside
Ni-NTA	nickel nitrilotriacetic acid
NSAID	non-steroidal anti-inflammatory drug
OD	optical density
OGNG	2,2-dihexylpropane-1,3-bis-β-D-glucopyranoside
PAH	p-aminohippuric acid
PBS	phosphate buffered saline
PCFT	proton-coupled folate transporter
PDB	protein data bank
PDC	protein-detergent complex
PEG	polyethylene glycol
PMF	proton motive force
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
POPG	1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol)

Pyranine	trisodium 8-hydroxypyrene-1,3,6-trisulfonate
RI	refractive index
RPM	revolutions per minute
S.E.M.	standard error of the mean
SCA	spinocerebellar ataxia
SCD1	stearoyl-coA desaturase 1
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
SEC-MALS	size-exclusion chromatography with multi angle light scattering
SGC	structural genomics consortium
SLC	solute carrier
SPT	serine palmitoyltransferase
STGD3	Stargardt-like macular dystrophy
TB	terrific broth
TCEP	tris(2-carboxyethyl)phosphine
TEA	tetraethylammonium
TER	trans-2-enoyl-CoA reductase
TES	tris-EDTA-sucrose buffer
TEV protease	tobacco etch virus nuclear-inclusion-a endopeptidase
TLC	TRAM, LAG1 and CLN8 homology domain
TM	transmembrane helix
T_m	melting temperature
UHPLC	ultra-high performance liquid chromatography
UMNG	2,2-diundecylpropane-1,3-bis-β-D-maltopyranoside
VLCFAs	very-long-chain fatty acids
VMAT	vesicular monoamine transporter
X-ALD	X-linked adrenoleukodystrophy
β-OG	n-octyl-β-D-glucoside

Table of contents

Abstract i

Acknowledgements iii

Authorship declaration iv

Publications resulting from this work v

List of abbreviations vi

1 Introduction 1

1.1 Secondary active transport in the Major Facilitator Superfamily 1

1.1.1 Membrane transport 1

1.1.2 Solute carriers (SLCs) 2

1.1.3 The Major Facilitator Superfamily 3

1.1.3.1 MFS topology 4

1.1.3.2 Alternating access mechanism in MFS transporters 5

1.1.4 MFSD10 9

1.2 Fatty acid elongation and sphingolipid biosynthesis 12

1.2.1 Sphingolipids 13

1.2.2. Ceramide synthesis 15

1.2.3 Ceramide synthases 17

1.2.4 Ceramide synthases are potential therapeutic targets 22

1.2.5 Known CerS inhibitors 24

1.2.6 Fatty acid elongation 26

1.2.7 3-keto acyl-CoA synthases (KCSs) 28

1.2.8 ELO elongases 28

1.2.9 ELOs: disease implications and therapeutic potential of their inhibition 29

1.3 Aims of this thesis 30

2 Materials and methods 32

2.1 General methods 32

2.1.1 Constructs 32

2.1.2 Protein expression 33

2.1.3 Large-scale protein purification 34

2.1.3.1 MFSD10 34

2.1.3.2 ELOVL7 36

2.1.3.3 Ceramide synthases 36

2.1.4 Protein reconstitution into proteoliposomes 38
2.1.5 NanoDSF label-free thermal unfolding assays 39
2.1.6 Denaturing intact protein LC-ESI-MS analysis of purified protein samples 40
2.1.7 SDS-PAGE 42
2.2 Methods specific to the MFSD10 project 42
2.2.1 Membrane isolation 42
2.2.2 Protein crystallisation 43
2.2.3 Preparation of monoolein:cholesterol mixture for optimised *in meso* crystallisation 43
2.2.4 Crystal screening and X-ray data collection 44
2.2.5 X-ray data processing and molecular replacement 44
2.2.6 Model building and refinement 44
2.2.7 Pyranine assays 45
2.2.8 Fluorescence and absorbance spectra of Biliverdin and purified MFSD10 46
2.3 Methods specific to the ELOVL7 project 46
2.3.1 SEC-MALS 46
2.3.2 Denaturing intact protein mass spectrometry 47
2.3.2.1 Analysis of the relative stability of the unmodified and modified forms of purified ELOVL7 47
2.3.2.2 Covalent intermediate identification 48
2.3.2.3 Evaluation of covalent acyl-enzyme intermediate formation using active site histidine mutants 48
2.3.2.4 Reaction of malonyl-CoA with the covalent intermediate 49
2.4 Methods specific to the ceramide synthases project 49
2.4.1 Detergent screening (3 mL purifications) 49
2.4.2 Small-scale purification screening 51
2.4.3 *In vitro* biotinylation 52
2.4.4 Nanobody expression and purification 53
2.4.4.1 Small-scale nanobody expression and purification for binding screening 53
2.4.4.2 Large-scale nanobody expression and purification 54
2.4.5 Bio-Layer Interferometry (BLI) 55
2.4.6 Purification of CerS6-nanobody complexes for structural studies 56
2.4.7 Cryo-EM sample preparation 56
2.4.8 Cryo-EM grid screening and small characterisation datasets 57
2.4.9 CerS6-c025 in GDN: data collection and processing 57

3 Structural insights into the transport mechanism of the human NSAID transporter MFSD10 59
3.1 Aims 59
3.2 Purification and crystallisation of human MFSD10 59
3.2.1 Large-scale purification of human MFSD10 59

3.2.2 NanoDSF screening of putative substrates 63
3.2.3 Crystallisation in the lipidic cubic phase (LCP) 65
3.3 Crystal structure of human MFSD10 66
3.3.1 Data collection, phasing, model building and refinement 66
3.3.2 Packing in the MFSD10 crystals 70
3.3.3 Overall structural features 71
3.3.4 Substrate-binding site 74
3.3.5 A-motif 78
3.3.6 Extracellular gate 79
3.3.7 Soaking studies with a putative substrate 82
3.3.8 Co-crystallisation studies with putative substrates 84
3.4 Insights into the transport mechanism 91
3.4.1 MFSD10 acts as a proton-coupled NSAID antiporter 91
3.4.2 Structural insights into the mechanism of proton coupling 94
3.5 Discussion 97

4 Mechanism of fatty acid elongation by human ELOVL7 102
4.1 Background leading to this work 102
4.1.1 Crystal structure of human ELOVL7 102
4.1.2 ELOVL7 structure contains a covalently-bound product analogue 103
4.1.3 Substrate-binding and active sites 106
4.1.4 Aims 107
4.2 Biochemical characterisation of the reaction mechanism of ELOVL7 108
4.2.1 Purification of wild-type ELOVL7 108
4.2.2 Purified ELOVL7 is a dimer in the crystal and in solution 108
4.2.3 The apo form of purified ELOVL7 is unstable and lost during purification 110
4.2.4 Substrate-binding arrangement suggests a ping-pong type mechanism 112
4.2.5 Identification of a covalent acyl-enzyme intermediate 114
4.2.6 Metals are not required for the transacylation step 117
4.2.7 Catalysis proceeds via an unusual acyl-imidazole intermediate 119
4.2.8 Condensation of the acyl-enzyme intermediate with malonyl-CoA 124
4.3 Discussion 128
4.3.1 Proposed reaction mechanism of human ELOVL7 128
4.3.2 Formation of the 3-keto acyl-CoA adduct observed in the crystal structure 131
4.3.3 Comparison of the mechanism of ELOs and FAE1-type KCSs 133

5 Structural studies of human ceramide synthases 135
5.1 Aims of this chapter 135
5.2 Establishing purification conditions for human ceramide synthases for structural studies. 136

5.2.1 Constructs designed 136
5.2.2 Expression screening 137
5.2.3 Detergent screening 139
5.2.4 Small-scale purification screening 142
5.2.5 Large-scale CerS5 purification 145
5.2.6 Large-scale CerS6 purifications 150
5.2.7 NanoDSF screening of known inhibitors 154
5.2.8 Design and purification of glycosylation mutants for crystallisation trials 155
5.2.9 Identification of a dimeric population in purified CerS6 157
5.3 Generation and validation of nanobodies against CerS5 and CerS6 159
5.3.1 Overview and strategy 159
5.3.2 Sample preparation for immunisation and library screening 160
5.3.3 Validation of unique binders generated 162
5.3.3.1 Small-scale expression and purification of unique nanobodies 163
5.3.3.2 Nanobody screening by Bio-Layer Interferometry 165
5.4 Characterisation of CerS6 using Cryo-EM 167
5.4.1 Optimisation of CerS6 grids 167
5.4.2 CerS6 Krios dataset 171
5.4.3 Purification of CerS6-nanobody complexes 175
5.4.4 CerS6-nanobody complexes: grid screening 176
5.5 Discussion 179

6 Conclusion 182
6.1 MFSD10 182
6.2 ELOVL7 186
6.3 Ceramide synthases 191
6.4 Closing remarks 194

Appendices 197
Appendix 1 – Full-length protein sequences 197
Appendix 2 – CerS1 construct screening 198
Appendix 3 – Nanobodies generated against CerS5 and CerS6 199
Appendix 4 – Nanobody screening results 201
Appendix 5 – Vector Maps 203

Bibliography 208

1 Introduction

1.1 Secondary active transport in the Major Facilitator Superfamily

1.1.1 Membrane transport

Polar small molecules and ions do not easily diffuse across the hydrophobic lipid bilayer of cellular membranes. Therefore, their translocation is dependent on protein carriers, which can be categorised into primary and secondary transporters depending on the energy source required for transport. Primary transporters use the energy stored in ATP to achieve solute translocation. These include ATP-binding cassette (ABC) transporters, which hydrolyse ATP to power the transport of small molecules, and ion pumps, responsible for establishing ion gradients across biological membranes [1]. One example of such an ion pump is the Na^+/K^+ ATPase, responsible for establishing steep sodium and potassium gradients across the plasma membrane in animals [2]. Likewise, proton gradients can be generated by vacuolar ATPases (V-ATPases), which use ATP to pump protons and are primarily involved in the acidification of intracellular compartments [3]. Other mechanisms to establish proton gradients exist across all kingdoms of life. Proton electrochemical gradients (Proton Motive Force, PMF) are required for ATP synthesis, and these can be generated by redox-driven proton pumps, present, for example, in mitochondria, chloroplasts and bacteria [4-7]. Overall, these processes generate electrochemical gradients across cellular membranes, which can be used to power secondary active transport by solute carriers (SLCs) and are also required for the rapid diffusion of ions down a concentration gradient through ion channels.

1.1.2 Solute carriers (SLCs)

SLCs are a functionally and structurally diverse class of membrane transporters (reviewed in [8]) which do not consume ATP to power transport. Although some SLCs act as ion channels, I will focus on describing SLCs that function as secondary active transporters.

SLCs can function as uniporters, symporters or antiporters. Uniporters carry a single substrate across the membrane, down its concentration gradient and therefore carry out passive transport; symporters co-transport two substrates in the same direction across the membrane, and antiporters transport two substrates in opposite directions (Fig. 1.1). Symporters and antiporters use the energy stored in the transmembrane electrochemical gradient of one substrate (typically, ions such as protons and sodium) to power the transport of the other.

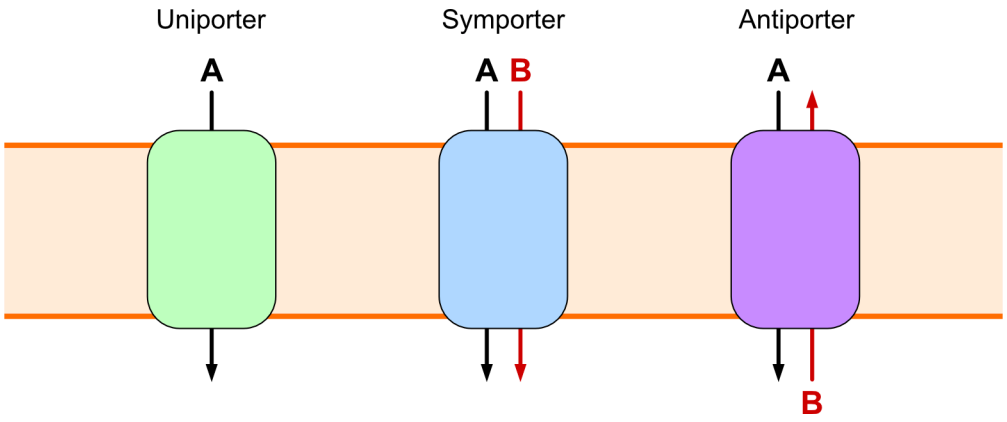


Figure 1.1. SLC transporters can act as uniporters, symporters and antiporters. In this schematic representation, the two substrates of symporters and antiporters are represented as A and B.

SLC transporters have long been proposed to function through an alternating access mechanism, whereby the transporter alternately exposes its substrate-binding site to either side of the membrane [9]. There are three different mechanisms by which SLCs achieve alternating access. SLCs of the Major Facilitator Superfamily are simplistically described as using a 'rocker-switch' mechanism (described in the next section). Transporters of the LeuT fold

use a 'rocking bundle' mechanism, whereby a scaffold bundle remains static while a mobile rocking bundle undergoes conformational changes, exposing the substrate-binding site to one side of the membrane while closing the gate to the other side [10]. Lastly, transporters using an elevator mechanism have a scaffold domain and a transport domain. The scaffold domain participates in oligomerisation and remains static, while the transport domain contains the substrate-binding site and slides perpendicularly to the membrane, carrying the substrate-binding site to either side of the membrane [8, 11].

1.1.3 The Major Facilitator Superfamily

The Major Facilitator Superfamily (MFS) is the largest family of membrane transporters and is present across all kingdoms of life [12, 13]. The Transporter Classification Database (TCDB) (https://tcdb.org) [14] classifies the MFS superfamily into 16 families and 89 subfamilies, but additional subfamilies have recently been classified into the MFS, suggesting a total of 105 subfamilies [15]. In humans, there are 129 MFS transporters classified in the TransportDB database (http://www.membranetransport.org/transportDB2/) [16]. SLCs of the MFS transport a wide range of substrates, including monosaccharides [17], nucleotide sugars [18], lipids [19, 20], peptides [21, 22], nucleosides [23], organic anions and organic cations [24]. In addition, MFS transporters play an important role as efflux pumps in bacteria, removing toxic compounds, including drugs, from the cell [25, 26]. These transporters are commonly multispecific, transporting multiple structurally diverse compounds and contributing to multidrug resistance in bacteria [26-28]. As a result, these are referred to as MFS-MDR transporters (MFS multidrug resistance transporters) and typically act as drug:H^+ antiporters [26].

1.1.3.1 MFS topology

The majority of MFS transporters adopt a canonical topology where the core of the transporter is formed by 12 transmembrane helices (TMs). However, some MFS transporters contain 14 TMs, including the drug:H^+ antiporter family 2 (DHA2; TC # 2A.1.3) [29] and some proton-coupled oligopeptide symporters [30]. In addition, there are rare examples that contain 11 TMs [31].

The 12-TM core of MFS transporters is structurally formed by two 6-TM bundles (also referred to as N- and C- domains) related by two-fold pseudosymmetry. A central cavity where the substrate-binding site is located is formed between these two domains (reviewed in [32, 33]). In addition, each 6-TM domain can be subdivided into two 3-TM inverted repeats (TM1-3 and TM4-6; TM7-9 and TM10-12), related to each other by C2 pseudosymmetry [34]. The overall topology of MFS transporters is displayed in Fig. 1.2, where the 3 helices constituting each 3-TM repeat are labelled respectively as A-helix (TM1, TM4, TM7, TM10), B-helix (TM2, TM5, TM8, TM11) and C-helix (TM3, TM6, TM9, TM12). A-helices are found in the centre of the transporter, B-helices play important roles in mediating contacts between the two 6-TM domains, and C-helices are located on the outside of the transporter [34]. With this arrangement, A- and B- helices typically line the central cavity and participate in substrate-binding [30, 33].

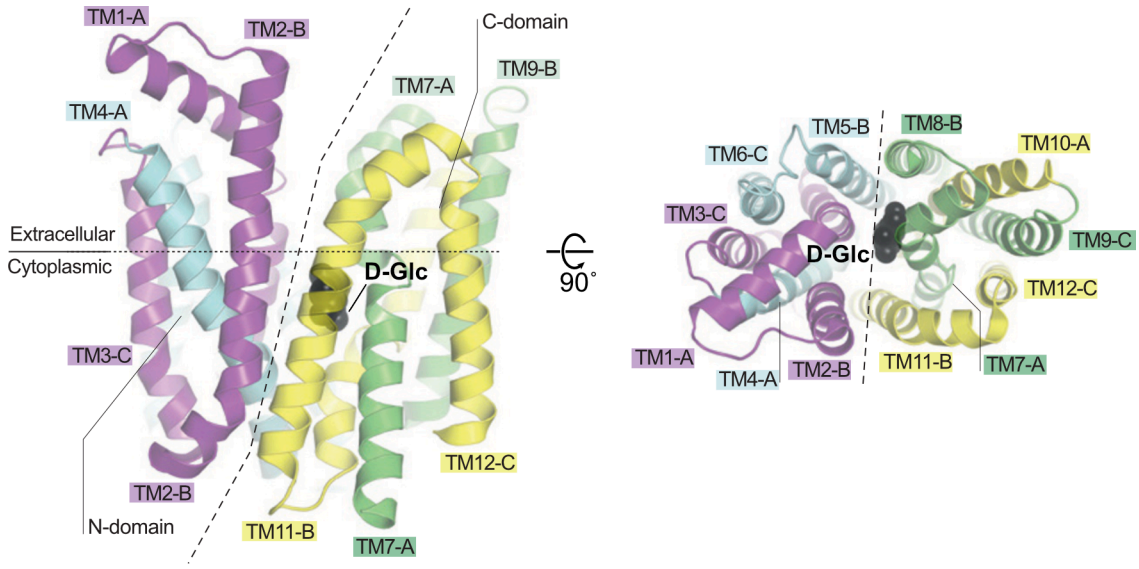


Figure 1.2. General topology of MFS transporters. MFS transporters have a canonical 12-TM core formed by two 6-TM domains (N- and C-domains), each containing two 3-TM inverted repeats. Each 3-TM unit is shown in a different colour and helices are numbered 1-12 and A-C, depending on their position with the 3TM unit. The structure shown corresponds to GLUT3 in an outward-occluded conformation (PDB ID: 4ZW9) and is viewed along the plane of the membrane (*left)* and from the extracellular side, perpendicular to the membrane (*right)*. The substrate (D-glucose) is displayed as black spheres. Adapted from [32].

1.1.3.2 Alternating access mechanism in MFS transporters

When the first crystal structures of MFS transporters (GlpT and LacY, from *E. coli*) were solved, a general 'rocker-switch' model was proposed to describe the conformational changes enabling alternating access during the transport cycle of MFS transporters [35-37]. In this model, the two 6-TM domains were proposed to rotate as rigid bodies, rocking against each other and thereby exposing the central substrate-binding site to either side of the membrane without ever allowing simultaneous access from both sides (Fig. 1.3). Importantly, the formation and breakage of intra- and inter-domain salt bridges is thought to play an essential role in regulating this conformational switch [13, 32, 33, 38-40].

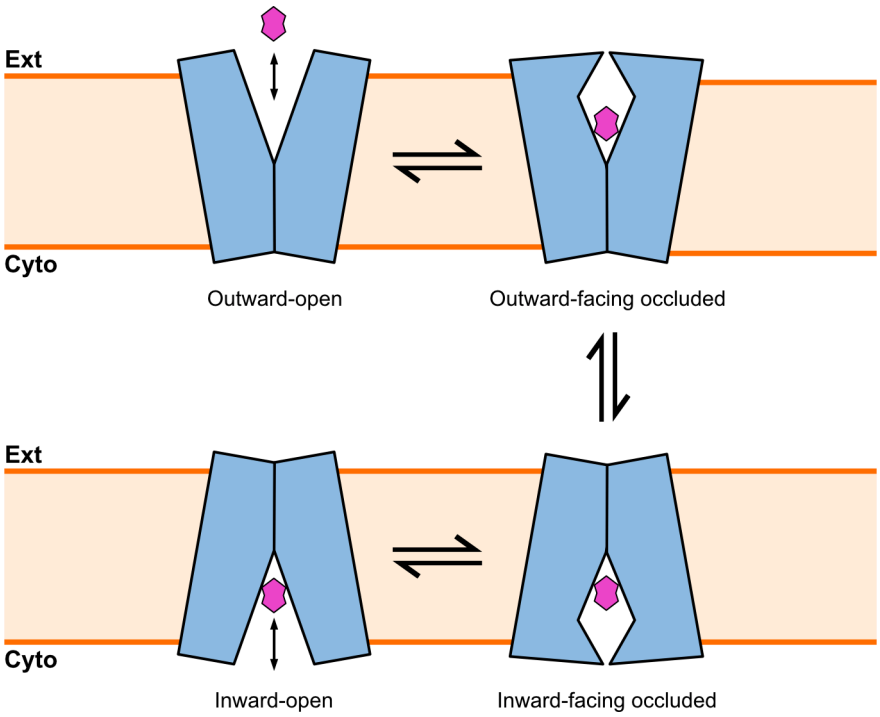


Figure 1.3. Alternating access in MFS transporters. In the alternating access model, the central cavity of the transporter is alternately accessible to either side of the membrane. In MFS transporters, the conformational changes in the transport cycle were initially described by a simplified rocker-switch mechanism in which the two 6-TM domains rotate symmetrically as rigid bodies. Substrate-binding or release triggers the structural changes. However, intermediate occluded states formed upon the closing of thin gates via local structural rearrangements exist between inward- or outward-facing states. Such local conformational changes were not accounted for in the original rocker-switch model. In this diagram, a uniporter is represented, but these principles are common to other transporters. Occluded conformations can occur either in substrate-bound states or in the absence of substrate, except in antiporters, where occluded conformations can only occur in the presence of a bound substrate. Substrate is shown as a magenta polygon.

However, as more structures were experimentally determined, the importance of local rearrangements deviating from rigid body motions became evident (reviewed in [32, 41]). Using the helix nomenclature described by Quistgaard and colleagues [Ref. 32] and above, A- and B-helices are typically involved in local gating rearrangements that regulate access to the central cavity. These local rearrangements of gating helices enable the occlusion of the substrate-binding site in the absence of the major domain rearrangements proposed in the rocker-switch model [30, 40-42]. Such observations led to an updated model to describe the structural changes in the MFS transport cycle [32, 41]. In this model, access to the central

cavity is controlled by 'thick' and 'thin' gates. The 'thick' gates represent the occlusion of the cavity caused by the major switch when the transporter is re-oriented from outward- to inward-facing and *vice-versa*. On the other hand, the 'thin' gate describes the local rearrangements of A- and B- helices. Thus, in this updated model, an outward-open to inward-open transition would comprise 1) closing of the thin extracellular gate, leading to an outward-facing occluded conformation; 2) rigid body movement of the N- and C- domains which approximates the conformational changes described in the rocker-switch model, leading to an inward-facing occluded conformation; 3) opening of the thin intracellular gate, finally allowing solvent access to the central cavity in the inward-open state.

There is an additional layer of complexity to this model, as these local gating rearrangements differ considerably in different MFS transporter families and can occur approximately symmetrically between both domains or can result from asymmetric conformational changes [33]. As an example, substrate-induced asymmetrical gating rearrangements in MFS sugar porters have been recently elegantly described by Drew and colleagues [Ref. 33], and I will briefly introduce this concept in Section 3.3.6. Also, there is a conserved cytoplasmic motif designated as the A-motif present in most MFS transporters [12, 25] which plays a crucial part in gating on the cytoplasmic side (described in Section 3.3.5).

As mentioned before, MFS transporters include uniporters, symporters and antiporters. In both symporters and uniporters, the transporter must be able to spontaneously reset itself in the absence of substrates [41]. In contrast, substrate binding triggers the conformational switches of MFS antiporters in both directions, meaning that empty occluded states of antiporters cannot exist [41, 43]. MFS symporters require the binding of both substrates to initiate the conformational switch. This can often be accompanied by an absolute requirement for sequential binding of the two co-transported substrates. For example, in the lactose per-

mease LacY, which acts as a proton-coupled symporter, proton binding from the extracellular side has been shown to be required for binding of the sugar, which then drives the switch to the inward-facing conformation [44]. A similar model has also been proposed for the proton-coupled oligopeptide transporters (which are peptide/H^+ symporters). In this model, protonation of a key acidic residue in TM7 would open the extracellular gate to enable peptide binding, which in turn triggers closure of the extracellular gate [22, 45-47]. However, in sugar porters, protonation and sugar binding appear to be independent but essential for closure of the extracellular gate [33]. On the other hand, MFS antiporters ensure that substrates are released before the transporter undergoes another conformational switch by typically coupling release of one substrate to the binding of the other, such that both substrates cannot bind simultaneously. As a result, in proton-coupled antiporters, substrate binding is thought to typically compete with protonation of the transporter [48-51].

Amongst MFS antiporters, particular interest has been given to the MFS-MDR transporters responsible for multidrug efflux in bacteria, which act as drug:H^+ antiporters (briefly introduced in Section 1.1.3). Given their central role in conferring antibiotic resistance to bacteria, MFS-MDR transporters are considered important drug targets to restore the efficacy of antibiotics [52]. However, MFS-MDR transport activity is not limited to prokaryotes, and multiple MFS-MDR transporters have been identified in yeast [53]. Furthermore, although ABC transporters appear to be primarily responsible for multidrug efflux activity in humans [54, 55], MFS transporters may also play a role [33]. For example, the human Major Facilitator Superfamily Domain-containing protein 10 (MFSD10) can confer resistance to apoptosis induced by non-steroidal anti-inflammatory drugs (NSAIDs) [56]. Also, MFS transporters of the SLC22 family are known to transport several drugs, and the organic cation transporter SLC22A18 is thought to have drug:H^+ antiport activity [57]. Interestingly, both

MFSD10 and SLC22A18 share sequence similarity with *E. coli* tetracycline resistance transporters [58, 59]. In the first part of this thesis, I aimed to elucidate the mechanism of human MFSD10.

1.1.4 MFSD10

The human *MFSD10* gene was first identified in 1993 by sequence similarity to an *E. coli* tetracycline efflux transporter [58]. Later, its overexpression was found to confer resistance to apoptosis induced by the non-steroidal anti-inflammatory drugs (NSAIDs) diclofenac and indomethacin in a human adenocarcinoma gastric cell line [56]. This resistance to NSAID-induced apoptosis could be reversed by siRNA knockdown of *MFSD10* expression, suggesting that MFSD10 may be acting as an NSAID efflux pump. Later, MFSD10 was demonstrated to transport organic anions, including NSAIDs, across the plasma membrane of HEK293 cells [60].

NSAIDs are one of the most widely prescribed classes of drugs and exert their anti-inflammatory effects by blocking prostaglandin biosynthesis by the cyclooxygenases COX-1 and/or COX-2. These enzymes convert arachidonic acid to prostaglandin G2, which is further processed to generate important inflammatory mediators [61]. However, NSAID use is associated with a range of gastrointestinal, cardiovascular and renal complications [62, 63]. For example, diclofenac, the most widely used NSAID [64], contributes to increased cardiovascular mortality and morbidity [65]. Given its proposed function as an NSAID efflux transporter, MFSD10 could influence the pharmacokinetics of this class of drugs. However, very little is known about its molecular mechanism of transport, despite its recognised importance.

MFSD10 belongs to the Major Facilitator Superfamily of membrane transporters, and has 12 putative transmembrane helices. Interestingly, it displays low sequence similarity to other human transporters of several SLC families. As a result, it is one of the 30 human atypical SLCs that had not previously been classified into existing SLC families [66]. Perland and colleagues [Ref. 66] carried out a phylogenetic analysis of the 28 atypical human SLCs belonging to the Major Facilitator Superfamily, and proposed that they could be classified into 15 families. According to this classification, MFSD10 could be included in the Atypical MFS Transporter Family 1 (AMTF1), along with MFSD9, MFSD14A and MFSD14B. At the same time, however, MFSD10 has been recently classified into the SLC22 family as SLC22A32, primarily on the basis of its proposed transporter function (https://slc.bioparadigms.org). Within the SLC22 family, MFSD10 is more closely related to the organic cation transporter SLC22A18 (22.2 % sequence identity), which likely acts as a drug:H^+ antiporter (see Section 1.1.3.2).

MFSD10 was initially studied in the context of the kidney and was identified in the luminal membrane of proximal tubules, where it was proposed to be the missing transporter involved in the last secretion step of organic anions into the urine [60]. However, *MFSD10* mRNA and protein expression have been found in a wide range of tissues, including the brain, immune cells, lymphoid tissues, gut, placenta and prostate (GTEx and ProteomicsDB databases [67, 68]). Therefore, MFSD10 is likely to have a physiological role in other tissues. Indeed, it has been identified as a surface marker for eosinophils [69, 70].

Furthermore, the subcellular localisation of MFSD10 is poorly understood. MFSD10 has been proposed to be localised to the plasma membrane [60, 69, 70], to the inner nuclear membrane [71], and to the ER membrane [72]. However, in each case, a different cell type was used in the localisation studies. For example, plasma membrane localisation has been observed in HEK293 cells [60] and eosinophils [69, 70], localisation to the inner nuclear

membrane was observed in mesenchymal cells [71], and localisation to the ER membrane was observed in a BioID proximity labelling map of HEK293 cells [72]. As a result, whereas the information available on substrate specificity from these studies suggests a role for MFSD10 in the efflux of toxic compounds, the uncertainty associated with its localisation means that its cellular function remains unclear, with suggestions ranging from a role in NSAID transport across the plasma membrane [60] to the removal of toxic compounds and/or xenobiotics from the nucleus [71].

Moreover, very little is also known about the transport mechanism of MFSD10. NSAID transport by MFSD10 does not depend on Na^+ or K^+ ions, but is sensitive to nigericin, a H^+/K^+ exchanger, suggesting it is coupled to the proton motive force [60]. In addition, examination of the MFSD10 primary sequence reveals the presence of the conserved C-motif (GX_3GPX_2GG, where Gly can often be substituted by Ala), also referred to as the antiporter motif. This motif is present in the TM5 helix of MFS antiporters, but not in uniporters or symporters [73]. Comparison of the inward- and outward-facing structures of the *E. coli* multidrug efflux antiporter MdfA, reveals that this motif leads to a local kinking and twisting of TM5 exclusively in the outward-facing state, and is therefore proposed to act as a hinge point that facilitates relative rotation between the N- and C-domains in MFS antiporters during the conformational switch [74]. In human MFSD10, this antiporter motif contains the sequence $G^{187}FTLG^{191}P\ ^{192}MLG^{195}A^{196}$ and is also located on TM5. Together, these observations suggest that MFSD10 could act as an NSAID:H^+ antiporter. Indeed, the Transporter Classification Database (https://www.tcdb.org) classifies MFSD10 into the MFS drug:H^+ antiporter family 1 (DHA1) (TC # 2.A.1.2). Most well-studied examples of the DHA1 family are bacterial multidrug resistance transporters, including MdfA, EmrD, LmrP, NorA and YajR, which act as efflux transporters for a large range of structurally diverse compounds [26, 75]. However, eukaryotic DHA1 transporters can also transport solutes

across intracellular membranes, against the H^+ electrochemical gradient. For example, human vesicular neurotransmitter transporters transport monoamines and acetylcholine into synaptic vesicles of pre-synaptic neurons using the H^+ electrochemical gradient established by V-ATPases [76-78].

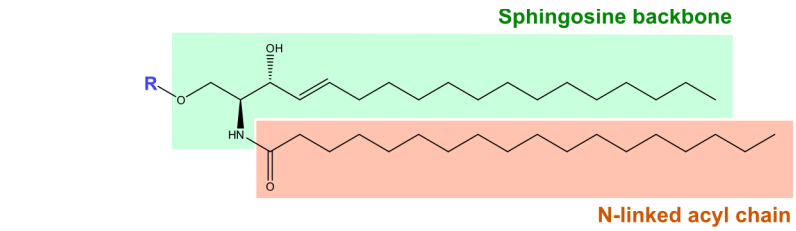
Overall, available information suggests that MFSD10 acts as a proton-coupled antiporter with multidrug efflux activity. However, while it has been proposed to transport a range of NSAIDs and other organic anions, its endogenous substrates are currently unknown. Also, to the best of my knowledge, although several structures of bacterial transporters of the DHA1 family have been determined, no eukaryotic DHA1 transporter structures have been solved to date. Moreover, despite the important biological function MFSD10, very little research has been carried out to understand its transport mechanism. Thus, I aimed to determine its structure using X-ray crystallography and to set up functional assays to probe its mechanism. This work is described in Chapter 3 of this thesis.

1.2 Fatty acid elongation and sphingolipid biosynthesis

In addition to the work on MFSD10, the second focus of this thesis falls on two families of human membrane enzymes that play key roles in determining the acyl chain specificity in fatty acid elongation (ELOVLs) and sphingolipid biosynthesis (ceramide synthases) and are regarded as attractive therapeutic targets. In the following sections, I will introduce the physiological roles of sphingolipids and very-long-chain fatty acids, the biosynthetic pathways responsible for their production, and the critical roles played by these two enzyme families.

1.2.1 Sphingolipids

Sphingolipids are key structural components of membranes, act as signalling molecules and carry out important regulatory roles in a wide range of cellular functions [79-82]. The basic building block of all sphingolipids is a sphingoid base such as sphingosine, sphinganine (dihydrosphingosine) or phytosphingosine. The sphingoid base is appended with an N-linked fatty acid and an additional optional head group linked to the hydroxyl group at position 1 (Fig. 1.4). Ceramides, the precursors for the synthesis of complex sphingolipids, do not contain a head group (R = H) and therefore consist only of a sphingosine backbone with an N-acyl group. Complex sphingolipids can be subsequently produced from ceramides by the addition of different head groups (Fig. 1.4).



Sphingolipid	R (head group)
Ceramide	H
1-O-acyl-ceramide	Fatty acid
Sphingomyelin	Phosphocholine
Cer 1-phosphate	Phosphate
Glycosphingolipids	Sugars

Figure 1.4. Sphingolipid structure overview. The main sphingoid base in mammals, sphingosine, is shown in green, but several other sphingoid bases are present in sphingolipids. Sphingolipids contain an acyl chain (shown in orange) linked to the primary amine of the sphingoid base. Head groups (R) are added to the hydroxyl group at position 1 of the sphingoid base. Examples of sphingolipids and their respective head groups are shown.

Ceramides are key bioactive lipids, playing important signalling roles, including in the regulation of fatty acid metabolism to lower the levels of free fatty acids in situations of lipid excess [83]. In this process, ceramides influence a number of pathways, including blocking

Akt (protein kinase B) activation [84, 85] and disrupting mitochondrial function by impairing the mitochondrial electron transport chain complexes and promoting mitochondrial fission [85-88]. Thus, elevated ceramide levels can lead to apoptosis, impaired insulin signalling, and contribute to the development of metabolic diseases [83, 89, 90].

Complex sphingolipids contain polar head groups and perform several unique cellular functions. For example, sphingomyelin, the most common complex sphingolipid, is an essential structural component of myelin. Sphingomyelin has also been found to play a role in cell proliferation and adhesion [91-93] and is involved in regulating vesicle trafficking by directly interacting with the transmembrane domain of p24, a regulator of COPI vesicle formation [94]. Furthermore, sphingomyelin can also act as a source of ceramides through the action of sphingomyelinases, and this additional layer of regulation enables ceramides to be rapidly targeted to a particular membrane location [82, 95].

Glycosphingolipids are primarily located in the outer leaflet of the plasma membrane, where they play signalling roles. For example, glycosphingolipids regulate the function of plasma membrane receptors and other proteins involved in signal transduction [96-98], participate in cell-cell interactions [99, 100], and are critical structural components of the myelin sheath in the nervous system [101]. Also, glucosylceramide levels have been found to be linked to cell proliferation [102-104].

Sphingolipids have been proposed to assemble in membrane microdomains enriched in specific sphingolipids, which influence membrane fluidity and protein recruitment and are thought to facilitate the compartmentalisation of cellular processes and the regulation of signalling pathways [82, 96, 105-108].

Importantly, the length of the N-linked acyl chain of sphingolipids plays a determinant role in their biological function. For example, very-long-chain fatty acids (fatty acids with 20 or

more carbons in their acyl chain; VLCFAs) become incorporated into very-long-chain ceramides and complex sphingolipids to perform unique cellular functions that cannot be carried out by shorter fatty acids, namely in myelin maintenance [109, 110], skin permeability barrier formation [111-113], spermatogenesis [114], retinal function [115] and liver homeostasis [109, 116]. In addition, the acyl chain length of sphingolipids has been shown to regulate the formation and function of membrane microdomains [96, 108]. Also, the role described above for sphingomyelin in regulating vesicle trafficking is exclusive to C18 sphingomyelin, as only this species interacts with p24, a transmembrane protein of COPI vesicles [94]. Lastly, the relative abundance of ceramides with different acyl chain lengths is a key determinant in many metabolic diseases (see Section 1.2.4). For these reasons, it is crucial to understand how the synthesis of ceramides with different acyl chains, and the production of the different acyl-CoA precursors, are regulated.

1.2.2. Ceramide synthesis

There are two main pathways for the synthesis of ceramides in mammalian cells, termed the *de novo* and salvage pathways. Here, I will focus primarily on the *de novo* synthesis of ceramides and the long and very-long-chain fatty acids required for the N-acylation of the sphingoid base in this pathway. A schematic representation of the *de novo* ceramide synthesis pathway, which takes place in the ER membrane, is displayed in Fig. 1.5. The initial rate-limiting step is carried out by serine palmitoyl transferase (SPT), which catalyses the condensation of L-serine and palmitoyl-CoA to generate 3-ketodihydrosphingosine [117, 118]. Then, 3-ketodihydrosphingosine is reduced to dihydrosphingosine (also known as sphinganine, d18:0) by 3-ketodihydrosphingosine reductase (KDSR) [119, 120]. This way, SPT and KDSR generate a pool of free sphinganine. Sphinganine is N-acylated with different acyl chains by ceramide synthases (CerS1-6) to produce dihydroceramides [81]. Finally,

these dihydroceramides are converted into ceramides by a dihydroceramide Δ4-desaturase (DES1), which introduces a double bond between C4 and C5 of the dihydrosphingosine moiety of dihydroceramides [121, 122]. Therefore, as introduced above, the ceramide generated as the final product of this pathway is an N-acylsphingosine.

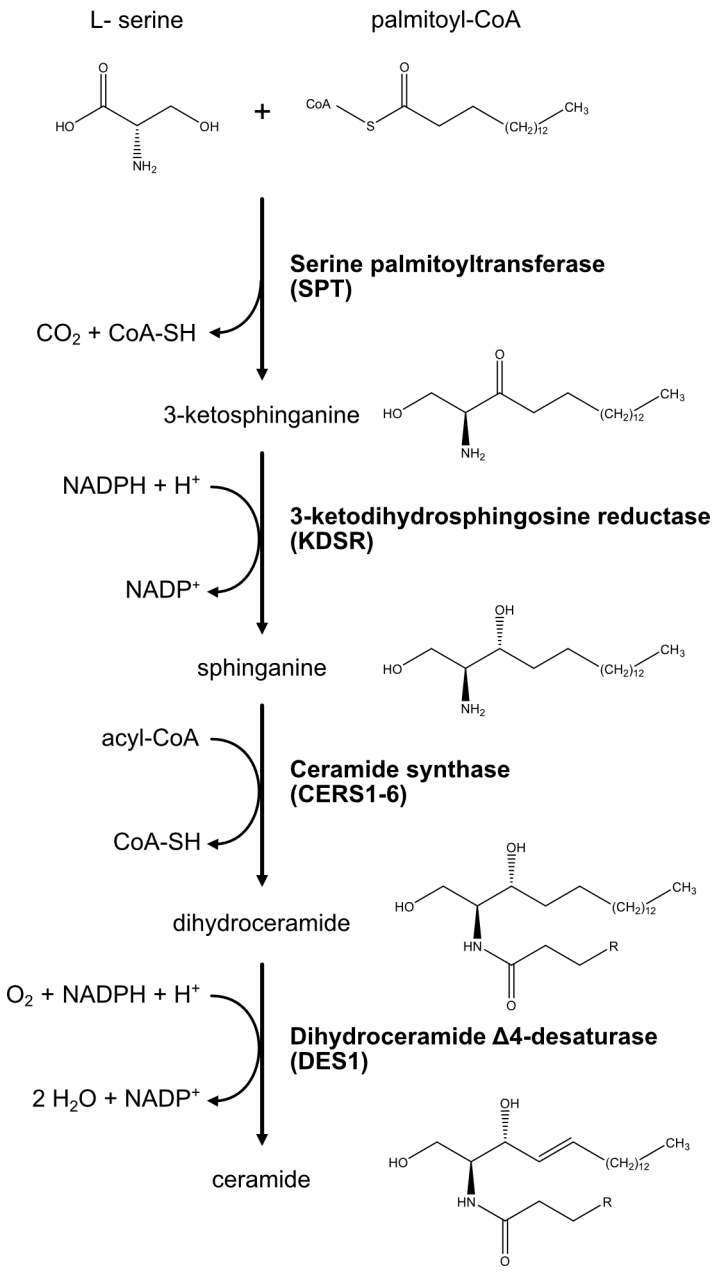


Figure 1.5. *De novo* ceramide synthesis pathway.

Briefly, the salvage pathway consists of the re-acylation of sphingosine to produce ceramides. In the salvage pathway, sphingomyelin is converted into ceramide by acid sphingomyelinase in lysosomes [123], while glucosylceramide is broken down to ceramide by the action of acid β-glucosidase 1 [124]. However, the resulting ceramide does not leave the lysosomes [125]. Instead, lysosomal ceramide needs to be deacylated by acid ceramidase [126, 127], yielding a free fatty acid and sphingosine, which are then able to exit the lysosomes [128]. Sphingosine generated through this pathway can then be directly re-acylated by ceramide synthases without the requirement for a dihydroceramide Δ4-desaturase step [129-131], or it can be converted into sphingosine-1-phosphate by sphingosine kinases [132]. For a review on ceramide salvage pathways, see [133]. Moreover, ceramides can be generated via a third route, in which sphingomyelin can be converted into ceramides in other membrane locations by sphingomyelinases [95, 134, 135].

1.2.3 Ceramide synthases

As described in the previous section, ceramide synthases (CerS) catalyse the transfer of the acyl chain of an acyl-CoA to the primary amine of a sphingoid base. CerS are a central component of sphingolipid biosynthesis, as they determine the acyl chain length that is added to the sphingoid base and therefore determine the properties of the ceramides produced in the *de novo* and salvage pathways [81].

In humans, there are 6 CerS enzymes (CerS1-6), which differ in tissue distribution and acyl chain length specificity. The substrate specificity of the 6 human CerS is summarised in Fig. 1.6. Human CerS enzymes display amino acid sequence similarities between 22 – 62 %, with CerS1 being more distantly related to the other 5 human CerS (22 – 27 % sequence identity) [136]. Perhaps unsurprising given their similar substrate specificity, CerS5 and CerS6 display the highest sequence identity (62 %) [136].

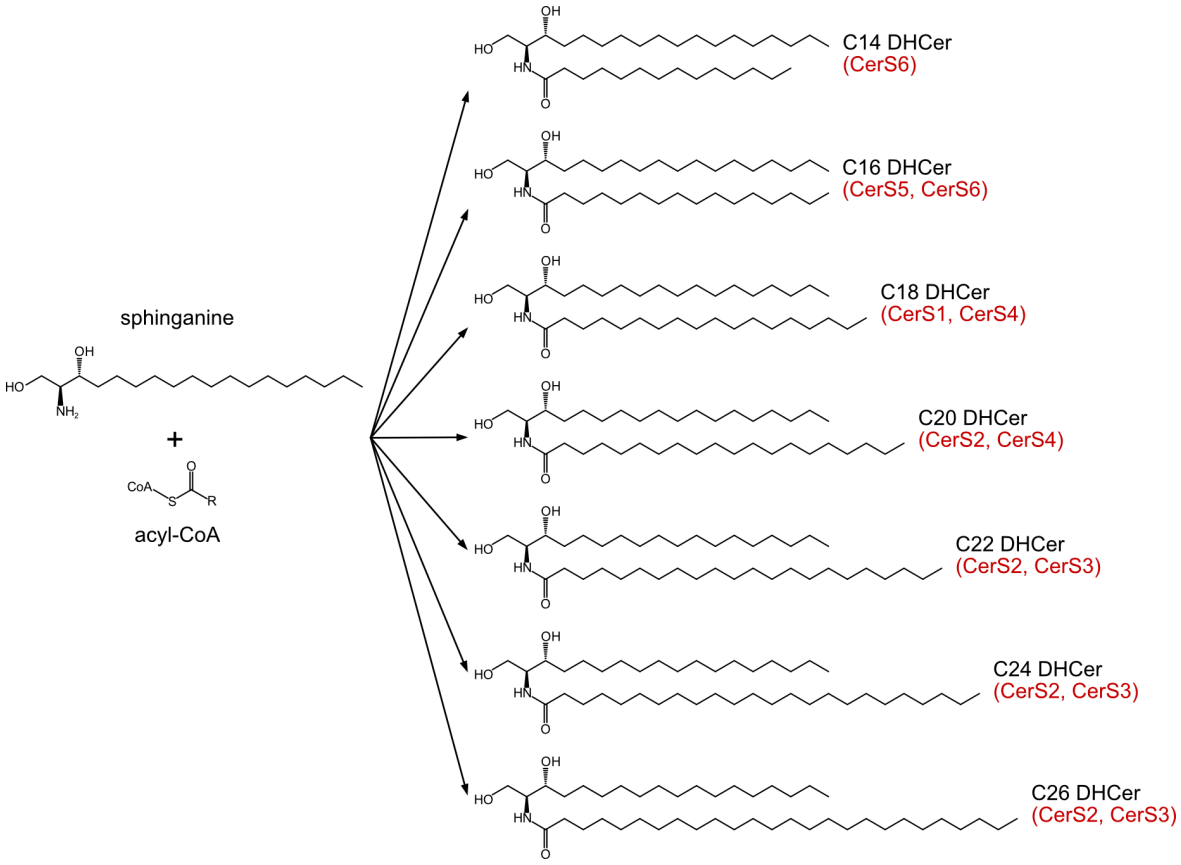


Figure 1.6. Substrate specificity of mammalian ceramide synthases. In the *de novo* ceramide synthesis pathway, ceramide synthases (CerS) catalyse the N-acylation of sphinganine (dihydrosphingosine, d18:0) with acyl-CoA-derived acyl chains containing from C14 to >C26, generating dihydroceramides (DHCer). The enzymes catalysing the production of C14 to C26 dihydroceramides are displayed in red. CerS3 can additionally catalyse the synthesis of C26-C32 dihydroceramides.

CerS are integral membrane enzymes and are located in the ER membrane. Before the advent of AlphaFold2 and RoseTTAFold for structure modelling [137, 138], there was great uncertainty regarding the number of transmembrane helices of CerS enzymes, with transmembrane helix prediction programs typically suggesting 5-7 transmembrane helices. Mizutani and colleagues [Ref. 139] identified that the N-terminus of mouse CerS2, CerS5 and CerS6 is glycosylated and must therefore face the ER lumen. In addition, the authors isolated intact organelles after expression of CerS6 with a C-terminal GFP fusion and found that proteinase K treatment of these organelles led to a proteinase K-resistant fragment corresponding to the GFP fusion, indicating that the C-terminus of CerS6 faces the cytoplasm

[139]. These findings suggest that CerS6 should have an odd number of helices, with its N- and C-termini facing opposite sides of the ER membrane. Later, additional evidence suggested that CerS2 and CerS5 also have an odd number of transmembrane helices [140] and that glycosylation was present in the N-terminus of human CerS2-CerS6, confirming that all these proteins are likely to share the same overall topology [141]. However, the exact number of transmembrane helices in the different CerS remained inconclusive [142]. Recently, Alphafold2 predicted 7 transmembrane helices for all human CerS, consistent with the observations that the N- and C-termini must be located on opposite sides of the membrane (Fig. 1.7).

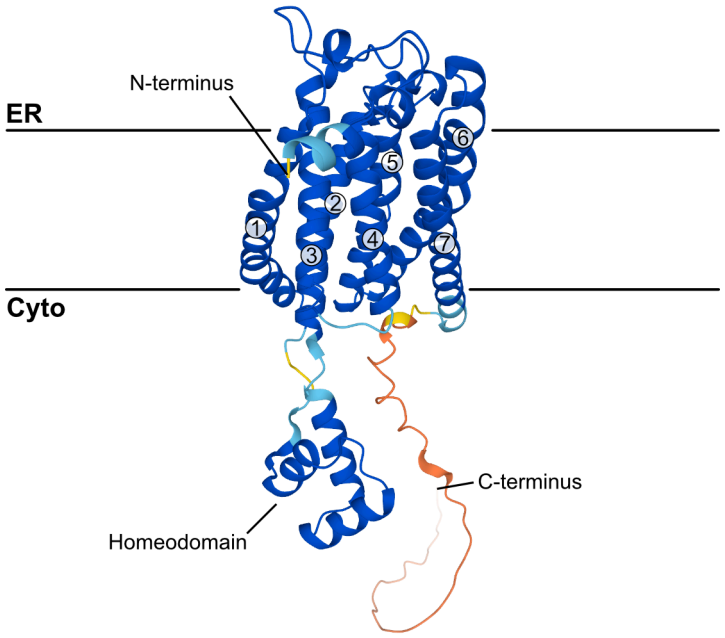


Figure 1.7. AlphaFold2 structural model of ceramide synthase 6. CerS6 has 7 transmembrane helices and a cytoplasmic homeodomain located between transmembrane helices 1 and 2. Its N-terminus is located in the ER lumen, whereas its C-terminus is predicted to be disordered and located in the cytoplasm. Other human CerS are also predicted to contain 7 transmembrane helices and a cytoplasmic homeodomain between TM1 and TM2, except CerS1, which lacks a homeodomain. Numbers in circles denote the 7 transmembrane helices.

CerS enzymes contain a TLC (TRAM, LAG1 and CLN8 homology) domain which is conserved in other lipid-binding proteins [143]. The TLC domain includes a Lag1p motif which contains conserved histidines and aspartates required for activity [144]. Importantly, acyl

chain specificity has been shown to be determined in the TLC domain [145], particularly in a short loop predicted to be exposed to the ER side [141]. Given that the acyl-CoA substrates are synthesised in the cytoplasm, it remains unclear how this loop could modulate acyl chain length specificity. However, it may act to limit the dimensions of the substrate-binding site, thus regulating how many carbon units in the acyl chain of the acyl-CoA substrate can fit [141].

In addition, CerS2-6 but not CerS1 contain a homeodomain N-terminal to the TLC domain, which faces the cytoplasmic side and is not involved in determining acyl chain specificity. This homeodomain was shown not to be required for CerS5 activity but is flanked by two arginines critical for ceramide synthase activity [146]. Initially, the role of this homeodomain was not well understood. However, CerS proteins were recently shown to be able to localise to the inner nuclear membrane, where they can bind DNA through their homeodomain and act as transcriptional regulators, namely by repressing the expression of several lipases [147, 148]. These observations suggest that CerS play a central role in sphingolipid metabolism, not only through their ceramide synthase activity, but also by acting as a transcriptional regulator, modulating gene expression in response to fatty acid and lipid levels [148]. In agreement with this proposed role, zebrafish CerS2b has been shown to regulate its own expression in response to sphingosine accumulation, in a feedback mechanism proposed to respond to elevated sphingosine levels by activating the salvage pathway [149].

CerS activity is subject to complex regulatory mechanisms that are currently still being unveiled. As an example, CerS2-6 activity is modulated by phosphorylation of their C-terminus (CerS2 and CerS4-6 are phosphorylated by CK2) [150]. While phosphorylation has a minor effect on the activity of CerS3-6, it has a large activating effect on the catalytic activity of CerS2 [150]. Furthermore, several CerS have been shown to form homo- and heterodimers [140]. Several studies have found that CerS2 co-immunoprecipitates with CerS4,

CerS5 and CerS6 [140, 151], suggesting a physical interaction. Notably, CerS5 or CerS6 co-expression stimulated CerS2 activity, and stimulation of ceramide synthesis, in turn, promoted CerS dimerisation [140]. In addition, CerS2, CerS5 and CerS6 formed homodimers, and co-expression of catalytically active and inactive forms of CerS5 inhibited the activity of the catalytically active monomer within this dimer. Laviad and colleagues [Ref. 140] also generated a constitutive CerS5 homodimer by introducing an additional transmembrane helix between the C-terminus of one monomer and the N-terminus of the other (which sit on opposite sides of the membrane). This constitutive homodimer displayed enhanced activity. Likewise, a constitutive CerS2-CerS5 heterodimer was generated, and CerS2 activity was dependent on the presence of a catalytically active form of CerS5. This observation again suggested that the increase in CerS2 activity upon CerS5 co-expression is due to a physical interaction between the two proteins. In both cases, the subcellular localisation of the CerS2 and CerS5 was not altered in the constitutive dimer forms, suggesting that the physical interaction is likely responsible for increased activity. Altogether, the authors proposed that, in the CerS dimers, substrate-binding to one protomer allosterically regulates the activity of the other [140]. In agreement with this model, a later study found that Bcl-2 family member Bcl2L13 interacts with CerS2 and CerS6, blocking complex formation between CerS2 and CerS6 and lowering their activity [152]. As a result, CerS dimerisation appears to play a regulatory role, particularly by enhancing CerS2 activity. However, other potential effects should not be ignored. Ceramide synthases are known to interact with other proteins involved in fatty acid or lipid metabolism [153, 154], and could form supramolecular complexes. Therefore, it is not surprising that different CerS isoenzymes co-immunoprecipitate, and the physical interactions between different members of the CerS family could influence dihydroceramide synthase activity by other mechanisms, such as through the recruitment of binding partners. Overall, the exact mechanisms by which CerS enzymes modulate their

activity through physical interactions (including dimerisation) remains poorly understood and further work is required to obtain additional insights into this complex regulation.

1.2.4 Ceramide synthases are potential therapeutic targets

Elevated levels of ceramides have been observed in a number of metabolic diseases, including obesity, diabetes [155], non-alcoholic fatty liver disease and steatohepatitis (NAFLD, NASH) [156]. NAFLD and NASH are characterised by the accumulation of triacylglycerides in hepatocytes and are commonly associated with obesity and diabetes [157], increased fatty acid synthesis and impaired β-oxidation [158]. Recent progress in understanding the role of different ceramides in the context of metabolic disease has demonstrated that the relative levels between ceramides of different acyl chain lengths play a critical role. For example, elevated levels of C18:0 ceramide, but not of total ceramides, in skeletal muscle were found to correlate with insulin resistance, suggesting a specific role for CerS1-mediated C18:0 ceramide production [159]. In addition, increased CerS6 expression and C16 ceramide levels were found in the adipose tissue of obese humans [160]. Therefore, given the different substrate specificities of the 6 human ceramide synthases, these enzymes have emerged as important potential therapeutic targets for the treatment of metabolic disorders [161].

Studies using knockout and disease mouse models have highlighted CerS6 as a potential target for the treatment of NASH, obesity and type 2 diabetes. Palmitate availability resulting from a high-fat diet (HFD) stimulates the production of C16 ceramide by CerS6 in mice, resulting in impaired mitochondrial β-oxidation in the liver and brown adipose tissue, leading to hepatosteatosis, obesity and insulin resistance [85, 160]. Likewise, in a HFD-induced mouse model of NAFLD, HFD led to increased levels of CerS6 and reduced levels of CerS2,

causing an increase in the production of C16 ceramide, which resulted in steatohepatitis caused by ER stress [162].

Importantly, deletion of CerS6 in mice granted protection against HFD-induced obesity and insulin resistance [88, 160]. Strikingly, this protective effect was shown to only occur in CerS6-deficient mice and not in CerS5-deficient mice [88], even though both CerS5 and CerS6 produce C16 ceramides, and the C16:0 ceramide content in the liver was reduced when either CerS5 or CerS6 was knocked-out. This finding contrasts with earlier reports that CerS5 knockout mice displayed improved glucose tolerance and lowered epididymal white adipose tissue inflammation after HFD [163]. This landmark study reveals that, specifically, CerS6-derived C16:0 ceramides accumulate in mitochondria-associated membranes and mitochondria in the liver of HFD mice, where these ceramides interact with the mitochondrial fission factor Mff and promote Mff-dependent mitochondrial fragmentation. These observations clearly demonstrate that despite producing ceramides with the same acyl chain length (C16), CerS5 and CerS6 activity can lead to drastically different physiological outcomes in response to palmitate excess and suggest CerS6 but not CerS5 as a possible therapeutic target against obesity and type 2 diabetes [88]. Additionally, the therapeutic potential of inhibiting CerS was recently further evaluated by knocking down CerS6 expression in an obese insulin resistance mouse model [164]. CerS6 knockdown was successful in reducing body fat and insulin resistance and in improving oral glucose tolerance [164], thus providing further support for the inhibition of CerS6 in the treatment of metabolic diseases such as obesity, type 2 diabetes and NASH [161, 164].

On the other hand, C24 ceramide production by CerS2 exerts a protective effect against glucose intolerance, hepatic insulin resistance and NAFLD [85, 162, 165]. Importantly, in both homozygous and heterozygous CerS2 knockout mice, lower C22 and C24 ceramide

levels are compensated by an increase in C16 ceramide synthesis, likely by CerS6, ultimately contributing to higher susceptibility to steatohepatitis and hepatic insulin resistance [85, 161, 166, 167]. In addition, homozygous null CerS2 knockout mice displayed abnormal liver function [116], reinforcing the notion that very-long-chain ceramides play key protecting roles and demonstrating the importance of preventing off-target CerS2 inhibition.

As described above, C18 ceramide production by CerS1 has been proposed to drive insulin resistance in skeletal muscle in humans [159]. Likewise, CerS1 and C18 ceramides have been shown to be elevated in obese mice, and genetic ablation of CerS1 in skeletal muscle of mice led to increased insulin sensitivity [168]. These observations suggested CerS1 as a candidate therapeutic target for the treatment of insulin resistance [168]. A CerS1-specific small-molecule inhibitor [169] successfully reduced C18 ceramide levels and stimulated mitochondrial fatty acid oxidation in skeletal muscle, but did not prevent insulin resistance in mice fed with HFD. However, CerS1 inhibition led to a 50 % decrease in C18 ceramide levels in skeletal muscle, compared to a 90 % decrease in knockout mice [168]. Therefore, further validation of CerS1 as a candidate therapeutic target is required [161].

1.2.5 Known CerS inhibitors

Despite the great interest in achieving isoform-specific inhibition of particular ceramide synthases, most known inhibitors to date are non-specific or display low potency. A few examples are described below.

One of the best-described CerS6 inhibitors is the fungal toxin Fumonisin B1 (FB1) (reviewed in [170]). FB1 is the most prevalent member of the fumonisin family of mycotoxins produced by *Fusarium* fungi. It is a carcinogenic compound and has also been proposed to lead to birth defects [171, 172]. Fumonisin B1 disrupts sphingolipid biosynthesis by directly

inhibiting ceramide synthesis. It is a potent non-selective inhibitor of all human CerS [169, 173] and appears to be competitive towards both sphinganine and acyl-CoA substrates [173].

In addition, the sphingosine-1-phosphate receptor agonist FTY720 (fingolimod), approved for the treatment of multiple sclerosis, has been shown to also act as a non-selective CerS inhibitor [169, 174]. Furthermore, a number of FTY720 derivatives that achieved a low level of potency (μM range) and selectivity have been described [175, 176], including derivatives of a chiral analogue of FTY720, AAL(S). These AAL(S) derivatives displayed modest selectivity for CerS1 [176]. Recently, Turner and colleagues [Ref. 169] described the development of a novel potent CerS1-specific inhibitor, P053, based on their previously described AAL(S) analogues [169]. This finding demonstrated for the first time the possibility of achieving isoform-specific inhibition of CerS isoenzymes.

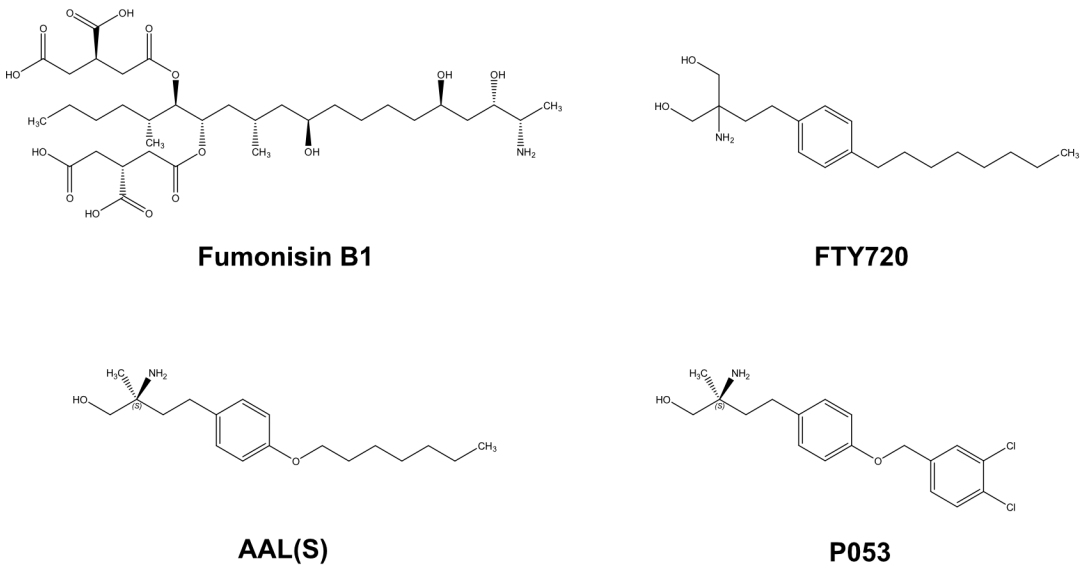


Figure 1.8. Chemical structures of known CerS inhibitors described in Section 1.2.5.

1.2.6 Fatty acid elongation

In the previous sections, I described the central role of ceramide synthases in determining the acyl chain composition of ceramides. However, one must also consider the role of the fatty acid elongation machinery, as it generates the different acyl-CoAs required for the N-acylation of sphingoid bases by the ceramide synthases. Therefore, it is fundamental to understand the molecular mechanisms underlying the synthesis of the different acyl-CoA precursors. For this reason, here, I will describe fatty acid elongation from the perspective of sphingolipid biosynthesis.

Very-long-chain fatty acids (VLCFAs) (with 20 or more carbons in the acyl chain) are incorporated into very-long-chain dihydroceramides via the N-acylation of dihydrosphingosine in the *de novo* ceramide synthesis pathway [177]. The resulting very-long-chain ceramides are capable of performing unique cellular functions that cannot be performed by long-chain ceramides, as described in Section 1.2.1. The majority of saturated or monounsaturated very-long-chain fatty acids are incorporated into sphingolipids [177], whereas polyunsaturated very-long-chain fatty acids are generally incorporated into glycerolipids. An exception to this is the presence of ceramides and sphingomyelin containing very-long-chain polyunsaturated fatty acids in germ cells of mammals [178], where they play a critical role during spermatogenesis [114]. Moreover, C20 polyunsaturated fatty acids can also be used as the precursors for the biosynthesis of eicosanoids [179].

In eukaryotes, fatty acids are synthesised and initially elongated to C16 or C18 in the cytoplasm [180, 181]. These are then converted to acyl-CoAs [180] and further elongated to VLCFAs by a four-enzyme cycle in the ER membrane that sequentially adds two carbon units to their acyl chain (Fig. 1.9).

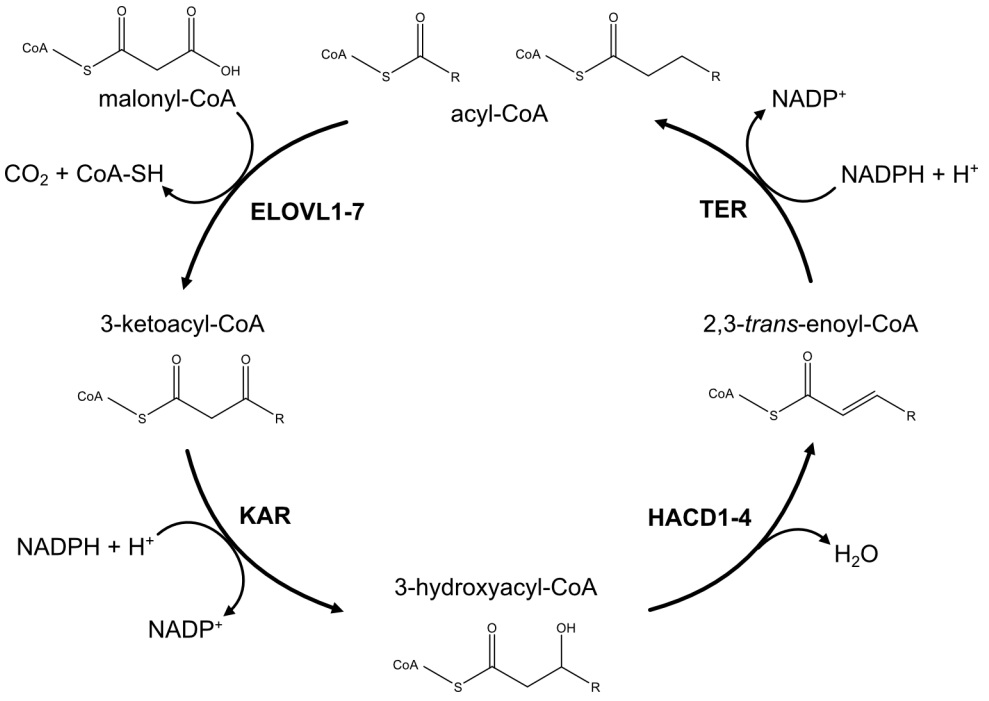


Figure 1.9. Fatty acid elongation cycle in the ER membrane. This cycle is conserved in eukaryotes, but the enzymes listed in this diagram correspond to the names of the mammalian enzymes.

The first rate-limiting step of this cycle is catalysed by 3-keto acyl-CoA synthases (KCSs; ELOVLs in mammals). These enzymes transfer a two-carbon unit from malonyl-CoA to an acyl-CoA through a condensation reaction that yields a 3-keto acyl-CoA product. This 3-keto acyl-CoA is then reduced to 3-hydroxyacyl-CoA by 3-keto acyl-CoA reductase (KAR in mammals) [182] and is subsequently dehydrated by 3-hydroxyacyl-CoA dehydratases (HACD1-4 in mammals), yielding a 2,3-*trans*-enoyl-CoA [183]. Lastly, the double bond of the 2,3-*trans*-enoyl-CoA is reduced by *trans*-2-enoyl-CoA reductase (TER in mammals) [182], resulting in an acyl-CoA with two additional carbons in the acyl chain. These acyl-CoAs can then re-enter the cycle through the action of KCSs for incorporation of more carbon units, or can exit the cycle and be used for other purposes, for example, in *de novo* ceramide biosynthesis.

1.2.7 3-keto acyl-CoA synthases (KCSs)

Despite catalysing the same addition of a two-carbon unit from malonyl-CoA to an acyl-CoA, two different enzyme families of KCSs participate in fatty acid elongation in eukaryotes. Plants and some protists use KCSs related to the *A. thaliana* fatty acid elongase 1 (FAE1) [184-186]. FAE1-type KCSs are anchored to the membrane but possess a soluble catalytic domain related to keto acyl-ACP synthase III and polyketide synthases, which relies on a conserved Cys-His-Asp triad for catalysis [186]. On the other hand, in mammals [187], yeast [188, 189], some protists [190-192] and other eukaryotes [184], this condensation reaction is catalysed by a family of integral membrane enzymes predicted to have 5 or 7 transmembrane helices, designated as ELOs (Elongation of Fatty Acids Proteins).

1.2.8 ELO elongases

ELOs lack a strictly conserved cysteine or other sequence motifs known to participate in decarboxylating Claisen-like condensation reactions (carbon-carbon bond formation between malonyl-CoA or malonyl-ACP and acyl thioesters of CoA or ACP [193]), and therefore likely perform a different catalytic mechanism [194]. Importantly, ELOs have a conserved HxxHH motif, which is required for catalysis, although its precise role remains unclear [194, 195]. In mammals, there are 7 ELOs, referred to as ELOVLs (Elongation of Very Long Chain Fatty Acids Proteins), with different tissue distribution and substrate specificity [153, 187, 196, 197] (Fig. 1.10). ELOVL elongases determine the acyl chain specificity in the fatty acid elongation cycle, thus regulating the availability of different acyl-CoA precursors (except palmitoyl-CoA) for ceramide synthesis.

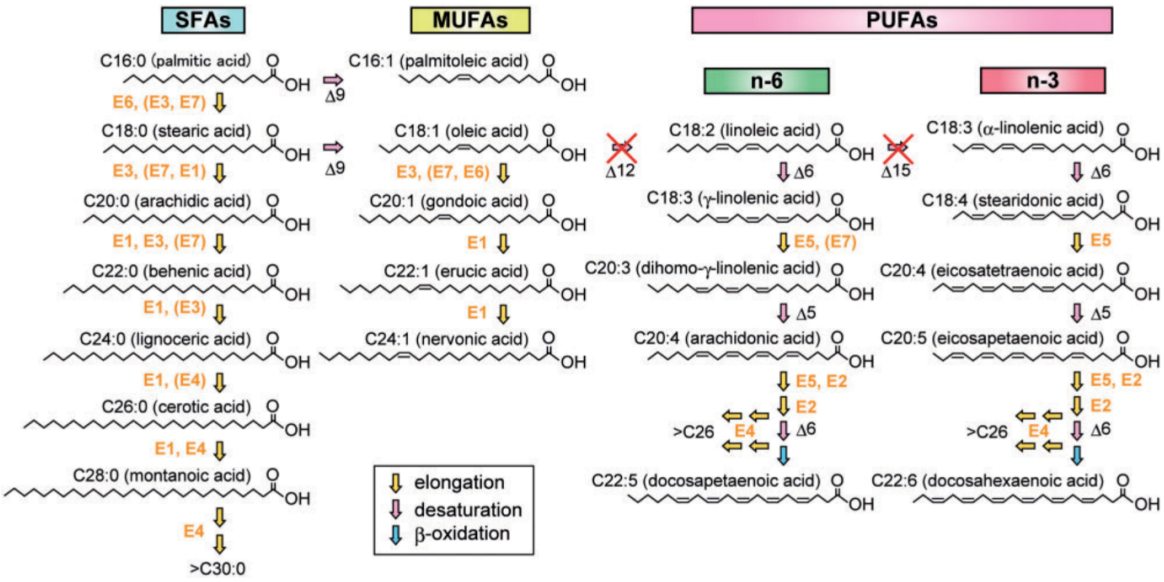


Figure 1.10 Human fatty acid elongation pathways. The ELOVL enzyme responsible for each elongation step is shown in orange (E1-7, representing ELOVL1-7); fatty acid desaturase enzyme numbers are shown as Δx, where x denotes the position where a double bond is introduced. Humans lack Δ12 and Δ15 desaturases (red crosses). SFAs = saturated fatty acids; MUFAs = monounsaturated fatty acids; PUFAs = polyunsaturated fatty acids. Reproduced from [177], with permission.

1.2.9 ELOs: disease implications and therapeutic potential of their inhibition

In humans, inherited mutations in ELOVL elongases are associated with genetic disorders including Stargardt-like macular dystrophy (STGD3) [198], spinocerebellar ataxia-34 and -38 (SCA34, SCA38) [199, 200] and neuro-ichthyotic disorders [201-204]. In addition, X-linked adrenoleukodystrophy (X-ALD), which is caused by mutations in the peroxisomal transporter ABCD1, leads to VLCFA accumulation due to elongation by ELOVL1 [205]. Also, mouse studies revealed that ELOVL1 ablation resulted in epidermal barrier defects [113] and ELOVL3 deletion led to hair loss and disruption of the skin water barrier [206].

Observations from mouse knockout models suggest that ELOVL activity plays a crucial role in the context of metabolic disease. Particularly, ELOVL5 knockout mice developed hepatic steatosis [207], whereas ELOVL6 ablation protected mice from obesity-induced insulin resistance [208], indicating that isoform-specific inhibition of ELOVLs could be a possible

route to treat certain metabolic diseases, including type 2 diabetes [208]. Furthermore, recent studies led to the proposal of ELOVL elongases as potential targets of novel cancer therapies [209, 210] and X-linked adrenoleukodystrophy [211].

Additionally, *de novo* fatty acid biosynthesis in trypanosomes primarily depends on enzymes of the ELO family of elongases, as they lack a cytoplasmic fatty acid synthase complex (FAS I) [190, 191, 212]. Trypanosome ELO elongases initiate fatty acid elongation from short C4-CoA units [190], whereas human ELOVLs only act on longer acyl-CoA substrates. Therefore, this pathway could be evaluated as a novel therapeutic target against these parasites [191, 195].

Overall, ELOVL elongases are responsible for determining the synthesis of VLCFAs with different acyl chain lengths and thus play a key role in regulating the availability of substrates for ceramide and sphingolipid biosynthesis. However, despite their crucial physiological role, very little is known about their molecular mechanisms at the enzyme level. The unusual presence of an essential histidine-rich motif in an enzyme catalysing a Claisen-like condensation reaction suggests ELOs likely use a novel mechanism [194]. Therefore, understanding the catalytic mechanism of this family of enzymes could enable developing novel inhibitors of fatty acid elongation.

1.3 Aims of this thesis

In this thesis, I focused on three families of human integral membrane proteins implicated in either disease or drug interactions, and aimed to obtain structural and/or mechanistic insights into their function. Importantly, no structural information was available for any of these proteins at the start of my DPhil project.

MFSD10 is annotated as a member of the drug:H^+ antiporter family 1 of MFS transporters, of which only bacterial members have been structurally and biochemically characterised in detail to date. ELOVLs and ceramide synthases are the two enzyme families responsible for acyl chain specificity in fatty acid elongation and sphingolipid biosynthesis, respectively, and have been proposed as attractive targets for the development of novel therapies for metabolic diseases. However, although their central roles in fatty acid and lipid metabolism have been the subject of extensive research over the last three decades, their molecular mechanisms remain poorly understood.

In Chapter 3, I solved the crystal structure of human MFSD10 and confirmed that it functions as an NSAID:H^+ antiporter using a fluorescence-based assay. The structure provides insights into the recognition of organic anions in the substrate-binding site of MFSD10 and suggests a potential mechanism of proton coupling. In Chapter 4, I followed up on the crystal structure of human ELOVL7 recently solved by others in the Integral Membrane Proteins (IMP) group of the SGC by characterising the mechanism of this enzyme using denaturing intact protein mass spectrometry. My results revealed a novel mechanism for this key enzyme in fatty acid biosynthesis, involving an unexpected histidine nucleophile. In Chapter 5, I optimised the recombinant expression and purification of human ceramide synthases 5 and 6, obtained and validated nanobodies against both proteins, and optimised cryo-EM conditions for CerS6 samples. This work provides an advanced position for future structural studies to obtain valuable three-dimensional information on these important drug targets. Overall, this thesis presents a combination of X-ray crystallography, mass spectrometry, electron microscopy and activity assays, highlighting the advantage of combining these powerful techniques to interrogate the molecular mechanisms of proteins of interest.

2 Materials and methods

2.1 General methods

2.1.1 Constructs

MFSD10 and ELOVL7 constructs were available prior to my arrival to the group. Some CerS1, CerS5 and CerS6 were likewise previously generated by others, and I designed additional construct for each of these CerS enzymes. Full details on which CerS constructs were designed by me and others are available in Section 5.2.1 and Appendix 2. Information relevant to all constructs is provided below, regardless of when they were generated.

The *Homo sapiens MFSD10* gene, encoding the MFSD10 protein (Uniprot ID: Q14728), was used to generate several N- and C-terminal truncations. For expression in *Sf*9 cells, the DNA coding for the MFSD10 crystallisation construct (Met1-Lys444) was cloned by the Biotechnology group of the SGC into the baculovirus transfer vector pFB-CT10HF-LIC (available through Addgene), containing C-terminal TEV-cleavable 10x His and FLAG tags for purification (Fig. A.3, Appendix 5). Likewise, for expression in mammalian cells, the DNA coding for this crystallisation construct was cloned into the baculovirus transfer vector pHTBV1.1-CTGFP-SIII-10H, containing a C-terminal TEV-cleavable GFP fusion followed by Twin-Strep and 10x His tags (Fig. A.4, Appendix 5). This expression vector, as well as pHTBV1.1-LIC and pHTBV1.1-CT10H-SIII-LIC, described below, were adapted from the BacMam vector pHTBV1.1 (provided by Frederick Boyce, Massachusetts General Hospital, Cambridge, MA), by the Biotechnology group of the SGC.

The *Homo sapiens ELOVL7* gene, encoding the full-length ELOVL7 protein (Uniprot ID A1L3X0), was cloned by the Biotechnology group of the SGC into the baculovirus transfer vector pFB-CT10HF-LIC, containing C-terminal TEV-cleavable 10x His and FLAG tags

for purification. The His150Ala and His181Ala mutants were generated by Dr Laiyin Nie using site-directed mutagenesis [213].

The *Homo sapiens CerS1*, *CerS5* and *CerS6* genes, encoding the CerS1, CerS5 and CerS6 proteins (Uniprot IDs: P27544, Q8N5B7 and Q6ZMG9, respectively), were used to generate several N- and C-terminal truncations. For each truncation designed, the DNA coding for the truncated constructs was cloned into either the baculovirus transfer vector pFB-CT10HF-LIC for expression in *Sf*9 cells, or into the baculovirus transfer vectors pHTBV1.1-LIC (containing C-terminal TEV-cleavable 10x His and FLAG tags) (Fig. A.5, Appendix 5) or pHTBV1.1-CT10H-SIII-LIC (containing C-terminal TEV-cleavable 10x His and Twin-Strep tags) (Fig. A.6, Appendix 5) for expression in mammalian cells. Avi-tagged CerS5 and CerS6, as well as the CerS6 glycosylation mutant, were synthesised and cloned into pHTBV1.1-CT10H-SIII-LIC by Twist Bioscience.

2.1.2 Protein expression

For expression in *Sf*9 or mammalian cells, baculoviral DNA was generated by transposition of DH10Bac with the baculovirus transfer vectors described in 2.1.1. Baculovirus was then produced by transfecting *Sf*9 cells (Thermo Fisher Scientific) with the baculoviral DNA, using the transfection reagent Insect GeneJuice (Merck Millipore). The virus was amplified twice by infecting mid-log *Sf*9 cells (2 x 10^6 cells mL^{-1}) in the presence of 2 % fetal bovine serum and incubating for 72 hours at 27 ºC. For large-scale protein expression in *Sf*9 cells, each litre of mid-log *Sf*9 cells grown in Sf-900 II serum-free medium (Thermo Fisher Scientific), in 2 L roller bottles, was infected with 5 mL L^{-1} of baculovirus and incubated in an orbital shaker for 72h at 27 ºC. Cells were harvested by centrifugation, washed with Phosphate Buffered Saline (PBS), pelleted again by centrifugation, flash-frozen in liquid N_2 and stored at -80 ºC.

For expression in Expi293F or Expi GnTI- cells (Thermo Fisher Scientific), 1 L of Expi293F cells at a density of 2 x 10^6 cells mL^{-1} (in 2 L roller bottles) in Freestyle 293 expression medium (Thermo Fisher Scientific) was transduced with 30 mL of baculovirus (third passage) and supplemented with 5 mM sodium butyrate (Sigma-Aldrich). Cells were incubated in a humidity-controlled orbital shaker at 37 ºC with 8 % CO_2 for 48h and harvested as described above.

2.1.3 Large-scale protein purification

2.1.3.1 MFSD10

For the purification of MFSD10 expressed in *Sf*9 cells, cells were thawed, resuspended in Buffer A (50 mM HEPES pH 7.5, 200 mM NaCl) supplemented with Roche protease inhibitor cocktail tablets to a total volume of 50 mL per L of original cell culture volume, and lysed by two passes at 15 kpsi through an Emulsiflex C5 homogeniser (Avestin). The lysate was solubilised by adding 1 % (w/v) Octyl Glucose Neopentyl Glycol (OGNG; Anatrace) and 0.1 % (w/v) cholesteryl hemisuccinate (CHS; Sigma-Aldrich) and incubating the mixture for 1h at 4 ºC with gentle rotation. Insoluble material was removed by centrifugation (1h at 35,000 x g). All subsequent steps were performed at 4 °C unless otherwise indicated. Solubilised protein was then purified from the supernatant by batch binding to Co^{2+} TALON resin (Clontech) for 2h, in the presence of 5 mM imidazole. From this point onwards, all buffers were supplemented with 0.12 % (w/v) OGNG and 0.012 % (w/v) CHS. The resin was transferred to a gravity column and washed with 30 column volumes of TALON Washing Buffer (Buffer A supplemented with 20 mM imidazole). The protein was eluted with TALON Elution Buffer (Buffer A supplemented with 250 mM imidazole) and immediately exchanged into Buffer A using PD-10 columns (GE Healthcare). The C-termi-

nal tag was cleaved overnight with TEV protease at a 1:10 weight ratio. The following morning, TEV protease and uncleaved protein were removed by binding to TALON resin for 1h. Cleaved protein was concentrated using a 20 mL 100 kDa cut-off concentrator (Corning) and applied onto a Superose 6 Increase 10/300 GL column (GE Healthcare) or onto a Superdex 200 Increase 10/300 GL column (GE Healthcare), pre-equilibrated in SEC buffer (20 mM HEPES pH 7.5, 200mM NaCl). Peak fractions were pooled and purity was assessed by SDS-PAGE. For crystallisation purposes, purified protein was processed immediately. For other applications, the protein was flash-frozen in liquid N_2 and stored at -80 ^{o}C.

Purification of MFSD10 expressed in Expi293F cells was carried out essentially as described above for *Sf*9-expressed MFSD10, with the following changes: Buffer A was replaced with Buffer B (50 mM HEPES pH 7.5, 200 mM NaCl, 5 % (v/v) glycerol) throughout the purification. Tagged protein was purified from total detergent-solubilised material by batch binding to Strep-Tactin XT Superflow resin (IBA Lifesciences) for 2h. The resin was transferred to a gravity column and washed with 30 column volumes of Strep Washing Buffer 1 (20 mM HEPES pH 7.5, 200 mM NaCl, 5 % (v/v) glycerol), followed by 30 column volumes of Strep Washing Buffer 2 (20 mM HEPES pH 7.5, 200 mM NaCl, 5 % (v/v) glycerol, 10 mM $MgCl_2$, 1 mM ATP). The bound protein of interest was then eluted with Strep Elution Buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 5 % (v/v) glycerol, 100 mM D-biotin (Fluorochem)). TEV protease was then added at a 1:10 weight ratio, and the C-terminal tag was cleaved overnight. All subsequent steps were carried out exactly as described above for the purification of *Sf*9-expressed protein.

2.1.3.2 ELOVL7

ELOVL7 was purified as described above for MFSD10 expressed in *Sf*9 cells. However, for mass spectrometry experiments requiring covalent intermediate identification, ELOVL7 purification was carried out only until the PD10 buffer exchange step that followed elution from the TALON resin, in order to preserve the presence of unmodified enzyme (discussed in Section 4.2.3).

2.1.3.3 Ceramide synthases

*Sf*9-expressed CerS6-c003 was purified as described for MFSD10, with the following changes: all buffers before IMAC elution were supplemented with 5 % (v/v) glycerol; lysate was solubilised with 1 % (w/v) Lauryl Maltose Neopentyl Glycol (LMNG; Anatrace) (with or without supplementation with 0.1 % (w/v) CHS); purification buffers used after the solubilisation step contained 0.003 % (w/v) LMNG (with or without supplementation with 0.0003 % (w/v) CHS).

Purification of CerS5-c009 and CerS6-c009 (containing TEV-cleavable C-terminal 10x His and FLAG tags) expressed in Expi293F or Expi GnTI- cells was carried out as follows. Cells were thawed and re-suspended in Buffer B (50 mM HEPES pH 7.5, 200 mM NaCl, 5 % (v/v) glycerol) supplemented with Roche protease inhibitor cocktail tablets, to a total volume of 50 mL per L of original cell culture volume. All subsequent steps were carried out at 4 ^{o}C unless otherwise indicated. The cells were directly solubilised by incubating the mixture for 1h with gentle rotation after adding either 1 % (w/v) OGNG and 0.1 % (w/v) CHS (for CerS5-c009) or 1 % (w/v) LMNG and 0.1 % (w/v) CHS (for CerS6-c009). Cell debris was removed by centrifugation (1h at 35,000 x g). Solubilised protein was then purified from the supernatant by batch binding to Co^{2+} TALON resin (Clontech) for 2h, in the presence of 5

mM imidazole. The resin was collected by centrifugation (5 min at 700 x g) and the supernatant was discarded. From this point onwards, all buffers were supplemented with either 0.1 % (w/v) OGNG, 0.01 % (w/v) CHS (CerS5-c009) or 0.003 % (w/v) LMNG, 0.0003 % (w/v) CHS (CerS6-c009), unless otherwise specified. TALON resin was washed and His-tagged proteins were eluted and subsequently buffer exchanged as described in 2.1.3.1 for the purification of *Sf*9-expressed MFSD10, with the exception that Buffer A was replaced with Buffer B. At this stage, due to the high background binding of contaminants from mammalian expression systems to TALON resin, anti-FLAG M2 gel (Sigma-Aldrich) was added, and the mixture was incubated for 1h with agitation before being transferred to a gravity column. The unbound flow-through was collected and the gel was washed with Buffer B before bound protein was eluted with Buffer B supplemented with 0.2 mg mL^{-1} 3x FLAG peptide. The TEV cleavage and reverse IMAC steps were carried out as described in 2.1.3.1 for MFSD10. Cleaved protein was concentrated using a 20 mL 100 kDa cut-off concentrator (Corning) and applied onto a Superdex 200 Increase 10/300 GL column (GE Healthcare), pre-equilibrated in SEC buffer. In the case of CerS6-c009, the detergent concentration was lowered at the SEC step to 0.002 % (w/v) LMNG, 0.0002 % (w/v) CHS.

Purification of CerS5-c017 and CerS6-c025 (containing TEV-cleavable C-terminal 10 x His and Twin-Strep tags) expressed in Expi293F or Expi GnTI- cells was carried out as described for CerS5-c009 and CerS6-c009, with a few exceptions. Note that, for each CerS5-c017 and CerS6-c025, detergent concentrations used were the same as described for CerS5-c009 and CerS6-c009, respectively. Solubilised tagged proteins were purified by batch binding to Strep-Tactin XT Superflow resin (IBA Lifesciences) for 2h, with gentle rotation. The resin was washed with 30 column volumes of Strep Washing Buffer 1, followed by 30 column volumes of Strep Washing Buffer 2. The bound protein of interest was then eluted with

Strep elution buffer and cleaved overnight with TEV protease, as described above. As before, a reverse IMAC step was included to remove uncleaved protein, free tags and His-tagged TEV protease. The cleaved protein was concentrated as before and further purified by SEC.

The CerS6-c025 GDN sample for cryo-EM was prepared similarly to the LMNG / CHS sample, with the following exceptions. After initial solubilisation using 1 % (w/v) LMNG, 0.1 % (w/v) CHS, no further LMNG / CHS was added. Instead, the protein was progressively exchanged into GDN (Anatrace) during all subsequent steps. Strep-Tactin XT Superflow resin (IBA Lifesciences) was washed, and the protein was eluted using Strep Washing and Strep Elution Buffers supplemented with 0.05 % (w/v) GDN. Likewise, the SEC buffer was supplemented with 0.01 % (w/v) GDN. The final sample was concentrated to 5 mg mL^{-1} using a 2 mL Vivaspin 100 kDa cut-off concentrator (Sartorius).

In all cases, purified CerS proteins were either processed immediately for crystallisation or electron cryomicroscopy, or were flash frozen in liquid N_2 and stored at -80 ºC for other applications.

2.1.4 Protein reconstitution into proteoliposomes

The lipids 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (POPG) were obtained as 25 mg mL^{-1} stocks in chloroform (Avanti Polar Lipids). These were combined in a round-bottomed flask (3:1 (w/w) POPE:POPG), and chloroform was evaporated using a rotary evaporator to obtain a thin lipid film. The lipids were then re-suspended in 20 mM HEPES pH 7.5, 200 mM NaCl, sonicated, and subjected to 5 freeze-thaw cycles (in liquid N_2) to form large

multilamellar vesicles (LMVs). These were then extruded 21x through a 400nm polycarbonate filter (Avestin) to yield large unilamellar vesicles (LUVs), which were then disrupted with 0.15 % (w/v) Triton X-100 (Sigma-Aldrich) (determined to enable membrane saturation with detergent by monitoring OD_{540}). To produce MFSD10 proteoliposomes for functional assays, purified protein was added at 1:50 (w/w) protein:lipid and the mixture was incubated for 30 minutes at 4 ºC. Triton was then removed through the sequential addition of 40 mg mL^{-1} Bio-Beads SM-2 (Bio-Rad). At every step, this mixture was incubated with gentle rotation, at 4 ºC. Initially, Bio-Beads were added and incubated for 2h. Then, the Bio-beads were replaced and the mixture was incubated overnight. The next morning, the Bio-Beads were again exchanged and the mixture was incubated for 2h. Proteoliposomes were collected by ultracentrifugation (1h at 100,000 x g), resuspended in the desired buffer (final concentration: 25 mg mL^{-1} lipid), flash-frozen in liquid N_2, and stored at -80ºC. Control liposomes were prepared in the same manner, by adding SEC buffer in place of protein.

For alpaca immunisation with CerS5 and CerS6 proteoliposomes for nanobody generation, proteoliposomes were prepared as described above, but the protein to lipid ratio used was 1:20 (w/w) and the proteoliposomes were resuspended to 40 mg mL^{-1} lipid after ultracentrifugation.

2.1.5 NanoDSF label-free thermal unfolding assays

Thermal unfolding experiments were carried out on a Prometheus NT.48 instrument (NanoTemper Technologies) using 5 µM of the relevant protein in SEC buffer (20 mM HEPES pH 7.5, 200 mM NaCl, supplemented with either 0.12 % (w/v) OGNG, 0.012 % (w/v) CHS (MFSD10) or 0.003 % (w/v) LMNG, 0.0003 % (w/v) CHS (CerS6)). Proteins were unfolded by increasing the temperature from 20 ºC to 95 ºC, at a rate of 1 ºC min^{-1}. Protein unfolding was monitored by tracking the ratio of fluorescence emission at 350 nm / 330 nm, and the

melting temperature was determined from the inflection point of the transition using the PR.ThermControl software (NanoTemper Technologies).

In thermostabilisation experiments, the proteins were incubated at 4 ºC without compounds or with 20 µM or 100 µM of the relevant compound for 1h, prior to being subjected to a temperature gradient. The temperature shifts (ΔT_m) were calculated as the difference between the melting temperatures obtained with and without compounds. For clarity, samples without compounds were incubated with the corresponding concentration of DMSO present in the samples incubated with compounds. In the case of putative MFSD10 substrates, compound stocks were prepared at a concentration of 50 mM in DMSO. In the case of CerS inhibitors, 10 mM stocks were prepared. All compounds used in nanoDSF experiments were purchased from Sigma-Aldrich.

2.1.6 Denaturing intact protein LC-ESI-MS analysis of purified protein samples

The intact masses of purified protein samples were analysed by denaturing intact protein mass spectrometry, conducted using an Agilent 1290 Infinity LC System in line with an Agilent 6530 Accurate-Mass quadrupole time-of-flight mass spectrometer (Agilent Technologies).

Reversed-phase chromatography was carried out in-line prior to MS analysis. The flow rate was set to 0.5 mL min^{-1}, and the solvent system consisted of 0.1 % Optima LC–MS grade formic acid (Fisher Chemical) in high-performance liquid chromatography electrochemical grade water (Fisher Chemical) (solvent A) and 0.1 % formic acid in Optima LC–MS grade methanol (Fisher Chemical) (solvent B). Typically, 1-4 µg of purified protein (at 1-2 mg mL^{-1}) was diluted to 60 µL in 30 % methanol in 0.1 % formic acid, and injected onto a ZORBAX StableBond 300 C3 HPLC column (5 µm particle size; 2.1 x 150 mm) (Agilent

Technologies). Initial conditions were 30 % B. After 1 min, a linear gradient from 30 % B to 95 % B was applied over 7 min to elute the samples. This was then followed by a final 2 min isocratic elution with 95 % B, before the system was re-equilibrated between samples with 30 % B. A blank injection was included between runs to prevent sample carry-over contamination.

The mass spectrometer was operated in positive electrospray ionisation (ESI+), 2 GHz mode (ESI+). MS parameters were capillary voltage of 4,000 V, fragmentor voltage of 250 V, and collision energy of 0 V. The drying gas was supplied at 350 ºC with a flow rate of 12 L min^{-1}. The nebulizer pressure was 60 psi.

Data were acquired between 100-3200 *m*/*z* and analysed using MassHunter Qualitative Analysis v.B.07.00 (Agilent Technologies) software. Peaks between 650-3200 *m*/*z* in the sum of the mass spectra obtained during protein elution were deconvoluted using the maximum entropy charge deconvolution algorithm. The nature of the deconvoluted mass peaks was assigned initially based on the theoretical average masses of purified protein constructs and subsequently based on mass shifts between deconvoluted mass peaks. The 'unmodified' ELOVL7 , CerS5 and CerS6 peaks (i.e. without further addition of post-translational modifications) corresponded to loss of the initiator methionine (theoretical −131.20 Da) and subsequent acetylation of the new N-terminus (theoretical +42.04 Da). Two 'unmodified' MFSD10 peaks were observed, with one corresponding to loss of the initiator methionine (without acetylation) (theoretical −131.20 Da) and the other corresponding to loss of the first 5 N-terminal residues. Experiments describing the formation and detection of ELOVL7 covalent intermediates are described in Sections 2.3.2.2 – 2.3.2.4. The interpretation of other mass shifts for all proteins is described throughout Chapters 3-5.

2.1.7 SDS-PAGE

For sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis of purified protein samples, typically 15 µl of the protein sample were mixed with 5 µl 4x NuPAGE LDS sample buffer (Novex) supplemented with 100 mM dithiothreitol (DTT). These samples were then loaded into NuPAGE 4-12% Bis-Tris Midi gels (Invitrogen), along with protein molecular weight markers (Precision Plus Protein Unstained Protein Standards (Bio-Rad)). The gels were run for 1 h at 145 V, using NuPAGE MES SDS running buffer (Invitrogen). At the end of the run, the protein bands were detected by staining the gels using InstantBlue Coomasie protein stain (Expedeon).

2.2 Methods specific to the MFSD10 project

2.2.1 Membrane isolation

During optimisation of MFSD10 crystallisation, MFSD10 was purified from isolated membranes (Section 3.3.8). To prepare membranes, MFSD10-overexpressing *Sf*9 cells were re-suspended in PBS supplemented with Roche protease inhibitor cocktail tablets and lysed by two passes at 15 kpsi through an Emulsiflex C5 homogeniser (Avestin). Cell debris was removed by centrifugation (10 min at 9,500 x g) and membranes were subsequently isolated by ultracentrifugation (1h 30 min at 160,000 x g). Membranes were washed twice by resuspension in a high-salt buffer consisting of 1 M potassium acetate, 20 mM HEPES pH 7.5 followed by ultracentrifugation at 160,000 x g for 1h, and once in a low-salt buffer consisting of 20 mM KCl, 20 mM HEPES pH 7.5. Washed membranes were re-suspended in PBS, flash-frozen in liquid N_2 and stored at -80 ^{o}C. On the day of purification, membranes were thawed, re-suspended in 50 mM HEPES pH 7.5, 200 mM NaCl and solubilised with 1 %

(w/v) OGNG, 0.1 % (w/v) CHS. MFSD10 was then purified as described in Section 2.1.3.1, starting from the solubilisation step.

2.2.2 Protein crystallisation

MFSD10 was crystallised using the lipidic cubic phase (LCP) technique. Initial crystals were obtained by mixing purified MFSD10 (expressed in *Sf*9 cells) at 20 mg mL^{-1} with monoolein (1-Oleoyl-rac-glycerol) (Nu-Chek Prep) in a 2:3 (w/w) protein:lipid ratio using a syringe-coupled device (Hamilton) to yield clear LCP. 50 nL of LCP were dispensed onto 96-well glass sandwich plates (Marienfield, Germany) and overlaid with 800 nL of crystallisation solution using a Mosquito LCP robot (TTP Labtech).

During attempts to obtain crystal structures of MFSD10-substrate complexes, MFSD10 was also crystallised *in meso* using a monoolein:cholesterol (9:1 (w/w)) mixture as the host lipid. Preparation of this lipid mixture is described in the next section. In all co-crystallisation attempts, the protein was incubated with substrates for at least 2 h at 4 °C prior to reconstitution in the lipidic cubic phase.

2.2.3 Preparation of monoolein:cholesterol mixture for optimised *in meso* crystallisation

The monoolein:cholesterol (9:1 (w/w)) mixture was prepared as follows. 100 mg of cholesterol from ovine wool (Avanti Polar Lipids) was measured into a glass vial and dissolved in 2 mL chloroform, to obtain a 50 mg mL^{-1} stock. Monoolein was pre-melted at 42 °C, and 720 mg were pipetted into a second glass vial. Then, 1600 µL of the 50 mg mL^{-1} cholesterol stock in chloroform (total 80 mg cholesterol) were added to the monoolein vial and mixed

by gentle swirling. The chloroform was subsequently evaporated under a gentle nitrogen stream to yield the final lipid mixture containing 9:1 (w/w) monoolein:cholesterol.

2.2.4 Crystal screening and X-ray data collection

Crystals were screened, and diffraction data were collected at 100 K in the I24 microfocus beamline (Diamond Light Source, UK) at a wavelength of 0.9687Å. Crystals were located by carrying out diffraction grid scans and evaluated by collecting three diffraction images 45º apart. For the datasets described in this work, diffraction data were collected for 360º, using 0.2º oscillations. A detailed description of the collection strategy is provided for the initial dataset in Section 3.3.1.

2.2.5 X-ray data processing and molecular replacement

For the initial dataset used to solve the MFSD10 structure, data were processed with XDS [214], scaled with XSCALE [214] and truncated and merged using Aimless [215]. The final 2.60 Å dataset contains 260º of data. Initial phases were obtained by molecular replacement using Phaser [216], using two separate search models corresponding to poly-alanine models of each 6TM bundle of the YajR structure [39] (PDB: 3WDO), generated by CHAINSAW [217].

Other MFSD10 datasets described in Sections 3.3.7 and 3.3.8 were processed in the same way, and initial phases were obtained by molecular replacement with the MFSD10 model.

2.2.6 Model building and refinement

For the initial dataset, initial electron density maps were sharpened with Parrot [218], and a starting auto-traced model was generated using Buccaneer [219]. The Buccaneer model was then manually rebuilt and completed by iterative cycles of manual model building in COOT

[220] and refinement using Refmac5 [221]. Two TLS groups (residues 18-88 and 89-443, determined using TLSMD [222]) were refined. The final model was refined with good geometry to R/R_{free} values of 20.6/25.7 %, and includes residues Gln18 to Gln443, 1 citrate ion, 2 monoolein molecules, and 60 solvent molecules.

For the analysis of the electron density in the substrate-binding site in other MFSD10 datasets described in Sections 3.3.7 and 3.3.8, initial models obtained after molecular replacement were also refined using Refmac5, as described above.

2.2.7 Pyranine assays

MFSD10 proteoliposomes were prepared as described in 2.1.4, resuspended in Internal Buffer (5 mM HEPES pH 6.8, 119 mM NaCl, 1 mM KCl, 2 mM $MgSO_4$) supplemented with 1 mM Pyranine, and subjected to three freeze-thaw cycles. Then, proteoliposomes were extruded 21x through first 400 nm, then 200 nm polycarbonate filters (Avestin), and unincorporated pyranine was removed using Bio-Spin 6 columns (Bio-Rad). For the assay, desalted proteoliposomes were diluted 1/25 into External Buffer (5 mM HEPES pH 6.8, 1 mM NaCl, 119 mM KCl, 2 mM $MgSO_4$, 1 mM pyranine), and a baseline was recorded by monitoring emission at 510 nm while alternating the excitation wavelength between 460 nm and 415 nm. Transport was initiated with 2 mM diclofenac (Sigma-Aldrich), and fluorescence was monitored for 2 minutes post-injection. Liposomes without protein were included as a control. To assess whether the liposomes generated are suitable for pyranine assays, their ability to prevent proton leakage and hold a pH gradient was assessed by recording fluorescence upon dilution into an external buffer with higher pH, in the absence of substrate (5 mM HEPES pH 7.5, 119 mM NaCl, 1 mM KCl, 2 mM $MgSO_4$). Fluorescence measurements were carried out in a 96-well plate format on a Clariostar plate reader (BMG Labtech), and

the results were analysed with Prism 9.0.2 (Graphpad). The fluorescence ratio (F_R) was calculated by dividing the fluorescence emission at 510 nm upon excitation at 460 nm by the fluorescence emission at 510 nm upon excitation at 415 nm. F_R was normalised to 1.0 at the time point of substrate addition to enable comparison between different experiments.

2.2.8 Fluorescence and absorbance spectra of Biliverdin and purified MFSD10

Purified MFSD10 at 6 mg mL^{-1} (125 µM) and 50 µM biliverdin hydrochloride (Sigma-Aldrich) were prepared in SEC buffer containing 0.12 % (w/v) OGNG, 0.012 % (w/v) CHS. Absorbance and fluorescence spectra were acquired using the cuvette port of a SpectraMAX M2 multi-mode microplate reader (Molecular Devices). Absorbance spectra were recorded from 280 nm to 720 nm. The wavelength for fluorescence excitation was 375 nm; fluorescence emission was recorded from 400 nm to 720 nm.

2.3 Methods specific to the ELOVL7 project

2.3.1 SEC-MALS

For SEC-MALS, 100 µL of purified ELOVL7 at 1 mg mL^{-1} was injected onto a Superdex 200 Increase 10/300 GL column (GE healthcare), pre-equilibrated in SEC buffer containing 0.1 % (w/v) OGNG, 0.01 % (w/v) CHS, using an OMNISEC RESOLVE system (Malvern Panalytical). Light scattering, refractive index and absorbance at 280 nm were monitored using an integrated OMNISEC REVEAL multi-detector module (Malvern Panalytical). Bovine serum albumin (100 µL at 1 mg mL^{-1}) was also injected and used for system calibration. Data were processed using OMNISEC v5.1 software. To enable calculation of the molecular weight of the protein fraction of the protein:detergent complex, the detergent dn/dc (the specific refractive index increment) was experimentally determined by generating a standard

curve of refractive index peak area versus injected detergent mass. This was achieved by injecting 100 µL of OGNG/CHS (10:1 (w/w)) at multiple concentrations (three replicates were carried out for each detergent concentration), determining the RI peak area for each sample and dividing it by the RI detector constant (K_{RI}) determined using the BSA standard sample. The slope of this standard curve, corresponding to the detergent dn/dc, was 0.127 mL g^{-1}. The protein dn/dc was assumed to be 0.185 mL g^{-1}.

2.3.2 Denaturing intact protein mass spectrometry

A detailed description of the setup used for denaturing intact protein mass spectrometry is available in Section 2.1.6. Sections 2.3.2.1-2.3.2.4 below refer to the 'stress test' and covalent intermediate formation experiments carried out for ELOVL7 in Chapter 3. All mass spectrometry experiments were conducted at least twice using distinct biological replicates, with similar results.

2.3.2.1 Analysis of the relative stability of the unmodified and modified forms of purified ELOVL7

LC-ESI-MS analysis of samples taken at various steps during ELOVL7 purification revealed that unmodified ELOVL7 enzyme (i.e. enzyme containing no covalent modifications) was lost at the end of the purification (Section 4.2.3), suggesting that the unmodified enzyme is inherently more unstable. As a result, a stress test was carried out to evaluate the relative stability of the unmodified and modified (i.e. containing the covalently linked 3-keto acyl-CoA adduct) species. ELOVL7 (solubilised in OGNG/CHS) obtained after the desalting step that followed IMAC elution was incubated for 1h at 37 ºC. Then, precipitated material was removed by centrifugation (20 min at 21,000 x g, at 4 ºC), and the soluble fraction was collected and diluted as described in Section 2.1.6 using 30 % methanol in 0.1 % formic

acid. Likewise, the pellet obtained after centrifugation (i.e. the precipitated material) was directly re-suspended in 30 % methanol in 0.1 % formic acid. Both samples were analysed by LC-ESI-MS. Comparison to the deconvoluted spectrum obtained for the initial material revealed that all the unmodified enzyme precipitated during this stress test, whereas the majority of the covalently modified enzyme remained in solution. This stress test was repeated with LMNG-solubilised ELOVL7, revealing that, in this detergent, the unmodified enzyme remained in solution at the end of the incubation period.

2.3.2.2 Covalent intermediate identification

Purified, tagged, post-IMAC wild-type ELOVL7 at 1.5 mg mL^{-1} was incubated for 2h at 37 ^{o}C with either no substrates, 100 µM C18:0-CoA (Avanti Polar Lipids) or 100 µM C18:3(n3)-CoA (Avanti Polar Lipids), in the presence or absence of 1 mM EDTA or 1 mM EGTA. At the end of the incubation period, the reaction was terminated by diluting 2 µL of each mixture into 30 % Methanol in 0.1 % formic acid. Acyl-enzyme intermediate formation was evaluated by LC-ESI-MS by monitoring the presence of additional desconvoluted mass peaks after incubation with the substrates, corresponding to covalent attachment at the substrate thioester carbonyl, resulting in thioester cleavage and loss of CoA. The theoretical mass additions on reaction with the substrates used are +266.47 Da for C18:0-CoA, and +260.42 Da for C18:3(n3)-CoA.

2.3.2.3 Evaluation of covalent acyl-enzyme intermediate formation using active site histidine mutants

To probe the site of covalent modification, the experiment described in 2.3.2.2 was repeated using the His150Ala and His181Ala mutants. However, these mutants displayed background peaks in the region of the deconvoluted mass spectra where the C18:0 intermediate would

be expected, and therefore this experiment was carried out using only C18:3(n3)-CoA. Covalent intermediate formation could be detected using the His181Ala mutant, but not using the His150Ala mutant, allowing identification of His150 as the nucleophile involved in covalent intermediate formation.

2.3.2.4 Reaction of malonyl-CoA with the covalent intermediate

ELOVL7 was purified as before, with the exception that it was initially solubilised in 1% (w/v) LMNG and purified using 0.002 % (w/v) LMNG. The LMNG-solubilised protein was used to probe the sequential nature of the reaction as the unmodified enzyme was more stable in this detergent, as assessed using the stress test described in Section 2.3.2.1. The purified, tagged, post-IMAC wild-type protein at 2 mg mL^{-1} was initially incubated for 15 min at 37 ^{o}C to form the covalent acyl-enzyme intermediate and any precipitated material at the end of this incubation was removed by centrifugation (30 min at 21,000 x g, at 4 ^{o}C). At this stage, the formation of the covalent acyl-enzyme intermediate was confirmed by denaturing intact protein mass spectrometry, and the protein was further incubated in the presence or absence of 200 µM malonyl-CoA (Sigma-Aldrich) for 3h at 4 ^{o}C. Analysis of the final mixtures revealed the loss of the covalent intermediate deconvoluted mass peak only in the sample containing malonyl-CoA.

2.4 Methods specific to the ceramide synthases project

2.4.1 Detergent screening (3 mL purifications)

For 3 mL-scale purification screening using 12 detergents, the cell pellets from 50 mL cultures were re-suspended in cold Buffer B (50 mM HEPES pH 7.5, 200 mM NaCl, 5 % (v/v) glycerol), supplemented with Roche protease inhibitor cocktail tablets. The re-suspended

cells were transferred to a row of a 24-well block (Greiner Bio-one) (1 construct per row) and lysed by sonication for a total time of 2 min (5 s on / 12 s off). All subsequent steps were performed at 4 ºC unless otherwise indicated. Each row of the 24-well block was transferred to a 96-well block (Greiner Bio-one) (2.5 mL in each well of the 24-well block were split equally between 2 wells in the 96-well block). For solubilisation, 1 % (w/v) of each detergent was added to each well (12 detergents per construct in columns 1-12 of the 96-well block). The block was sealed with a silicon mat and incubated with gentle rotation for 1h. Then, insoluble material was removed by centrifugation (35 min at 3,200 x g), and the supernatant was transferred into a new 96-well block. 50 µL Strep-Tactin XT Superflow resin (IBA Lifesciences) was added to each well of the new block, and the block was sealed as before and incubated with gentle rotation for 2h. The resin/supernatant mix was then transferred to a 2 mL 96-well filter block (Porvair) placed on top of a waste block. The flow-through containing unbound material was allowed to drip through the filter block and discarded. Then, the resin was washed with 2.5 mL (50 column volumes) of wash buffer (Buffer B supplemented with 3 x the critical micelle concentration (CMC) of the relevant detergent). Leftover wash buffer was removed by centrifugation (2 min at 300 x g), and the protein was eluted with 50 µL Strep elution buffer (50 mM HEPES pH 7.5, 200 mM NaCl, 5 % (v/v) glycerol, 100 mM D-biotin, supplemented with 3 x CMC of the relevant detergent). During this elution step, the resin was initially incubated with elution buffer for 20 min on an orbital shaker at room temperature. Due to the low sample volume, the eluates do not drip through the block during this incubation. Eluates were then collected on a 96-well V-bottom microtiter plate (Greiner Bio-one) by centrifugation (3 min at 500 x g). Samples were analysed by SDS-PAGE, and the identity of the bands suspected to correspond to the protein of interest was confirmed by in-gel tryptic digestion followed by LC-MS/MS, carried out by the Biotechnology group of the SGC.

Similar testing was carried out for constructs expressing C-terminal 10x His and FLAG tags. In this case, 50 µL Co^{2+} TALON resin was used per well, the resin was washed with Buffer B supplemented with 20 mM Imidazole-HCl pH 8.0, and proteins were eluted with Buffer B supplemented with 250 mM Imidazole-HCl pH 8.0. All other steps were carried out as described in the previous paragraph.

2.4.2 Small-scale purification screening

Small scale purifications were carried out for several CerS5 / CerS6 constructs and detergent combinations simultaneously. These CerS constructs contained C-terminal 10x His and FLAG tags and were expressed in Expi293F cells. Cells overexpressing each construct were thawed, re-suspended in Buffer B supplemented with Roche protease inhibitor cocktail tablets to a total volume of 50 mL per L of original cell culture volume, and mechanically lysed by two passes at 15 kpsi using an Emulsiflex C5 homogeniser (Avanti). All subsequent steps were performed at 4 °C unless otherwise indicated. The lysate was split so that each small scale purification corresponded to 250 mL of original cell culture volume (4 purifications per L of cell culture). The lysates were solubilised by adding 1 % (w/v) of the relevant detergent (and optionally 0.1 % (w/v) CHS) and incubating for 1h, with gentle rotation. Insoluble material was removed by centrifugation (1h at 35,000 x g). Solubilised protein was then purified from the supernatant by incubation with 250 µL Co^{2+} TALON resin (Clontech) for 1h 30min, with gentle rotation. The resin was collected by centrifugation (5 min at 700 x g). The supernatant was discarded, and the resin was transferred to a 2 mL 96-well filter block (Porvair), placed on top of a waste block. The resin was washed with 30 column volumes (7.5 mL) of Buffer B supplemented with 20 mM Imidazole-HCl pH 8.0. Tagged proteins were eluted with 2 x 250 µL Buffer B supplemented with 250 mM Imidazole pH 8.0 and buffer exchanged into Buffer B without imidazole using PD MidiTrap G-25 desalting columns (Cytiva). The C-terminal tags were cleaved overnight by incubating with 1:10 (w/w)

TEV:CerS. The following morning, uncleaved protein, cleaved tags and His-tagged TEV protease were removed by incubation with Co^{2+} TALON resin (Clontech). The flow-through containing unbound cleaved protein was concentrated to 50 μL using a 2 mL Vivaspin 50 kDa cut-off concentrator (Sartorius). The concentrated proteins were applied onto an SRT SEC-300 column (Sepax Technologies) pre-equilibrated in SEC buffer (typically containing 3 x CMC of the relevant detergent) using a Dionex HPLC system (Thermo Fisher Scientific). Fractions were collected and, when sufficient amounts of protein could be recovered, the melting temperature was determined by nanoDSF (described in Section 2.1.5).

2.4.3 *In vitro* biotinylation

For the *in vitro* biotinylation, Avi-tagged CerS5 and CerS6 constructs (CerS – GGGS – Avi – TEV – 10xHis – Twin-Strep) were purified from Expi GnTI- cells until the Strep elution step. As proteins were eluted from the Strep-TactinXT superflow resin (IBA Lifesciences) using 100 mM D-biotin (Fluorochem), *in vitro* biotinylation was carried out at this step of the purification protocol, at the same time as C-terminal tag cleavage. Thus, protein obtained after elution from the resin was supplemented with 15 mM $MgCl_2$, 15 mM ATP-NaOH pH 7.5, 50 mM bicine pH 8.3, 1:15 (w/w) BirA:CerS (for *in vitro* biotinylation) and 1:5 (w/w) TEV:CerS, 0.5 mM TCEP-NaOH pH 7.5 (for C-terminal tag cleavage). This mixture was incubated overnight at 4 °C, with gentle rotation, and the His-tagged BirA and TEV protease were removed the following morning using TALON resin. The remaining steps of the purification were carried out as described in Section 2.1.3.3. Biotinylation efficiency was estimated by denaturing intact protein mass spectrometry. BirA enzyme was kindly provided by Tina Bohstedt (Biotechnology group, SGC).

2.4.4 Nanobody expression and purification

2.4.4.1 Small-scale nanobody expression and purification for binding screening

For binding screening by Bio-Layer Interferometry, nanobodies were expressed in 50 mL scale. WK6 cells were freshly transformed with each of the nanobody plasmids, plated in LB-Agar plates supplemented with 100 μg mL^{-1} ampicillin and incubated overnight at 37 ºC. The following day, a single colony was inoculated into 1 mL 2x LB medium supplemented with 0.1 % (w/v) glucose, 1 mM $MgCl_2$ and 100 μg mL^{-1} ampicillin. The following day, 500 μL of this overnight culture was added to 50 mL terrific broth (TB) medium supplemented with 0.1 % (w/v) glucose, 1 mM $MgCl_2$ and 100 μg mL^{-1} ampicillin. This culture was incubated at 37 ºC in an orbital shaker (170 rpm) until the OD_{600} reached 2.0. Then, nanobody expression was induced with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG), the temperature was dropped to 28 ºC, and the cultures were incubated overnight. The following morning, the cells were harvested by centrifugation (20 minutes at 4,000 x g), flash-frozen in liquid N_2 and stored at -20 ºC.

For small-scale nanobody purification, cell pellets from 50 mL expressions were initially thawed, re-suspended in 5 mL ice-cold TES buffer (200 mM Tris pH 8.0, 0.5 mM EDTA, 500 mM sucrose) and left shaking on ice for 1h (220 rpm). The re-suspended cells were then diluted with 35 mL of ice-cold 10 mM $MgCl_2$ and were again left shaking on ice for 1h (220 rpm). The suspension was then centrifuged (1h at 15,000 x g, 4 ºC), the supernatant containing the periplasmic extract was collected, supplemented with 0.25 mL Ni-Sepharose 6 FastFlow resin, and incubated at room temperature for 1h on a rotating wheel (10 rpm). Then, the resin was collected by centrifugation (5 minutes at 700 x g), and the supernatant was discarded. The resin was then washed with 7.5 mL (30 column volumes) of Phosphate Buffer 1 (50 mM NaH_2PO_4 / Na_2HPO_4 pH 7.0, 1 M NaCl) and collected as before. The resin

was then washed further with 5 mL (20 column volumes) of Phosphate Buffer 2 (50 mM NaH_2PO_4 / Na_2HPO_4 pH 6.0, 1 M NaCl), collected by centrifugation and transferred to a 2 mL 96-well filter block (Porvair). The resin was washed once more with 1 mL Phosphate Buffer 2, residual buffer was removed by centrifugation (2 min at 300 x g), and the nanobodies were eluted with 250 µL EXB300 (50 mM HEPES pH 7.5, 300 mM KCl, 300 mM Imidazole pH 8.0). Eluted nanobodies were immediately buffer exchanged into SEC buffer (matching that of purified CerS5 or CerS6, depending on which protein the nanobodies were raised against) using a 96-well PD MultiTrap G-25 desalting block (Cytiva). The purified nanobodies were then analysed by SDS-PAGE, aliquoted, flash-frozen in liquid N_2 and stored at -80 ºC.

2.4.4.2 Large-scale nanobody expression and purification

For nanobody expression in large scale, single colonies were inoculated into 10 mL 2x LB medium supplemented with 0.1 % (w/v) glucose, 1 mM $MgCl_2$ and 100 µg mL^{-1} ampicillin. The following day, the 10 mL overnight culture was added to 1 L TB medium supplemented with 0.1 % (w/v) glucose, 1 mM $MgCl_2$ and 100 µg mL^{-1} ampicillin. This culture was incubated at 37 ºC in an orbital shaker (180 rpm) until the OD_{600} reached 2.0. Nanobodies were then expressed and cells were harvested as described in the previous section.

For the large-scale (1 L culture) purification of nanobodies for structural studies, cell pellets were thawed in a beaker containing 50 mL ice-cold TES buffer, with gentle stirring. Then, the re-suspended cells were gently stirred for 1h at 4 ºC (100 mL total volume). After 1h, the suspension was diluted with 100 mL ice-cold 5 mM $MgCl_2$ and the resulting mixture was stirred for a further 2h (at 4 ºC). The suspension was then centrifuged (40 min at 15,000 x g, 4 ºC) and the His-tagged nanobodies were purified from the supernatant containing the periplasmic extract by batch binding to 2 mL Ni-Sepharose 6 FastFlow resin (2h rotating at

room temperature). The resin was then collected by centrifugation (5 min at 700 x g), the supernatant was discarded, and the resin was washed with 120 mL (60 column volumes) Phosphate Buffer 1 and collected by centrifugation as before. This resin wash step was repeated with 120 mL Phosphate Buffer 2. Then, the resin was transferred to a gravity column, the nanobodies were eluted with 10 x 1 mL EXB300, and peak fractions were immediately exchanged into SEC buffer (containing the appropriate detergent) using a PD10 desalting column (Cytiva) and left at 4 ºC overnight.

The desalted nanobodies were then concentrated to 1 mL using a 6 mL 5 kDa cut-off concentrator (Corning) and applied onto a Superdex 200 Increase 10/300 GL column (GE healthcare). Two SEC runs were carried out per nanobody (500 µL injections). The peak fractions from both SEC runs were combined, and the nanobodies were concentrated to 20-25 mg mL^{-1} using a 2 mL Vivaspin 5 kDa cut-off concentrator (Sartorius), aliquoted, flash-frozen in liquid N_2 and stored at -80 ºC

2.4.5 Bio-Layer Interferometry (BLI)

BLI experiments were performed using biotinylated CerS5 and CerS6 in SEC buffer containing either 0.1 % (w/v) OGNG, 0.01 % (w/v) CHS (CerS5) or 0.003 % (w/v) LMNG, 0.0003 % (w/v) CHS (CerS6). Nanobodies used in these experiments were prepared in the same buffer as the target protein to avoid buffer mismatch. BLI experiments were carried out using an Octet Red 384 system (Sartorius). Biotinylated protein at 40 µg mL^{-1} in SEC buffer was loaded onto Streptavidin SA biosensors (Sartorius) for 180 s, and excess unbound protein was removed by dipping the sensors in buffer without protein for 120 s. For nanobody screening, a 60 s baseline was initially recorded in SEC buffer. This was then followed by a 120 s association step (500 nM nanobody) and a 120 s dissociation step in SEC buffer. All experiments were repeated using reference sensors without bound CerS proteins. Data

were analysed using the Octet Data Analysis 11.0 software and fit with a 1:1 binding model. Raw data were reference subtracted, aligned on the y axis to the average of the last 5 s of the baseline, inter-step corrected, and filtered with a Savitzky–Golay filter. Dissociation rates were recorded and used to prioritise nanobodies for structural studies.

2.4.6 Purification of CerS6-nanobody complexes for structural studies

CerS6-nanobody complexes were purified as described in Section 2.1.3.3 for GDN-solubilised CerS6-c025, with the following exception. Prior to the SEC step, purified nanobody was added to the sample at a 2:1 nanobody:CerS6 molar ratio. The mixture was incubated on ice for at least 1h and subsequently further purified by SEC. The SEC peak corresponding to the CerS6 dimer with bound nanobody was pooled, concentrated to 5 mg mL^{-1} and immediately processed for cryo-EM.

2.4.7 Cryo-EM sample preparation

CerS6 samples were purified as described in 2.1.3.3. LMNG/CHS-solubilised protein was concentrated to 1.2 mg mL^{-1} or 2.8 mg mL^{-1}, whereas GDN-solubilised protein was concentrated to 5 mg mL^{-1}. CerS6-nanobody complexes, solubilised in GDN, were likewise concentrated to 5 mg mL^{-1}.

All grids were freshly glow discharged for 30 s prior to sample application. LMNG/CHS samples were frozen on Quantifoil 300 mesh Au R1.2/1.3 grids; GDN samples were frozen on Quantifoil 200 mesh Au R1.2/1.3 grids. Plunge freezing was carried out using a Vitrobot Mark IV (Thermo Fisher Scientific) (chamber set to 4 ^{o}C and 100 % humidity). For LMNG/CHS samples, blotting times between 1-3 s were sampled. For GDN samples, a wait time of 20 s was set prior to blotting and blotting times between 5-9 s were sampled.

2.4.8 Cryo-EM grid screening and small characterisation datasets

Grid screening was carried out at the Oxford Particle Imaging Centre (Oxford, UK) in a Glacios microscope (Thermo Fisher Scientific) operating at 200 keV and equipped with a Falcon III direct electron detector (Thermo Fisher Scientific). Likewise, all small characterisation datasets were collected on this microscope, in fast linear mode. Automated data acquisition was set up through EPU (Thermo Fisher Scientific). Dose-fractioned movies were collected at 92,000 x nominal magnification (1.55 Å pixel^{-1}), using a defocus range between -1.5 μm and -3.0 μm. The total exposure dose was 52 e^{-} Å^{-2}, over 11 fractions.

2.4.9 CerS6-c025 in GDN: data collection and processing

The CerS6-c025 GDN dataset was collected at the electron Bio-Imaging Centre (Didcot, UK) on a Titan Krios microscope (Thermo Fisher Scientific) operating at 300 keV and equipped with a K3 direct electron detector (Gatan). Automated data acquisition was set up using EPU (Thermo Fisher Scientific). 5,972 dose-fractionated movies were collected at 130,000 x nominal magnification, in binned super-resolution mode (0.653 Å pixel^{-1}), using a defocus range between -0.8 μm and -2.4 μm. The total exposure dose was 50.9 e^{-} Å^{-2}, over 50 fractions.

Motion correction and CTF estimation were carried out in cryoSPARC (version 3.3.1). Micrographs with significant drift or bad CTF fitting were excluded. Initially, 1,838,601 particles were picked from 5,906 micrographs using the block picker and classified in multiple rounds of reference-free 2D classification to exclude bad and noise particles. A representative set of 20,000 particles was selected to train a model for particle picking using topaz [223]. 1,207,719 particles were then picked and subjected to reference-free 2D classification as before. The ratio of monomer and dimer particles was estimated by generating two *ab-*

initio 3D reconstructions. For downstream processing, five *ab-initio* models were generated. Dimeric particles were further classified by heterogeneous refinement using 5 reference classes corresponding to *ab-initio* monomer (2 classes) and dimer (3 classes) models. Monomeric particles were discarded, and the final dimer particle set obtained after heterogeneous refinement (78,859 particles) was used for 3D refinement in cryoSPARC, initially using homogeneous refinement and subsequently using the non-uniform refinement function [224], with C2 symmetry applied. A further masked local refinement of the protein region was also carried out with C2 symmetry.

3 Structural insights into the transport mechanism of the human NSAID transporter MFSD10

3.1 Aims

As introduced in Section 1.1.4, available evidence suggests MFSD10 could act as a proton-coupled efflux transporter for NSAIDs, organic anions and other unknown compounds**.** However, despite the biological significance of this proposed role, very little is known about the transport properties of MFSD10. In order to obtain initial insights into its transport mechanism, this project aimed to determine a crystal structure of human MFSD10. The citrate-bound crystal structure obtained provides the first molecular understanding of the recognition of organic anions by MFSD10. Also, as the NSAID:H^+ antiporter function had not been directly demonstrated, a fluorescence-based assay is implemented in this work to confirm the coupling of NSAID transport to proton antiport.

3.2 Purification and crystallisation of human MFSD10

3.2.1 Large-scale purification of human MFSD10

For the large-scale purification of human MFSD10 for structural studies, I expressed a MFSD10 construct (Met1-Lys444) truncated at the C-terminus to exclude the C-terminal 11 residues, predicted to be disordered, in *Sf*9 insect cells. The protein was fused on its C-terminus to a TEV protease cleavage site, followed by 10x His and 1x FLAG tags. Previous work carried out in the lab by Dr Andrew Quigley had established that this construct could be successfully expressed in *Sf*9 and purified in milligram amounts.

To purify sufficient amounts of protein for crystallisation, typically 6 L of *Sf*9 cells expressing MFSD10 were used. Briefly, cells were mechanically lysed, membrane proteins were extracted with 1 % (w/v) OGNG and 0.1 % (w/v) CHS, insoluble material was removed by centrifugation, and the tagged protein was captured from the soluble fraction using TALON metal affinity resin. This resin has complexed cobalt ions, making it more specific to His-tagged proteins than nickel-based resins such as Ni-NTA, for use in immobilized metal affinity chromatography (IMAC). After binding, the resin was washed, the tagged protein was eluted with imidazole, and the C-terminal tags were cleaved overnight using TEV protease. Uncleaved MFSD10 and the His-tagged TEV protease were removed in a reverse IMAC step using TALON resin, and the cleaved protein was concentrated and further purified by size-exclusion chromatography (SEC). As expected, analysis by SDS-PAGE reveals a slight shift of the MFSD10 band after tag cleavage. The SEC profile shows a single monodisperse peak and SDS-PAGE analysis of the SEC peak fractions reveals a pure final sample (Fig. 3.1A,B).

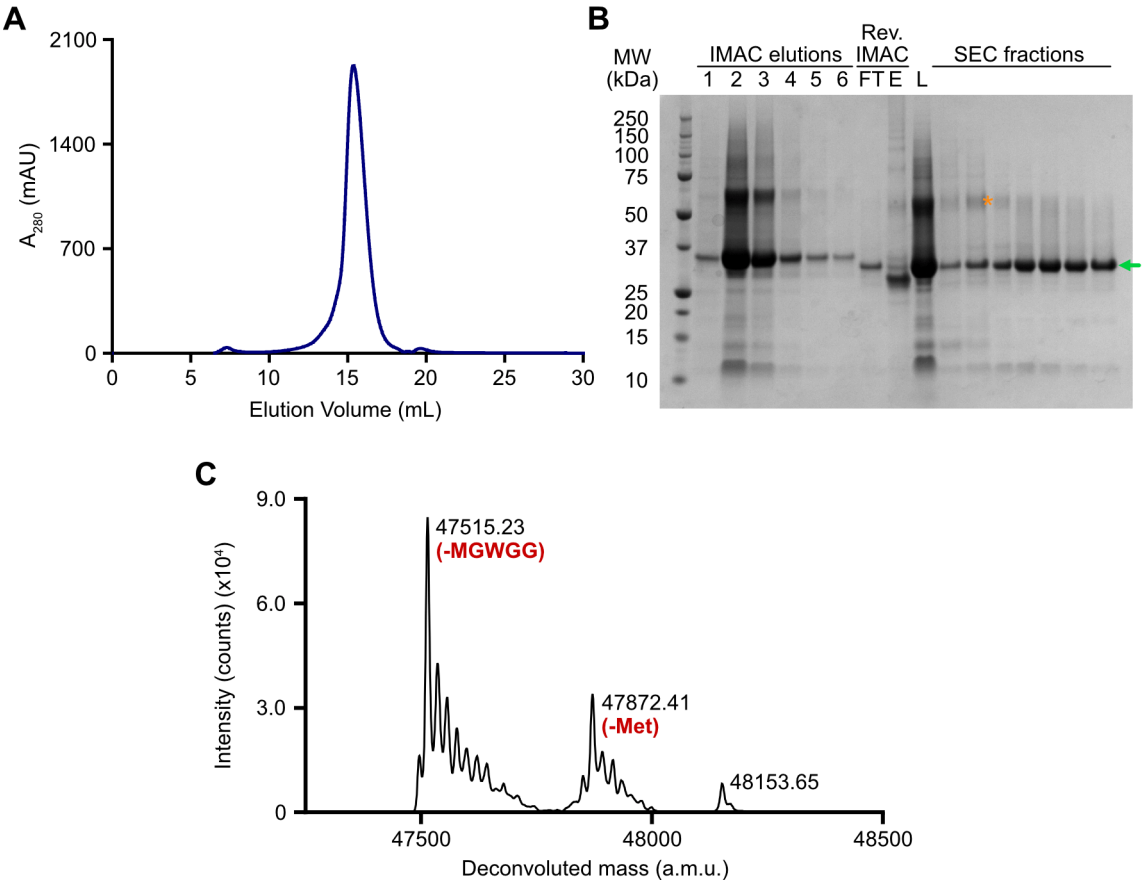


Figure 3.1. Purification of MFSD10 expressed in *Sf*9 cells. A. Elution profile of MFSD10 from a Superose 6 Increase 10/300 SEC column. **B.** SDS-PAGE analysis of samples taken during various steps of MFSD10 purification. IMAC elutions 1-6 correspond to 2.5 mL fractions collected during the elution of tagged protein from the TALON resin. Reverse IMAC flow-through (Rev. IMAC FT) corresponds to the MFSD10 protein cleaved overnight with TEV protease, which no longer contains the C-terminal tags and therefore did not bind to TALON resin in the reverse IMAC step. Reverse IMAC elution (Rev. IMAC E) corresponds to the elution of tagged proteins that were bound to the TALON resin during the reverse IMAC step and contains primarily the His-tagged TEV protease and residual uncleaved MFSD10. SEC load (L) corresponds to the concentrated MFSD10 sample applied onto the SEC column. SEC fractions correspond to the peak fractions from (A). The band corresponding to purified MFSD10 is marked with a green arrow. An orange asterisk marks a higher molecular weight band confirmed to also correspond to MFSD10 by tryptic digest MS/MS analysis. **C.** LC-ESI-MS intact mass analysis of purified MFSD10.

The purified protein was analysed by denaturing intact protein mass spectrometry, which reveals the presence of two short N-terminal truncations (Fig. 3.1C). The expected mass of the purified protein monomer is 48002.72 Da. The major species observed (47515.23 Da) corresponds to the loss of the 5 N-terminal residues (MGWGG) (predicted 47514.16 Da), whereas the second deconvoluted mass peak (47872.41 Da) corresponds to the loss of the

initiator methionine from the protein N-terminus (predicted 47871.52 Da). An additional deconvoluted mass peak could be observed (48153.65 Da), but its identity could not be determined. This likely represents an unknown addition of 638.42 Da to the base (loss of the first 5 N-terminal residues) peak. Alternatively, this peak could correspond to a 281.24 Da addition to the peak corresponding to the loss of the initiator methionine. However, as no ~281 Da addition to the base peak (loss of the first 5 N-terminal residues) is observed, this would require that the unknown ~281 Da addition occurs in one of the 4 amino acids following the initiator methionine (GWGG). This is highly unlikely as the only side-chain is that of a tryptophan.

Surprisingly, the purified protein exhibits an unexpected dark-green colour (Fig. 3.2A). This could be due to a bound compound that co-purified with the protein. One possible candidate is biliverdin, a product of haem catabolism. Interestingly, the absorption and fluorescence spectra of purified MFSD10 share similarities with those of biliverdin (Fig. 3.2B,C). Upon excitation at 375 nm, both MFSD10 and biliverdin display an emission maximum at 455 nm. In addition, the absorption spectrum of biliverdin displays a peak at ~375 nm, accompanied by a second lower intensity peak at ~660 nm, whereas MFSD10 displays a similar peak at ~375 nm, but the second, longer wavelength peak is blue-shifted by ~30 nm in comparison to that of biliverdin. The biliverdin absorption spectrum displays characteristic peaks of tetrapyrroles (such as porphyrins and bilins), referred to as the Soret (~ 380 nm) and Q- (~ 500-700 nm) bands. The Q-band of biliverdin can undergo a blue-shift upon protein binding [225], and therefore it is possible that the peaks observed in the absorption spectrum of MFSD10 could correspond to bound biliverdin or to a related compound. However, further work is required to determine the identity of the bound green compound.

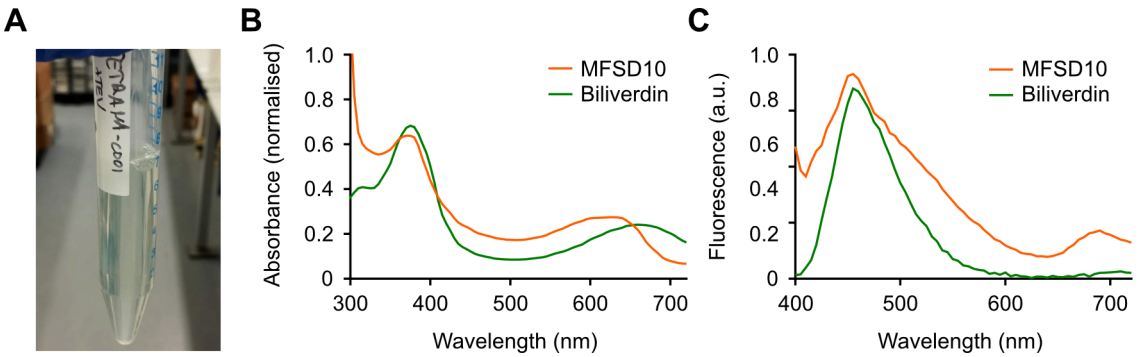


Figure 3.2 Spectroscopic properties of purified MFSD10. **A.** Picture of purified MFSD10 sample at 2 mg mL^{-1}, before being concentrated for SEC. **B.** Absorption spectra of purified MFSD10 and biliverdin. The absorption spectrum of biliverdin displays Soret- (~375 nm) and Q-bands (broad peak with maximum at ~660 nm), characteristic of tetrapyrroles. **C.** Fluorescence emission spectra of purified MFSD10 and biliverdin upon excitation at 375 nm.

3.2.2 NanoDSF screening of putative substrates

A range of NSAIDs, including proposed MFSD10 substrates such as indomethacin and diclofenac, were screened for thermostabilisation of the purified protein by nanoDSF in order to identify suitable candidate compounds for co-crystallisation experiments. Other compounds such as tetracycline and doxycycline were also included in the set, given the low homology of MFSD10 to *E. coli* tetracycline transporters [58]. The organic anions p-aminohippuric acid (PAH) and fluorescein have also been suggested to potentially act as MFSD10 substrates [60] and were also included, whereas the organic cation tetraethylammonium (TEA) was included as a negative control, since it is not a substrate of MFSD10 and no stabilisation was expected. Finally, biliverdin and bilirubin were included due to the observed green colour in the sample and spectroscopic properties of purified MFSD10.

To run nanoDSF, purified protein (5 µM) was incubated with 100 µM of each compound for 1 h and subsequently subjected to a temperature gradient of 1 °C min^{-1}, enabling the determination of its mid-point melting temperature (T_m) in the presence of the various added

compounds (Fig. 3.3). The T_m of as-purified MFSD10 in the absence of ligands was determined to be 57.80 ± 0.05 °C. Amongst the NSAIDs tested, thermostabilisation was observed for diclofenac (+1.83 ± 0.14 °C), indomethacin (+1.88 ± 0.04 °C) and sulindac (+2.45 ± 0.10 °C), whereas other NSAIDs had a minimal effect or de-stabilised the protein (in the case of celecoxib, -1.77 ± 0.02 °C). Fluorescein had a larger stabilising effect than any of the NSAIDs tested (+4.03 ± 0.10 °C). Notably, biliverdin demonstrated a significant stabilising effect (+12.04 ± 0.13 °C). On the other hand, bilirubin displayed a more modest thermostabilisation (+1.69 ± 0.04 °C).

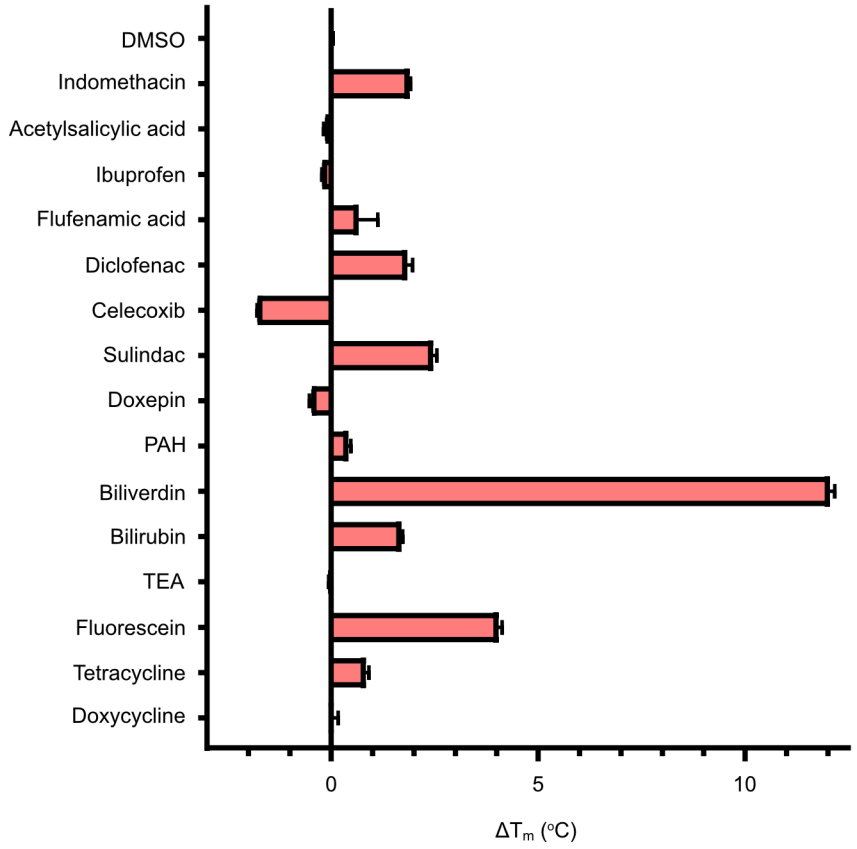


Figure 3.3. NanoDSF screening of putative MFSD10 substrates. Purified MFSD10 was incubated with 100 µM of each compound for 1 h before being subjected to a temperature gradient of 1 °C min^{-1}. Protein unfolding was monitored by recording tryptophan fluorescence at 330 nm and 350 nm, and the melting temperature (T_m) was determined from the inflection point of the calculated fluorescence ratio (350 nm / 330 nm). Melting temperature shifts (ΔT_m) were calculated as the difference between T_m values obtained with and without added compounds. Values shown are the average of three technical replicates, and the positive and negative error bars correspond to ± S.E.M.

Importantly, these results showed that a number of NSAIDs, including the proposed substrates diclofenac and indomethacin, had a stabilising effect on the protein, suggesting that they are interacting with the protein. Therefore, these compounds could be used for co-crystallisation experiments as well as functional assays to interrogate the mechanism of MFSD10. In addition, the considerable stabilisation by biliverdin was somewhat unexpected, and the basis of its interaction with MFSD10 should be investigated further to determine whether biliverdin could be transported by or inhibit MFSD10 and whether it corresponds to the small molecule that co-purifies with MFSD10 and gives rise to its dark-green colour.

3.2.3 Crystallisation in the lipidic cubic phase (LCP)

Purified MFSD10 was concentrated to 20 mg mL^{-1} and subjected to crystallisation screening using the lipidic cubic phase method [226, 227] using monoolein as the host lipid. Several sparse matrix screens were used, including the MemMeso and MemGoldMeso screens, MemGold2 (diluted by 30 % in-house for *in meso* crystallisation purposes), and a range of other screens developed in-house at the Structural Genomics Consortium (SGC). In one of these crystallisation screens developed in-house, thin large (~40-50 μm) plate-shaped crystals were obtained after 2 days in a condition containing 0.1 M ADA pH 6.5, 0.2 M lithium citrate, 30 % (v/v) PEG 400 (Fig. 3.4), and reached full size in 7 days. All crystals were mounted in cryo-loops, cryo-cooled in liquid nitrogen and taken to the micro-focus beamline I24 (Diamond Light Source, UK) for screening and data collection.

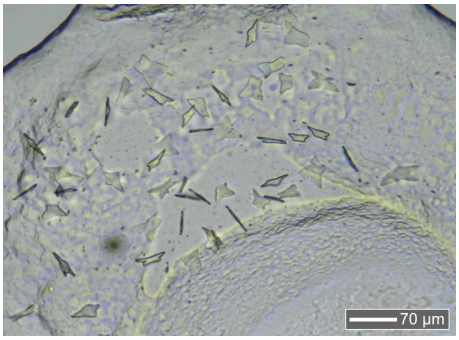


Figure 3.4. Representative MFSD10 LCP crystals in 0.1 M ADA pH 6.5, 0.2 M lithium citrate, 30 % (v/v) PEG 400. These crystals yielded the initial MFSD10 high-resolution dataset used to determine the structure described in Section 3.3.

3.3 Crystal structure of human MFSD10

3.3.1 Data collection, phasing, model building and refinement

MFSD10 crystals obtained in 0.1 M ADA pH 6.5, 0.2 M lithium citrate, 30 % (v/v) PEG 400 reached full size in 7 days and were then mounted in cryo-loops. Diffraction data were collected at 100 K at the micro-focus I24 beamline (Diamond Light Source, UK). Given the opaque nature of the lipidic cubic phase in the cryo-loop, the crystals were located inside the loop by running diffraction grid scans (Fig. 3.5). Three testing diffraction images were taken 45° apart, and a data collection strategy was devised using iMosflm [228]. A full dataset was collected from a single crystal using a beam size of 20 x 20 µm^{2}. Data were collected for 360°, using 0.2^{o} oscillations, at a wavelength of 0.9687Å.

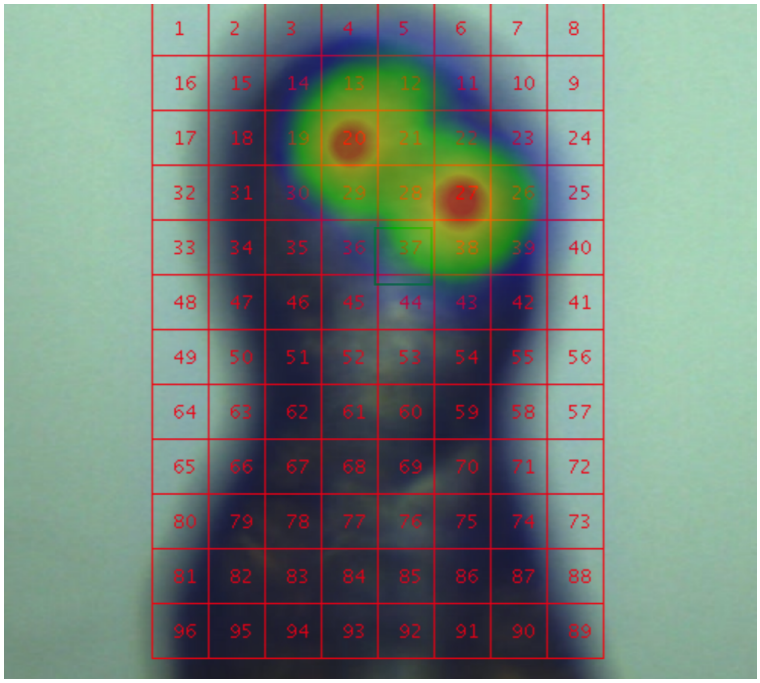


Figure 3.5 Grid scan conducted at the I24 beamline (Diamond Light Source). Grid scans enabled identifying the position of mounted MFSD10 crystals within the opaque lipidic cubic phase.

The initial 260° wedge of data did not suffer from radiation damage and were processed using XDS and scaled with XSCALE [214]. Unmerged scaled data were truncated and merged using AIMLESS [215]. The data were mildly anisotropic, being limited to 3.0 Å in the worst direction and 2.4 Å in the best direction (as determined using STARANISO [229] for $CC_{1/2} > 0.5$, where $CC_{1/2}$ is the Pearson half-dataset correlation coefficient). Data truncated isotropically in AIMLESS to 31.74 - 2.40 Å was used for downstream phasing, model building and refinement (nominal overall resolution determined to be 2.58 Å based on $CC_{1/2} > 0.5$ limit). These crystals belonged to the P 2_1 space group, and the Matthews coefficient suggests they contain one copy of the protein per asymmetric unit ($V_m = 2.38$ A^3 / Da).

Phasing was carried out by molecular replacement using PHASER [216]. Due to the low sequence similarity of MFSD10 to other MFS transporters with known structures, it was not straightforward (pre-Alphafold era) to identify suitable templates. As MFS transporters have a conserved overall fold, molecular replacement was attempted with known MFS structures

with 12 transmembrane helices. For each structure, full-atom models and poly-alanine models generated using CHAINSAW [217] were used in the search. In addition, for several MFS structures (and derived poly-alanine models), templates corresponding to only the transmembrane regions were also generated by truncating loop regions. However, none of these approaches were successful.

Next, a number of known MFS structures were split into two separate search models, corresponding to their N- and C-terminal 6-TM bundles. As before, poly-alanine models were also generated. As the long linker between the two 6-TM bundles of MFS transporters can adopt different conformations, this region was excluded from the search templates. Molecular replacement was attempted by searching for one copy of each of the two templates, and a single solution was found using poly-alanine models derived from the two 6-TM bundles of the structure of the *E. coli* DHA1 family transporter YajR (PDB: 3WDO, 19.6 % sequence identity to MFSD10) [39]. These two poly-alanine models corresponded to Glu10-Trp184 (N-terminal bundle) and Leu215-Ser386 (C-terminal bundle) in the YajR structure.

Initial phase estimates were improved using Parrot [218], and a starting auto-traced model was generated using Buccaneer [219]. This starting model contained several gaps in the tracing of the protein backbone, namely in the linker region between the two six-transmembrane bundles and in the long loop between TM1 and TM2. The model was then rebuilt manually in COOT [220] with iterative cycles of refinement using Refmac5 [221]. Two TLS groups were defined for the restrained refinement (residues 18-88 and 89-443, determined using TLSMD [222], which significantly improved refinement statistics. In the final stage of model building, 60 water molecules, a citrate ion from the crystallisation solution and two monoolein molecules were modelled. The final model was refined with good geometry to

R/R_{free} values of 20.9/25.8 %. The quality of the final model was analysed using the MolProbity server [230]. Coordinates and structure factors were deposited in the Protein Data Bank (PDB) with accession code 6S4M.

Table 3.1: X-ray data collection and refinement statistics.

Data collection	
Space group	P 2_1
Cell dimensions:	
***a, b, c* (Å)**	55.99, 70.38, 63.44
α, β, γ (°)	90.00, 115.45, 90.00
Resolution (Å)*	31.74-2.40 (2.49-2.40)
R_{meas}*	0.234 (1.853)
R_{pim}*	0.104 (0.822)
I/σI*	5.7 (1.0)
$CC_{1/2}$*	0.99 (0.40)
Completeness (%)*	98.7 (98.2)
Multiplicity*	4.9 (5.0)
Refinement	
Resolution (Å)	30.00-2.40
No. reflections (free)	16,449 (831)
R_{work} / R_{free}	0.2087 / 0.2583
No. atoms:	
Protein	3,169
Water	60
Lipids / citrate	60
B-factors (Å²):	
Protein	38.40
Water	37.56
Lipids / citrate	66.88
r.m.s. deviations:	
Bond lengths (Å)	0.0066
Bond angles (°)	1.46
Ramachandran plot:	
Preferred (%)	98.58
Allowed (%)	1.42
Outliers (%)	0.00

* values in parentheses are statistics for the highest resolution shell

3.3.2 Packing in the MFSD10 crystals

Membrane protein crystals grown *in meso* typically display type I packing, where the protein molecules are arranged as planar sheets, forming several layers [226]. Here, unlike crystals obtained by vapour diffusion crystallisation of protein in detergent micelles, extensive crystal contacts can occur between transmembrane domains of protein molecules, typically leading to tighter packing and often stronger diffraction [226, 231]. Also, as the individual protein molecules are reconstituted into the lipid bilayer without a preferred orientation, these layers may be formed of proteins aligned in the same direction or may contain individual protein molecules in different orientations [226]. In the planar layers of the present MFSD10 crystal, each protein molecule is surrounded by and contacts four other copies oriented in the opposite direction in the membrane (Fig. 3.6).

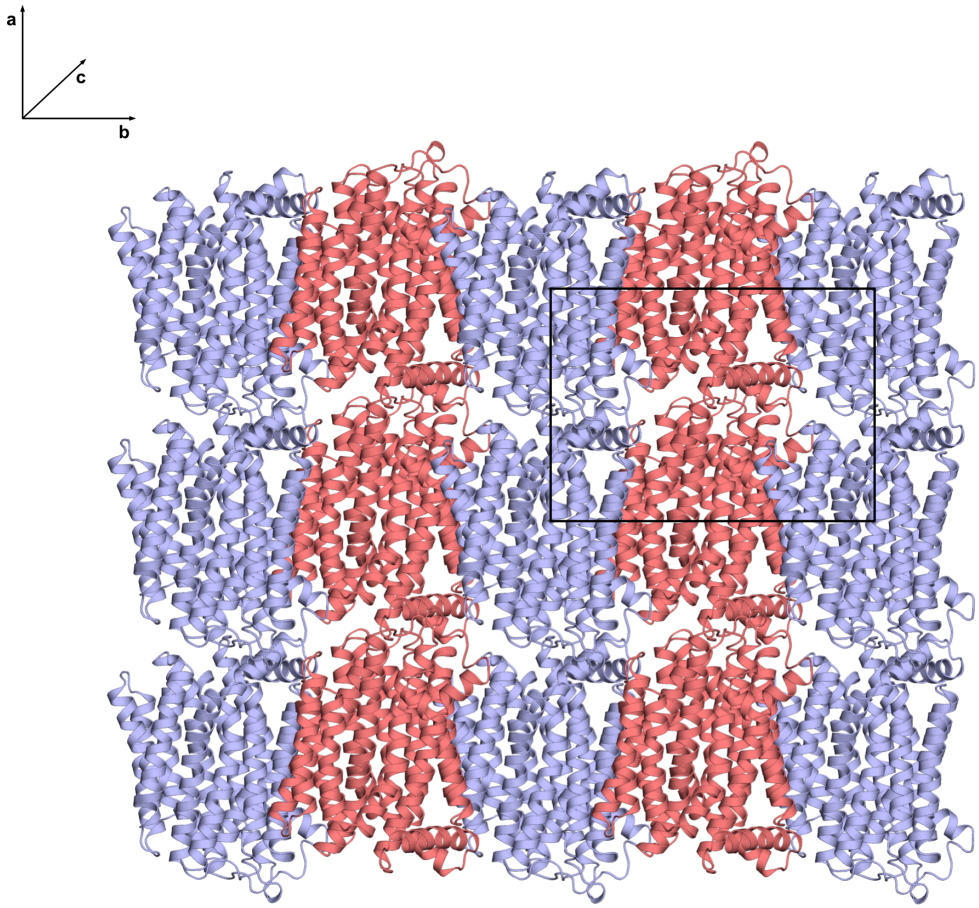


Figure 3.6. Packing in the MFSD10 crystal. MFSD10 LCP crystals display type I packing, with the protein molecules arranged as planar sheets. This figure shows molecules from three layers. The MFSD10 monomers are coloured in red or purple, depending on their orientation (red: intracellular side up; purple: extracellular side up). The figure is viewed face-on to the ab plane (and therefore along the c axis). The unit cell is displayed in black.

3.3.3 Overall structural features

The crystal structure of human MFSD10 determined in this study reveals the canonical MFS fold, with 12 transmembrane helices arranged as two six-transmembrane helix bundles with internal pseudo-two-fold symmetry. The two 6-TM bundles are connected by an intracellular linker region which, in MFSD10, is partially embedded in the membrane and forms three short helices (ICH1-3) (Fig. 3.7).

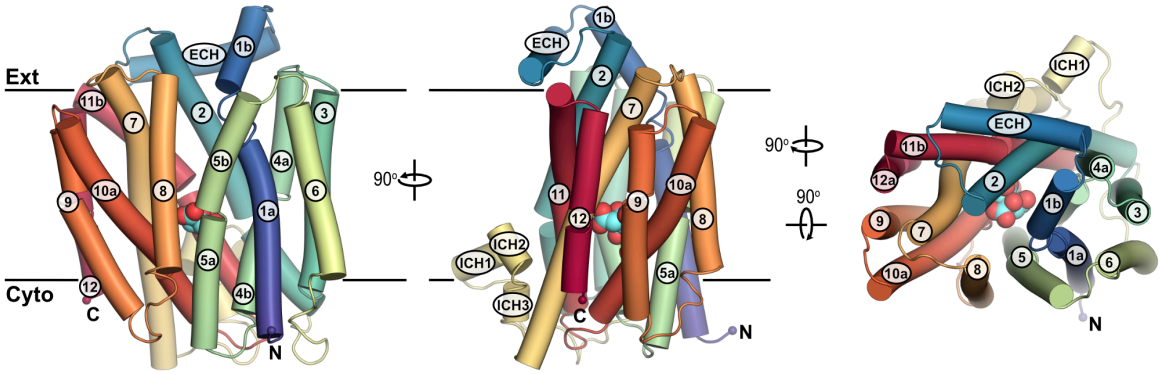


Figure 3.7. Overall structure of MFSD10. Cartoon representation of the crystal structure of human MFSD10 viewed parallel to the membrane plane (*left and middle)* and from the extracellular side (*right*). The structure reveals the typical 12-transmembrane helix architecture of MFS transporters, with the two 6-TM domains connected by a long linker which in MFSD10 forms three intracellular helices (ICH1-3). An unusual extracellular amphipathic helix (ECH) is present between TM1 and TM2. The bound citrate ion is shown in space-filling form.

In the structure, MFSD10 was captured in an outward-facing partially occluded conformation. Clear density corresponding to a citrate ion from the crystallisation solution was observed in the substrate-binding site (see Section 3.3.4), indicating that the transporter was captured in a conformation corresponding to the step of the MFS transport cycle prior to substrate release to the extracellular side.

A 35 Å-long central cavity is located between the two 6-TM bundles and is lined by the first and second helices of each 3-TM repeat (TM1a/b, TM2, TM4a/b and TM5 from the N-terminal bundle and TM7, TM8, TM10 and TM11 from the C-terminal bundle). The extracellular gate, formed by TM1-2 and TM7-8, is partially open on the extracellular side, forming a narrow entrance that enables solvent access to the central cavity. On the other hand, the intracellular gate, composed of TM4-5 and TM10-11, is fully closed, preventing access to the central cavity from the cytoplasm.

Interestingly, the loop between TM1 and TM2 forms an amphipathic extracellular helix (ECH), which sits on the extracellular face of the membrane with its hydrophobic side-chains buried in the bilayer (Fig. 3.8). In addition, two monoolein lipid molecules can be

observed clustered next to TM3 and TM4a, and the well-defined electron density confirmed the relative positioning of the ECH and the lipid head groups. This observation clearly demonstrates that the side-chains of the ECH hydrophobic facet are anchored in the bilayer.

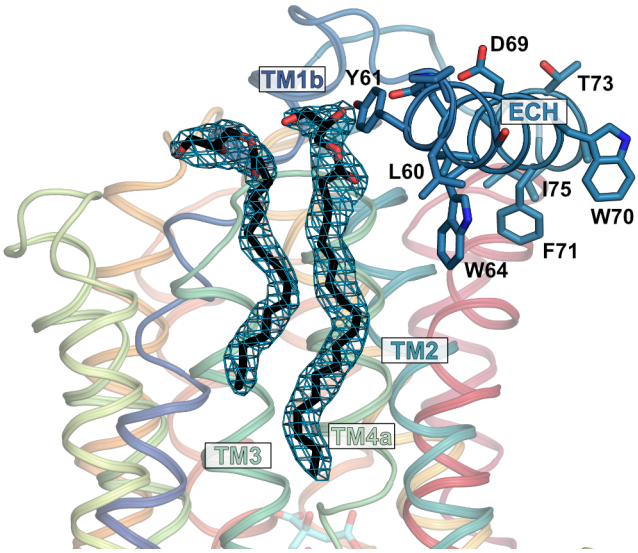


Figure 3.8. The TM1-2 linker of MFSD10 forms an amphipathic helix that sits on the membrane. View of the amphipathic extracellular helix (ECH) along the plane of the membrane. The ECH side-chains are shown as sticks, indicating that the membrane-facing face contains hydrophobic side-chains, whereas the outward-facing face is formed by polar residues. The hydrophobic face side-chains are buried in the bilayer, below the polar lipid head groups, as demonstrated by their positioning relative to two monoolein molecules (black / red, shown as sticks) bound next to TM3 / TM4a. Electron density is shown for the two modelled monoolein molecules (final REFMAC 2mFo-DFc: blue mesh, contoured at 1σ).

3.3.4 Substrate-binding site

A putative substrate-binding site can be observed at the bottom of the central cavity (Fig. 3.9A). As with other multidrug resistance antiporters of the MFS, it is located centrally between the N- and C-terminal bundles, being predominantly lined by residues from TM5, TM7, TM10 and TM11 (the first helix in each 3-TM repeat).

The putative binding site displays a dual polar / hydrophobic character, as it stretches from a positively charged polar pocket, lined by Tyr279 (TM7), Arg401 (TM11) and Arg408 (TM11) of the C-terminal domain, towards a long hydrophobic face formed by TM4 and TM5 of the N-terminal domain. Interestingly, there is clear electron density adjacent to the polar pocket, which can be accounted for by a citrate ion from the crystallisation solution (Fig. 3.9B). The citrate ion sits at the bottom of the cavity, where one of its terminal carboxylates fits in the polar pocket, forming hydrogen bonds to Tyr279, Arg401 and Arg408 (Fig. 3.9C,D). In addition, the citrate's second carboxylate interacts with Arg408, and the hydrocarbon backbone of the citrate ion makes non-polar contacts with Val157 (TM4b), Phe184 (TM5), Leu283 (TM7), Val377 (TM10) and Gly404 (TM11).

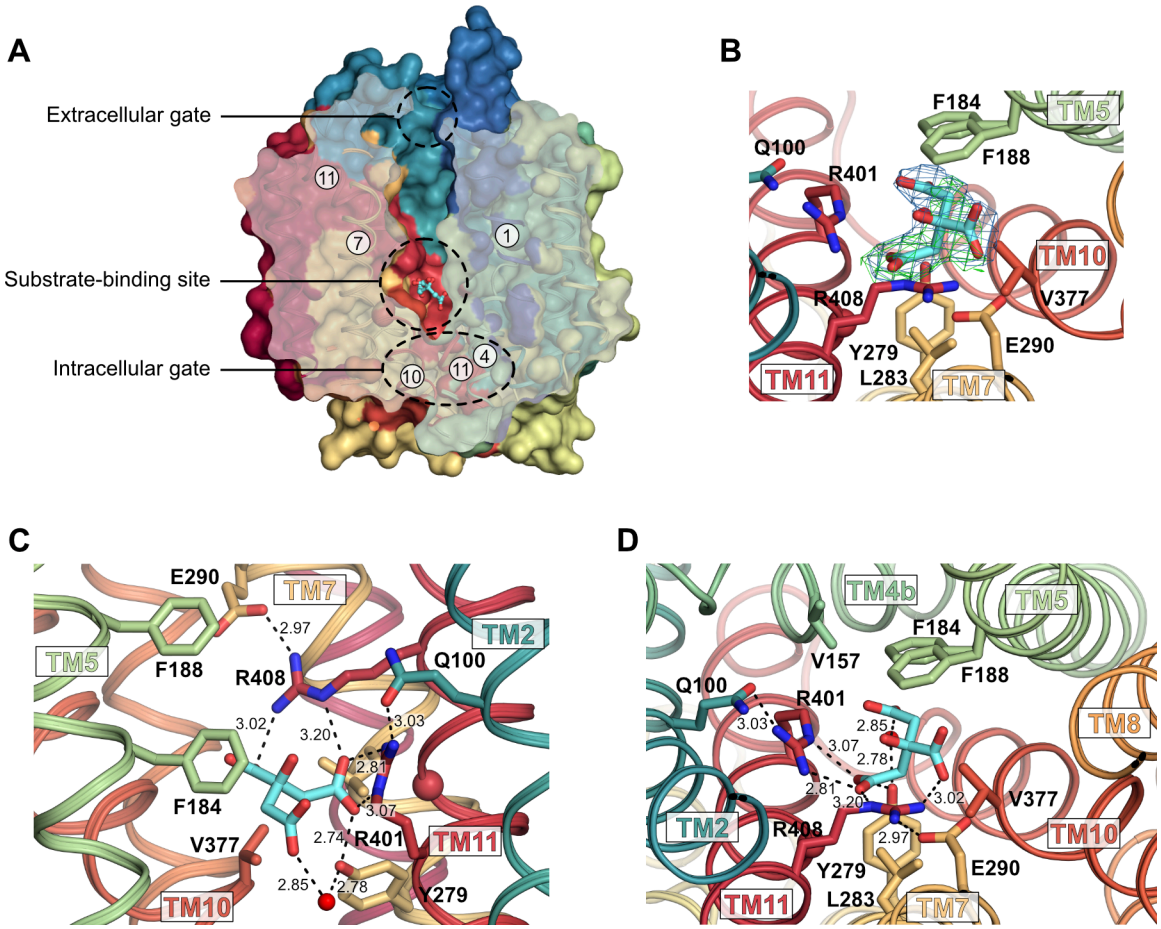


Figure 3.9. Citrate ion bound in the putative substrate-binding site. A. Cutaway molecular surface view of the central cavity in the MFSD10 structure. The structure is viewed parallel to the membrane plane. Numbers in circles denote transmembrane helices. **B.** Electron density in the substrate-binding site in the region around the citrate ion. The final REFMAC5 2mFo-DFc electron density map is shown as a blue mesh (contoured at 1σ); Buster omit mFo-DFc electron density map is shown as a green mesh (contoured at 2.5 σ). **B.** Substrate-binding site viewed parallel to the membrane plane. A water molecule within hydrogen-bonding distance of the citrate ion is shown as a red sphere. The Cα of Gly404 (TM11) is shown as a large sphere. **C.** Substrate-binding site viewed from the extracellular side. In all panels, the citrate ion is shown in stick representation (cyan / red). In panels C and D, hydrogen bonds are shown as black dashed lines, and hydrogen bond distances (in Å) are displayed.

As will be described in Sections 3.3.7 and 3.3.8, the presence of this bound citrate ion likely prevented successful co-crystallisation with substrates, but it nevertheless provides useful insights into the likely binding mode of MFSD10 substrates, such as the NSAIDs indomethacin and diclofenac. These NSAIDs contain a carboxylate group equivalent to that of the bound citrate (Fig. 3.10).

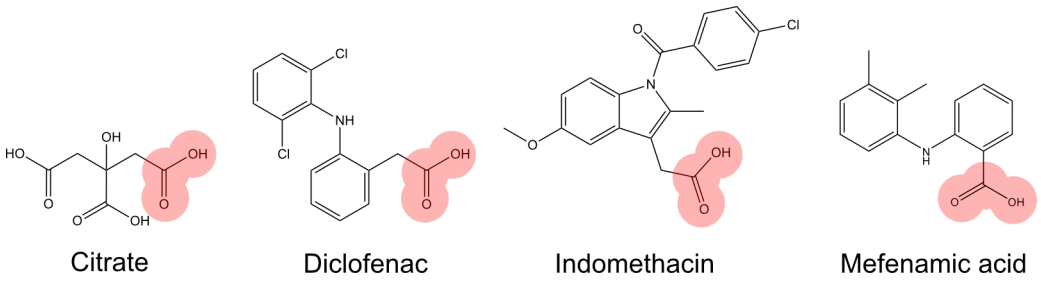


Figure 3.10. Chemical structures of citrate and proposed MFSD10 substrates. Similarly to citrate, proposed MFSD10 NSAID substrates contain a carboxylate group (highlighted in red). Additionally, these NSAIDs are largely hydrophobic and typically contain multiple aromatic rings.

Therefore, the fortuitous co-crystallisation with a citrate ion reveals a likely binding site for the carboxylate groups of MFSD10 substrates in the polar pocket between TM7 and TM11. Besides this acidic group, the remainder of the NSAID scaffold is variable but possesses a hydrophobic character with at least one aromatic ring. This could be accommodated by non-specific non-polar contacts across the long hydrophobic face formed by TM4 and TM5, which lies adjacent to the carboxylate binding site and stretches along the N-terminal domain side of the central cavity (Fig. 3.11). In addition, Arg408 lies on the C-terminal domain side of the cavity, opposite the aromatic rings of Phe184 and Phe188 (Fig. 3.9C,D). Together, they create a planar slot shape for this section of the central cavity, which matches the planar nature of the aromatic moieties of the proposed substrates. However, while it is plausible that the aromatic rings of the NSAID substrates could form stacking interactions in this section of the central cavity, additional work is required to determine the exact mode of binding of NSAIDs in the MFSD10 binding site.

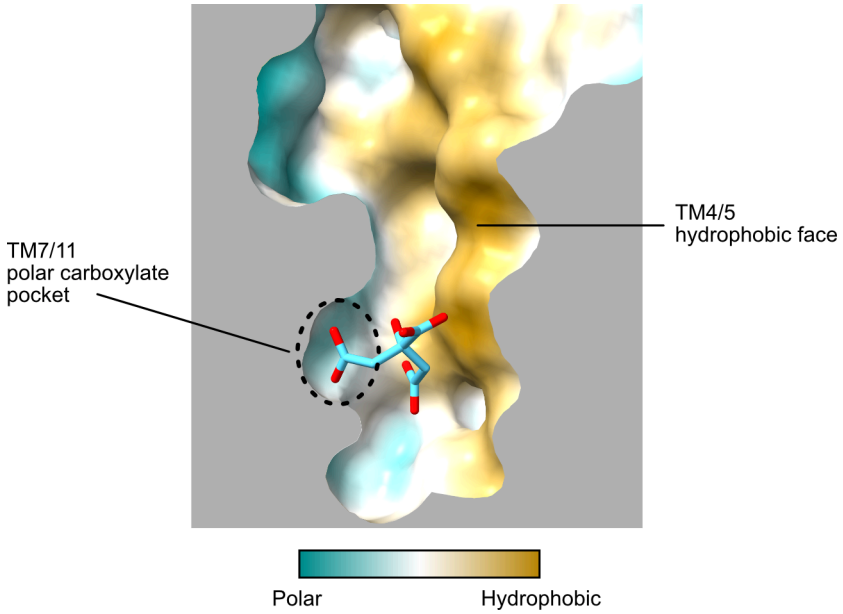


Figure 3.11. Dual polar-hydrophobic character of the substrate-binding site. Cutaway molecular surface view of the substrate-binding site, formed between the N- and C-terminal domains. Here, a terminal carboxylate of the bound citrate ion binds in a polar carboxylate pocket formed by TM7 and TM11 of the C-terminal domain. A large hydrophobic face contributed by TM4 and TM5 of the N-terminal domain is positioned adjacent to the polar carboxylate pocket and stretches along the N-terminal side of the central cavity towards the extracellular side. This large hydrophobic face could potentially enable non-polar interactions with hydrophobic groups of MFSD10 substrates. The substrate-binding site is viewed from the plane of the membrane, with the extracellular side of the central cavity facing up. Surfaces are coloured by lipophilicity, calculated by ChimeraX [232].

Substrates of MFS-MDR transporters are typically hydrophobic in nature and contain aromatic rings [26]. As a result, to account for the structural heterogeneity of their substrates, these transporters typically possess large substrate-binding cavities and achieve substrate recognition primarily via non-specific interactions such as non-polar contacts [26, 33]. Overall, the MFSD10 substrate-binding site appears to reveal these typical characteristics of MFS-MDR transporters, likely achieving substrate recognition via a combination of polar interactions at a localised pocket in the C-terminal domain and non-specific non-polar interactions along the hydrophobic face formed by the N-terminal domain helices TM4 and TM5.

Importantly, Arg408 is an integral part of the MFSD10 substrate-binding site, where it contributes to forming the carboxylate binding site. Arg408 is involved in a salt bridge with and stabilises the negative charge of the only acidic residue in the central cavity, Glu290. This

glutamate is a likely candidate to be protonated as part of the transport mechanism, and therefore this interaction suggests how MFSD10 might couple substrate binding to the protonation state of this residue (see Sections 3.4.2 and 3.5).

3.3.5 A-motif

Many MFS transporters contain a conserved sequence 'A-motif' (G-XXX-(D/E)-(R/K)-X-G-[X]-(R/K)-(R/K)) in their intracellular side, between TM2/3 and/or TM7/8 [12, 25]. This A-motif plays an essential role in modulating intra- and inter-domain gating in the cytoplasmic side, forming different interactions in the outward-facing, occluded and inward-facing states [32], and has been described in detail for a number of proteins, including YajR [39] and XylE [32, 40, 233].

In MFSD10, this motif contains the sequence $G^{108}ATSD^{112}CLGR^{116}RP$, and is located between TM2 and TM3. Here, a Gly-Gly interaction between Gly108 (TM2) and Gly394 (TM11) contributes to inter-domain packing between TM2 and TM11 (Fig. 3.12). This A-motif Gly-Gly interaction has been previously demonstrated to play a crucial role in stabilising the outward-open conformation of the *E. coli* multidrug transporter YajR [39]. Also, Asp112 forms a charge-helix dipole capping interaction with the N-terminal end of TM11, further stabilising the position of TM11 and the interface between the N- and C-domains. As in other MFS transporters, this capping interaction is modulated by a charge relay system involving Asp112, Arg116 and the conserved Asp166 in the C-terminal end of TM4b. The formation of this charge relay network also appears to close the intra-N-domain gate between TM2 and TM4b in the outward-facing MFSD10 structure (Fig. 3.12), and the equivalent charge relay system in the sugar porter XylE has been demonstrated to be essential for transport function [233].

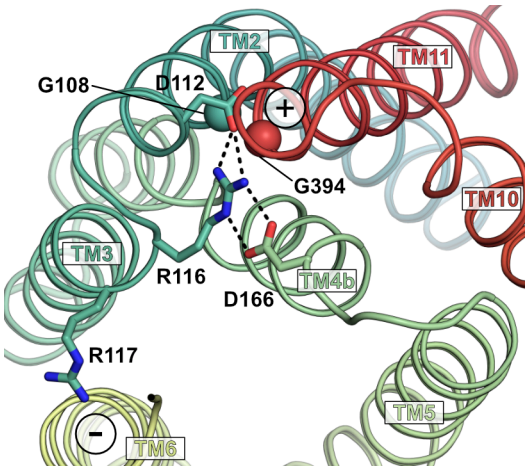


Figure 3.12. A-motif interactions in the outward-facing MFSD10 structure. The structure is viewed from the cytoplasmic side. A-motif residues contribute to packing between TM2 and TM11 through a Gly-Gly interaction between Gly108 (TM2) and Gly394 (TM11) (Gly Cα shown as spheres) and a charge-helix dipole interaction between Asp112 (TM2) and the N-terminal end of TM11. The characteristic A-motif charge-relay system (dashed lines) is created between Asp112 (TM2), Arg116 (TM3) and Asp166 (TM4b), closing the intra-N-domain gap between TM2, TM3 and TM4b in this outward-facing conformation.

Overall, the well-known conserved A-motif is present in the cytoplasmic side of most MFS transporters, where it plays a central part in gating and stabilisation of structural features required for the conformational transitions that occur during the transport cycle. In MFSD10, this motif can be identified in the TM2-TM3 loop and mediates a number of interactions previously described to play a critical structural role in the transport cycle of MFS transporters.

3.3.6 Extracellular gate

In the present outward-facing partially occluded MFSD10 conformation, the extracellular gate, formed by TM1, TM2, TM7 and TM8, is partially open, forming a narrow 7.5 Å x 18 Å entrance that enables solvent access to the central cavity (Fig. 3.13A). TM2 stretches over

the central cavity and interacts with TM7 through several hydrophobic contacts, one hydrogen bond and one salt bridge. The hydrogen bond, located closer to the substrate-binding site, is formed between the backbone carbonyl oxygen of Val86 (TM2) and the phenolic hydroxyl group of Tyr291 (TM7), whereas the salt bridge is formed between Glu80 (TM2) and His298 (TM7) (Fig. 3.13B,C). The pKa of His298 is calculated to be 6.85 (PROPKA 3.4 [234], available at https://server.poissonboltzmann.org/pdb2pqr), and therefore His298 is likely to be predominantly protonated under the crystallisation condition (pH 6.5), forming a salt bridge with Glu80. The charged state of this histidine *in vivo* will, of course, depend on the physiological pH that this 'extracellular' side of the protein is exposed to since, as described in Section 1.1.4, there is significant uncertainty in relation to the subcellular localisation of MFSD10. In contrast, no direct interactions between TM2 and TM8 can be observed. As a result, the inter-domain extracellular gap at the top of TM8 is only partially closed by a hydrogen bond between Arg82 (TM2) and a backbone carbonyl oxygen in the TM7-TM8 linker (Fig. 3.13B,C).

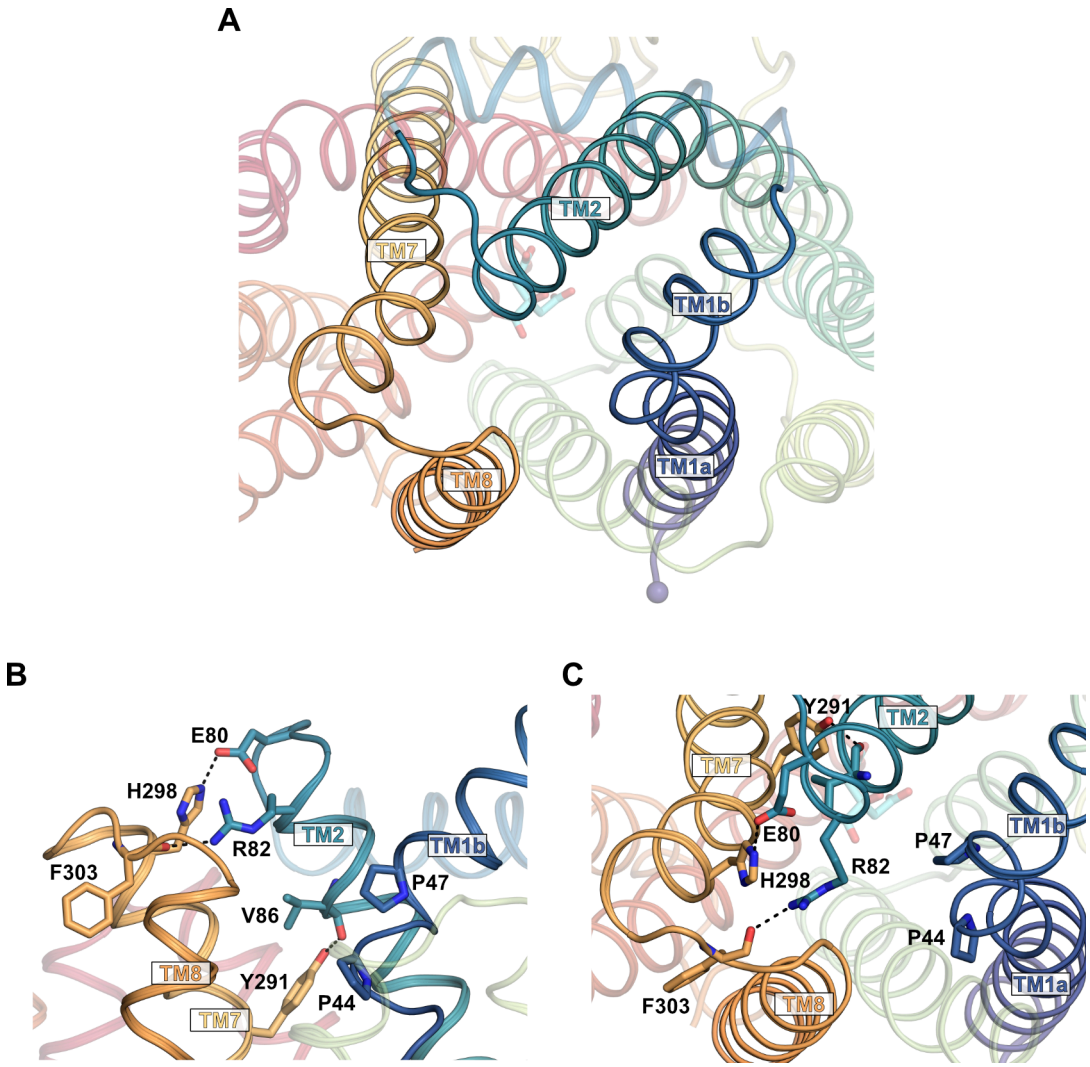


Figure 3.13 Extracellular gate in the MFSD10 structure. A. The extracellular gate of MFSD10 appears to be formed by TM1, TM2, TM7 and TM8. In this outward-facing partially occluded state, the extracellular gate is closed between TM2 and TM7 / TM8 but is partially open between TM8 and TM1, allowing access to the central cavity. **B,C.** Polar gating interactions between TM2 and TM7 / TM8, viewed from **B.** the plane of the membrane and **C.** the extracellular side. Hydrogen bonds and salt bridges (black dashes) formed between the N-terminal end of TM2 and the C-terminal end of TM7 and the TM7-TM8 linker enable partial closing of the extracellular inter-domain gate. In contrast, TM1 does not contact TM8 in the present structure, thereby enabling solvent access to the central cavity.

On the other hand, TM1 does not contact TM8, and this inter-domain gap enables solvent access to the central cavity via a narrow entrance. TM1 is broken at the top of the central cavity, and TM1b is bent away from TM8. This is likely facilitated by the presence of two

prolines: the first is located between TM1a and TM1b (Pro44) and the second at the beginning of TM1b (Pro47) (Fig. 3.13B,C). However, the significance of this break in TM1 is not immediately evident. Broken helices have been proposed to suggest a role in gating, where they may undergo a conformational rearrangement to regulate access to the central cavity [33]. This is a common feature of MFS sugar porters such as GLUT transporters [235, 236] and *Pf*HT1[237], where TM7 is clearly broken and TM7b, which forms an integral part of the substrate-binding site, moves to contact TM1 to close access to the cavity upon substrate binding.

In contrast, in MFS-MDR transporters, the conformational changes regulating access to the substrate-binding site are thought to generally occur more symmetrically between the two domains [33]. The present structure suggests this may also be the case for MFSD10, as the substrate-binding site of MFSD10 appears to be formed evenly by both domains. Moreover, TM1b is connected to the ECH via a very short linker, and therefore the flexibility for TM1b repositioning may be restricted, suggesting that a gating mechanism involving a significant local rearrangement of TM1b is perhaps unlikely. Therefore, it appears more plausible that access to the central cavity could be regulated by more symmetric rearrangements closing the gap between the top of TM1a and TM8. Nevertheless, additional structures in occluded or inward-facing states are required to describe the conformational changes required for closing the extracellular gate between TM1 and TM8.

3.3.7 Soaking studies with a putative substrate

The large, well-diffracting crystals described in Section 3.2.3 were selected for soaking experiments with the NSAID substrate indomethacin. For these soaking experiments, a 200 mM Indomethacin stock solution was prepared in DMSO and subsequently diluted to prepare a soaking reservoir solution containing 5 mM indomethacin (0.1 M ADA pH 6.5, 0.2

M Lithium citrate, 30 % (v/v) PEG400, 5 mM indomethacin, 2.5 % (v/v) DMSO). Then, LCP glass sandwich plate wells containing the crystals to be soaked were opened on one corner, and the reservoir solution was replaced with 1.5 µL of the freshly prepared indomethacin-containing reservoir solution. The wells were re-sealed with a small piece of coverslip and vacuum grease, and the plates were incubated to allow ligand soaking into the crystals in a plastic bag with moist tissue to maintain hydration. Crystals were subsequently harvested after soaking for 1.5h, 3h or 7h (crystals soaked for 7h were visibly affected by the long soaking) and subsequently screened. Crystals soaked for 7h were limited to a resolution of ~4 Å, and therefore data were collected for crystals soaked for 3h, which diffracted to ~2.9 Å. A full dataset was collected and processed to 2.99 Å (Table 3.2), but electron density in the substrate-binding site did not reveal the presence of bound substrate (Fig. 3.14).

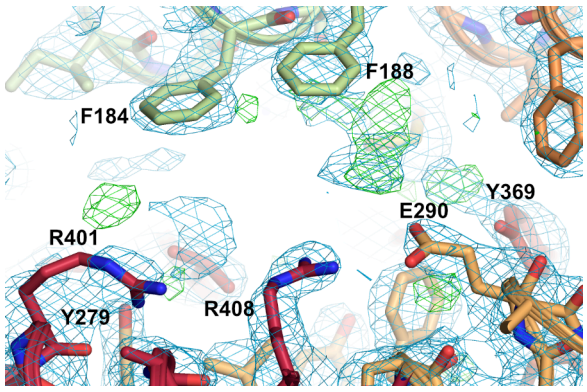


Figure 3.14. Electron density in the MFSD10 substrate-binding site after soaking with indomethacin. Electron density maps are overlaid on the MFSD10 model. The 2mFo-DFc map (contoured at 1 σ) is shown as a blue mesh; the mFo-DFc map (contoured at 2.5 σ) is shown as a green mesh.

Table 3.2 – Data collection statistics for indomethacin soak dataset

Dataset description	*Sf*9 expression 5 mM indomethacin soak (3h)
Space group	P 2_1
Cell dimensions:	
***a, b, c* (Å)**	56.02, 70.69, 63.30
α, β, γ (°)	90.00, 116.07, 90.00
Resolution (Å)*	40.99-2.99 (3.04-2.99)
R_{meas}*	0.247 (1.508)
R_{pim}*	0.096 (0.576)
I/σI*	7.7 (1.4)
$CC_{1/2}$*	0.99 (0.47)
Completeness (%)*	99.9 (100.00)
Multiplicity*	6.5 (6.8)

* Values in parentheses are for the highest-resolution shell

Given the partially occluded conformation of MFSD10 in the crystals (for a detailed description, see Sections 3.3.3 and 3.3.6), it may be difficult for the substrate to successfully diffuse into the central cavity of MFSD10. Therefore, efforts were shifted towards co-crystallisation attempts with a number of proposed NSAID substrates.

3.3.8 Co-crystallisation studies with putative substrates

Initially, the electron density in the binding site of the MFSD10 structure described in earlier sections was not immediately interpreted as corresponding to a citrate ion from the crystallisation condition. The presence of citrate was only acknowledged at a later stage after several datasets had been collected from crystals grown in citrate-containing conditions, which consistently revealed electron density in the binding site that could be nicely accounted for by a citrate ion. As a result, my attempts to co-crystallise MFSD10 with bound substrates

initially focused primarily on citrate-containing conditions, as these were the only conditions that yielded crystals of MFSD10 at first. In this section, I will describe how the collection of datasets from MFSD10 crystals grown in the presence of citrate and different NSAID substrates led me to finally pursue crystallisation in conditions that do not contain citrate.

At first, MFSD10 crystallisation success varied considerably between different purifications and was made more difficult in the presence of putative NSAID substrates (and likely DMSO, in which the substrate stocks were prepared). After some persistent trials, crystals were obtained in the presence of 10 mM sulindac in an optimisation screen well containing 0.1 M citrate pH 4.5, 20 % (v/v) PEG 300 (Fig. 3.15A). These plate-shaped crystals belonged to the P 2_1 space group and diffracted to high resolution (Table 3.3). Solving the structure revealed ambiguous electron density in the central cavity (Fig. 3.15B). It is likely that citrate prevents successful binding of the substrate, and therefore the observed electron density appears to correspond to a bound citrate ion from the crystallisation condition. Additional residual density stretching throughout the central cavity could not be attributed to sulindac.

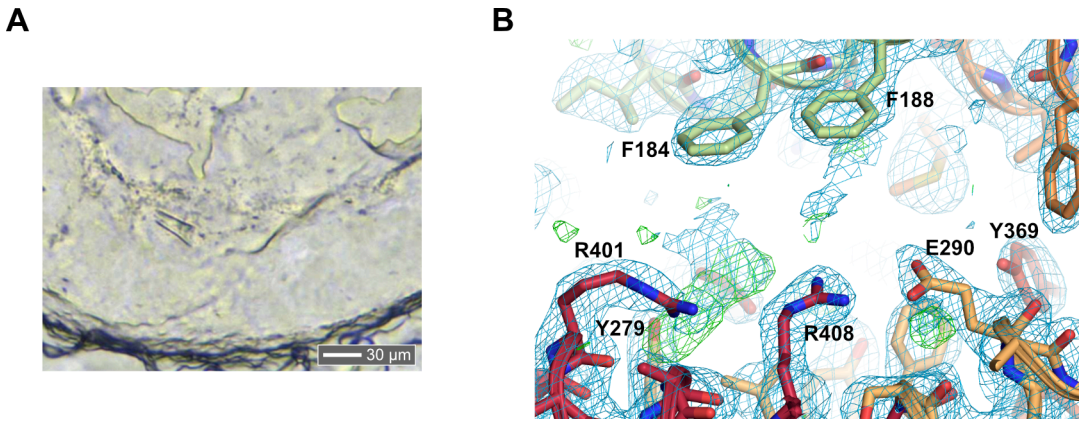


Figure 3.15 MFSD10 co-crystallisation attempt with sulindac. A. MFSD10 LCP crystals in 0.1 M citrate pH 4.5, 20 % (v/v) PEG 300, 10 mM sulindac. **B.** Electron density maps overlaid on the MFSD10 model. The 2mFo-DFc map (contoured at 1 σ) is shown as a blue mesh; the mFo-DFc map (contoured at 2.5 σ) is shown as a green mesh.

Due to the difficulty in obtaining MFSD10 crystals in the presence of substrates, additional strategies were devised. First, the purification protocol of MFSD10 was modified to include a membrane isolation step prior to purification (detailed in Section 2.2.1), with the rationale that this additional step enables the separation of membrane and soluble proteins prior to purification, thus increasing the purity of the final sample [238]. MFSD10 was purified from the prepared membranes to high purity, and the resulting protein was subjected to crystallisation attempts in LCP.

One additional variable was introduced at this stage: besides attempting to co-crystallise MFSD10 with substrates using monoolein (MAG9.9) as the host lipid, crystallisation was also attempted using the homologous monoacylglycerols MAG7.7, MAG7.8 and MAG 7.9. However, this change of host lipid did not yield any crystals.

Crystallisation was also attempted by supplementing monoolein with 10 % (w/w) cholesterol. The use of cholesterol as a supplement has been a popular option for the *in meso* crystallisation of GPCRs [227, 239] but has also been a successful strategy for the crystallisation of other membrane proteins, including SLCs [240]. This approach successfully generated well-diffracting MFSD10 crystals in conditions containing 10 mM indomethacin. Medium-sized (~40 μm) plate-shaped crystals could be obtained in 0.1M ADA, 0.1M lithium citrate, 25 % (v/v) PEG 400 (using an optimisation screen) (Fig. 3.16A). A full dataset was collected and processed to 2.75 Å resolution (Table 3.3). However, these crystals were again obtained in a condition with a high concentration of citrate (100 mM), which likely prevents the binding of indomethacin (Fig. 3.16B). Additional attempts were carried out to obtain crystals in conditions without citrate, but were not successful using this approach.

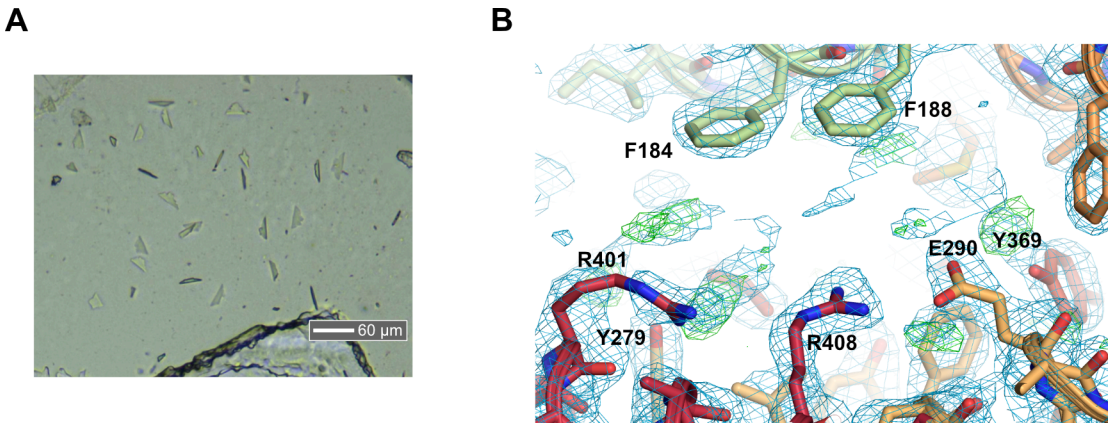


Figure 3.16 MFSD10 co-crystallisation attempt with indomethacin. A. MFSD10 LCP crystals in 0.1M ADA, 0.1M lithium citrate, 25 % (v/v) PEG 400, 10 mM indomethacin. **B.** Electron density maps overlaid on the MFSD10 model. The 2mFo-DFc map (contoured at 1 σ) is shown as a blue mesh; the mFo-DFc map (contoured at 2.5 σ) is shown as a green mesh.

The need to obtain crystals of MFSD10 in the absence of citrate was now evident, but all attempts carried out up until this point had been unsuccessful. Next, I attempted to use MFSD10 protein expressed in Expi293F cells. Recombinant expression of human membrane proteins in mammalian expression systems exposes the protein to a more native lipid environment and can also result in different post-translational modifications, often affecting the stability of the purified protein [241, 242]. In the case of MFSD10, I hypothesised that this change in expression system could potentially affect crystallisation success rates. The MFSD10 construct used corresponds to the same truncation used for expression in *Sf*9 cells (Section 3.2.1), but was fused on its C-terminus to a TEV protease cleavage site followed by GFP and Twin-Strep and 10x His tags. Purification was carried out as described in Section 2.1.3.1, yielding pure mono-dispersed protein (Fig. 3.17A,B). Denaturing intact protein mass spectrometry analysis reveals the same short N-terminal truncations as observed for *Sf*9-expressed MFSD10, corresponding to loss of the initiator methionine or of the first 5 N-terminal residues (MGWGG) (Fig. 3.17C). However, additional deconvoluted mass peaks are observed in Expi293F-expressed protein, corresponding to palmitoylation at one site (theoretical mass addition: 238.41 Da) of each of the two N-terminal truncations. The first

of these peaks (47752.92 Da) corresponds to a 238.29 Da addition to the MFSD10 truncation lacking the 5 initial N-terminal residues (47514.63 Da). Likewise, the second additional peak (48109.72 Da) corresponds to a 237.77 Da addition to the MFSD10 species lacking only its N-terminal methionine. Overall, these results show that a fraction of the Expi293F-expressed MFSD10 protein is palmitoylated at one site.

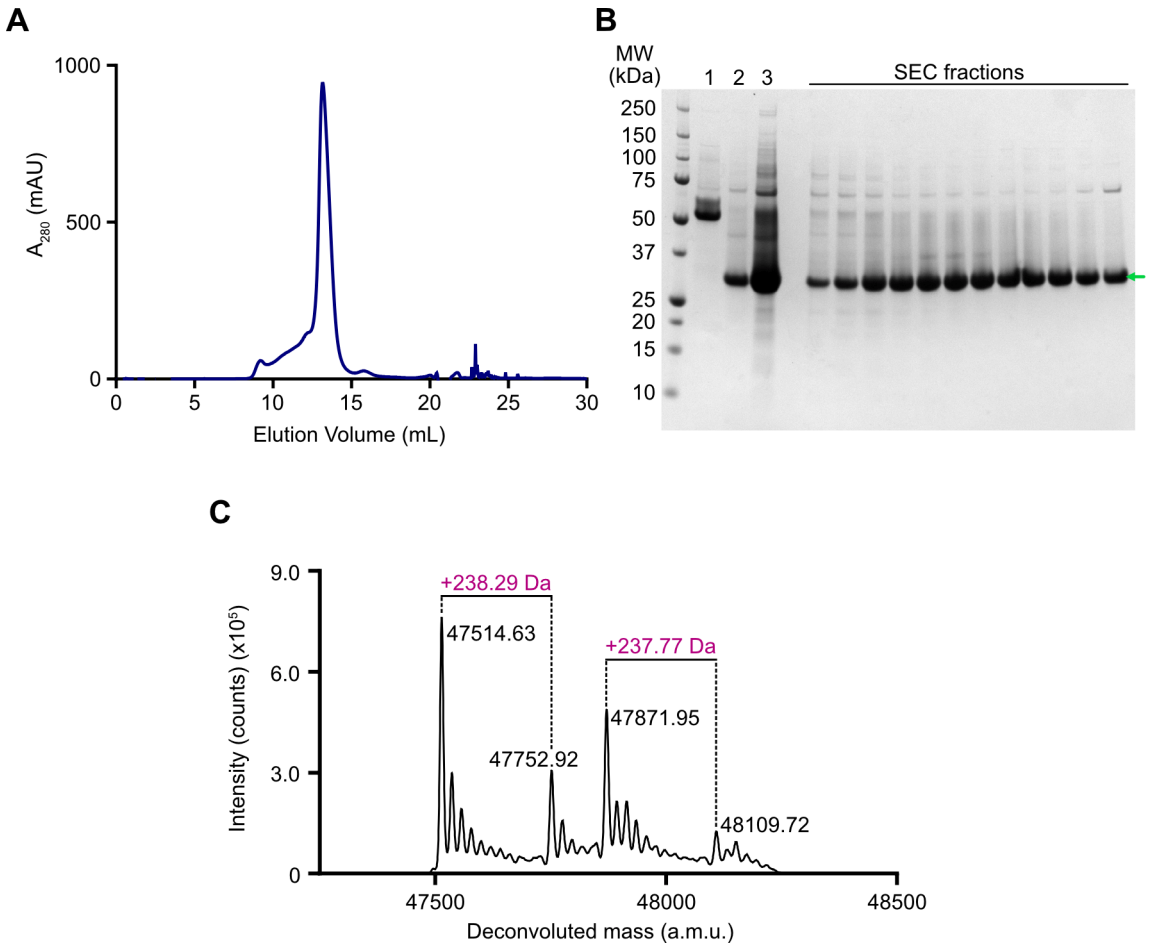


Figure 3.17. Purification of MFSD10 expressed in Expi293F cells. **A.** Elution profile of MFSD10 from a Superdex 200 Increase 10/300 SEC column. **B.** SDS-PAGE analysis of samples taken during MFSD10 purification. *1* – Eluate from Strep-Tactin XT resin (higher molecular weight due to C-terminal GFP – Twin-Strep – 10x His fusion). *2* – Flow-through from the reverse IMAC step used to capture uncleaved MFSD10 and His-tagged TEV. This flow-through sample contains uncleaved MFSD10 protein. *3* – concentrated MFSD10 sample loaded onto the SEC column. *SEC fractions* – peak fractions from (A). **C.** LC-ESI-MS intact mass analysis of purified MFSD10 expressed in Expi293F cells. MFSD10 expressed in Expi293F cells is palmitoylated at one site, as observed by the presence of +238.29 Da and +237.77 Da mass shifts relative to the truncated species previously described (theoretical mass addition for palmitoylation at one site: +238.41 Da).

Crystallisation attempts were carried out with the Expi293F-expressed protein, using monoolein supplemented with 10 % (w/w) cholesterol as the host lipid. In the absence of substrates, extensive screening of protein concentrations and crystallisation conditions in sparse-matrix screens was conducted. However, crystals were again only obtained in similar citrate-containing conditions. Medium-sized plate-shaped crystals were obtained in 0.1 M citrate pH 5.5, 35 % (v/v) PEG300 (Fig. 3.18A). These crystals again belonged to the P 2_1 space group, and a full dataset was collected to 2.80 Å resolution (Table 3.3) as a comparison to the original dataset obtained for *Sf*9-expressed protein. As expected, the determined structure corresponds to the same conformation of MFSD10 as before, with electron density suggesting that citrate is bound in the central cavity (Fig. 3.18B).

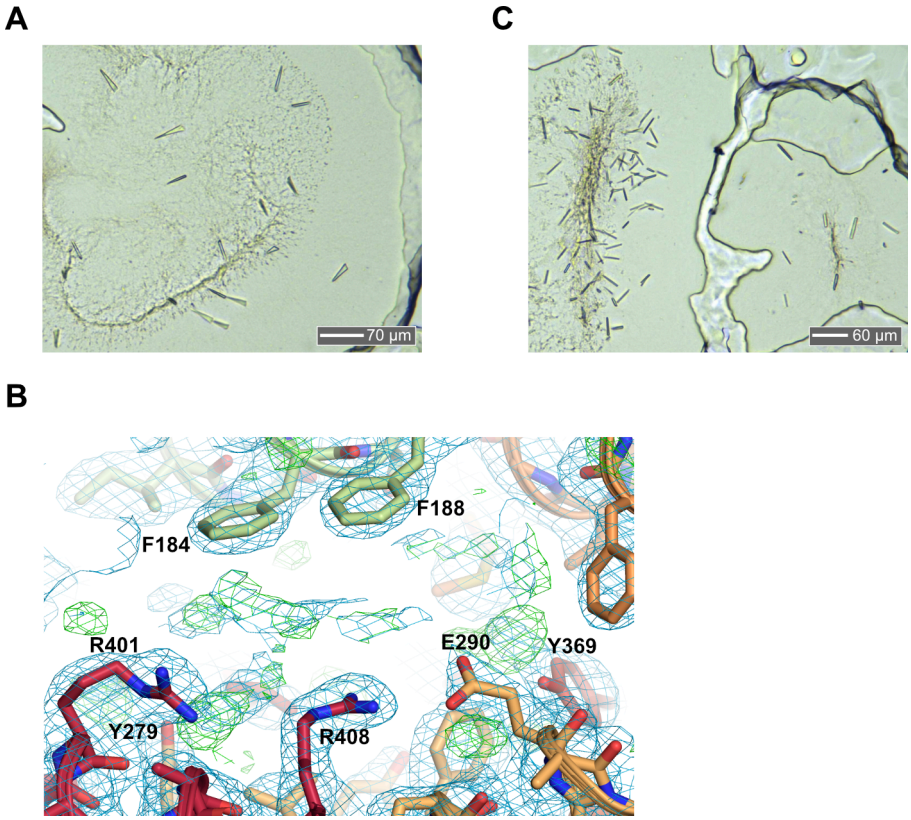


Figure 3.18 Crystallisation attempts with Expi293F-expressed MFSD10. A. MFSD10 LCP crystals in 0.1 M citrate pH 5.5, 35 % (v/v) PEG300. **B.** Electron density in the substrate-binding site of MFSD10 in the crystals shown in panel A. Electron density maps are overlaid on the MFSD10 model. The 2mFo-DFc map (contoured at 1 σ) is shown as a blue mesh; the mFo-DFc map (contoured at 2.5 σ) is shown as a green mesh. **C.** MFSD10 LCP crystals in 0.1 M ADA pH 6.5, 0.16 M Lithium sulfate, 0.004 M Strontium chloride hexahydrate, 40 % (v/v) PEG 400, 10 mM indomethacin.

Crystallisation was also attempted in conditions without citrate, using the Expi293F-expressed protein in the presence of a number of substrates such as diclofenac, indomethacin, and biliverdin. Encouragingly, crystals were obtained in the presence of 10 mM indomethacin, in the MemGoldMeso screen condition F07 (0.1 M ADA pH 6.5, 0.16 M Lithium sulfate, 0.004 M Strontium chloride hexahydrate, 40 % (v/v) PEG 400), yielding long (up to ~45 μm) rod-shaped crystals (Fig. 3.18C) after 8 days. However, these crystals were weakly diffracting, and their resolution was limited to ~9 Å.

Overall, as the presence of citrate in the binding site was not immediately clear, significant efforts were spent optimising crystallisation systems in citrate-containing conditions. Once multiple datasets were obtained, it became evident that citrate was occupying the substrate-binding site, thus out-competing the added substrates. At this stage, crystals in conditions without citrate were pursued, leading to initial poorly-diffracting crystals being obtained. Unfortunately, these early crystals could not be optimised within the time scale of the project. Furthermore, additional strategies were devised to obtain MFSD10 structures with bound NSAIDs in the future and will be discussed in Section 6.1.

Table 3.3 – Data collection statistics for additional MFSD10 datasets

Dataset ID	d005	d010	d014
Dataset description	*Sf*9 expression No membrane prep 5 mM Sulindac	*Sf*9 expression Membrane prep 10 mM Indomethacin	Expi293F expression No membrane prep No compound
Space group	P 2_1	P 2_1	P 2_1
Cell dimensions:			
***a, b, c* (Å)**	55.89, 70.06, 63.40	55.85, 70.12, 63.35	55.81, 70.42, 63.58
α, β, γ (°)	90.00, 115.82, 90.00	90.00, 115.42, 90.00	90.00, 115.80, 90.00
Resolution (Å)*	35.03-2.64 (2.74-2.64)	49.94-2.75 (2.90-2.75)	57.24-2.80 (2.95-2.80)
R_{meas}*	0.2004 (1.016)	0.340 (1.106)	0.375 (1.560)
R_{pim}*	0.0986 (0.534)	0.134 (0.442)	0.143 (0.595)
I/σI*	6.1 (1.2)	4.4 (1.7)	4.0 (1.2)
$CC_{1/2}$*	0.99 (0.52)	0.99 (0.63)	0.67 (0.48)
Completeness (%)*	99.37 (97.96)	99.9 (99.7)	100.00 (100.00)
Multiplicity*	4.0 (3.3)	6.3 (6.1)	6.9 (6.8)

* Values in parentheses are for the highest-resolution shell

3.4 Insights into the transport mechanism

3.4.1 MFSD10 acts as a proton-coupled NSAID antiporter

As introduced in Section 1.1.4, MFSD10 is classified as a member of the drug:proton (H^+) antiporter family 1 (DHA1) of MFS-MDR transporters, which uses the outwardly-directed proton electrochemical gradient to drive the efflux of multiple structurally diverse substrates.

In order to confirm the proposed drug:H^+ antiport activity of MFSD10, proton movement during the transport cycle was directly monitored *in vitro* using a pyranine-based assay. Pyranine is a membrane-impermeable ratiometric pH-sensitive dye that can be encapsulated into proteoliposomes and used to track changes in their internal pH [243, 244]. By alternately

exciting pyranine at 415 nm and 460 nm, while monitoring fluorescence at 510 nm, the ratio of fluorescence emission upon excitation at 460 nm / 415 nm (F_R) increases with increasing pH. In this way, changes in the internal pH of the proteoliposomes can be monitored in real time upon the addition of substrates to the outside buffer.

To set up the assay, purified MFSD10 expressed in *Sf*9 insect cells was first reconstituted into POPE:POPG (3:1 (w/w)) proteoliposomes at 1:50 protein:lipid (w/w) ratio, and the resulting proteoliposomes were loaded with pyranine through 3 freeze-thaw cycles (see Section 2.2.7). As this method requires that no protons leak across the membrane of the (proteo)liposomes during the timescale of the assay, the ability of these POPE:POPG liposomes to hold a pH gradient (pH 6.8 inside / pH 7.5 outside) was evaluated. Pyranine fluorescence was monitored for 120 s, demonstrating that the liposomes produced are capable of holding a gradient of 0.7 pH units (Fig. 3.19).

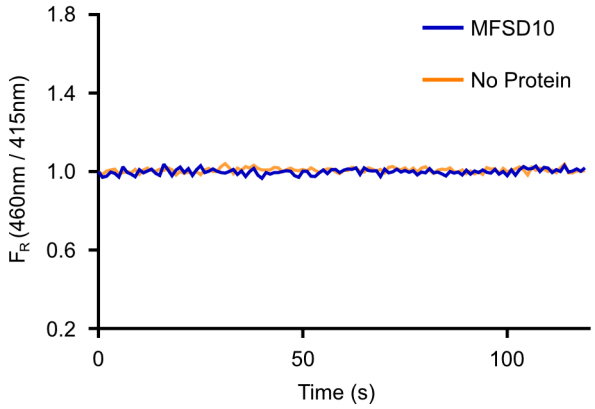


Figure 3.19 Prepared liposomes can hold a pH gradient. POPE:POPG MFSD10 proteoliposomes and no-protein control liposomes were prepared with an internal pH of 6.8. These were then diluted into a pH 7.5 external buffer. Pyranine fluorescence was monitored for 120 s (emission 510nm; excitation 460nm and 415nm), and the F_R (ratio of fluorescence at 510nm when excited at 460 nm / 415 nm) is plotted. No F_R change could be observed during the timescale of the assay, indicating that both MFSD10 proteoliposomes and control liposomes are able to hold a gradient of 0.7 pH units.

Having demonstrated that these liposomes do not leak protons under the assay conditions, MFSD10-reconstituted and protein-free liposomes (internal buffer: 5 mM HEPES pH 6.8, 119 mM NaCl, 1 mM KCl, 2 mM $MgSO_4$, 1 mM pyranine) were used to monitor diclofenac-driven proton flux (Fig. 3.20A). In this experiment, the liposomes were diluted 1/25 in a pH 6.8 external buffer (5 mM HEPES pH 6.8, 1 mM NaCl, 119 mM KCl, 2 mM MgSO4, 1 mM pyranine), thus resulting in no pH gradient across the liposome membrane. Pyranine fluorescence was recorded to monitor changes in the internal pH of the liposomes. Using this experimental setup, I observed that the addition of diclofenac to the outside of the proteoliposomes results in alkalinisation of the lumen (Fig. 3.20B-C). Control experiments with no-protein liposomes show that this is a robust system and that the recorded increase in internal pH is a direct result of diclofenac transport by MFSD10. In summary, this method enabled the indirect tracking of substrate (diclofenac) transport by directly monitoring the coupled proton movement and can be applied to the study of other putative substrates, including other NSAIDs. These diclofenac-driven proton flux experiments demonstrated that protons flow in the opposite direction of diclofenac, confirming earlier reports that diclofenac is a substrate of MFSD10 and that MFSD10 acts as an NSAID:H^+ antiporter.

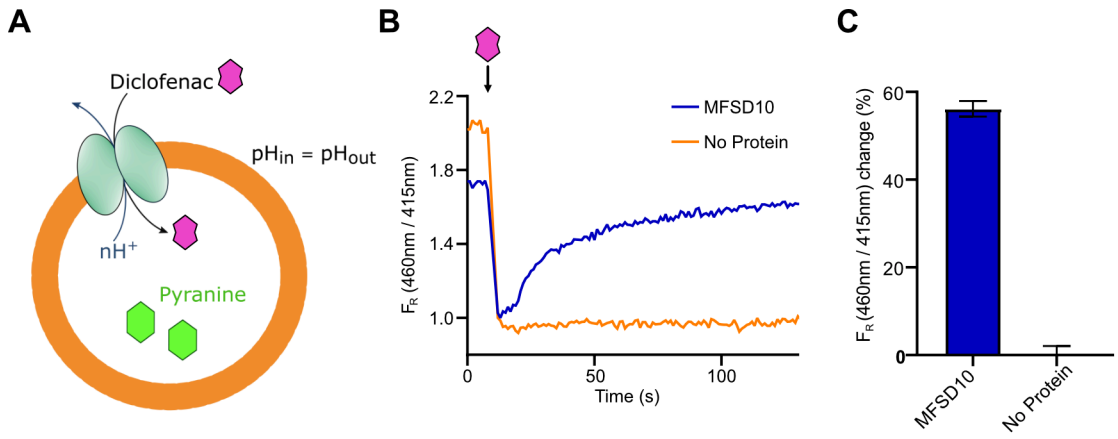


Figure 3.20 Pyranine-based assay to demonstrate MFSD10-dependent, NSAID:H^+ antiport. A. Schematic representation of the experimental setup for diclofenac-driven proton flux experiments.

MFSD10 (light green ovals) was reconstituted into liposomes (orange ring), which were then loaded with the pH-sensitive dye pyranine (green hexagon). Addition of diclofenac (2 mM; magenta polygon) drives proton efflux from the liposomes. This results in alkalinisation of the liposome lumen, which alters pyranine fluorescence. **B.** MFSD10 activity was monitored using this experimental setup. Pyranine fluorescence was monitored (emission 510 nm; excitation 460 nm and 415 nm), and the F_R (ratio of fluorescence at 510 nm when excited at 460 nm / 415 nm) is plotted. Addition of 2mM diclofenac (marked by an arrow/magenta polygon) causes an initial drop in F_R and subsequently leads to alkalinisation of MFSD10-containing proteoliposomes (observed as an increase in F_R; blue trace). In control liposomes without protein, F_R remains unchanged after the initial drop caused by diclofenac interference (orange trace). **C.** Bar chart showing the relative change in F_R in MFSD10 proteoliposomes and control liposomes without protein. Data shown are the average from triplicate experiments; Positive and negative error bars correspond to ± S.E.M.

3.4.2 Structural insights into the mechanism of proton coupling

As introduced in Section 1.1.3, MFS antiporters couple the dissociation of one of the substrates to binding of the second substrate [41, 43]. In the case of MFS drug:H^+ antiporters, substrate release to the extracellular environment must occur prior to protonation of the transporter, such that the transporter does not prematurely re-orient itself to the inward-facing state. This can be achieved by modulating the environment of protonatable residues such that substrate binding favours their unprotonated state.

MFSD10 contains an acidic residue in the central cavity, Glu290, located on TM7. Importantly, this residue is conserved across species, and an acidic residue at this position is conserved in several DHA1 transporters (see Section 3.5). In the outward-facing conformation, Glu290 is solvent-accessible from the extracellular environment and forms a salt bridge with Arg408 (TM11), likely stabilising its negative charge (Fig. 3.21A,B). This close proximity to Arg408 likely precludes Glu290 protonation, thus blocking the conformational shift to the inward-facing state. Since MFS H^+-coupled antiporters couple substrate release to protonation, and because Arg408 forms an integral part of the substrate-binding site (Fig. 3.21B), it is possible that displacement of the Arg408 side-chain upon substrate release could act as the trigger for protonation of Glu290.

Furthermore, there are two well-ordered water molecules within hydrogen-bonding distance of Glu290 (Fig. 3.21B). One of these water molecules (Wat1) is located in a small pocket, within hydrogen-bonding distance of Glu290 and Tyr369. The second water molecule (Wat2) is part of an ordered water network leading to Glu290. This water network extends in the direction of the narrow extracellular entrance to the central cavity, where two additional water molecules can be observed (Fig. 3.21A). The observation that Glu290 is in close proximity to a network of ordered water molecules in a solvent-filled cavity further suggests it is ideally placed to be protonated in the outward-facing conformation. Altogether, these features are consistent with a potential mechanism to effectively couple substrate release to protonation (see Section 3.5).

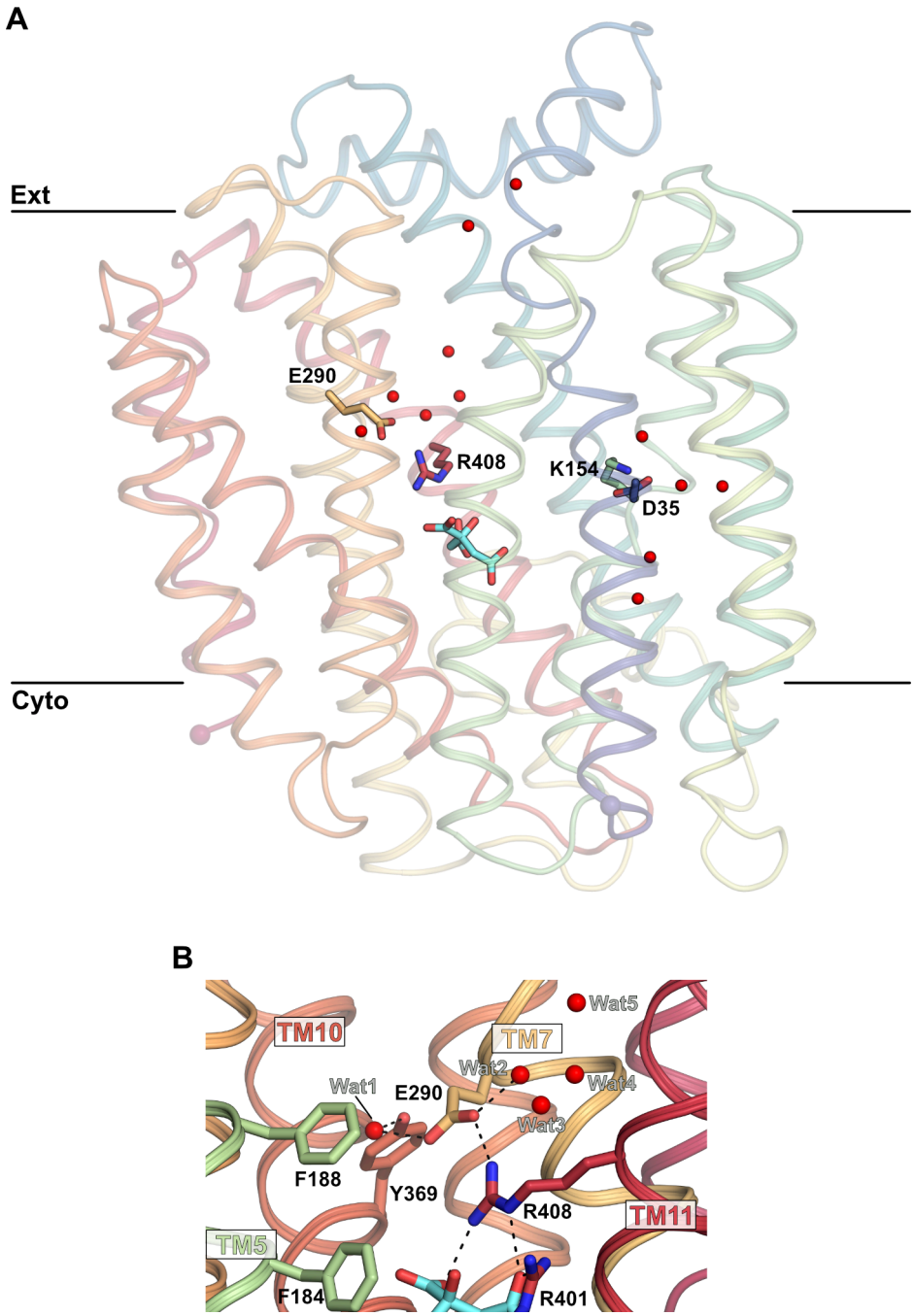


Figure 3.21. Acidic residues in and adjacent to the central cavity of MFSD10. A. Position of the Glu290-Arg408 and Asp35-Lys154 pairs in the MFSD10 structure, viewed parallel to the membrane plane. Glu290 (TM7) is located in the central cavity above the substrate-binding site, sits at the bottom of a solvent-filled cavity and is within hydrogen-bonding distance of two water molecules (Wat1 and Wat2, see (B); displayed as red spheres). Therefore, it is in a suitable position to be protonated from the extracellular environment. Asp35 (TM1) is located adjacent to but inaccessible from the central cavity in this outward-facing conformation and is also surrounded by water molecules (red spheres). **(B)** Close-up view of the Glu290-Arg408 pair viewed parallel to the membrane plane. Glu290 forms hydrogen bonds with two water molecules, Wat1 and Wat 2 (red spheres), and is therefore ideally placed to be protonated. Glu290 forms a salt bridge with Arg408 (TM11), which likely precludes premature protonation of Glu290 by stabilising its negative charge. Arg408

is an integral part of the substrate-binding site, participating in forming the polar pocket where a citrate carboxylate is bound (citrate ion: cyan / red, shown in stick representation).

One additional acidic residue, Asp35 (TM1), is located in close proximity to the central cavity and forms a salt bridge with Lys154 (TM4) (Fig. 3.21A). Asp35 is not solvent-accessible from the central cavity in the outward-facing conformation, but might become accessible during the conformational shift to the inward-facing state. As a result, Asp35 could also play a role in the proton transfer pathway, and this possibility should be tested biochemically in the future.

3.5 Discussion

The atypical SLC MFSD10 has been shown to provide resistance to NSAID-mediated apoptosis in a human adenocarcinoma gastric cell line [56]. Later, it was demonstrated to transport a range of organic anions, including NSAIDs, and this transport activity was shown to be dependent on the proton motive force [60]. As a result, and due to the presence of an antiporter motif in its primary sequence, MFSD10 is annotated as a drug:H^+ antiporter, with multidrug resistance activity. Therefore, there is a significant interest in obtaining insights into its transport mechanism. Here, I solved the crystal structure of human MFSD10, captured in an outward-facing partially occluded conformation with a citrate ion from the crystallisation solution bound in the putative substrate-binding site.

Extensive efforts were carried out to obtain substrate-bound structures of MFSD10 to determine the basis of substrate recognition. However, the presence of a bound citrate ion was not immediately recognised and therefore soaking and co-crystallisation efforts were initially largely focused on citrate-containing conditions. Importantly, the presence of high citrate concentrations in these conditions likely prevents substrate binding to MFSD10. Thus, this led to efforts being shifted towards co-crystallisation in conditions without citrate. After

extensive screening, initial crystals were obtained using protein expressed in Expi293F cells and monoolein supplemented with 10 % (w/w) cholesterol as the host lipid. However, their diffraction was limited to ~9 Å. These crystals could not be optimised within the time-frame of the project but could be followed up in the future. Other approaches that could enable obtaining substrate-bound structures of MFSD10 are described in Section 6.1. Nevertheless, the presence of a bound citrate ion in the structure provides initial insights into how organic anions are recognised by MFSD10.

The present structure also provides initial suggestions into the proton-coupling mechanism of MFSD10. MFS antiporters are known to couple the release of one substrate to the binding of the other, such that both substrates do not bind simultaneously [41, 43]. Thus, proton-coupled antiporters require the coupling of substrate release to protonation, thus avoiding premature protonation. This way, substrate binding is thought to impact the pKa of protonatable residues in the central cavity, either directly (through overlapping binding sites) or indirectly [43, 49]. Such a mechanism has been clearly demonstrated for example for tetraphenylphosphonium (TPP^+) transport by MdfA, where substrate binding indirectly competes with protonation of the transporter [48, 245]. On the other hand, the structure of the DHA1 family transporter NorA (recently published by Da-Neng Wang's group) shows overlapping binding sites for substrates and protons, suggesting these could directly compete for binding to the transporter [51]. The citrate-bound MFSD10 structure reveals the presence of an acidic residue in the central cavity in the outward-facing conformation (Glu290, TM7), which forms a salt bridge with a neighbouring arginine (Arg408, TM11). This salt bridge likely stabilises the negative charge of Glu290, precluding its protonation from the extracellular side. Arg408 forms an integral part of the substrate-binding site, thus suggesting a potential route for modulation of the pKa of Glu290 upon substrate binding / release. An additional acidic residue (Asp35, TM1) is located adjacent to the central cavity but is not directly

accessible in this partially occluded outward-facing conformation. However, this residue could face the central cavity, for example, during the shift to the occluded conformation and could participate in proton transport. Notably, acidic residues located in TM7 of several bacterial multidrug resistance transporters of the DHA1 family have been proposed to participate in proton coupling [51, 246-248]. In addition, human H^+-coupled vesicular monoamine transporters VMAT1 and VMAT2 (SLC18A1 and SLC18A2, respectively), which also belong to the DHA1 family, contain essential acidic residues in TM1 and TM7, proposed to play a role in proton transport [249, 250]. Moreover, acidic residues have also been identified in TM1 of bacterial DHA1 transporters such as MdfA and MdtM. In MdfA, these have been shown to be protonated as part of the transport mechanism [48], and a similar role has been proposed for the corresponding residues in MdtM [50]. Asp35 (TM1) of MFSD10 is conserved in human VMAT1 and VMAT2, and the acidic residues of MdfA and MdtM are located in adjacent positions in TM1 (Fig. 3.22A). Strikingly, an acidic residue in the position corresponding to Glu290 (TM7) of MFSD10 is conserved in human VMAT1 and VMAT2, *E. coli* EmrD, *L. lactis* LmrP, and *S. aureus* NorA (Fig. 3.22B). Together, these observations further suggest these residues are likely involved in proton transport by MFSD10.

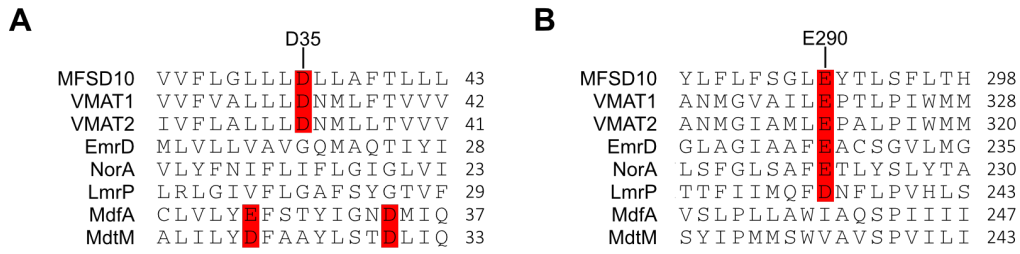


Figure 3.22 Sequence alignment of selected human and bacterial DHA1 transporters. A. Alignment is shown for the region of TM1 around Asp35 of MFSD10. **B.** Alignment is shown for the region of TM7 around Glu290 of MFSD10. Acidic residues are shown in red. The sequence alignment was generated using PROMALS3D [251].

Furthermore, nanoDSF experiments with proposed NSAID substrates confirmed that these can interact with the purified protein and suggested candidate substrates for functional assays and co-crystallisation efforts. Given that the proposed drug:H^+ antiporter function of MFSD10 has not been directly demonstrated, a pyranine-based fluorescence assay was then employed to indirectly monitor substrate transport by directly monitoring changes to the internal pH of MFSD10 proteoliposomes. The addition of diclofenac to the outside of the proteoliposomes caused proton efflux from the MFSD10 proteoliposomes, as observed by an increase in their internal pH. This observation validates diclofenac as a substrate and confirms the drug:H^+ antiporter function of MFSD10. Furthermore, this assay can be employed in the future to evaluate the proposed roles of Asp35 and Glu290 in proton transport (see Section 6.1). In addition, although this method can enable the calculation of the substrate:H^+ stoichiometry [243] by determination of the reversal potential (the potential at which no net proton movement is achieved), this requires a transmembrane potential to be established, typically through the use of valinomycin and potassium ion gradients. Additionally, valinomycin can be employed to investigate whether transport can be solely driven by membrane potential ($\Delta\Psi$) [252] and consequently whether the transporter is electrogenic. Unfortunately, the MFSD10-containing proteoliposomes and control liposomes here described did not withstand the simultaneous addition of valinomycin and NSAID substrates such as diclofenac and indomethacin, which resulted in proteoliposome lysis, as observed by a rapid equilibration of the luminal and external pH. As a result, the current setup does not enable the calculation of the substrate:H^+ stoichiometry in the transport cycle of MFSD10 or the evaluation of whether membrane potential can drive transport, but could be optimised in the future, for example by altering the lipid composition of the liposomes.

However, given that MFSD10 substrates are negatively charged and are moved in the opposite direction of protons, this transport process would necessarily be electrogenic if no other unknown compounds / ions are co-transported.

Another pending question regarding the biology of MFSD10 is the nature of its natural substrates. Purified MFSD10 displays a dark-green colour, likely due to the presence of a co-purified compound. If this compound could be identified (for example, using mass spectrometry-based approaches, discussed in Section 6.1), it could potentially suggest novel candidate substrates of MFSD10. Considering the observed colour of the purified protein and the properties of known MFSD10 substrates, biliverdin (a breakdown product of haem catabolism) is a possible candidate. The spectroscopic properties of purified MFSD10 and biliverdin display significant similarities, suggesting further work should be conducted to investigate if biliverdin can interact with MFSD10 (discussed in Section 6.1). Encouragingly, nanoDSF experiments reveal that the addition of biliverdin leads to a large stabilisation effect on purified MFSD10, suggesting a direct specific interaction. However, due to the uncertainty in the subcellular localisation of MFSD10, it is difficult to suggest a physiological link between MFSD10 and biliverdin. Biliverdin is typically generated from haem in the cytoplasm [253] (or mitochondria, from where it is exported to the cytoplasm by ABCB10 and perhaps other unknown transporters [254]). The majority of cytoplasmic biliverdin is converted to bilirubin by biliverdin reductase [255]. However, biliverdin has also been shown to be transported across the plasma membrane of hepatocytes, into the bile, by an unknown transporter [256]. Given the current need to determine the subcellular localisation of MFSD10 in different cell types, several collaborations have been launched. These are discussed in Section 6.1.

4 Mechanism of fatty acid elongation by human ELOVL7

4.1 Background leading to this work

As introduced in Sections 1.2.7 and 1.2.8, the ELOVL family of fatty acid elongases determine which acyl-CoAs enter the fatty acid elongation cycle in the ER, and therefore are a key player in determining the levels of VLCFAs with different chain lengths. Obtaining structural and mechanistic insights into their function could help understand how fatty acid elongation could be modulated in the context of metabolic disease. To this end, I and others in the integral membrane protein (IMP) group of the SGC have focused on elucidating the mechanism of one member of this family of elongases, human ELOVL7.

ELOVL7, the most recently discovered member of ELOVL elongases in humans [196], preferentially elongates C18 acyl-CoA substrates but also displays activity toward C16 and C20 acyl-CoAs [197]. VLCFA synthesis by ELOVL7 has been shown to promote membrane permeabilisation during necroptosis, a form of regulated cell death, whereas ELOVL7 knockdown using shRNA protected against necroptosis [257]. In addition, ELOVL7 has been shown to be implicated in prostate and gynaecological cancer [196, 258-260], and single-nucleotide polymorphisms in *ELOVL7* are associated with early-onset Parkinson's disease [261].

4.1.1 Crystal structure of human ELOVL7

Before I started working on this project, Dr Laiyin Nie and Dr Ashley Pike from the SGC-IMP group solved the crystal structure of human ELOVL7 to 2.6 Å resolution [213]. The structure revealed an inverted dimer arrangement, which is likely an artefact from purification (see Section 4.2.2). Each ELOVL7 monomer has 7 transmembrane helices (TMs), with

TM2-7 forming a 6-TM barrel around a narrow central tunnel. In this 6-TM barrel, TM2-4 and TM5-7 form inverted-topology repeats (Fig. 4.1B). The structure shows that the central tunnel is open on the cytoplasmic side, but it is closed on the ER face by disulfide-linked loops (Fig. 4.1C,D).

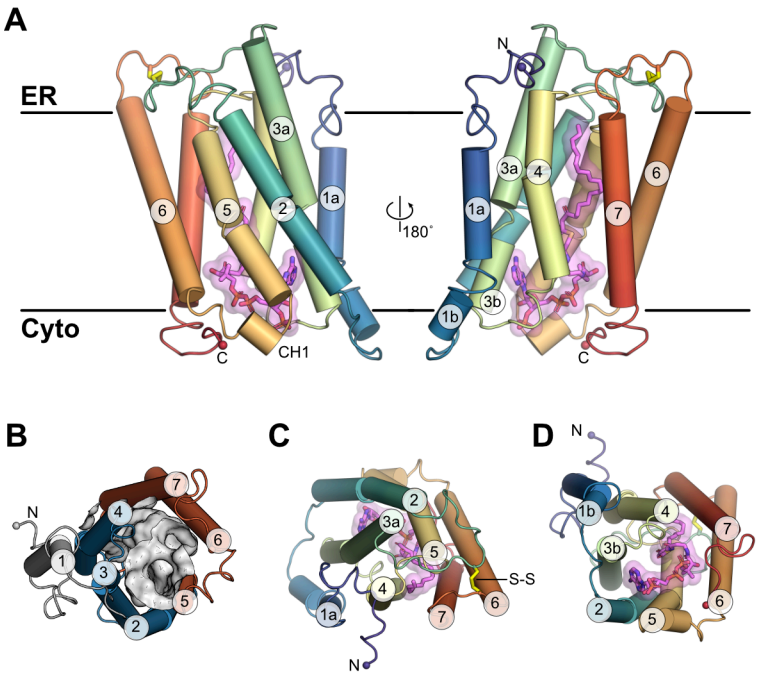


Figure 4.1. Crystal structure of human ELOVL7. A. Overall structure of human ELOVL7, viewed from the plane of the membrane. **B.** The six-transmembrane-helix barrel surrounding the narrow central tunnel (tunnel is shown in surface representation) is formed by two inverted-topology repeats. **C,D.** ELOVL7 structure viewed from **C.** the ER side and **D.** the cytoplasmic side. The co-purified product analogue (described in the next section) is shown in stick and molecular surface representations, in pink. Adapted from Nie et al. (2021) [Ref. 213].

4.1.2 ELOVL7 structure contains a covalently-bound product analogue

Interestingly, the ELOVL7 structure revealed a long hooked non-protein density in the central tunnel, spanning the whole length of the tunnel, from the cytoplasmic entrance to the occluded end (Fig. 4.2A). Furthermore, this density is connected to the protein, suggesting covalent attachment (Fig. 4.2B) (described below). As a result, Dr Laiyin Nie analysed the final purified protein sample by denaturing intact protein mass spectrometry, with assistance

from Mr Tiago Moreira from the Biotechnology Group of the SGC, to identify covalent adducts. Strikingly, analysis of the deconvoluted mass spectra revealed that virtually all purified protein contains a +1073.66 Da adduct (Fig. 4.2C), which corresponds to the mass addition for a 3-keto eicosanoyl-CoA (3-keto C20:0-CoA) covalently attached at two sites (theoretical: +1074.02 Da). Given that ELOVL7 displays preference for C18-CoA substrates [197], this putative 3-keto C20:0-CoA species would correspond to the product of C18:0-CoA elongation by the ELOVL7 enzyme. Dr Ashley Pike modelled this product analogue in the central tunnel, showing remarkable agreement with the electron density (Fig. 4.2B). Notably, the electron density also clearly supports the covalent attachment of this product analogue at two positions: its C2 carbon is linked to the Nε2 of His181, while its C5 carbon is attached to the Nε2 of His150 (Fig. 4.2). This binding configuration is intriguing as it suggests that the bound ligand would likely lead to the covalent inactivation of the enzyme.

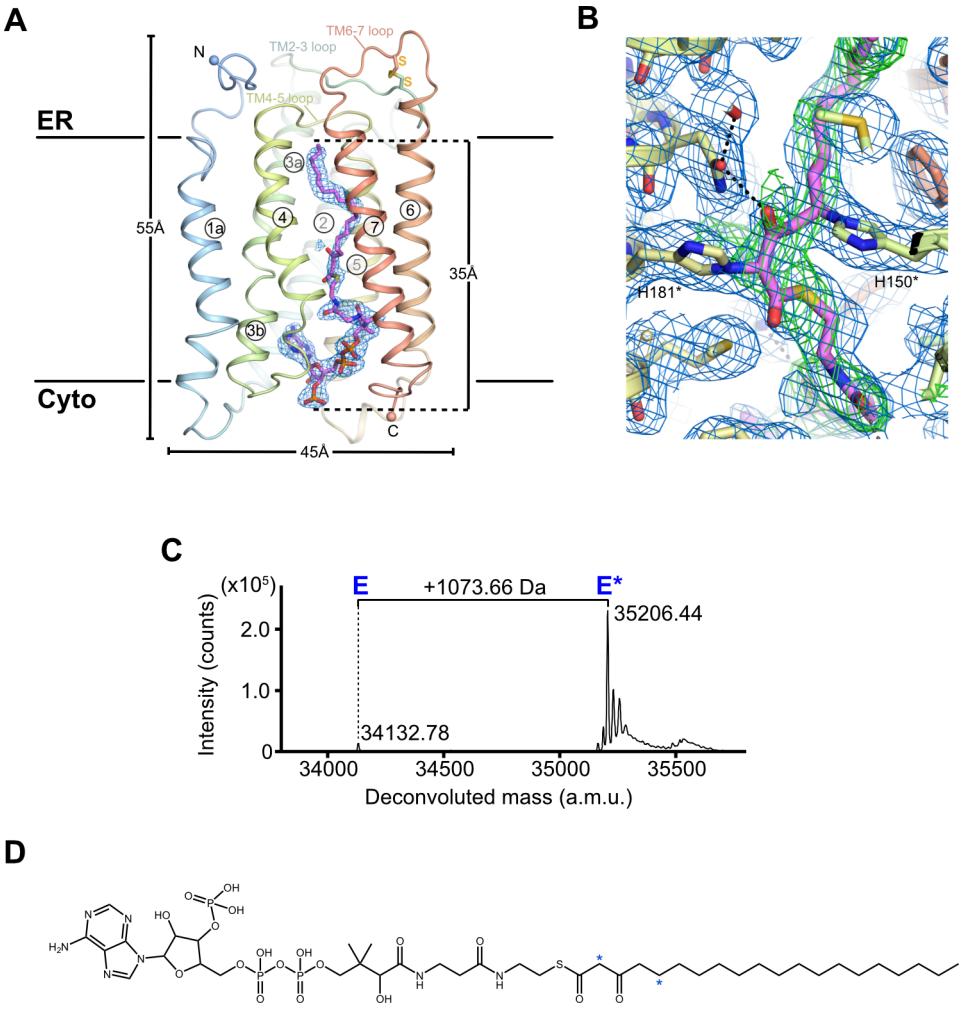


Figure 4.2. Covalently linked co-purified 3-keto acyl-CoA. A. The modelled 3-keto acyl-CoA in the ELOVL7 structure, viewed from the plane of the membrane. The mFo–DFc omit electron density map (contoured at 3σ) is shown as a blue mesh. **B.** Electron density in the active site indicates covalent attachment of the 3-keto acyl-CoA product analogue to His150 and His181. The mFo–DFc omit electron density map obtained in the absence of the modelled analogue is shown as a green mesh (contoured at 2.5σ). The final BUSTER 2mFo–DFc map (with the compound modelled and refined) is shown as a blue mesh (contoured at 1σ). In panels A and B, the covalently attached 3-keto acyl-CoA is shown in stick representation (pink carbon atoms) **C.** Denaturing intact protein mass spectrometry analysis of purified ELOVL7, revealing two species. The first, minor species, corresponds to unmodified ELOVL7 (*E*) with loss of the initiator methionine and acetylation of the new N-terminus (theoretical: 34133.22 Da). The larger deconvoluted mass peak, resulting from a mass addition of +1073.66 Da, corresponds to ELOVL7 covalently modified with 3-keto acyl-CoA (*E**) (theoretical mass addition after the covalent attachment at two sites observed in panel B: +1074.02 Da). Panels A-C are adapted from Nie et al. (2021) [Ref. 213]. **D.** Chemical structure of the 3-keto eicosanoyl-CoA. Carbon atoms covalently linked to the protein in the crystal structure are labelled with blue asterisks.

4.1.3 Substrate-binding and active sites

While the underlying mechanism for the formation of the unusual covalent attachments to the 3-keto eicosanoyl-CoA is not fully understood (see Discussion Section 4.3.2), the fortuitous co-purification of this product analogue likely stabilised the enzyme during its purification and crystallisation. In addition, it provides valuable insights into the binding mode of acyl-CoAs to ELOVL7 (Fig. 4.3A) and potentially to other ELOs as this is the first known structure for the family. The structure reveals that the CoA moiety binds at the cytoplasmic entrance of the central tunnel, where it is specifically recognised by hydrogen bonds and salt bridges [213]. In contrast, the acyl chain binds in the occluded end of the narrow tunnel and, as expected, is lined primarily by hydrophobic sidechains from TM4 and TM7 [213].

Furthermore, the positioning of the product analogue thioester reveals that the active site is located within the membrane and is lined by residues conserved across ELO elongases [213], including the conserved ^{147}HxxHH^{151} motif (TM4) previously demonstrated to be required for the activity of ELOVL7 [197] and other ELOs [194, 195] (Fig. 4.3B). HxxHH and other histidine box motifs can be found in metal-binding proteins, including acyl-CoA [262] and lipid [263] desaturases. For example, in the human stearoyl-coenzyme A desaturase 1 (SCD1), three histidine box motifs coordinate two zinc ions in the catalytic site [264]. In the ELOVL7 structure, the geometry of the HxxHH histidines is not compatible with metal ion binding (Fig. 4.3B), and no density corresponding to metal ions could be observed, suggesting a different role for the HxxHH motif in the mechanism of ELO elongases [213].

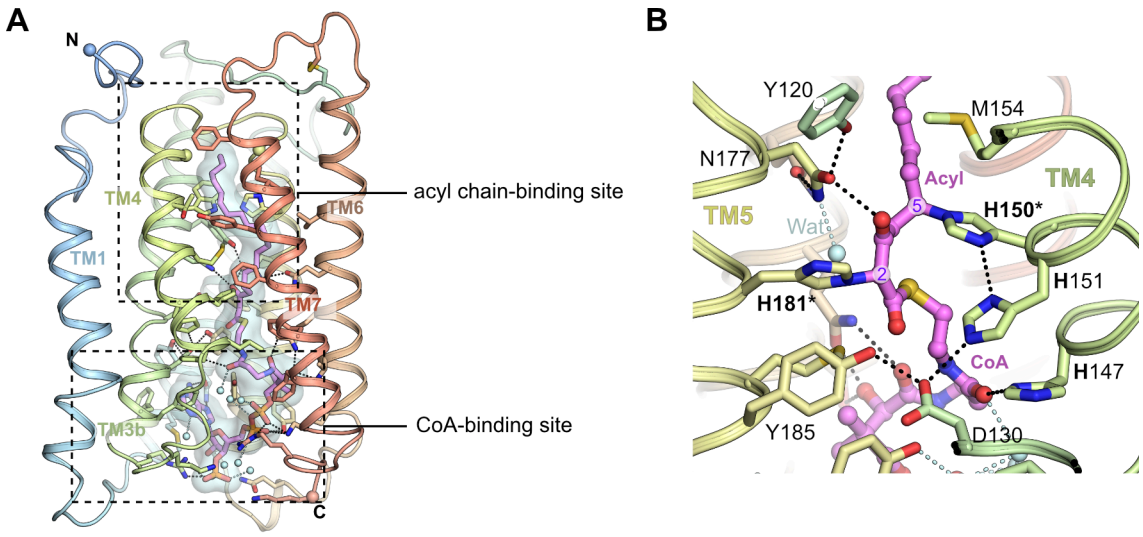


Figure 4.3. ELOVL7 substrate-binding and active sites. A. Binding sites for the acyl chain and CoA moieties in the ELOVL7 structure. The structure is viewed parallel to the membrane plane. The modelled 3-keto acyl-CoA is shown in stick (pink carbon atoms) and molecular surface (green) representations. Waters are shown as pale blue spheres. **B.** Active site viewed parallel to the membrane plane. Histidine residues covalently linked to the 3-keto acyl-CoA are labelled with an asterisk. The 3-keto acyl-CoA is shown in ball-and-stick representation. A water molecule (Wat) is shown as a pale blue sphere. Hydrogen bonds are shown as black dashes. Adapted from Nie et al. (2021) [Ref. 213].

4.1.4 Aims

The crystal structure of ELOVL7, solved by Dr Laiyin Nie and Dr Ashley Pike in the group, provided valuable insights into the architecture of the substrate-binding and active sites, which constitutes a useful starting point to interrogate the mechanism of this enzyme. Here, I aimed to characterise the mechanism of human ELOVL7 using denaturing intact protein mass spectrometry. My experimental results are reported in Sections 4.2.1 – 4.2.8. Together with the structural data previously obtained, this leads to the proposal of a ping-pong mechanism involving an unexpected acyl-imidazole intermediate [213].

4.2 Biochemical characterisation of the reaction mechanism of ELOVL7

4.2.1 Purification of wild-type ELOVL7

Wild-type full-length human ELOVL7 was expressed in *Sf*9 cells fused on its C-terminus to a TEV protease cleavage site, followed by 10x His and FLAG tags, and purified as described in Section 2.1.3.2 [213]. Briefly, the protein was solubilised with 1 % (w/v) OGNG / CHS and initially purified using TALON metal affinity resin. The C-terminal tags were cleaved overnight with TEV protease, and the cleaved protein was further purified by SEC. The SEC elution profile displays a single monodisperse peak, and SDS-PAGE analysis reveals a pure final sample (Fig. 4.4A,B).

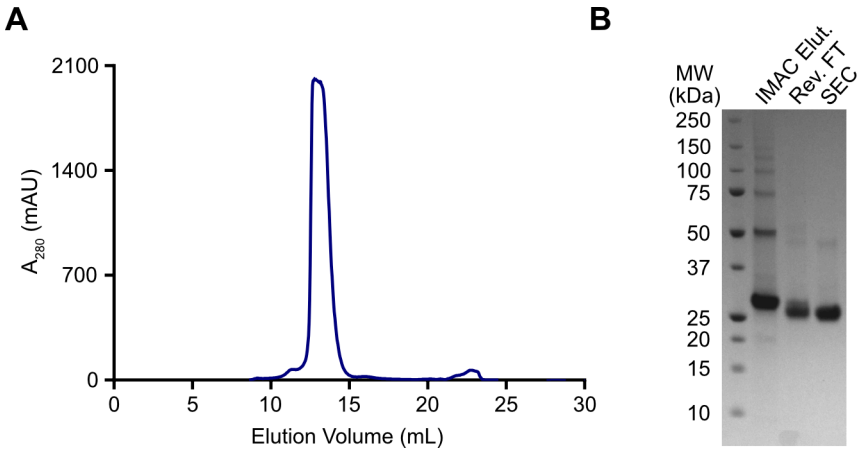


Figure 4.4. ELOVL7 purification. A. Elution profile of ELOVL7 from a Superdex 200 Increase 10/300 SEC column. **B.** SDS-PAGE analysis of samples taken during various steps of ELOVL7 purification. *IMAC elution* corresponds to the eluate from the TALON resin. Reverse IMAC flow-through (*Rev. FT*) corresponds to the ELOVL7 protein cleaved overnight with TEV protease, which no longer contains the C-terminal tags and therefore did not bind the TALON resin during the reverse IMAC step. *SEC* corresponds to the pooled peak fractions from (A).

4.2.2 Purified ELOVL7 is a dimer in the crystal and in solution

As described above, the crystal structure of ELOVL7 reveals an inverted dimer. Given that ELOVL7 substrates access the central tunnel from the cytoplasmic side, this inverted dimer

arrangement is likely a crystallisation or purification artefact. To determine whether ELOVL7 dimers are formed prior to crystallisation, I aimed to determine the oligomeric state of purified ELOVL7 in solution using size exclusion chromatography with multiangle light scattering (SEC-MALS).

In this method, molecular weight determination relies on the specific refractive index increment as a function of concentration (dn/dc) [265]. For soluble proteins, this is straightforward, as the dn/dc of proteins can be approximated as 0.185 mL g^{-1}. However, membrane proteins solubilised in detergents are surrounded by a detergent belt that adds significant molecular weight. Therefore, the determination of the molecular weight of membrane proteins within the protein:detergent complex (PDC) by SEC-MALS requires the determination of the molecular weight fraction of this complex that corresponds to the protein. To achieve this, one must initially determine the dn/dc of the detergent [266]. Therefore, I calculated the dn/dc of OGNG / CHS (10:1 (w/w), as used for the purification of ELOVL7). This was done by generating a standard curve of refractive index peak area versus mass (grams) of detergent injected [267] (Fig. 4.5A). The detergent dn/dc corresponds to the slope of this curve, and was calculated to be 0.127 mL g^{-1}. Using this setup, I calculated the molecular weight of the PDC to be 149,676 Da, where the protein molecular weight was calculated to be 81,873 Da (Fig. 4.5B). Given the molecular weight of the expressed ELOVL7 construct (theoretical monomer molecular weight: 34222.38 Da), this reveals that OGNG-solubilised ELOVL7 exists as a dimer in solution.

ELOVL7 crystals were obtained using protein treated with iodoacetamide (IAM) to block reduced cysteines. Therefore, I repeated the SEC-MALS experiment using IAM-treated protein and obtained similar results. As a result, it appears likely that the dimeric arrangement observed in the crystal structure arises when the protein is extracted from the membrane with detergent. However, this experiment only determined that this detergent-solubilised

species is dimeric and does not provide information on the relative orientation of the protomers in solution.

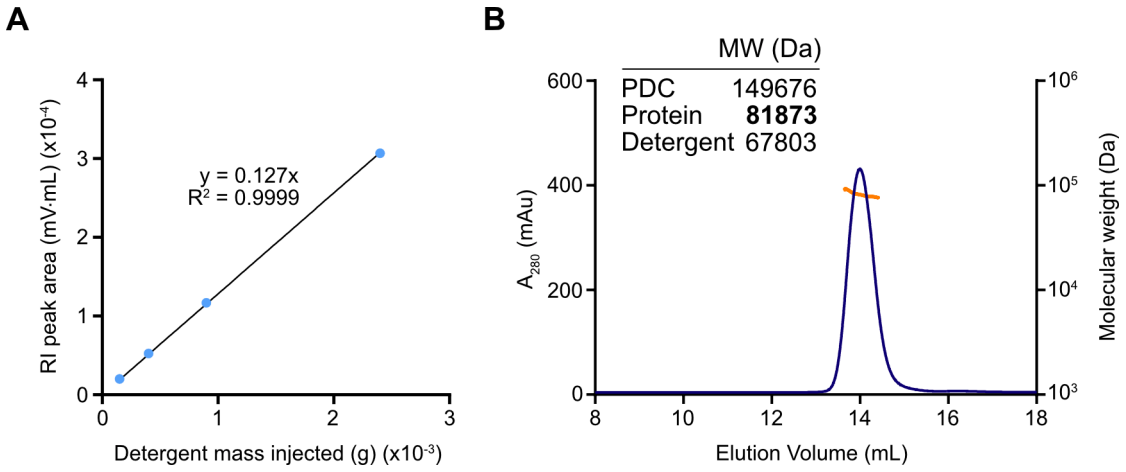


Figure 4.5. SEC-MALS analysis of purified ELOVL7. A. Standard curve of Refractive Index (RI) peak area versus injected OGNG/CHS (10:1 w/w) mass (100 µL injections). The dn/dc of OGNG/CHS was obtained from the slope of the standard curve (0.127 mL g^{-1}). Plotted RI peak areas correspond to $RI_{area,\ observed}$ / K_{RI} (where K_{RI} is the RI detector response factor). **B.** SEC-MALS analysis of OGNG-solubilised ELOVL7. The blue trace corresponds to the elution profile (absorption at 280 nm) of ELOVL7 from a Superdex 200 Increase 10/300 SEC column; the orange line corresponds to the molecular weight (MW) of ELOVL7 within the protein:detergent complex (PDC). The calculated molecular weights of the protein and detergent fractions of the PDC are listed. Data were analysed using OMNISEC v.5.1 software. (B) is adapted from Nie et al. (2021) [Ref. 213].

4.2.3 The apo form of purified ELOVL7 is unstable and lost during purification

As described in Section 4.1.2, the final purified ELOVL7 sample is almost completely covalently modified with a 3-keto eicosanoyl-CoA product analogue. To investigate this further, I analysed protein samples obtained during the course of purification and found that, after the initial IMAC step, both unmodified and modified proteins were present, although there is more of the former (Fig. 4.6). However, at the end of the 2-day purification (after tag cleavage and SEC), only the modified protein is present, independently of IAM treatment (Fig. 4.6).

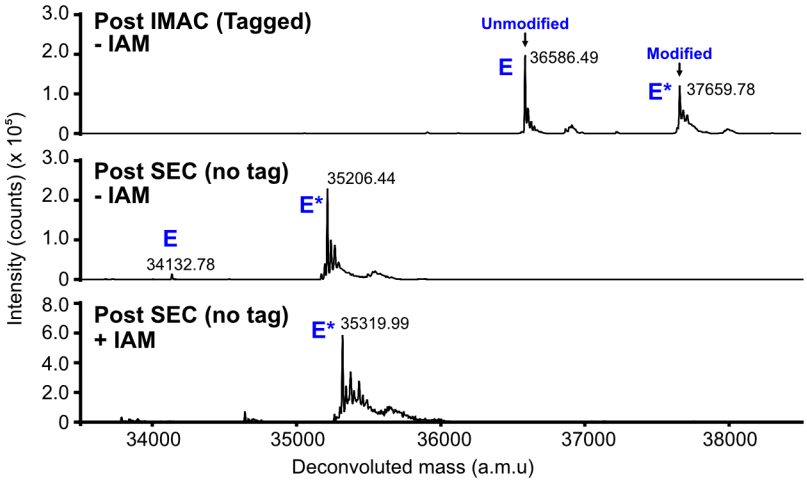


Figure 4.6. Denaturing intact LC-ESI-MS analysis of ELOVL7 samples during purification. Deconvoluted intact mass spectra are shown for protein obtained immediately after IMAC and at the end of the purification (after tag cleavage and SEC). The expected molecular weight of the unmodified enzyme is 36,675.86 Da prior to C-terminal tag cleavage and 34,222.38 Da after tag cleavage. The observed unmodified enzyme peaks (*E*) correspond to unmodified ELOVL7 with loss of the initiator methionine and acetylation of the new N-terminus (theoretical MW: 36586.67 Da tagged; 34,133.22 Da untagged). The deconvoluted mass peaks corresponding to modified enzyme (*E**) arise from the covalent attachment of a 3-keto eicosanoyl-CoA product analogue (theoretical: +1073.66 Da). A further mass addition of 113.55 Da in the +IAM sample indicates that IAM treatment resulted in the modification of two cysteines (theoretical: +57.07 Da per site modified). Adapted from Nie et al. (2021) [Ref. 213].

This observation suggests that the unmodified protein, originally isolated from the pool of recombinant ELOVL7, is lost during the course of purification due to precipitation / aggregation. In this case, all the covalent modification would take place *in cellula*, and the attached product analogue would stabilise the protein, enabling its purification and crystallisation.

To demonstrate that the unmodified species of purified ELOVL7 is inherently more unstable, I carried out a 'stress test' of OGNG-solubilised material obtained after the IMAC step. As can be seen in Fig. 4.7, this initial sample contained both the modified and unmodified ELOVL7 species. The sample was subjected to incubation at 37 ºC for 1h, and then the soluble and insoluble fractions were separated by centrifugation and analysed by denaturing

intact protein mass spectrometry. The results of this experiment clearly indicated that the modified / adducted material remained in solution while the unmodified material precipitated and was found exclusively in the centrifuged pellet (Fig. 4.7).

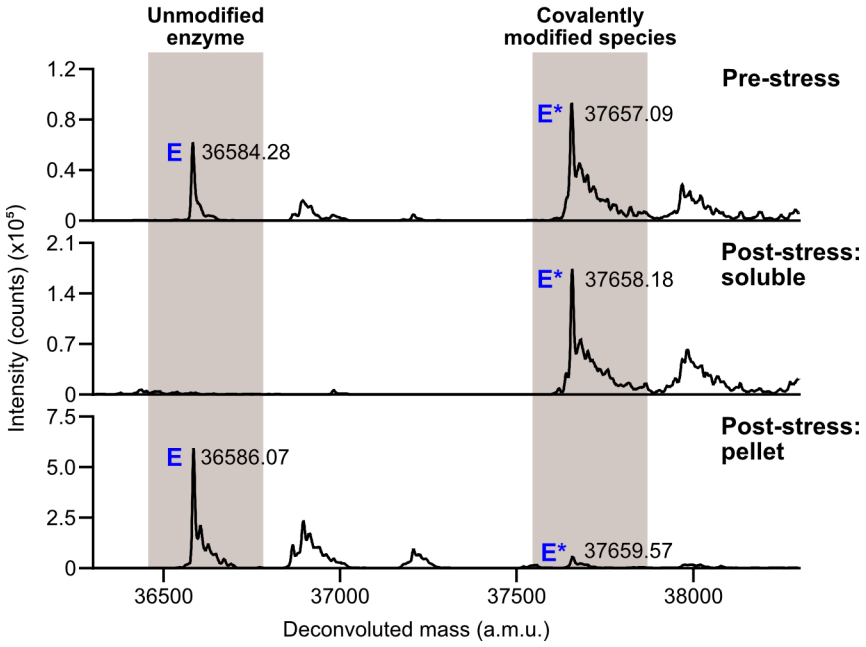


Figure 4.7 'Stress test' of purified ELOVL7. Protein obtained after the IMAC step was subjected to a 'stress test' (1h incubation at 37 °C), and precipitated material was collected by centrifugation. LC-ESI-MS analysis of the soluble and precipitated (pellet) samples obtained after this 'stress test' reveals that the unmodified enzyme (*E*) readily precipitated and was found in the pellet, whereas the covalently modified enzyme (*E**) remained in the soluble fraction.

4.2.4 Substrate-binding arrangement suggests a ping-pong type mechanism

Acyltransferases generally catalyse acyl transfer via either a ternary complex or ping-pong type mechanism. In a ternary complex mechanism, both substrates bind at the same time, and the enzyme catalyses the direct transfer of the acyl chain from one substrate to another. In contrast, in a ping-pong type mechanism, the substrates bind to and react with the enzyme

sequentially, such that the first substrate-turned-product must leave before the second substrate can bind. As a result, this type of mechanism involves the formation of a covalent acyl-enzyme intermediate.

The crystal structure of ELOVL7 revealed that substrates would bind in a very narrow tunnel (Fig. 4.8). As a result, there is not sufficient space for both substrates to bind at the same time, excluding a ternary complex mechanism for this enzyme.

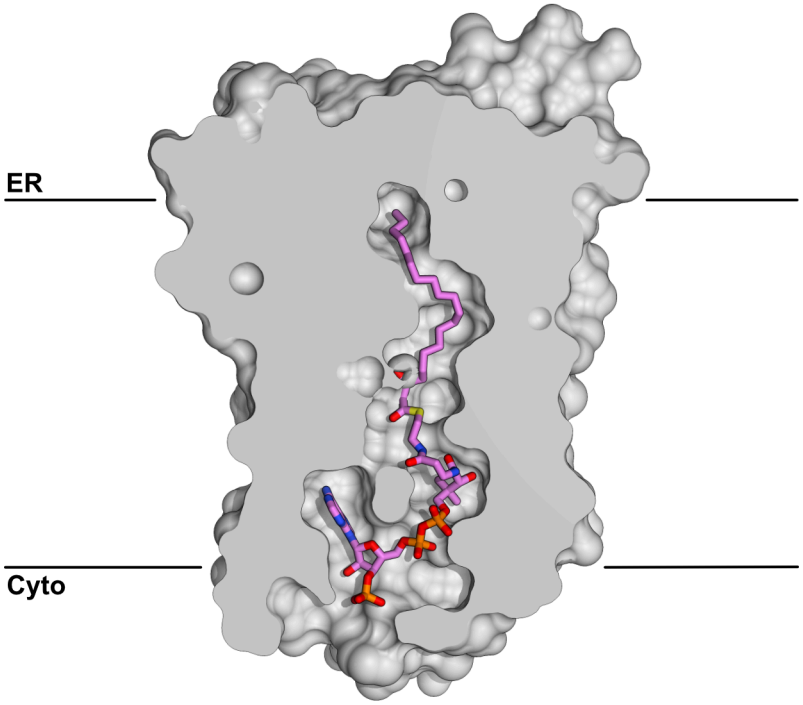


Figure 4.8. Cutaway molecular surface view of the narrow substrate-binding tunnel of ELOVL7. The co-purified 3-keto acyl-CoA product analogue is shown in stick representation. Adapted from Nie et al. (2021) [Ref. 213].

However, the architecture of the substrate-binding tunnel could enable a ping-pong type mechanism, where the first product, CoA, must leave before the second substrate, malonyl-CoA, can bind. In this scenario (Fig. 4.9), the first substrate (acyl-CoA) would bind and react with the enzyme, leading to thioester cleavage and release of the first product (CoA). This

first transacylation step would result in the formation of an acyl-enzyme intermediate, where the acyl chain derived from the acyl-CoA substrate would remain buried deep in the ER end of the narrow central tunnel. In the second condensation step, the second substrate, malonyl-CoA, would enter the central tunnel, and its CoA moiety would occupy the now empty CoA-binding site. This would be followed by decarboxylation and carbon-carbon bond formation, releasing the product, 3-keto acyl-CoA, from the enzyme.

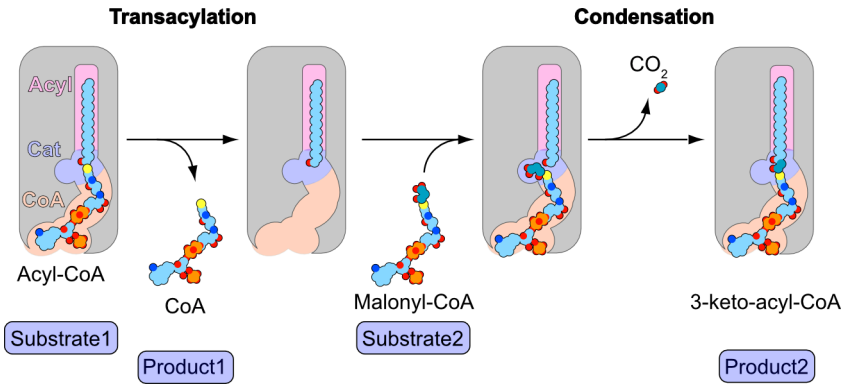


Figure 4.9 Schematic representation of proposed ping-pong mechanism. Transacylation reaction with the first substrate (acyl-CoA) would lead to the formation of an acyl-enzyme intermediate and release of the first product (CoA). Then, in the condensation step, the second substrate (malonyl-CoA) would bind to the enzyme, undergo decarboxylation and react with the acyl-enzyme intermediate to form the 3-keto acyl-CoA product. Adapted from Nie et al. (2021) [Ref. 213].

4.2.5 Identification of a covalent acyl-enzyme intermediate

The ELOVL7 structure suggests a ping-pong type reaction mechanism, which would imply the formation of a covalent intermediate at the end of the first step of the reaction. Therefore, I investigated whether reaction of ELOVL7 with substrate acyl-CoAs results in the formation of a covalent acyl-enzyme intermediate *in vitro*. This was carried out using the denaturing intact protein LC-ESI-MS method established in Section 4.2.3. However, given that at the end of the purification >90 % of the purified enzyme was already covalently modified

with the 3-keto acyl-CoA, I used protein obtained after the IMAC step in these mass spectrometry experiments.

Full experimental details are provided in Sections 2.1.6 and 2.3.2.2. Briefly, I incubated the purified enzyme post-IMAC with either C18:0-CoA or C18:1(n3)-CoA, previously demonstrated to act as ELOVL7 substrates [197], and assessed covalent intermediate formation by inspecting mass additions in the deconvoluted mass spectra. If the enzyme reacts with these acyl-CoA substrates in the proposed manner (thioester cleavage, loss of CoA and covalent attachment of the substrate acyl chain), then the expected mass additions upon intermediate formation would correspond to the mass of the substrate acyl chain (Fig. 4.10A). Indeed, incubation with C18:0-CoA leads to the formation of a +266.51 Da adduct (theoretical: 266.47 Da) (Fig. 4.10B,C). Likewise, incubation with the unsaturated substrate C18:3(n3)-CoA resulted in a mass addition of +260.74 Da (theoretical: +260.42 Da) (Fig. 4.10D). Overall, these results provide the first evidence for a covalent acyl-enzyme intermediate in the catalytic mechanism of ELO elongases.

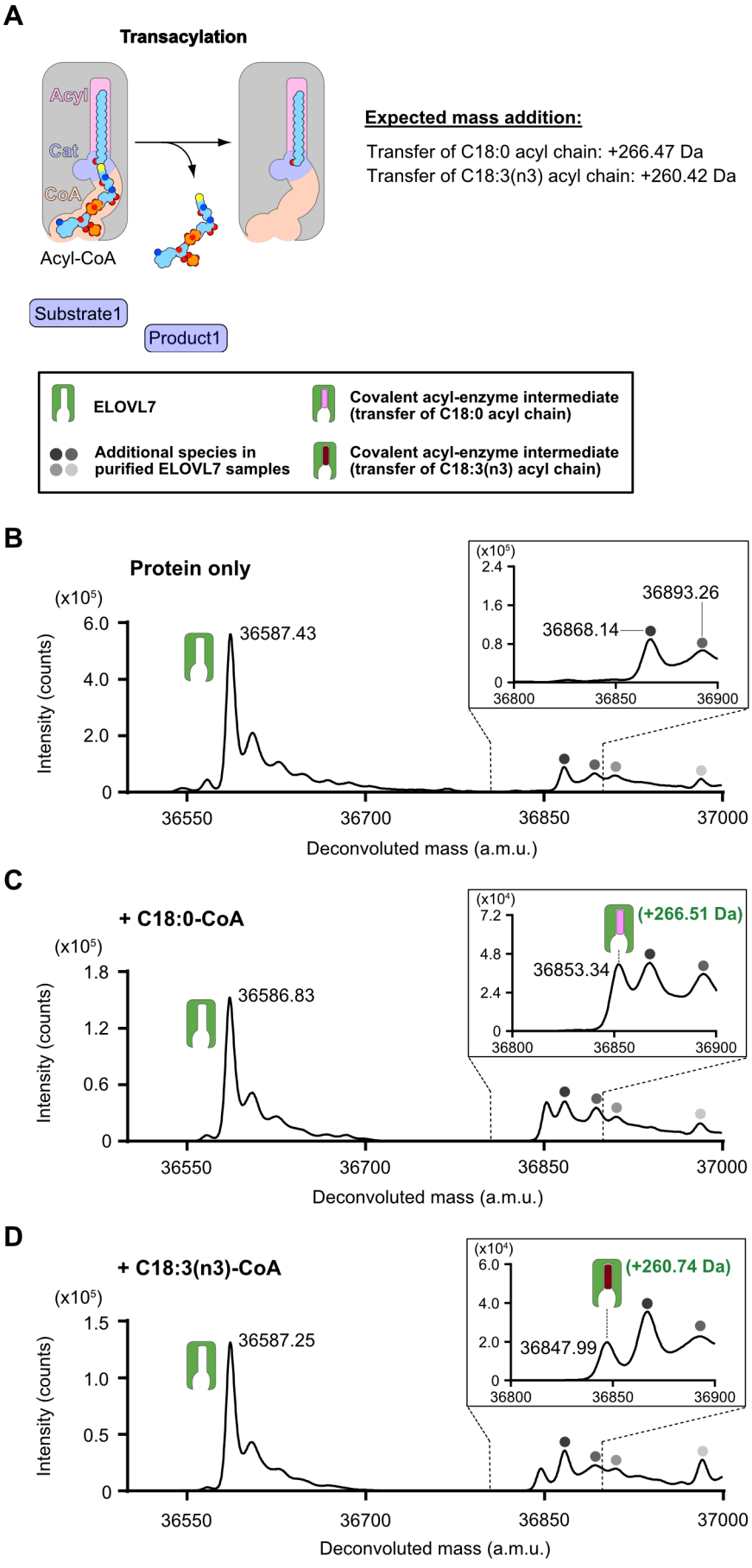

A
Transacylation
Acyl
Cat
CoA
Acyl-CoA
Substrate1
Product1
Expected mass addition:
Transfer of C18:0 acyl chain: +266.47 Da
Transfer of C18:3(n3) acyl chain: +260.42 Da
ELOVL7
Additional species in purified ELOVL7 samples
Covalent acyl-enzyme intermediate (transfer of C18:0 acyl chain)
Covalent acyl-enzyme intermediate (transfer of C18:3(n3) acyl chain)
B
Protein only
(x10⁵)
Intensity (counts)
36587.43
36868.14
36893.26
Deconvoluted mass (a.m.u.)
C
+ C18:0-CoA
(x10⁵)
(x10⁴)
36586.83
36853.34
(+266.51 Da)
Deconvoluted mass (a.m.u.)
D
+ C18:3(n3)-CoA
(x10⁵)
(x10⁴)
36587.25
36847.99
(+260.74 Da)
Deconvoluted mass (a.m.u.)

Figure 4.10. Identification of a covalent acyl-enzyme intermediate of ELOVL7. A. Schematic representation of the proposed transacylation step in the reaction mechanism of ELOVL7. Expected mass additions are shown for the transfer of C18:0 and C18:3(n3) acyl chains to the protein. **B-D.** Deconvoluted mass spectra obtained for purified (post-IMAC) and tagged ELOVL7 protein incubated **B.** in the absence of substrates, **C.** with 100 µM C18:0-CoA and **D.** with 100 µM C18:3(n3)-CoA. The inset panels are zoomed-in views of the mass range where the covalent intermediates are expected. Additional species in purified ELOVL7 samples (dots in shades of grey) refer to background species present in all traces. Adapted from Nie et al. (2021) [Ref. 213].

4.2.6 Metals are not required for the transacylation step

As described in Section 4.1.3, no metal ions could be observed in the structure, and the geometry of the conserved HxxHH motif of ELOVL7 is not suitable for metal binding. Instead, the second histidine of this motif, His150, is covalently linked to the co-purified 3-keto acyl-CoA. Overall, structural evidence suggested that the HxxHH motif of ELOVL7 is not involved in metal binding and that metal-assisted catalysis is unlikely. To confirm this, I assessed the formation of the covalent intermediate in the presence of the metal chelators EDTA or EGTA. As expected, the addition of these chelating agents did not interfere with covalent intermediate formation (Fig. 4.11A-D). Together, the structural and biochemical evidence conclusively demonstrate that the essential HxxHH motif of ELO elongases is not involved in metal ion binding, previously suggested by others [195].

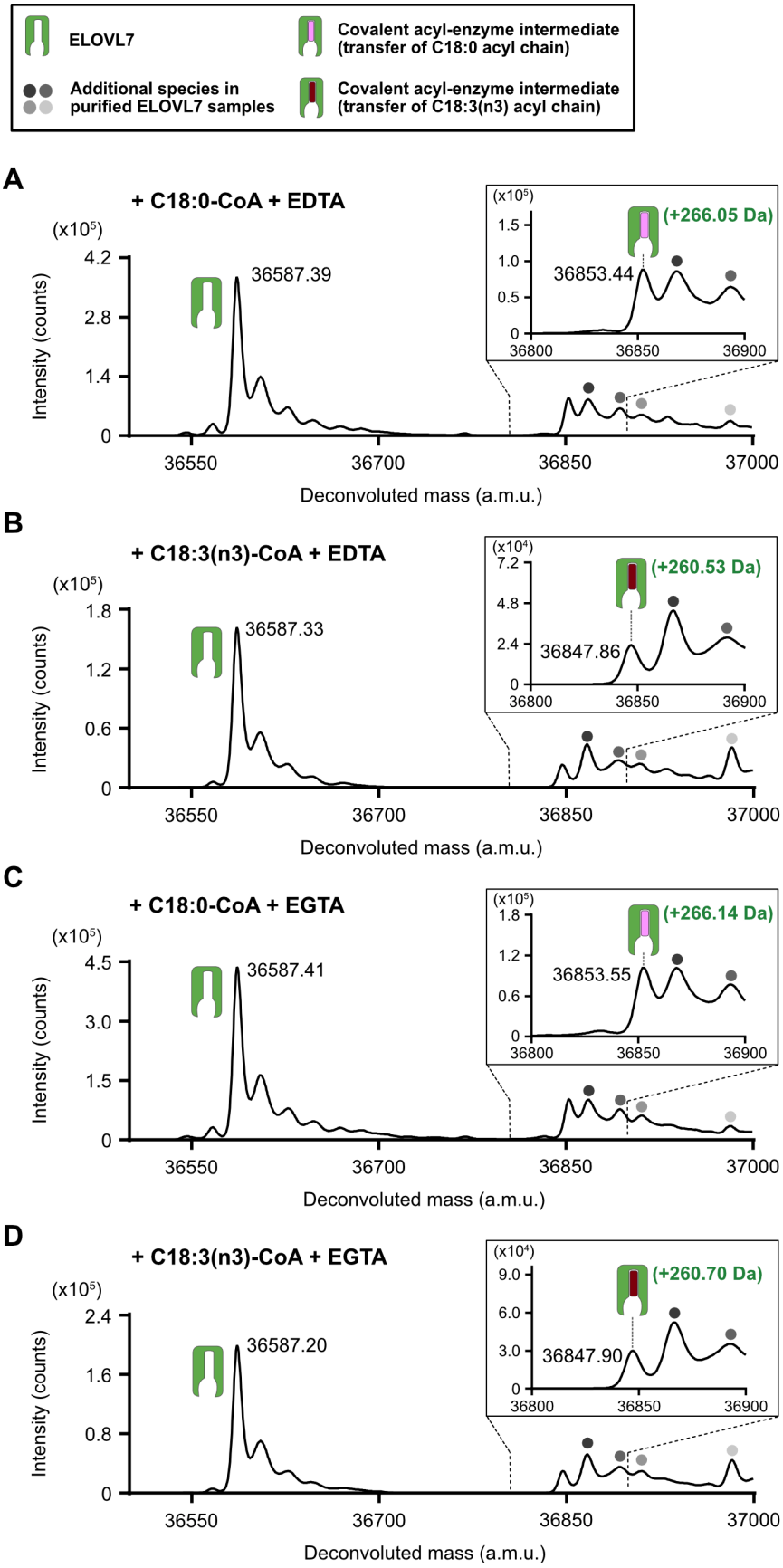

ELOVL7
Covalent acyl-enzyme intermediate (transfer of C18:0 acyl chain)
Additional species in purified ELOVL7 samples
Covalent acyl-enzyme intermediate (transfer of C18:3(n3) acyl chain)
A
+ C18:0-CoA + EDTA
(x10⁵)
Intensity (counts)
36587.39
(+266.05 Da)
36853.44
Deconvoluted mass (a.m.u.)
B
+ C18:3(n3)-CoA + EDTA
(x10⁵)
(x10⁴)
36587.33
(+260.53 Da)
36847.86
Deconvoluted mass (a.m.u.)
C
+ C18:0-CoA + EGTA
(x10⁵)
36587.41
(+266.14 Da)
36853.55
Deconvoluted mass (a.m.u.)
D
+ C18:3(n3)-CoA + EGTA
(x10⁵)
(x10⁴)
36587.20
(+260.70 Da)
36847.90
Deconvoluted mass (a.m.u.)

Figure 4.11. Metal-chelating agents do not affect covalent intermediate formation. Deconvoluted mass spectra obtained for purified (post-IMAC) and tagged ELOVL7 protein incubated in the presence of **A.** 100 µM C18:0-CoA and 1 mM EDTA, **B.** 100 µM C18:3(n3)-CoA and 1 mM EDTA, **C.** 100 µM C18:0-CoA and 1mM EGTA, **D.** 100 µM C18:3(n3)-CoA and 1mM EGTA. The inset panels are zoomed-in views of the mass range where the covalent intermediates are expected. Adapted from Nie et al. (2021) [Ref. 213].

4.2.7 Catalysis proceeds via an unusual acyl-imidazole intermediate

The catalytic mechanisms of fatty acid synthases / elongases, lipid synthases and other acyl-transferases often involve the formation of covalent acyl-enzyme intermediates on cysteines or serines [186, 268, 269]. For example, plant FAE1-type 3-keto acyl-CoA synthases and polyketide synthases such as chalcone synthases contain a Cys-His-Asn catalytic triad, which acts to activate the catalytic cysteine residue of this triad as a nucleophile, enabling the formation of a covalent acyl-enzyme intermediate with a thioester linkage [186, 270]. On the other hand, as introduced in Section 1.2.8, ELOs (including mammalian ELOVLs) lack known catalytic triad motifs in their amino acid sequences and instead possess a histidine-rich HxxHH motif [194]. In addition, the structure of ELOVL7 revealed that there are no cysteines or serines in the active site (Fig. 4.3B). Instead, two histidines (His150 and His181) were covalently linked to the co-purified 3-keto acyl-CoA (see Sections 4.1.2 and 4.1.3). Importantly, His150 belongs to the conserved HxxHH motif.

In addition, Dr Ashley Pike modelled a substrate acyl-CoA bound to ELOVL7, based on the binding mode observed for the co-purified product analogue (Fig. 4.12). I further inspected this model to understand the role played by His150, as this conserved residue would lie adjacent to the thioester carbonyl of the acyl-CoA, and would therefore be ideally placed to act as the nucleophile in the first (transacylation) step of the reaction. Importantly, in the ELOVL7 structure, His150 is involved in a hydrogen-bonding network with His151, Asp130 and Tyr185 (Fig. 4.3B; Fig. 4.12). I reason that this hydrogen-bonding network

could enhance its nucleophilicity by ensuring Nε2 is unprotonated, or could otherwise contribute to positioning His150 for catalysis. Indeed, the observation that its Nε2 is covalently linked to the C5 of the 3-keto acyl-CoA in the structure suggests His150 can act as a nucleophile (for a potential mechanism for the formation of this displaced bond in the structure, see Discussion Section 4.3.2).

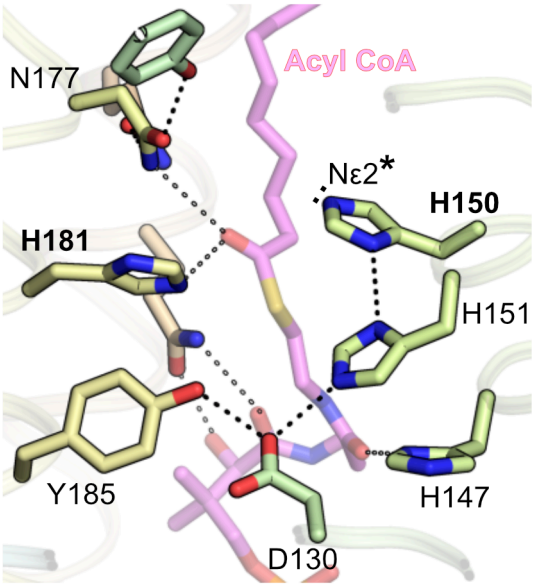


Figure 4.12. Structural modelling suggests His150 could act as the nucleophile during the transacylation step. Modelling of the co-purified 3-keto acyl-CoA product analogue in the ELOVL7 structure suggests H150 could lie adjacent to the thioester carbonyl of a substrate acyl-CoA. Hydrogen bonds are shown as black dashes. Adapted from Nie et al. (2021) [Ref. 213].

Overall, the structural and modelling data suggested a role for His150 as the primary nucleophile during catalysis. To evaluate this hypothesis, I assessed whether the covalent acyl-enzyme intermediate could be formed when the two key histidine residues forming the covalent attachments in the structure were each substituted by alanines. Therefore, the His150Ala and His181Ala mutant proteins post-IMAC were obtained as described above for wild-type protein, incubated with C18:3(n3)-CoA, and the presence of the covalent in-

termediate was evaluated by denaturing intact protein mass spectrometry. Strikingly, no covalent intermediate could be detected when the His150Ala mutant protein was incubated with C18:3(n3)-CoA (Fig. 4.13A,B), but the expected adduct species was still observed for the His181Ala mutant (theoretical: 260.42 Da; observed: +261.97 Da) (Fig. 4.13C,D). I was unable to carry out this experiment using C18:0-CoA due to the presence of background peaks in the deconvoluted mass spectra (theoretical mass after addition of C18:0 acyl chain: 36787.08 Da) (Fig. 4.13A-D).

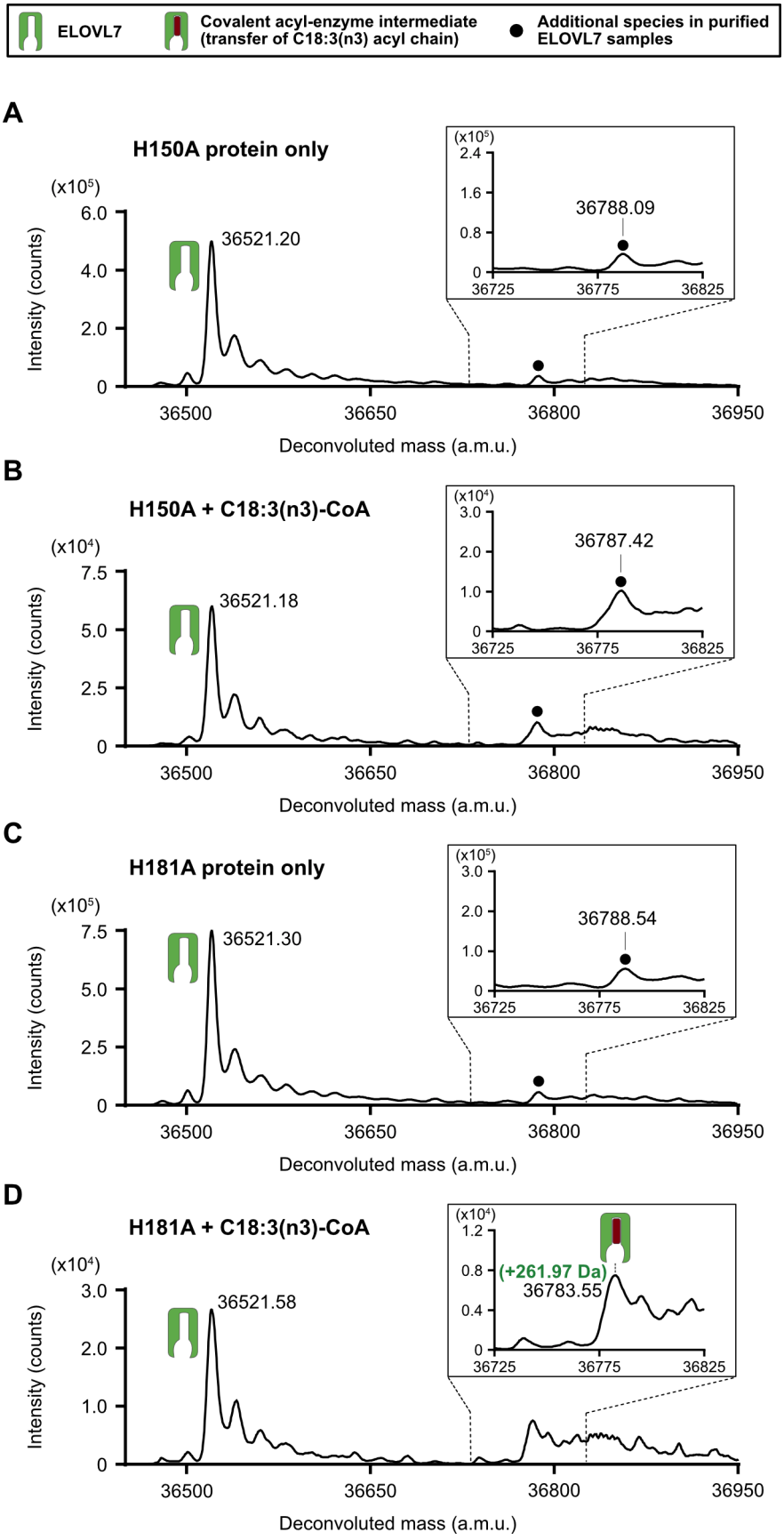

ELOVL7
Covalent acyl-enzyme intermediate (transfer of C18:3(n3) acyl chain)
Additional species in purified ELOVL7 samples
A
H150A protein only
(x10⁵)
6.0
4.0
2.0
0
36521.20
Intensity (counts)
36500
36650
36800
36950
Deconvoluted mass (a.m.u.)
(x10⁵)
2.4
1.6
0.8
0
36788.09
36725
36775
36825
B
H150A + C18:3(n3)-CoA
(x10⁴)
7.5
5.0
2.5
0
36521.18
Intensity (counts)
36500
36650
36800
36950
Deconvoluted mass (a.m.u.)
(x10⁴)
3.0
2.0
1.0
0
36787.42
36725
36775
36825
C
H181A protein only
(x10⁵)
7.5
5.0
2.5
0
36521.30
Intensity (counts)
36500
36650
36800
36950
Deconvoluted mass (a.m.u.)
(x10⁵)
3.0
2.0
1.0
0
36788.54
36725
36775
36825
D
H181A + C18:3(n3)-CoA
(x10⁴)
3.0
2.0
1.0
0
36521.58
Intensity (counts)
36500
36650
36800
36950
Deconvoluted mass (a.m.u.)
(x10⁴)
1.2
0.8
0.4
0
(+261.97 Da)
36783.55
36725
36775
36825

Figure 4.13. The covalent linkage of the acyl-enzyme intermediate is formed at His150. The active site histidines (His150 / His181) were each substituted by alanines, and covalent intermediate formation was assessed using LC-ESI-MS on protein samples post-IMAC. Deconvoluted mass spectra are shown for **A.** His150Ala mutant incubated in the absence of substrates. **B** His150Ala mutant incubated with 100 µM C18:3(n3)-CoA. Using this mutant, no covalent intermediate could be detected. **C.** His181Ala mutant incubated in the absence of substrates. **D.** His181Ala mutant incubated with 100 µM C18:3(n3)-CoA. The His181Ala mutation did not affect covalent intermediate formation. The inset panels are zoomed-in views of the mass range where the covalent intermediates are expected. Adapted from Nie et al. (2021) [Ref. 213].

Furthermore, I analysed the denaturing intact mass spectra of these mutant proteins at the end of the purification (post-SEC) to understand if they co-purified with additional covalently linked species (Fig. 4.14). On the one hand, for the His150Ala mutant, no covalent adducts could be observed. On the other hand, analysis of the His181Ala mutant showed that this mutant protein contains adducts in the 1030-1060 Da range instead of the +1074 Da adduct expected for the 3-keto acyl-CoA product analogue found in the wild-type protein. However, the nature of these additions to the His181Ala mutant protein could not be identified.

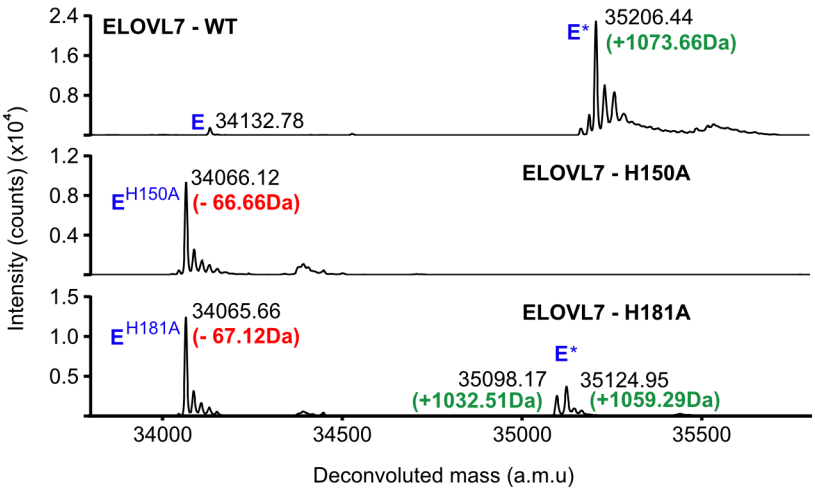


Figure 4.14. Denaturing intact LC-ESI-MS analysis of active site histidine mutants. Deconvoluted mass spectra are shown for untagged ELOVL7 wild-type, His150Ala and His181Ala proteins at the end of the purification (post-SEC). The deconvoluted masses of the unmodified (*E*) histidine mutants correspond to the expected molecular weight (theoretical mass shift for a His to Ala mutation: -66.06 Da). The wild-type spectrum is reproduced from Fig. 4.2C and shows the presence of the

modified species (*E**) with a covalently bound 3-keto acyl-CoA. In contrast, the introduction of the His150Ala mutation led to protein without any covalent modifications. For the His181Ala mutant, different covalent modifications could be observed in the 1030-1060 Da range. Adapted from Nie et al. (2021) [Ref. 213].

Overall, the findings that the purified His150Ala mutant protein does not contain any covalent adducts and is not capable of covalent modification with a substrate acyl-CoA, together with structural modelling, point to His150 acting as the nucleophile during the first step of the reaction.

4.2.8 Condensation of the acyl-enzyme intermediate with malonyl-CoA

In the proposed ping-pong type mechanism, the second step of the reaction would involve a Claisen-like condensation, where malonyl-CoA (the second incoming substrate) would be combined with the acyl-enzyme intermediate bound in the active site tunnel. This would initially involve malonyl-CoA binding, decarboxylation (to form a nucleophilic enolate carbanion), and carbon-carbon bond formation (either in a concerted or sequential fashion) to form the 3-keto acyl-CoA product. The reaction of the malonyl-CoA-derived carbon nucleophile with the acyl-enzyme intermediate would involve a nucleophilic attack at the carbonyl carbon, thus kicking out the covalently linked His150 and regenerating the unmodified enzyme.

To confirm this proposed ping-pong type reaction mechanism and to evaluate whether the observed covalent addition of the substrate acyl chain to His150 corresponds to the real acyl-enzyme intermediate formed during catalysis, I aimed to investigate whether this intermediate could subsequently react with the second substrate, malonyl-CoA (Fig. 4.15).

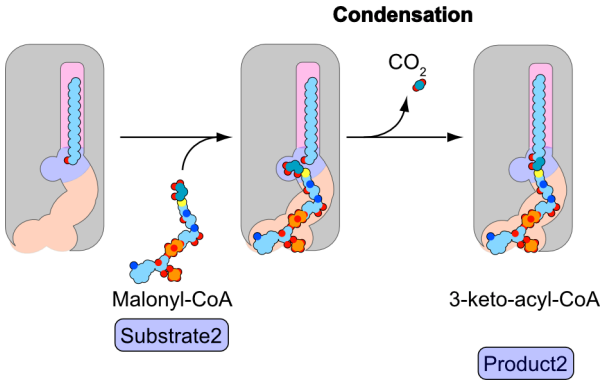


Figure 4.15. Proposed condensation step of the ELOVL7-catalysed reaction. Here, condensation of the second substrate (malonyl-CoA) with the acyl-enzyme intermediate would yield the 3-keto acyl-CoA product and regenerate the unmodified enzyme. Adapted from Nie et al. (2021) [Ref. 213].

This experiment required that the detergent-solubilised enzyme could remain active and in solution at the end of the initial transacylation reaction with acyl-CoA so that it could react further with malonyl-CoA in the second condensation step. As observed in the 'stress test' described in Section 4.2.3, the OGNG/CHS-solubilised unmodified apo enzyme precipitates during prolonged (1h) incubation at 37 ºC. Therefore, I investigated the relative stability of the unmodified enzyme in different detergents and observed that LMNG-solubilised unmodified enzyme remained in solution after 1h incubation at 37 ºC (Fig. 4.16). As a result, LMNG-solubilised ELOVL7 was used to monitor the reaction of the covalent acyl-enzyme intermediate with malonyl-CoA.

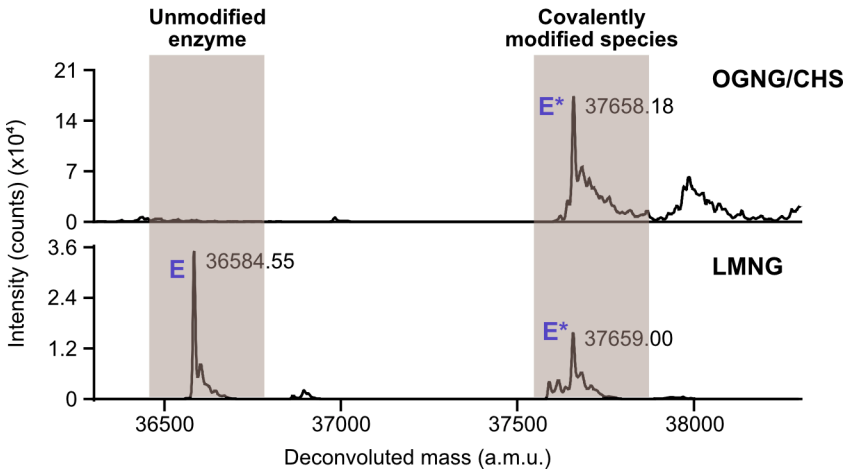


Figure 4.16. Comparison of unmodified enzyme stability in different detergents. Purified ELOVL7 (after the IMAC step), solubilised in either OGNG/CHS or LMNG was subjected to a stress test (as in Fig. 4.7). The soluble material after stress was analysed by denaturing intact protein mass spectrometry, revealing that LMNG-solubilised unmodified protein remained in solution, whereas all unmodified material was lost when the protein was solubilised in OGNG/CHS. The OGNG/CHS spectrum is reproduced from Fig. 4.7.

In this two-step sequential reaction, wild-type ELOVL7 was initially incubated with the first substrate (C18:0-CoA), leading to the formation of the covalent acyl-enzyme intermediate at the end of the transacylation step (Fig. 4.17A,B). Then, the protein was further incubated with the second substrate (malonyl-CoA) for the condensation step. As observed in Fig. 4.17C, analysis of the final sample after the reaction reveals that incubation with malonyl-CoA resulted in complete loss of the acyl-enzyme intermediate species, which is consistent with the two-step reaction having gone to completion. Analysis of control samples incubated in the absence of malonyl-CoA shows that the loss of the covalent intermediate deconvoluted mass peak occurs as a direct result of the reaction with malonyl-CoA (Fig. 4.17D).

These results confirm the sequential nature of the reaction, supporting the proposed ping-pong type mechanism. However, the exact mechanism of malonyl-CoA decarboxylation to form the nucleophilic enolate that attacks the acyl-enzyme intermediate remains elusive (see Discussion Section 4.3.1).

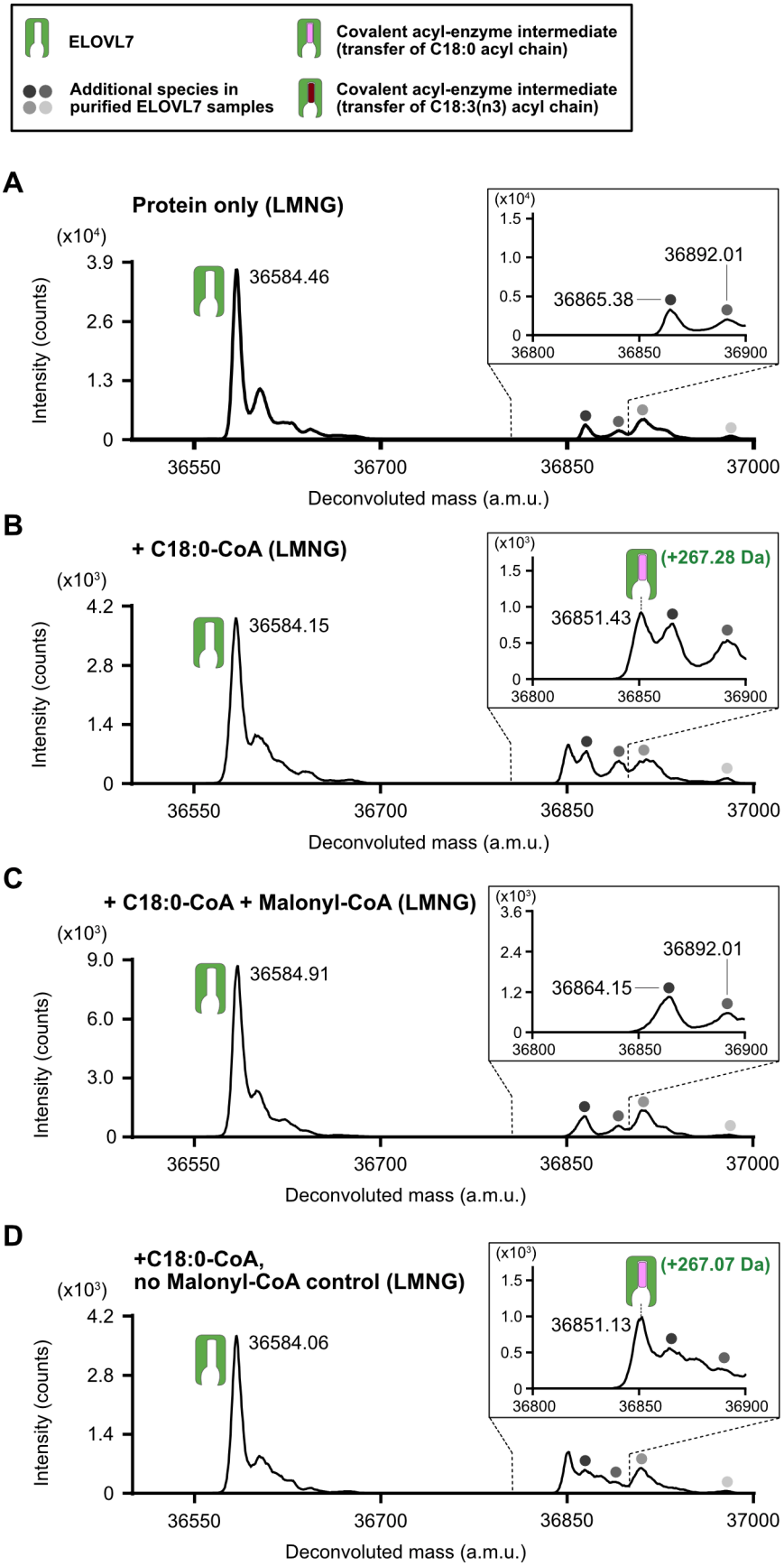


Figure 4.17. Sequential reaction of ELOVL7 with acyl-CoA and malonyl-CoA. Purified (post-IMAC) and tagged ELOVL7 protein solubilised in LMNG was incubated **A.** in the absence of substrates or

B. in the presence of 100 µM C18:0-CoA. Precipitated material was removed by centrifugation, and the soluble material was analysed by denaturing intact protein mass spectrometry (A-B). Then, the covalent intermediate obtained in panel B was further incubated **C.** with 200 µM malonyl-CoA or **D.** without malonyl-CoA. The deconvoluted mass spectra obtained for the final samples reveal that incubation with malonyl-CoA specifically resulted in the complete loss of the peak corresponding to the covalent intermediate. Adapted from Nie et al. (2021) [Ref. 213].

4.3 Discussion

4.3.1 Proposed reaction mechanism of human ELOVL7

Overall, together with the structural data obtained by Dr Laiyin Nie and Dr Ashley Pike, the biochemical characterisation carried out in my thesis work shed light on a ping-pong reaction mechanism for ELOVL7 [213] (Fig. 4.18). In the first step of the reaction (transacylation), the acyl-CoA substrate initially binds with the acyl chain buried deep in the narrow central tunnel and the CoA moiety on the cytoplasmic entrance. Then, His150 nucleophilically attacks the thioester carbonyl of the acyl-CoA substrate, resulting in thioester cleavage, formation of an unusual acyl-imidazole intermediate and subsequent loss of CoA. His150 is likely activated as a nucleophile by a hydrogen-bonding network involving His151, Asp130 and Tyr185, which likely ensures its Nε2 nitrogen is unprotonated. Mutational studies of the EloA elongase of the slime mould *Dictyostelium discoideum* showed that the corresponding residues, including the residue corresponding to Asp130, are essential for activity [194], further supporting the proposed role of this hydrogen-bonding network. However, mutation of the tyrosine to phenylalanine did not affect activity [194], suggesting that only the His-His-Asp triad is required for catalysis.

While histidine nucleophiles are unusual, several examples have been described in the literature. Histidine nucleophiles are most commonly used by enzymes whose mechanism involves the formation of phosphoenzyme intermediates, such as nucleotidyltransferases [271, 272], histidine-dependent acid phosphatases [273] and the phospholipase D superfamily

[274, 275]. However, histidine nucleophiles are also present in some acyltransferases, where they attack the thioester of the substrate to form an acyl-imidazole intermediate. Examples include HlyC, an acyltransferase that activates *E. coli* hemolysin [268, 276], and complement components C3 and C4B, which catalyse internal acyl transfer via a histidine to enable covalent attachment to the surface of pathogens [277, 278]. In addition, artificial enzyme systems employing nucleophilic histidines for acyl transfer have also been described [279-281].

In the second reaction step (Claisen-like condensation), malonyl-CoA would initially bind in the tunnel, with its CoA moiety likely occupying the same CoA-binding site observed in the structure. The malonyl moiety would then undergo decarboxylation to form a nucleophilic enolate carbanion which would attack the acyl-enzyme intermediate, forming the 3-keto acyl-CoA product and regenerating the unmodified enzyme.

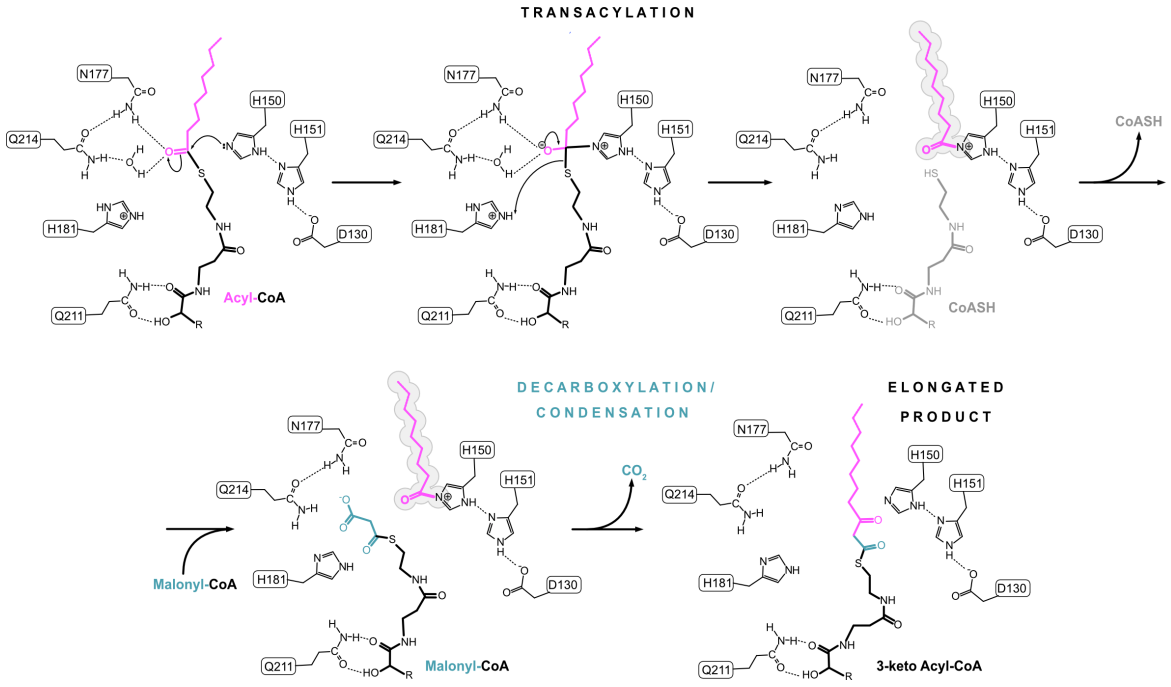


Figure 4.18. Proposed ping-pong mechanism of ELOVL7. In the first step (transacylation), the acyl-CoA binds and the acyl chain is transferred to the nucleophilic imidazole of His150. In the second step, malonyl-CoA is condensed with the acyl-enzyme intermediate to yield the 3-keto acyl-CoA

product. In this diagram, protonation of the leaving CoA thiol in the condensation step is proposed to be carried out by His181. However, this proposed role for His181 requires further investigation. Adapted from Nie et al. (2021) [Ref. 213].

While the condensation step remains poorly understood, Dr Ashley Pike proposed that the malonyl unit could bind in a side pocket adjacent to the central tunnel, where it would be well placed to react with the acyl-enzyme intermediate (Fig. 4.19A,B) [213]. This side pocket is lined by conserved residues (Asn177, Gln211, Gln214, as well as His181). Indeed, the corresponding residues of *Dictyostelium discoideum* EloA have been previously demonstrated to be essential for activity [194]. However, additional evidence is required to understand the mechanistic details of this Claisen-like condensation step. In particular, we still do not know whether ELOVLs prevent premature malonyl-CoA decarboxylation by requiring prior acyl-enzyme intermediate formation, as observed for some but not all KCSs [186, 282, 283].

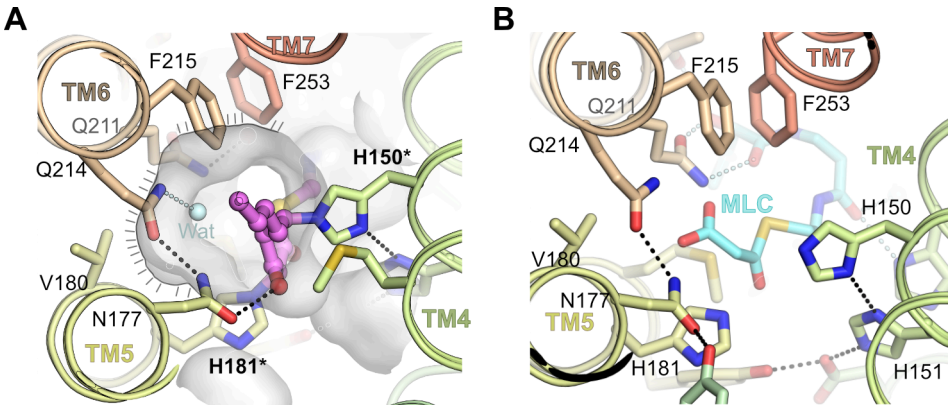


Figure 4.19. Putative malonyl binding pocket proposed by Dr Ashley Pike. A. Putative malonyl binding pocket (molecular surface shown in grey) located adjacent to the central tunnel in the ELOVL7 structure. The structure is viewed perpendicular to the membrane, from the ER side. The co-purified 3-keto acyl-CoA is shown in ball and stick representation (pink carbon atoms) **B.** Modelling of malonyl-CoA in the ELOVL7 structure suggests that the malonyl moiety could bind in this pocket. The putative malonyl binding pocket is lined by Asn177, Gln211, Gln214 and His181, which are known to be essential for catalysis. Hydrogen bonds are shown as dashes. No hydrogen bonds are shown for malonyl-CoA, as it was modelled into the structure, and its exact binding mode has not been determined experimentally. Adapted from Nie et al. (2021) [Ref. 213].

4.3.2 Formation of the 3-keto acyl-CoA adduct observed in the crystal structure

While the exact origin and underlying mechanism of formation of the 3-keto acyl-CoA adduct observed in the structure remain unclear, the presence of this covalently attached product analogue provided valuable insights into the binding mode of acyl-CoAs to ELOVL7, which enabled the proposal of the ping-pong type mechanism that I subsequently characterised using mass spectrometry-based assays. As the crystal structure of ELOVL7 solved by Dr Laiyin Nie and Dr Ashley is the first structure of an ELO elongase, the binding mode of acyl-CoAs to these enzymes was previously unknown.

In the structure, both His150 and His181 are covalently bonded to C5 and C2 of the product analogue, respectively, as supported by the electron density (Fig. 4.2, Section 4.1.2). These exact covalent attachments are not expected in the proposed catalytic mechanism of ELOVL7 (Fig. 4.18) and therefore likely lead to a dead-end complex, in which ELOVL7 is inhibited by the crosslinking of this product analogue at two positions in the active site.

In this work, I demonstrated that His150 acts as a nucleophile during the first step of the reaction and is regenerated upon reaction with the second substrate, malonyl-CoA. The observation of its nucleophilic properties enables us to speculate how this residue could become covalently attached at this position [213]. Briefly, if ELOVL7 reacted with an acyl-CoA containing a Cα-Cβ double bond, His150 could become covalently attached at the Cβ position via a conjugate addition to this α,β-unsaturated carbonyl compound. This could occur if a 2,3-*trans*-enoyl-CoA served as a substrate of ELOVL7. In this case, the α,β double bond would likely be placed adjacent to His150 at the end of the reaction, and the proposed conjugate addition would result in the displaced covalent bond observed in the structure (Fig. 4.20). 2,3-*trans*-enoyl-CoAs are an intermediate compound in the fatty acid elongation cycle, acting as the substrates of *trans*-2-enoyl-CoA reductase (TER) [182], which catalyses

the final step in the cycle. Therefore, it is possible that overexpression of recombinant ELOVL7 could result in the accumulation of this metabolite *in vitro*. Notably, the enzymes of this elongation cycle are thought to form a supramolecular elongation complex in the ER membrane, which is proposed to facilitate substrate channelling [153, 183, 195, 284], thus likely preventing the reaction of native ELOVL7 with this intermediate metabolite. Overall, the proposed origin of the dead-end complex seen in the structure, while plausible, requires additional validation. Specifically, we do not know whether 2,3-*trans*-enoyl-CoAs can act as substrates of ELOVLs. Unfortunately, 2,3-*trans*-enoyl-CoAs are not commercially available, and therefore I was not able to investigate their ability to act as substrates and whether this reaction would lead to covalent modification of the enzyme.

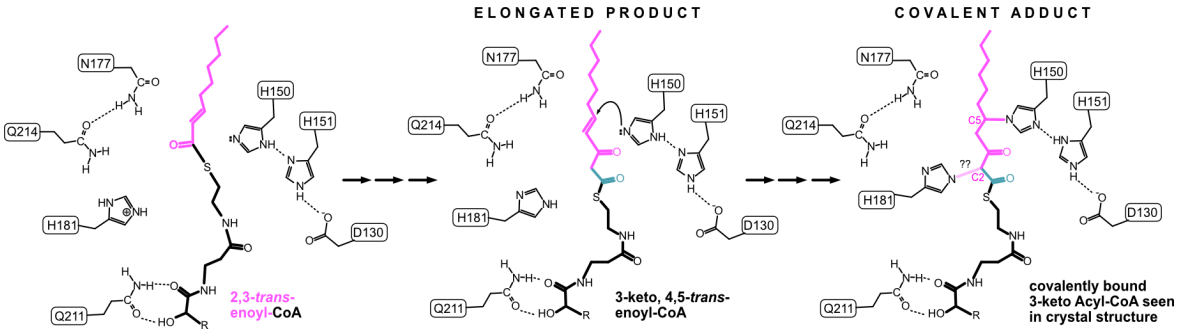


Figure 4.20. Proposed mechanism for the formation of the covalent linkage involving His150 in the crystal structure. If a 2,3-*trans*-enoyl-CoA (an intermediate of the fatty acid elongation cycle) binds in the central tunnel, it could initially be elongated to a 3-keto-4,5-*trans*-enoyl-CoA. Then, the covalent linkage to His150 would be formed via conjugate addition of the nucleophilic histidine to the α,β-unsaturated carbonyl. How His181 could become attached to C2 of this product analogue remains unclear. Adapted from Nie et al. (2021) [Ref. 213].

Furthermore, while a covalent linkage between His181 and C2 of the 3-keto acyl-CoA is also clearly supported by the electron density (Fig. 4.2, Section 4.1.2), it is essential to note that nitrogen and carbon atoms cannot be easily distinguished in the electron density maps. Therefore, the histidine side-chain orientation could be flipped, and the modelled N-C bond could potentially in fact be a C-C bond involving the His181 Cε1 atom. In both cases, these

linkages are not easily rationalised, and therefore, at present, the exact nature of this linkage cannot be determined, and a suitable mechanism for their formation could not be proposed.

4.3.3 Comparison of the mechanism of ELOs and FAE1-type KCSs

As described in Sections 1.2.6 - 1.2.8, in eukaryotes, the first step of the fatty acid elongation cycle is catalysed by two distinct families of 3-keto acyl-CoA synthases. In plants and some protists, this reaction is catalysed by FAE1-type KCSs [184-186], whereas other eukaryotes such as mammals, yeast, and parasites use ELO elongases [187-192]. Despite catalysing a similar decarboxylative Claisen-like condensation, where a malonyl-CoA-derived C_2 unit is transferred to an acyl-CoA in the fatty acid elongation cycle in the ER, FAE1-type KCSs and ELOs do not share sequence or structural similarities. FAE1-type KCSs are predicted to be anchored to the ER membrane at their N-terminus by 1 or 2 transmembrane helices, and their sequences contain a conserved Cys-His-Asn catalytic triad, where the Cys residue is thought to act as the nucleophile during transacylation, leading to the formation of a thioester acyl-enzyme intermediate [186, 283, 285]. In contrast, ELO elongases contain only a HxxHH motif essential for catalysis [194, 195, 197]. The ELOVL7 structure solved by Dr Laiyin Nie and Dr Ashley Pike reveals a 7TM architecture for this family of enzymes and shows that the active site is lined by histidines, including the HxxHH motif, previously only known to participate in metal binding. Using denaturing intact protein mass spectrometry, I demonstrated that catalysis is independent of metal ions and proceeds via a ping-pong mechanism, where an unusual covalent acyl-imidazole intermediate is formed on His150 of the ^{147}Hxx**H**H^{151} motif during the transacylation step. The nucleophilicity of His150 is likely enhanced by a His-His-Asp hydrogen-bonding network in the active site, where both histidines belong to the HxxHH motif. Together, these results reveal an unexpected mechanism

for ELO elongases where the conserved histidine-rich motif plays a central role in forming the catalytic triad.

5 Structural studies of human ceramide synthases

5.1 Aims of this chapter

As discussed in earlier sections, ceramides with different acyl chains have different physiological roles, and the relative distribution of these different ceramides is of central importance in the context of metabolic diseases. As a result, there is a great interest in achieving selective inhibition of CerS6 and perhaps CerS1 and CerS5. Elucidating the structural basis of substrate specificity within the CerS family would constitute the first step towards the design of isoform-specific inhibitors. In addition, structural information could provide useful insights into the little-understood mechanism of dihydroceramide synthesis by these enzymes. Therefore, in this chapter, I focused on establishing the expression and purification of CerS constructs for structural studies and obtained nanobody binders against CerS5 and CerS6 that are currently being used as fiducial markers to assist in the structural determination of these enzymes by electron cryomicroscopy (cryo-EM). This work aimed to establish a tractable system that would enable the determination of structures of ceramide synthases in complex with known non-specific inhibitors. The initial cryo-EM efforts described in this chapter focused on the CerS6 protein, given its importance as a target for the development of novel therapies.

5.2 Establishing purification conditions for human ceramide synthases for structural studies

5.2.1 Constructs designed

CerS1, CerS5 and CerS6 constructs for recombinant expression were designed with different N- and C-terminal truncations of the full-length polypeptide. These were then sub-cloned by the Biotechnology Group (SGC) using ligation-independent cloning [286]. Construct boundaries and vector choices were designed by me or others in the SGC (see below). N- and C-terminal truncations up until the first or last predicted transmembrane helix, respectively, were made. In addition, several constructs with different truncations along the disordered C-terminus were designed for crystallisation purposes. The full-length sequences of CerS1, CerS5 and CerS6 are available in Appendix 1, and detailed lists of truncations generated are available in Table 5.1 (CerS5), Table 5.2 (CerS6) and Appendix 2 (CerS1).

As no experimental structures are available for CerS and homologues, transmembrane helix and disorder prediction servers were used to decide on construct boundaries. To this end, transmembrane helices were predicted using TMHMM2.0 [287], and disordered regions were predicted using PrDOS [288] and GlobPlot [289]. Also, additional secondary structure and disorder predictions were obtained using the PSIPRED protein analysis workbench [290] (available at http://bioinf.cs.ucl.ac.uk/psipred/). One should mention that recent advances in protein structure prediction such as RoseTTAFold [138] and AlphaFold2 [137] were not yet available at the time CerS constructs were designed.

Also, prior to my arrival at the lab, several CerS5 and CerS6 constructs had been generated and screened for expression on a 3 mL scale. CerS5 constructs c001-c006 and CerS6 constructs c001-c006 (*Sf*9 expression) were designed by Dr Simon Bushell; CerS5 constructs c007-c014 and CerS6 constructs c007-c014 (HEK293 expression) were designed by Dr Jon

Elkins. However, at that stage, only CerS6-c003 could be purified, but precipitated heavily during purification and displayed a low melting temperature (35.9 ºC) (discussed in Section 5.2.6). No other constructs could be purified when expressed in *Sf*9 cells, and constructs designed for mammalian HEK293 expression had not been evaluated beyond initial 3 mL expression tests (described in the next section). Nevertheless, these preliminary results provided invaluable information for further construct design and contributed to establishing priorities for purification screening.

5.2.2 Expression screening

All designed constructs were initially screened for expression by the Biotechnology Group (SGC). This expression test was carried out in *Sf*9 or Expi293F cells (depending on the construct) on a 3 mL scale in 24-well blocks. Briefly, after cells were harvested, they were lysed by sonication, solubilised in 1 % (w/v) DDM / 0.1 % (w/v) CHS and the soluble material was bound to either TALON or Strep-Tactin XT resin (depending on the C-terminal tags). Bound proteins were eluted and analysed by SDS-PAGE. The identity of the bands predicted to correspond to the protein of interest was confirmed by in-gel tryptic digestion followed by MS/MS analysis carried out by the Biotechnology Group (SGC). Expression screening results are summarised in Table 5.1 (CerS5), Table 5.2 (CerS6) and Appendix 2 (CerS1). At this stage, I focused on CerS5 and CerS6 for downstream work, but the data generated for CerS1 could be useful for future work.

This approach enables assessing the yield of the membrane proteins that can be extracted in DDM / CHS, and has been optimised in the SGC, providing an initial indication that the target membrane protein construct is expressed in its folded form and can be extracted in this detergent. However, an obvious limitation of the approach of using a single detergent mixture for extraction is the fact that false negatives are expected, as membrane proteins

expressed in their folded form may fail to be extracted by DDM / CHS. However, from an operational perspective (for the SGC), this provides a useful compromise that enables accelerating the expression tests for a large number of constructs, where downstream detergent screening for purification can be carried out on positive hits. An alternative approach that would limit the number of false negatives would be to carry out the initial expression tests in a small number of different detergents. This approach was not employed for the CerS constructs but could perhaps constitute a valuable strategy in cases where a small number of positive hits is obtained in initial tests carried out using DDM / CHS.

Table 5.1. Expression levels of available CerS5 constructs. The expression level was judged by the intensity of the CerS5 SDS-PAGE bands obtained after a 3 mL purification using DDM / CHS as the extraction detergent. Constructs marked with * were designed in the SGC prior to my arrival.

Construct ID	N-terminus	C-terminus	Expression system	Purification tags	Expression level
CerS5-c001*	M1	E392	*Sf*9	10x His - FLAG	Medium
CerS5-c002*	M1	P368	*Sf*9	10x His - FLAG	None
CerS5-c003*	M1	V352	*Sf*9	10x His - FLAG	None
CerS5-c004*	L23	E392	*Sf*9	10x His - FLAG	Medium
CerS5-c005*	L23	P368	*Sf*9	10x His - FLAG	None
CerS5-c006*	L23	V352	*Sf*9	10x His - FLAG	None
CerS5-c007*	M1	E392	Mammalian	10x His - FLAG	Medium
CerS5-c008*	M1	P368	Mammalian	10x His - FLAG	None
CerS5-c009*	M1	V352	Mammalian	10x His - FLAG	Medium
CerS5-c010*	L23	E392	Mammalian	10x His - FLAG	None
CerS5-c011*	L23	P368	Mammalian	10x His - FLAG	Medium
CerS5-c012*	L23	V352	Mammalian	10x His - FLAG	High
CerS5-c013*	D31	V352	Mammalian	10x His - FLAG	Medium
CerS5-c014*	Y41	V352	Mammalian	10x His - FLAG	Medium
CerS5-c016	M1	E392	Mammalian	10 x His - Twin Strep	None
CerS5-c017	M1	V352	Mammalian	10 x His - Twin Strep	High
CerS5-c018	M1	K343	Mammalian	10 x His - Twin Strep	High
CerS5-c019	M1	K337	Mammalian	10 x His - Twin Strep	High
CerS5-c020	L11	E392	Mammalian	10 x His - Twin Strep	Low
CerS5-c021	L11	V352	Mammalian	10 x His - Twin Strep	High
CerS5-c022	L11	K343	Mammalian	10 x His - Twin Strep	High
CerS5-c023	L11	K337	Mammalian	10 x His - Twin Strep	High
CerS5-c024	L23	V352	Mammalian	10 x His - Twin Strep	High
CerS5-c025	L23	K343	Mammalian	10 x His - Twin Strep	Low
CerS5-c026	G40	V352	Mammalian	10 x His - Twin Strep	Medium
CerS5-c027	G40	K343	Mammalian	10 x His - Twin Strep	Medium
CerS5-c028	G40	K337	Mammalian	10 x His - Twin Strep	Low

Table 5.2. Expression levels of available CerS6 constructs. The expression level was judged by the intensity of the CerS6 SDS-PAGE bands obtained after a 3 mL purification using DDM / CHS as the extraction detergent. Constructs marked with * were designed in the SGC prior to my arrival.

Construct ID	N-terminus	C-terminus	Expression system	Purification tags	Expression level
CerS6-c001*	M1	D392	*Sf*9	10x His - FLAG	None
CerS6-c002*	M1	A370	*Sf*9	10x His - FLAG	None
CerS6-c003*	M1	D350	*Sf*9	10x His - FLAG	Medium
CerS6-c004*	A22	E392	*Sf*9	10x His - FLAG	Medium
CerS6-c005*	A22	A370	*Sf*9	10x His - FLAG	None
CerS6-c006*	A22	D350	*Sf*9	10x His - FLAG	None
CerS6-c007*	M1	D392	Mammalian	10x His - FLAG	None
CerS6-c008*	M1	A370	Mammalian	10x His - FLAG	Medium
CerS6-c009*	M1	D350	Mammalian	10x His - FLAG	Medium
CerS6-c010*	A22	D392	Mammalian	10x His - FLAG	Low
CerS6-c011*	A22	A370	Mammalian	10x His - FLAG	Medium
CerS6-c012*	A22	D350	Mammalian	10x His - FLAG	Low
CerS6-c013*	E28	D350	Mammalian	10x His - FLAG	None
CerS6-c014*	P33	D350	Mammalian	10x His - FLAG	None
CerS6-c023	M1	D392	Mammalian	10 x His - Twin Strep	None
CerS6-c024	M1	A370	Mammalian	10 x His - Twin Strep	High
CerS6-c025	M1	D350	Mammalian	10 x His - Twin Strep	High
CerS6-c026	M1	K334	Mammalian	10 x His - Twin Strep	High
CerS6-c027	A22	A370	Mammalian	10 x His - Twin Strep	Medium
CerS6-c028	A22	D350	Mammalian	10 x His - Twin Strep	Low
CerS6-c029	A22	K334	Mammalian	10 x His - Twin Strep	Low

5.2.3 Detergent screening

Once expressing constructs were identified, they were then expressed in a 50 mL scale for detergent screening. The method employed is outlined in Section 2.4.1 and closely resembles that of the expression tests described in Section 5.2.2, with the exception that 12 different detergents are being screened simultaneously per construct (so, in 12 x 3 mL of original cell culture volume). This 12-detergent screening strategy enables the identification of suitable detergents (with or without added CHS) for the extraction of the protein of interest and provides a useful starting point for downstream small-scale purification optimisation (using different detergents identified in this 12-detergent screening step).

Using the setup described in this work, 12 detergents are screened simultaneously per construct. This includes 11 potentially useful detergents for downstream work and a harsh positive control, fos-choline 12, which has a high extraction efficiency but is known to extract membrane proteins in misfolded states [291]. Different detergent combinations can be selected for testing at this stage, depending on the intended downstream application. For example, for crystallisation attempts, one may want to screen short-chain detergents such as OGNG and β-OG, which form smaller micelles, leading to a smaller shielding effect on crystal contacts. However, short-chain detergents such as β-OG and LDAO often display a lower extraction efficiency, and, in that case, the protein of interest may be exchanged into these detergents at a later stage after extraction in another detergent. Certain detergents such as LMNG have been shown to enable successful crystallisation in the lipidic cubic phase. In addition, LMNG is a suitable choice for cryo-EM studies due to its low critical micelle concentration (CMC, the concentration of the surfactant above which micelles form spontaneously). On the other hand, for protein reconstitution into liposomes by rapid dilution followed by dialysis, it would be beneficial to screen detergents with a high critical micelle concentration (CMC) such as DM. Overall, the choice of detergents to be screened at this stage will vary depending on the intended application.

High-expressing CerS5 and CerS6 constructs identified in the 3 mL test expressions described in Section 5.2.2 were selected for detergent extraction screening. Representative results obtained for selected constructs are shown in Fig. 5.1A,B and demonstrate that certain constructs are preferentially extracted in certain detergents and could not be extracted in others. In this case, we can observe that both CerS5-c009 and CerS6-c009 could be extracted in DM, LMNG / CHS, OGNG / CHS, C12E10 and CYMAL-5 (Fig. 5.1A). For both constructs, the addition of CHS led to a dramatic increase in extraction efficiency when LMNG was used as the extraction detergent.

Lessons learnt from initial detergent screens resulted in an updated detergent screening list. These were used for screening additional constructs, such as CerS5-c023 and CerS5-c024 (Fig. 5.1B). For both these constructs, the majority of the detergents screened successfully extracted the protein of interest, with the exception of β-OG. In addition, we can observe that the addition of CHS enabled increasing the extraction efficiency when using OGNG.

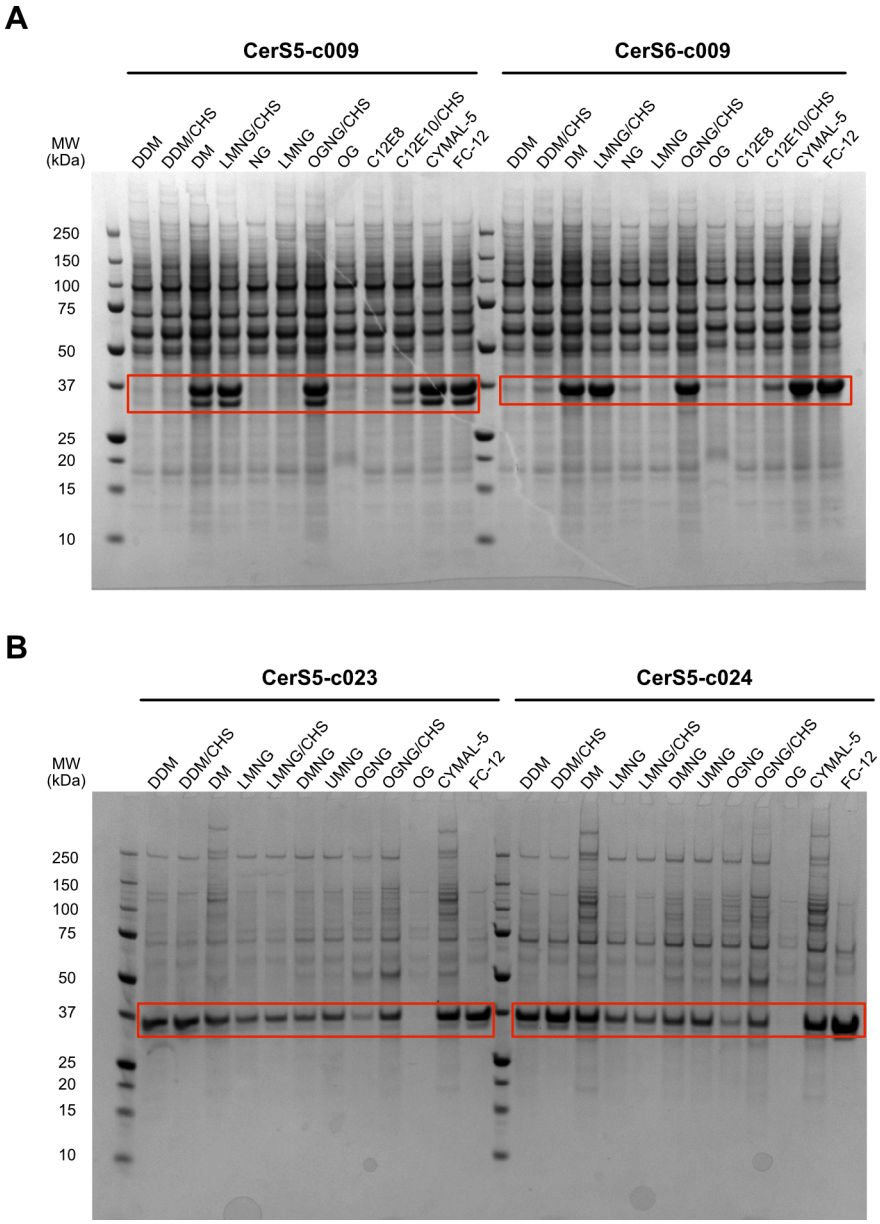


Figure 5.1. Representative detergent screening of CerS5 and CerS6 constructs. A. Detergent extraction screening of CerS5-c009 (Met1-Val352) and CerS6-c009 (Met1-Asp350), based on SDS-PAGE analysis of IMAC elution samples from 3 mL-scale purifications. The double-banding pattern

observed for CerS5-c009 arises from glycosylation at one site, identified in Section 5.2.5. **B.** Detergent extraction screening of CerS5-c023 (Leu11-Lys337) and CerS5-c024 (Leu23-Val352), based on SDS-PAGE analysis of Strep elution samples from 3 mL-scale purifications. CerS5 / CerS6 protein bands are highlighted by red boxes.

5.2.4 Small-scale purification screening

Following the identification of successful detergents in an extraction assay, these were then screened in small-scale purifications followed by analysis by size-exclusion chromatography (SEC) and SDS-PAGE. Briefly, the cells were thawed, mechanically lysed and later split for solubilisation in the various detergents tested. Typically, 250 mL of original cell culture volume of Expi293F cells were used for each detergent / condition tested, and membrane proteins were extracted with 1 % (w/v) of the selected detergent. Then, the protein of interest was bound to either TALON or Strep-Tactin XT resin (depending on the C-terminal tags of the construct), eluted, and the C-terminal tags were cleaved using TEV protease overnight. The following day, TEV protease and cleaved tags were removed through a reverse IMAC step; the cleaved protein was concentrated and analysed by size-exclusion chromatography using an SRT SEC-300 column.

This approach enabled, for the first time, the identification of suitable conditions for the purification of CerS5. The success in optimising the purification of CerS5-c009 is described here as an example of the work carried out. A number of detergents that successfully extracted the protein as analysed by SDS-PAGE from a 3 mL TALON purification (Fig. 5.1A) were selected for small-scale purifications (LMNG / CHS, OGNG / CHS, CYMAL-5, DM). In most of the detergents tested, the protein either eluted at the void volume of the SEC column indicating aggregation (CYMAL-5) or yielded broad and asymmetrical peaks, often

with poor separation from larger aggregates (LMNG / CHS and DM). However, when purified using OGNG / CHS, CerS5-c009 elutes in a large mono-disperse peak from SEC (Fig. 5.2A), indicating that this detergent is suitable for large-scale purification attempts.

Similar work was conducted to screen suitable purification conditions for selected CerS6 constructs. As a representative example, Fig. 5.2B shows the elution profiles obtained from small-scale purifications of CerS6-c008, CerS6-c009 and CerS6-c011 carried out using LMNG / CHS. Here, mono-disperse peaks were obtained for CerS6-c008 and CerS6-c009 but not for CerS6-c011. In addition, when sufficient material could be recovered from the SEC fractions, melting curves were obtained by nanoDSF to compare the thermostability of the protein in different detergents or of different constructs in the same detergent, thus adding another dimension to the data generated. In this example, comparing the melting temperatures of purified CerS6-c008 (43.1 ºC) and CerS6-c009 (54.9 ºC) in LMNG / CHS (Fig. 5.2C) led to the prioritisation of CerS6-c009 for downstream work.

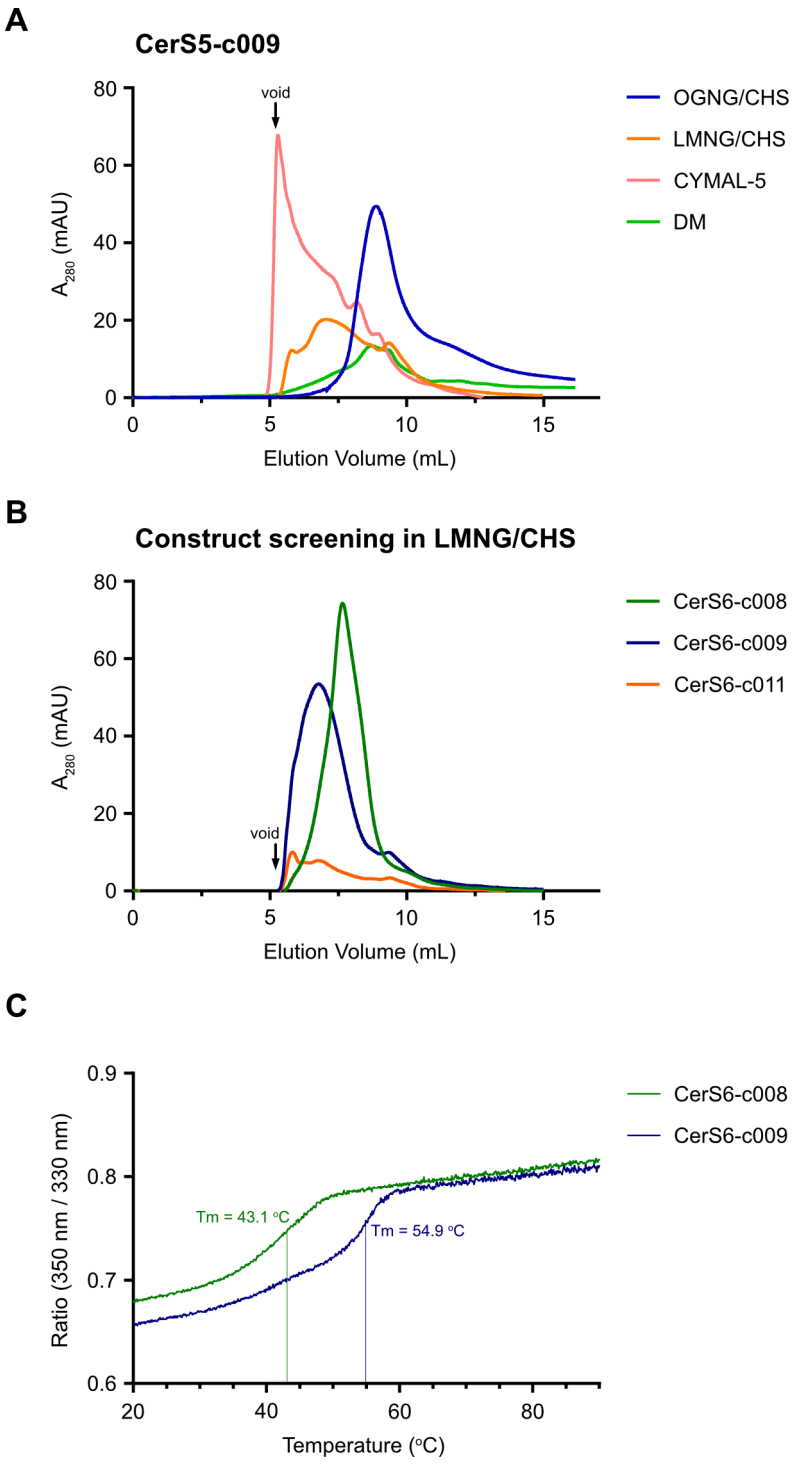


Figure 5.2. Small-scale purification screening of CerS constructs. A. SEC traces from small-scale purifications of CerS5-c009 in selected detergents identified in Section 5.2.3. **B.** SEC traces from small-scale purifications of CerS6-c008, CerS6-c009 and CerS6-c011 in LMNG/CHS. **C.** Melting curves of purified CerS6-c008 and CerS6-c009 obtained from small-scale purification trials (n=3 technical replicates). Protein unfolding was monitored by recording tryptophan fluorescence at 330 nm and 350 nm, and the melting temperature (Tm) was determined from the inflection point of the calculated fluorescence ratio (350 nm / 330 nm) using PR.ThermControl software (NanoTemper Technologies).

One should mention that a common alternative approach is to express the proteins with a C-terminal GFP fusion and evaluate monodispersity by fluorescence-detection size exclusion chromatography (FSEC) [292] upon identification of detergents that successfully extract the protein of interest. This strategy enables increasing the screening throughput, as the protein does not need to be purified upon detergent extraction and can be loaded onto the SEC column directly. By recording GFP fluorescence during SEC, the signal obtained is specific to the GFP-tagged protein and enables monodispersity screening in the selected conditions without prior purification. However, this strategy can lead to more false positives, and downstream purification optimisation of FSEC-positive conditions is still required. On the other hand, the more laborious approach to follow up detergent extraction assays with small-scale purifications, as described in this work, provides a more accurate comparison to the conditions of later larger scale purifications at the cost of a reduction in throughput. For these reasons, a FSEC step was not included in the CerS purification screening strategy, and therefore the designed constructs did not include a GFP fusion.

5.2.5 Large-scale CerS5 purification

The CerS5-c009 construct was subsequently expressed in large scale in Expi293F cells. Typically, 4L of Expi293F cells were used in each purification. In short, cells were mechanically lysed, solubilised with 1 % (w/v) OGNG / CHS and insoluble material was removed by centrifugation. In this construct, the truncated CerS5 protein (Met1 - Val352) is fused on its C-terminus to a TEV protease cleavage site, followed by 10x His and FLAG tags and, therefore, the tagged protein was initially purified using TALON metal affinity resin. However, given the known high background binding of contaminants from this mammalian expression system to TALON resin, the TALON step was followed by binding to anti-FLAG M2 affinity gel. Following elution of the tagged protein, the C-terminal tags were cleaved

overnight with TEV protease. Uncleaved protein and the His-tagged TEV protease were removed via a reverse IMAC step using TALON resin and the cleaved CerS5 protein was concentrated and further purified by SEC (Fig. 5.3A). The SEC profile shows a single monodisperse peak, corresponding to pure CerS5, as analysed by SDS-PAGE and confirmed by LC-MS/MS following digestion of the gel bands with trypsin (Fig. 5.3B).

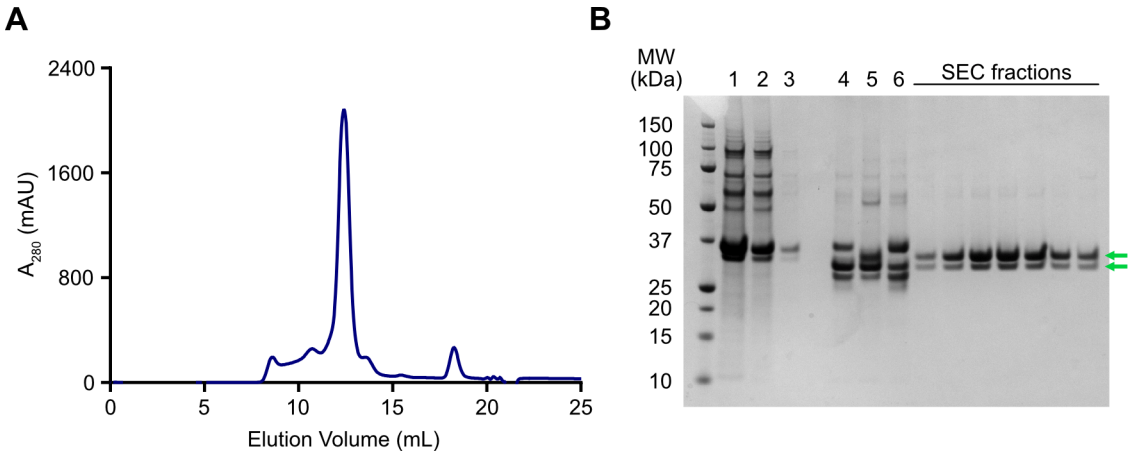


Figure 5.3. Large-scale purification of CerS5 expressed in Expi293F cells. A. Elution profile of CerS5-c009 (Met1-Val352) from a Superdex 200 Increase 10/300 SEC column. **B.** SDS-PAGE analysis of samples taken during purification. *1* – Eluate from TALON resin; *2* – Flow-through from the anti-FLAG M2 affinity gel; *3* – FLAG resin wash; *4* – overnight TEV cleavage, showing uncleaved CerS5 (top band), cleaved CerS5 (middle band) and TEV protease (bottom band); *5* – Reverse IMAC flow-through containing TEV-cleaved CerS5; *6* – Eluate from reverse IMAC step, containing uncleaved CerS5 and TEV protease. *SEC fractions* – peak fractions from (A). The two bands marked with green arrows were confirmed to correspond to CerS5. This double-banding pattern is a consequence of N-linked glycosylation.

This purification strategy was employed to prepare Cers5 protein for nanobody generation (see Section 5.3.2). However, given the lower yields and increased cost resulting from the need to include the additional anti-FLAG gel step, this CerS5 truncation (Met1 - Val352) was re-cloned into the pHTBV1.1-SIII-LIC vector, which contains C-terminal TEV cleavable Twin-Strep and 10x His tags. Further work was carried out using this Twin-Strep tagged construct (CerS5-c017).

In addition, the final purified samples were analysed by denaturing intact protein mass spectrometry to identify post-translational modifications. When expressed in Expi293F cells, CerS5-c017 appears palmitoylated and is N-glycosylated at one site. The expected mass of CerS5-c017 after tag cleavage is 42409.86 Da. The observed unmodified protein peak (42319.41 Da) constitutes a loss of 90.45 Da, corresponding to the loss of the initiator methionine followed by acetylation of the new N-terminus (theoretical mass shift: -89.16 Da). Palmitoylation could be identified by the presence of a +238.22 Da mass shift (theoretical: +238.41 Da). However, because palmitoyl-CoA is the preferred substrate of CerS5, if the reaction proceeds via a covalent intermediate, this mass addition could also correspond to the addition of the acyl chain in the active site. Therefore, the site for this palmitoylation is unknown and could happen either in a surface-exposed cysteine as a post-translational modification or in the active site corresponding to a trapped reaction intermediate. Glycosylation could also be identified by the addition of 1242.12 Da, likely corresponding to GlcNAc-Man_3-GlcNAc-Fuc-GlucNAc (theoretical mass addition: +1242.14 Da). Additional deconvoluted mass peaks at higher molecular weights seem to occur with a repeating pattern of +203 Da addition, likely corresponding to the further addition of multiple hexose (GlcNAc or GalNAc) units to this glycan (Fig. 5.4A).

The presence of this complex glycosylation pattern would lead to sample heterogeneity that could hamper successful crystallisation. Therefore, CerS5-c017 was next expressed in Expi GnTI- cells, a HEK293 line which lacks the gene encoding for N-acetylglucosaminyltransferase I and therefore do not produce complex N-linked glycans. Protein was purified as before and analysed by denaturing intact protein mass spectrometry (Fig. 5.4B). Using this expression system, glycosylation is mostly limited to the core N-linked glycan Man_5-GlcNAc$_2$, identified by the presence of a +1217.05 Da mass shift (theoretical: +1217.09 Da), and to the further addition of a fucose (observed: +145.58 Da; theoretical: 146.14 Da), which

is expected to be added to the initial GlcNAc. In support of this last statement, cleavage of the core glycan with EndoH, which cleaves after the initial N-linked GlcNAc, exclusively leads to mass shifts corresponding to the initial GlcNAc (observed: +203.11 Da; theoretical: +203.19 Da) with or without an additional fucose (observed: +146.07 Da; theoretical: +146.14 Da) (Fig. 5.4C). In all of the above cases, CerS5-c017 continues to display palmitoylation at a single site.

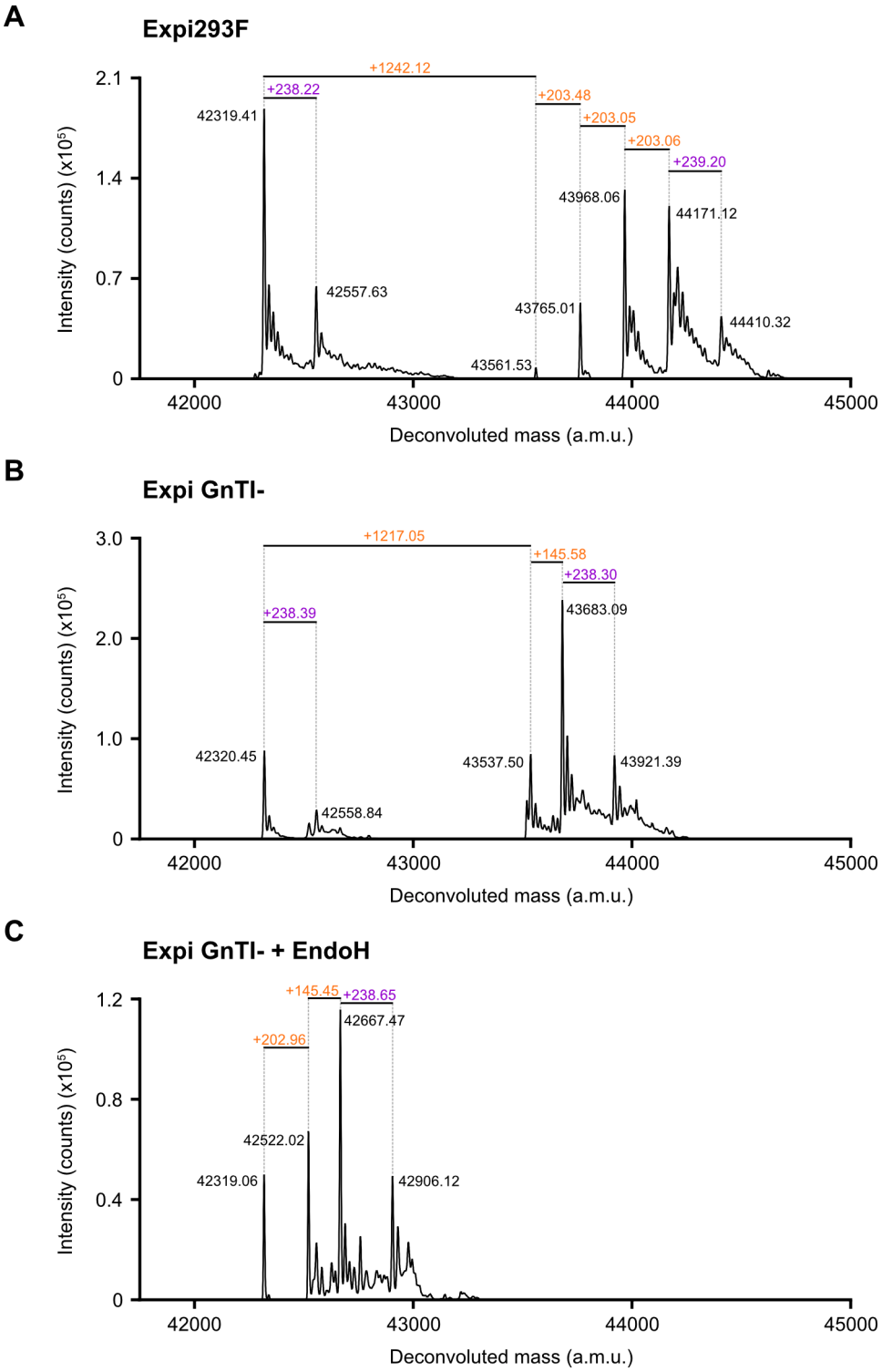


Figure 5.4. Denaturing intact LC-ESI-MS of purified CerS5 samples. A. CerS5-c017 purified from Expi293F cells. The unmodified protein peak (42319.41 Da) corresponds to the mass of the purified construct with loss of methionine and acetylation of the new N-terminus. Observed mass shifts displayed in orange text correspond to a core N-linked glycan (1242.12 Da) and additional HexNac (GlcNAc or GalNAc) additions (theoretical: +203.19). In addition, palmitoylation at one site can be observed (mass shifts displayed in purple text; theoretical: +238.41 Da). **B.** CerS5-c017 purified from Expi GnTI- cells, which lack N-acetylglucosaminyltransferase. In these cells, the purified protein displays mass additions corresponding to a core N-linked glycan (observed: + 1217.05 Da) and

an additional fucose (observed: +145.58 Da), linked to the initial Asn-linked GlcNAc, as evidenced by the results displayed in panel C. As before, there is palmitoylation at one site (mass shifts in purple text). **C.** CerS5-c017 purified from Expi GnTI- cells and cleaved with EndoH. In this sample, glycosylation is limited to the initial GlcNAc (observed: +202.96 Da) and an additional fucose (observed: +145.45 Da).

Overall, the complex N-linked glycosylation identified on CerS5-c017 leads to a heterogeneous protein sample. For crystallisation efforts, I demonstrated that this N-linked glycosylation can be limited to the initial GlcNAc ± fucose through treatment with EndoH. For other applications, such as the nanobody generation described in Section 5.3, the protein was expressed in Expi GnTI- cells and the core N-linked glycan was not cleaved.

5.2.6 Large-scale CerS6 purifications

Similarly to CerS5, conditions for CerS6 purification identified in small-scale screening were followed up with large scale expressions and purifications. In addition, data generated by Dr Andrew Quigley prior to my arrival suggested that CerS6-c003 (Met1 - Asp350) could be successfully expressed in *Sf*9 cells and purified in milligram amounts. As a result, this construct was initially selected for large scale purifications for structural studies. CerS6-c003 purification was carried out as described in Section 2.1.3.3, using LMNG as the extraction and purification detergent, and yielded pure monodisperse protein (Fig. 5.5A,B). However, CerS6-c003 readily precipitates during purification, leading to a lower yield for the final purified sample. In addition, purified CerS6-c003 displayed a low melting temperature of 35.9 ± 0.1 ºC (Fig. 5.5E). Therefore, additional CerS6 constructs were generated and screened to obtain a final purified CerS6 sample with increased stability for structural studies. The extraction and purification screening described in Sections 5.2.3 and 5.2.4 identified that the addition of CHS during purification of CerS6-c003 enabled increasing its

melting temperature to 43.7 ± 0.2 ºC (Fig. 5.5E). In addition, screening of constructs expressed in Expi293F cells identified CerS6-c025 as a suitable candidate to be scaled up for structural studies. Interestingly, this construct corresponds to the same truncation as CerS6-c003 (Met1 - Asp350). However, when expressed in Expi293F cells and purified using LMNG / CHS, the protein displayed a melting temperature of 54.2 ± 0.2 ºC (Fig. 5.5C-E). Overall, this systematic screening of expression and purification conditions improved the thermostability of this CerS6 truncation by 18.3 ºC, making it a more attractive candidate for structural studies.

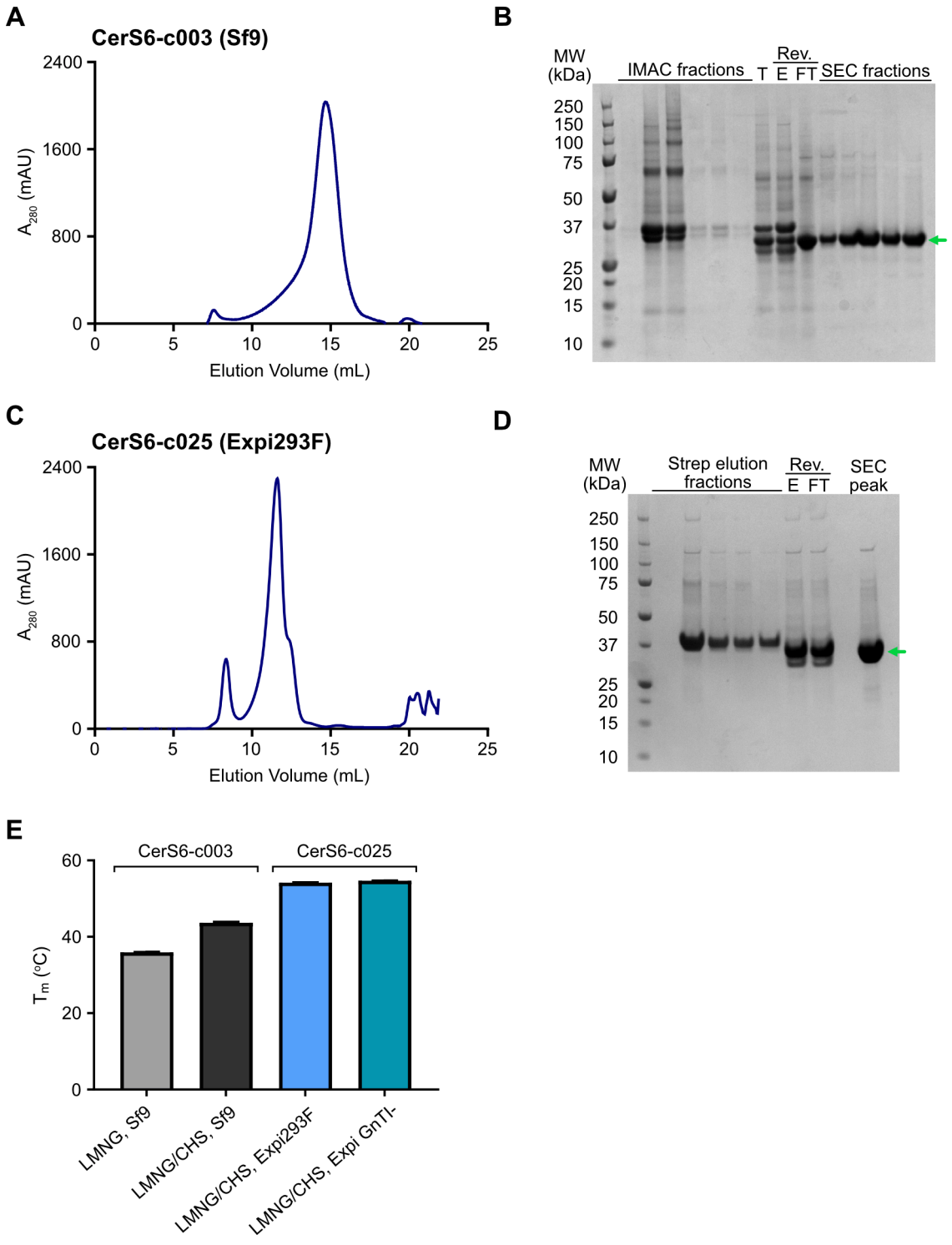


Figure 5.5. Large-scale purifications of CerS6. A. Elution profile of CerS6-c003 (Met1-Asp350, expressed in *Sf*9 cells) from a Superose 6 Increase 10/300 SEC column. **B.** SDS-PAGE analysis of samples taken during CerS6-c003 purification. *IMAC fractions* correspond to 2.5 mL fractions collected during elution of tagged protein from the TALON resin. Sample *T* corresponds to the sample after incubation with TEV protease overnight and before reverse IMAC. Reverse IMAC elution (*Rev. IMAC E*) corresponds to the elution of tagged proteins that bound to the TALON resin during the reverse IMAC step and contains primarily the His-tagged TEV protease and residual uncleaved CerS6-c003. Reverse IMAC flow-through (*Rev. IMAC FT*) corresponds to the cleaved CerS6-c003 sample that did

not bind the TALON resin during the reverse IMAC step. *SEC fractions* correspond to the peak fractions from (A). The band corresponding to purified CerS6 (confirmed by tryptic digest LC-MS/MS) is marked with a green arrow. **C.** Elution profile of CerS6-c025 (Met1-Asp350, expressed in Expi293F cells) from a Superdex 200 Increase 10/300 SEC column. **D.** SDS-PAGE analysis of samples taken during CerS6-c025 purification. *Strep elution fractions* correspond to the samples collected during elution from the Strep-Tactin XT resin. Reverse IMAC elution (*Rev. IMAC E*) corresponds to the elution of tagged proteins that bound to the TALON resin during the reverse IMAC step and contains primarily the His-tagged TEV protease and residual uncleaved CerS6-c025. Reverse IMAC flow-through (*Rev. IMAC FT*) corresponds to TEV-cleaved CerS6-c025 which did not bind the TALON resin during the reverse IMAC step. *SEC peak* corresponds to the pooled fractions from (C). The band corresponding to purified CerS6 (confirmed by tryptic digest LC-MS/MS) is marked with a green arrow. **E.** Melting temperatures of purified CerS6 samples reveal the thermal stabilisation conferred by the supplementation of LMNG with cholesteryl hemissucinate (CHS) during purification and by expressing this CerS6 truncation (Met1-Asp350) in mammalian cells (Expi293F or Expi GnTI-). The values shown are the average of three technical replicates, and error bars correspond to ± S.E.M.

The purified CerS6 samples were next analysed by denaturing intact protein mass spectrometry. As observed for the CerS5-c017 sample, CerS6-c025 is palmitoylated at one site and N-glycosylated at a single site. Expression in Expi GnTI- limits this glycosylation to only the core N-linked glycan and was selected as the preferred strategy for structural studies (Fig. 5.6).

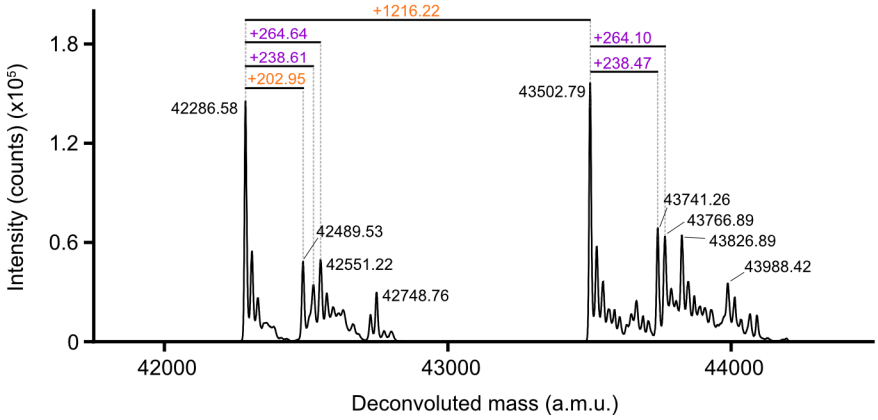


Figure 5.6. LC-ESI-MS intact mass analysis of purified CerS6-c025 expressed in Expi GnTI- cells. The lower molecular weight deconvoluted mass peak corresponds to the mass of the purified CerS6-c025 construct with loss of the initiator methionine followed by acetylation of the new N-terminus. Additional peaks observed correspond to post-translational modifications. N-linked glycosylation (orange text) is limited to the core N-linked glycan (observed: +1216.22 Da; theoretical: +1217.05 Da) or to only the initial N-linked GlcNAc (observed: +202.95 Da; theoretical: +203.19 Da). In addition, palmitoylation at one site can be observed (theoretical: +238.41 Da). One additional

modification can be observed (+264.64 Da / +264.10 Da), but its identity could not be assigned. Palmitoylation and this additional +264 Da modification are shown in purple text.

As described for CerS5 in Section 5.2.5, the site of palmitoylation is unknown and could either correspond to the modification of a surface-exposed cysteine or to a covalent reaction intermediate, as palmitoyl-CoA is also the preferred substrate of CerS6. Also, an additional modification can be observed (+264.64 Da / +264.10 Da). While these mass shifts do not provide sufficient evidence to confidently assign its identity, they match those expected for the covalent addition of a C18:1 acyl chain. However, further work is required to determine the nature of this modification and the site of attachment to the protein. This is discussed in Section 6.3.

5.2.7 NanoDSF screening of known inhibitors

Thermostability screening was carried out using purified CerS6-c025, in order to assess the ability of this C-terminally truncated construct to bind the non-selective inhibitors FTY720 and Fumonisin B1 and to prioritise inhibitors to be used in crystallisation efforts, in the first instance. This relies on the assumption that a change in T_m of a protein in the presence of a ligand is an indicator of its binding [293].

Purified protein (5 μM) was incubated with 20 μM or 100 μM of each compound for 1h and subsequently subjected to a temperature gradient, enabling the determination of its melting temperature in the presence of the various added compounds (Fig. 5.7). The T_m of as-purified CerS6-c025 in the absence of ligands was 54.67 ± 0.04 ºC. FTY720 stabilised the protein moderately by +0.62 ± 0.05 ºC at 20 μM and +0.75 ± 0.18 ºC at 100 μM. On the other hand, Fumonisin B1 gave rise to a larger thermal stabilisation (+1.55 ± 0.04 ºC at 20 μM; +2.41 ± 0.56 ºC at 100 μM).

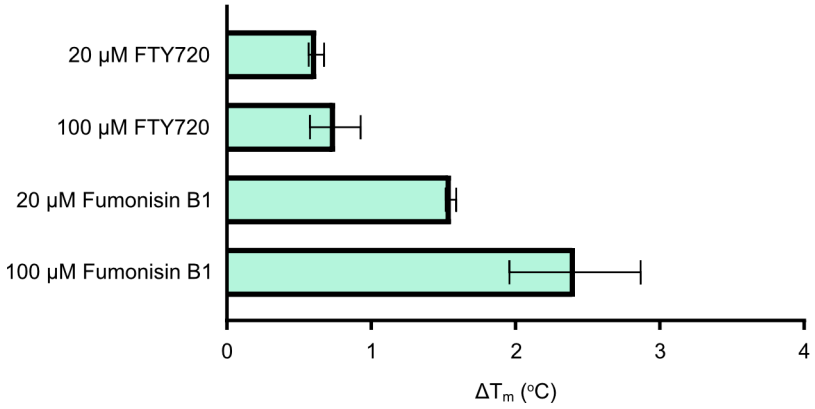


Figure 5.7. NanoDSF screening of commercially available non-selective CerS inhibitors for thermostabilisation of CerS6. Purified CerS6 was incubated with 20 µM or 100 µM of each compound for 1h before being subjected to a temperature gradient of 1 °C min^{-1}. Protein unfolding was monitored by recording tryptophan fluorescence at 330 nm and 350 nm, and the melting temperature (T_m) was determined from the inflection point of the calculated fluorescence ratio (350 nm / 330 nm). Melting temperature shifts (ΔT_m) were calculated as the difference between T_m values obtained with and without added compounds. Values shown are the average of three technical replicates, and the error bars correspond to ± S.E.M.

5.2.8 Design and purification of glycosylation mutants for crystallisation trials

Initially, crystallisation of CerS5-c017 and CerS6-c025 was attempted both by vapour diffusion and using the lipidic cubic phase method, with the aim towards structure determination by X-ray crystallography. However, these initial attempts were not successful as no crystals could be obtained even when the N-linked glycan was cleaved using EndoH. While crystallisation efforts continued with this EndoH-treated protein, site-specific variants were designed to completely remove this N-linked glycosylation. Denaturing intact protein mass spectrometry analysis of the purified proteins (Sections 5.2.5 and 5.2.6) suggested that there is only 1 N-linked glycosylation site in each protein. Using the N-linked glycosylation prediction server NetNGlyc-1.0 [294] (available at https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0), this site was predicted to be Asn26 for Cers5 and Asn18 for CerS6. These predictions are consistent with earlier reports that CerS5 and CerS6 are N-glycosylated in these N-terminal Asn residues [139, 141]. Therefore, these positions were mutated

to Gln residues, which preserved the chemistry in that position but are no longer suitable for N-linked glycosylation. Importantly, this N-linked glycosylation is not essential for CerS6 function [139].

At this stage, crystallisation efforts were focused on the CerS6 Asn18Gln mutant protein. This mutant corresponds to the same truncation as CerS6-c025 (i.e. Met1 – Asp350) and will be designated as CerS6-c036. CerS6-c036 was expressed in Expi293F cells and purified in milligram amounts following the same method used for CerS6-c025. As expected, the Asn18Gln mutation leads to loss of N-linked glycosylation, as determined by denaturing intact protein mass spectrometry (Fig. 5.8). This strategy confirmed that human CerS6 is glycosylated exclusively at Asn18 when expressed in Expi293F cells and enabled producing protein lacking N-linked glycosylation for crystallisation studies.

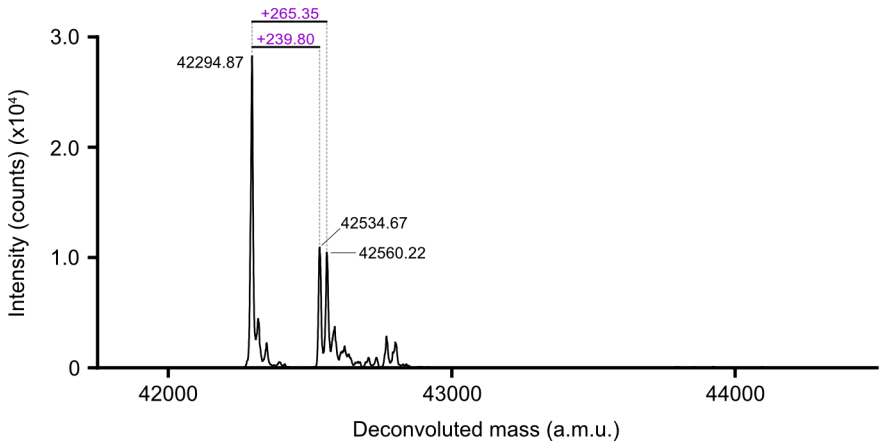


Figure 5.8. Denaturing intact LC-ESI-MS analysis of purified CerS6-c036 (Met1 – Asp350; Asn18Gln). Introduction of the Asn18Gln mutation led to the loss of N-linked glycosylation, confirming this position as the site of N-linked glycosylation and making CerS6-c036 a suitable candidate for crystallisation trials. Post-translational modifications observed correspond to palmitoylation (observed: +239.80 Da) and to an unknown modification (+265.35 Da), also observed in CerS6-c025.

Crystallisation trials were carried out for the CerS6-c036 protein using the lipidic cubic phase method, as LMNG / CHS is not preferred for vapour diffusion crystallisation attempts

due to the shielding effect of the LMNG micelle. However, no crystals could be obtained despite extensive screening of conditions and concentrations.

5.2.9 Identification of a dimeric population in purified CerS6

Ceramide synthases have been proposed to form homo- and heterodimers that are physiologically relevant for their functions [140] (described in Section 1.2.3). If these dimeric species exist and could be purified, the molecular weight for a CerS6-c025 homodimer would be 84 kDa and, therefore, would reach the size threshold commonly regarded for structure determination by cryo-EM.

To evaluate the presence of homodimers from recombinant expression, CerS6 constructs with different C-terminal tags were co-expressed in Expi GnTI- cells (Fig. 5.11A). This differential tag strategy enables separating dimers from monomers by selectively purifying CerS6 dimers formed by two protomers with different tags (expected to correspond to ~50 % of the total dimer population (Fig.5.9A)). The constructs selected for this experiment were CerS6-c009 (C-terminal 10x His and FLAG tags) and CerS6-c025 (C-terminal Twin-Strep and 10x His tags). Co-expression was carried out by co-infection with baculoviruses for each of the constructs, and successful expression of each construct was confirmed with a 3 mL test expression purification using either Strep-Tactin XT resin or Anti-FLAG M2 affinity gel. Then, purification of the dimeric protein was carried out by initially capturing the Strep-tagged protein (CerS6-c025), followed by isolation of FLAG-tagged proteins in the Strep-Tactin XT eluate, thus corresponding to the differentially tagged dimeric species (Fig. 5.9B).

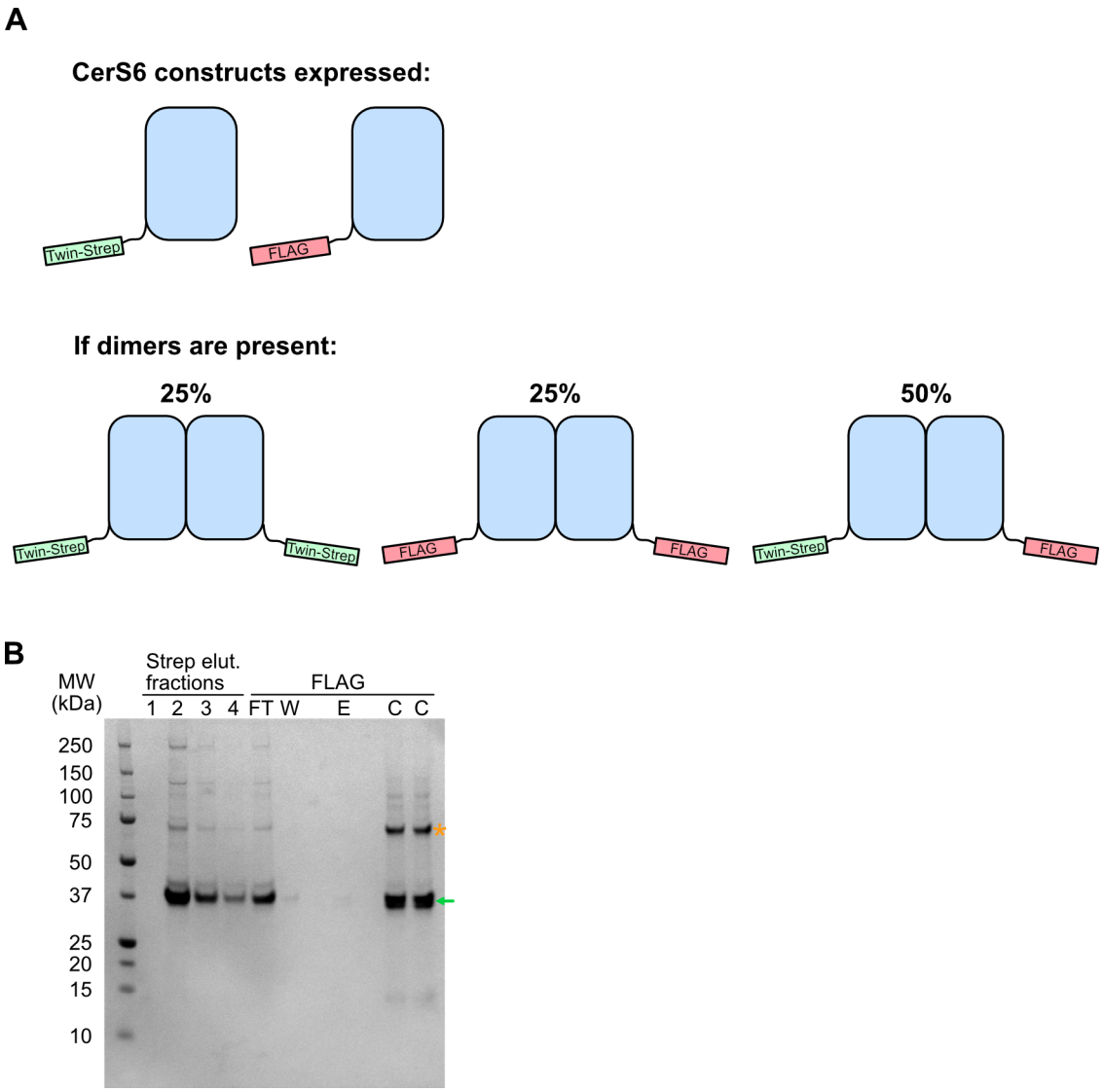


Figure 5.9. Purification of CerS6 dimers. A. Schematic representation of the strategy employed to selectively purify CerS6 dimers. CerS6-c009 (Met1-Asp350; C-terminal 10x His and FLAG tags) and CerS6-c025 (Met1-Asp350; C-terminal Twin-Strep and 10x His tags) were co-expressed in Expi293F cells. If CerS6 dimers exist, then ~50 % of the dimeric population would be expected to be differentially tagged, with a Twin-Strep tag in one monomer and a FLAG tag in the other. Differential purification using first Strep-Tactin XT resin, followed by anti-FLAG M2 affinity gel would therefore enable evaluating the presence of these stable dimers. **B.** SDS-PAGE analysis of the CerS6 dimer purification. CerS6 was successfully captured using Strep-Tactin XT (observed in the Strep elution fractions 2-4). Strep-tagged monomers and homodimers do not bind to the anti-FLAG M2 affinity gel and are therefore detected in the FLAG flow-through (*FLAG FT*). Flag resin was washed (*FLAG W*), and Flag-tagged CerS6 dimers were eluted (*FLAG E*) and subsequently concentrated (*FLAG C*). This strategy enabled the identification of CerS6 in the concentrated FLAG elution sample, confirming that a fraction of the expressed CerS6 protein is present as a stable dimer. The identity of the CerS6 band (green arrow) in the final sample was identified by tryptic digestion of the gel band, followed by MS/MS analysis. A higher molecular weight species (orange asterisk) was also confirmed to be CerS6 and therefore likely corresponds to a CerS6 dimer that was not resolved by SDS-PAGE. The concentrated FLAG elution sample was loaded on the gel twice to provide additional material for tryptic digest MS/MS analysis.

The observation that these truncated CerS6 constructs can form dimers encouraged efforts to attempt the structural determination of CerS6-c025 by cryo-EM. In the first instance, this would require investigating what fraction of the purified protein population exists in a dimeric state, which will be described in later sections.

5.3 Generation and validation of nanobodies against CerS5 and CerS6

5.3.1 Overview and strategy

As initial crystallisation attempts failed to yield any crystals of CerS5 or CerS6, and because I also aimed to attempt to solve the structure of CerS6 by cryo-EM, nanobody binders were generated against these proteins to assist in structural determination. Nanobodies are single-domain fragments of the heavy-chain-only antibodies present in camelids. These can be used as crystallisation chaperones, facilitating crystal contacts and increasing the chances of successful crystallisation of the target protein. In addition, the use of binders, including nanobodies, as fiducial markers for cryo-EM provides useful features for particle alignment to enable the structural determination of small proteins. Furthermore, when the small size of nanobodies (~14-15 kDa) is not sufficient to assist in cryo-EM structural determination, these nanobodies can be fused to a larger rigid scaffold to increase their molecular weight [295]. Also, additional approaches have been developed which use nanobody-binding Fab fragments [296] or a combination of a nanobody-binding Fab fragment and a nanobody-binding protein A fragment fused to MBP [297] to increase the size of the fiducial marker complex. In all these approaches, their success relies on the previous generation of suitable nanobodies that specifically bind the protein of interest.

Nanobodies directed against recombinant CerS5 or CerS6 proteins were generated to facilitate structure determination by either X-ray crystallography or cryo-EM. To this end, the

selected approach was to immunise alpacas with recombinant CerS5 and CerS6 reconstituted into proteoliposomes. These samples were generated in-house and supplied to the Nanobody Service Facility (University of Zurich) as a collaboration. There, the proteoliposomes containing the protein antigen were injected into alpacas, and lymphocyte RNA was collected from a small sample of the animal's blood. The nanobody open reading frames were then cloned into a phage display vector, and phages were enriched before ELISA screening of enriched clones. ELISA-positive clones were then sequenced, and unique nanobody binders were identified. These selected unique nanobody binders can then be recombinantly expressed in *E. coli* and further validated for *in vitro* binding to the target protein [298].

The next section will describe the preparation of the protein samples required for alpaca immunisation (CerS5 and CerS6 proteoliposomes), as well as for subsequent phage-display library panning and ELISA screening carried out by the Nanobody Service Facility (purified biotinylated CerS5 and CerS6, in detergent). Biotinylated proteins were also required for later validation of selected unique binders by Bio-Layer Interferometry (BLI) (Section 5.3.3.2).

5.3.2 Sample preparation for immunisation and library screening

CerS5 and CerS6 proteoliposomes were generated as described in Section 2.1.4. POPE:POPG (3:1 w/w) was selected as the lipid mixture. In short, large unilamellar vesicles (LUVs) were initially disrupted with 0.15 % Triton X-100 (determined to be the Triton X-100 concentration at which the liposomes are saturated with detergent). Then, purified protein in detergent was added to the mixture at a 1:20 protein:lipid (w/w) ratio and detergent was subsequently removed by several rounds of incubation with biobeads. Proteoliposomes were then collected by ultracentrifugation and resuspended in 58 mM Na_2HPO_4, 17 mM

NaH_2PO_4, 68 mM NaCl. Protein incorporation was estimated by SDS-PAGE and proteoliposomes were diluted to obtain a final protein concentration of approximately 1 mg mL^{-1}.

Once an immune nanobody library was obtained, target-specific nanobodies were enriched by phage display and individual clones obtained after enrichment were validated in an ELISA. In both these steps, biotinylated CerS proteins were used to enable the non-denaturing immobilisation of these target proteins. I designed CerS5 and CerS6 constructs with a C-terminal Avi-tag (GLNDIFEAQKIEWHE) preceding the TEV cleavage site and linked to the protein by a flexible 'GGGS' linker (CerS – GGGS – Avi – TEV cleavage site – Twin-Strep – 10x His). These constructs correspond to the best truncations identified in previous sections (CerS5: Met1 to Val352; CerS6: Met1 to Asp350). However, after tag cleavage with TEV protease, the final proteins now contain a C-terminal GGGS linker followed by an Avi-tag.

CerS5 and CerS6 proteins were biotinylated *in vitro*. The Avi-tagged constructs were expressed in Expi GnTI- and purified as previously described for CerS5-c017 and CerS6-c025, respectively, using Strep-Tactin XT resin. However, upon elution from the resin with 100 mM D-biotin, the Avi-tagged protein samples were biotinylated overnight as described in Section 2.4.3. This enabled the simultaneous C-terminal tag cleavage and *in vitro* biotinylation of the protein samples. It should be noted that a much lower concentration of D-biotin could be used, but as elution from the Strep-Tactin XT resin is carried out using 100 mM D-biotin, this high concentration was preserved for *in vitro* biotinylation.

Biotinylation of the final purified samples was confirmed by denaturing intact protein LC-ESI-MS. Both CerS5 and CerS6 purified samples were completely biotinylated, as identified by the presence of an expected mass shift of ~226 Da (Fig. 5.10A,B)

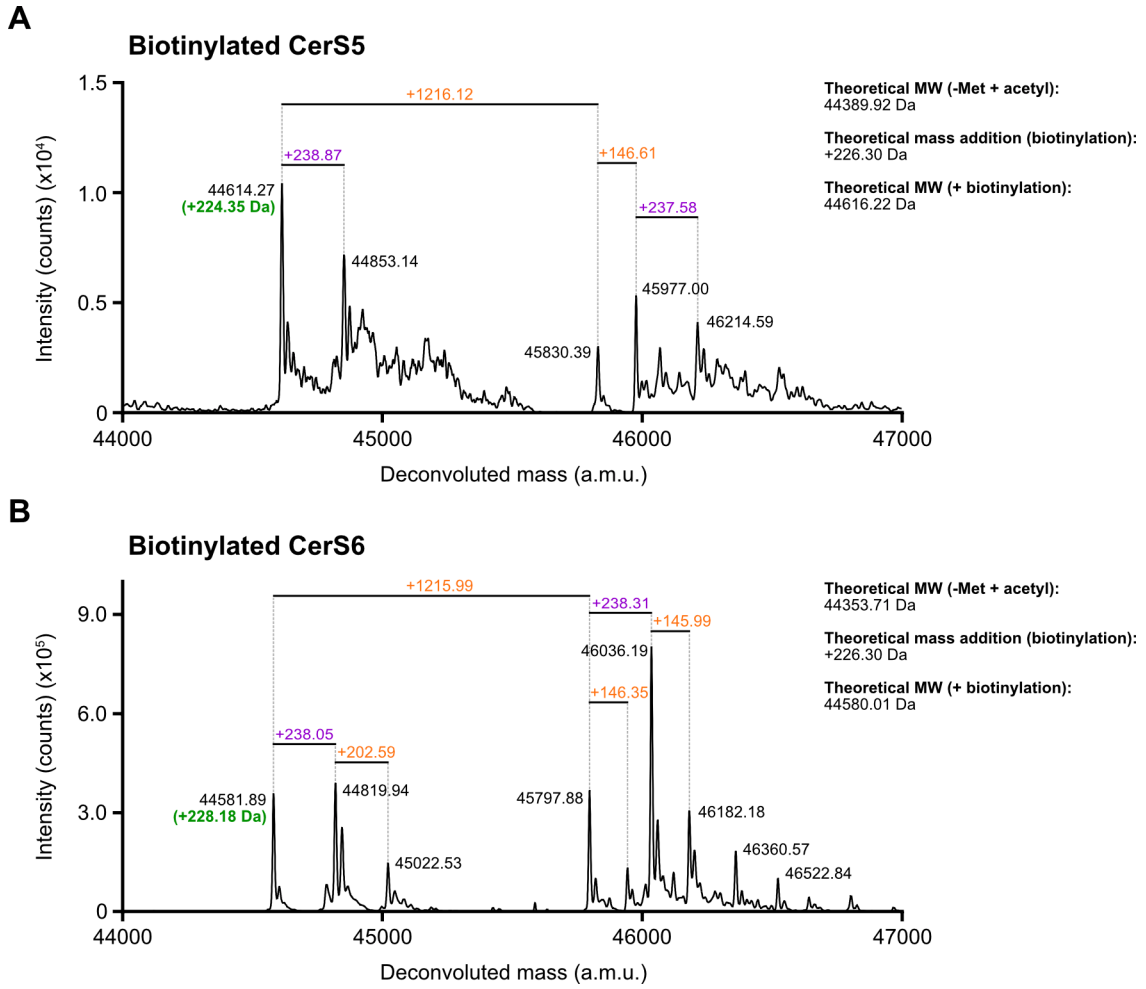


Figure 5.10. *In vitro* biotinylation of avi-tagged CerS5 and CerS6. Avi-tagged CerS5 and CerS6 proteins were expressed in Expi GnTI- cells, purified and biotinylated *in vitro* during purification. Denaturing intact LC-ESI-MS spectra of the final purified protein samples confirms successful biotinylation (green text; theoretical mass addition: +226.30 Da) of **A.** CerS5 and **B.** CerS6. Other post-translational modifications can be observed. For the assignment of these deconvoluted mass peaks, see Fig. 5.4B (CerS5) or Fig. 5.6 (CerS6).

5.3.3 Validation of unique binders generated

Phage-display libraries were screened by the Nanobody Service Facility, and target-specific binders were enriched. Then, 2 x 96-well plates of enriched clones were selected for ELISA screening and, upon completion of this ELISA screening step, 96 clones were selected and sequenced. At this stage, I received the isolated plasmids from the 96-well sequencing plates.

Sequencing of ELISA-positive clones identified 35 unique nanobodies raised against CerS5 and 42 unique nanobodies against CerS6 (Appendix 3). At this stage, all nanobodies were selected for additional biophysical validation, as slight differences in the nanobody sequence, such as single amino acid variations in the CDRs, could potentially lead to significant changes in binding affinity.

Overall, a general strategy was devised for the validation of the nanobodies received. This involved the small-scale expression in *E. coli* and purification of all unique binders (Section 5.3.3.1), followed by validation using BLI and determination of off-rates, to rank and prioritise binders for structural studies (Section 5.3.3.2).

5.3.3.1 Small-scale expression and purification of unique nanobodies

The nanobodies received were in the phage display vector pDX [299], which is an adapted version of the commonly used phage display vector pMESy4 [298] (Fig. A.7, Appendix 5). In this vector, the nanobodies are fused to a PelB leader peptide on their N-terminus (which directs the protein to the bacterial periplasm) and to 6x His and EPEA (C-tag) tags on their C-terminus (PelB – Nanobody – 6x His – EPEA). These tags are followed by an amber stop codon (TAG), which is in turn followed by an HA-tag and the filamentous phage fd gene III, coding for the gene-3-protein (G3P). G3P is a coat protein of filamentous bacteriophages and mediates the initial contact between the bacteriophage and the F pilus on the bacterial host surface [300]. During phage display, the nanobodies are expressed fused to the N-terminus of the G3P protein [301-303] (in an amber codon suppression strain) so that phages displaying nanobodies that bind to the immobilised biotinylated protein of interest can be enriched in each selection round [298].

However, for downstream work requiring the recombinant expression of the selected nanobodies, including for structural studies, the nanobodies must be expressed as a C-terminal fusion to only the purification tags, while excluding the G3P coat protein. This can be achieved by using cells that do not suppress the amber stop codon, such as the *E. coli* strain WK6. In this way, translation terminates at the amber stop codon, and the nanobodies can be successfully expressed as C-terminal 6x His – EPEA fusions.

Nanobodies were expressed in 50 mL scale in WK6 cells and purified using a single IMAC step using Ni^{2+} Sepharose FF resin as described in Section 2.4.4.1. Typically, 24 nanobodies were purified in a day. In short, the purification procedure involves an initial periplasmic extraction step, followed by an IMAC step and finally buffer exchange of the purified nanobodies to ensure no buffer mismatch between the nanobodies and the purified biotinylated CerS proteins in downstream BLI experiments. To facilitate parallel purification of a large number of nanobodies, Ni^{2+} resin washing, nanobody elution and buffer exchange were carried out in 96-well format. This approach resulted in the expression and purification of 32 out of 35 CerS5 binders and 36 out of 42 CerS6 binders. SDS-PAGE analysis from representative purification runs of 11 nanobodies is shown in Fig. 5.11.

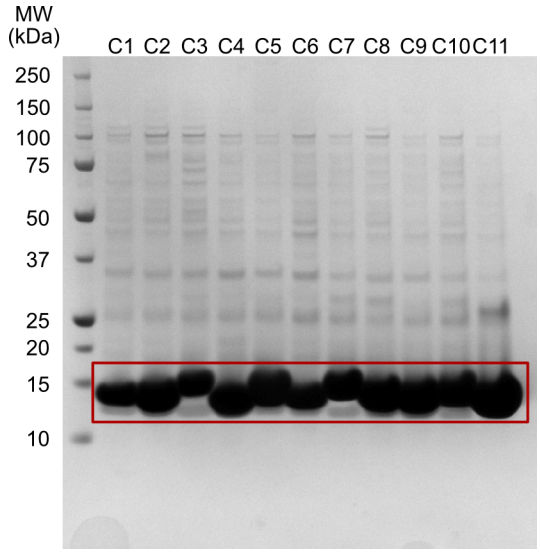


Figure 5.11. Representative purified CerS5 nanobodies from 50 mL expressions. SDS-PAGE analysis of the purified nanobodies C1-C11 (raised against CerS5 – Appendix 3). Samples loaded correspond to the final purified sample obtained after a single IMAC step using Ni^{2+} Sepharose FF resin and buffer exchanged into 20 mM HEPES pH 7.5, 200 mM NaCl, 0.1 % (w/v) OGNG / 0.01 % (w/v) CHS for BLI experiments. Bands corresponding to the nanobodies are highlighted by a red box.

5.3.3.2 Nanobody screening by Bio-Layer Interferometry

Validation of nanobody binding to the target proteins was carried out using Bio-Layer Interferometry (BLI). BLI experiments were conducted using biotinylated CerS5 and CerS6 immobilised on Strepavidin-SA biosensors (for a detailed description, see Section 2.4.5). Nanobodies were allowed to associate for 120 s, which was followed by a 120 s dissociation step. The nanobodies were screened at 500 nM concentration, which is above the useful K_D range for the intended structural studies, therefore ensuring all binders can be identified. Using this setup, 21 out of the 32 purified nanobodies raised against CerS5 were confirmed to bind the target protein. Likewise, 32 out of 36 purified CerS6 nanobodies could be successfully validated. Representative BLI sensograms are displayed in Fig. 5.12A,B.

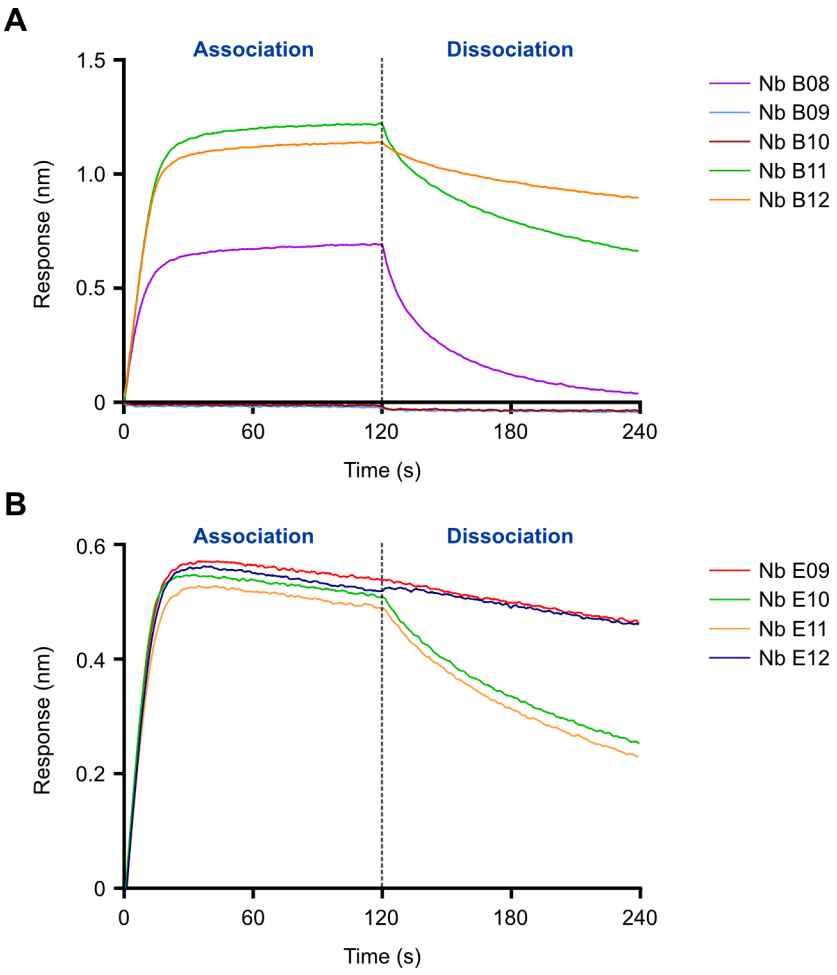


Figure 5.12. Validation of purified nanobodies by Bio-Layer Interferometry (BLI). Binding experiments using immobilised CerS5 or CerS6 evaluated all 77 nanobodies and, importantly, determined the off-rate of the interactions. After an initial baseline is recorded in buffer (not shown), the nanobodies (500 nM) are then allowed to associate for 120s. This is then followed by a dissociation step (120s) in buffer without nanobody, enabling nanobody ranking based on their off-rate. **A.** Representative BLI sensograms of 5 selected CerS5 nanobodies (B08 - B12). Of these 5 nanobodies, 3 successfully bind CerS5 (B08, B11, B12) and dissociate at different rates. **B.** Representative BLI sensograms of 4 selected CerS6 nanobodies (E09 - E12). All 4 nanobodies bind, although nanobodies E09 and E12 exhibit slower off-rates.

In addition, this approach provides valuable insights into the dissociation rates (k_{off}) of these interactions. Using this strategy, CerS5 and CerS6 binders were ranked [304] so that the nanobodies with the slowest off-rates were selected as initial candidates for structural studies (Appendix 4).

5.4 Characterisation of CerS6 using Cryo-EM

5.4.1 Optimisation of CerS6 grids

The initial stages of cryo-EM studies involved the optimisation of grid preparation. This optimisation included the evaluation of suitable grid parameters, protein concentration, detergent choice and grid blotting time, and the initial parameters selected were advised by Dr Gamma Chi from the IMP group of the CMD, who also supervised me during grid preparation.

As described in Section 5.2.6, LMNG mixed with CHS was used as the detergent formulation during CerS6 extraction and purification. LMNG has a very low CMC (~0.001 %) and is therefore compatible with maintaining a low detergent concentration in the final concentrated sample. As a result, and due to its mild character, LMNG is a popular detergent choice for cryo-EM [305]. In addition to LMNG / CHS, CerS6 grids were also prepared using protein extracted in LMNG / CHS and later exchanged into GDN. GDN is a synthetic, low-CMC substitute for the mild detergent digitonin [306] and has recently become one of the most popular detergent choices for cryo-EM applications because of its mildness and low CMC [305]. This parallel screening of two different surfactants increased the parameters in initial CerS6 screening for faster follow-up optimisation.

For grids prepared using protein in LMNG / CHS, CerS6 was purified in 0.002 % (w/v) LMNG / 0.0002 % (w/v) CHS and concentrated to 1.2 or 2.8 mg mL^{-1}. 3 µL of the purified protein sample were applied onto 300 gold mesh Quantifoil holey carbon R 1.2/1.3 µm grids. Grids were prepared as detailed in Section 2.4.7, and blot times were sampled between 1-3 s. Grids were screened starting with the lower protein concentration (1.2 mg mL^{-1}), revealing a good sample distribution in the holes (Fig. 5.13A). A small characterisation dataset (278 movies) was collected on this sample using a Glacios microscope (Thermo Fisher Scientific)

operating at 200 keV at the Oxford Particle Imaging Centre (OPIC) (Oxford, UK). 2D classification of particles picked using the blob picker in CryoSPARC [307] clearly reveals the presence of a dimeric species, as expected. However, this analysis also demonstrated that the particles display a severe orientation bias, as mostly top / bottom views can be observed (Fig. 5.13B).

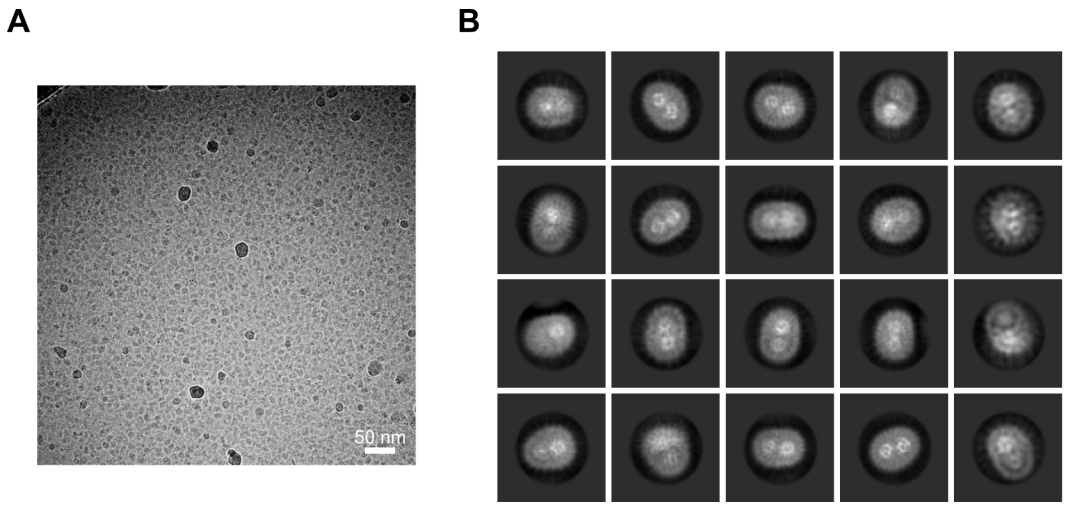


Figure 5.13. Initial CerS6 grid optimisation - protein purified using LMNG / CHS. A. Representative Glacios micrograph acquired using a Falcon III direct electron detector (92,000 x magnification; approximately -2.4 µm defocus). **B.** Representative 2D classes from a small Glacios linear collection (278 micrographs). A dimeric assembly can be clearly observed, but the particles display severe orientation bias.

The other CerS6 samples in GDN were prepared essentially as described for the purification using LMNG / CHS, with the exception that, after extraction in LMNG / CHS and binding to the Strep-Tactin XT resin, CerS6 was exchanged into GDN by substituting LMNG / CHS by GDN in all subsequent steps (for a detailed description, see Section 2.1.3.3). Interestingly, two peaks can be observed in the SEC elution profile of CerS6 purified in GDN (Fig. 5.14A). These peaks are partially overlapping and both contain pure CerS6 protein as de-

tected by SDS-PAGE (Fig. 5.14B). Therefore, these likely correspond to two distinct oligomeric states. The peak corresponding to the putative dimer was pooled, concentrated and used to make grids.

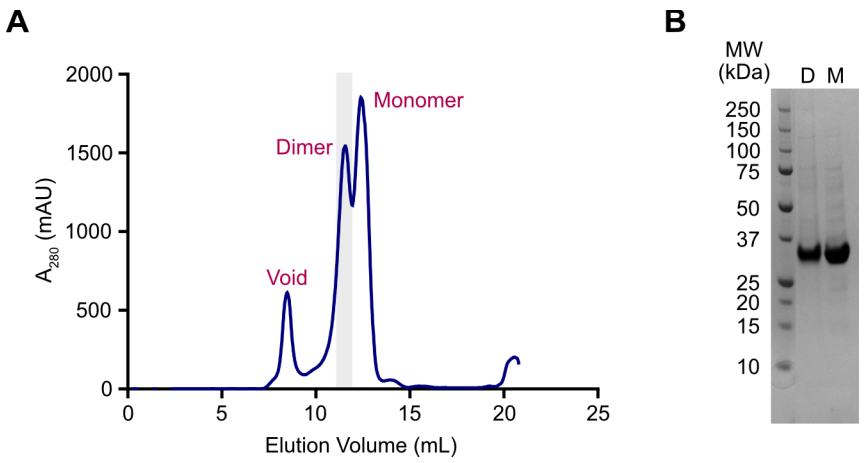


Figure 5.14. Purification of CerS6-c025 using GDN. A. Elution profile from a Superdex 200 Increase 10/300 SEC column. Two non-void peaks were obtained, likely corresponding to dimeric and monomeric species, respectively. The putative dimer peak (shaded in grey) was pooled and concentrated to make grids. **B.** SDS-PAGE analysis of the final purified dimeric (D) and monomeric (M) samples.

For the CerS6 GDN sample, I used 200 gold mesh Quantifoil holey carbon R 1.2/1.3 µm grids. GDN has been previously observed in the IMP group of the SGC to typically result in a sharp ice gradient where the protein is excluded from thinner ice. As a result, one would collect data on slightly thicker grid squares (compared to LMNG), which have fewer holes available for collection. Therefore, the larger grid squares in 200 mesh grids simplify setting up data collection at a later stage. Grids were prepared as described for the LMNG / CHS sample, with the following changes: protein concentration was 5 mg mL^{-1}, blot time was sampled between 4-7 s, and a 5 s waiting time delay was included after sample application, prior to blotting. Screening of these grids revealed that, as expected, particle distribution was heavily dependent on ice thickness, with particles being excluded from thinner ice and thus clustering closer to the edge of the hole (Fig. 5.15A,B). A small characterisation dataset

was collected (560 movies) at OPIC (Oxford, UK), using a Glacios microscope operating at 200 keV. 2D classification of the particles again reveals the presence of a dimeric species, and, encouragingly, the use of GDN led to an improved orientation distribution (Fig. 5.15C). Two small protrusions from the micelle could be identified in some 2D classes, likely corresponding to the homeodomains predicted to be located between TM1 and TM2 of each CerS6 monomer (Fig. 5.15C). Overall, when CerS6 is purified in GDN, the severe particle orientation bias observed for the LMNG / CHS sample was avoided, and the dimeric state could be preserved. Therefore, protein in GDN was selected for further work. However, given the partial overlap of the monomeric and dimeric SEC peaks, this dimeric sample is likely contaminated with CerS6 monomers, which could make downstream data processing difficult.

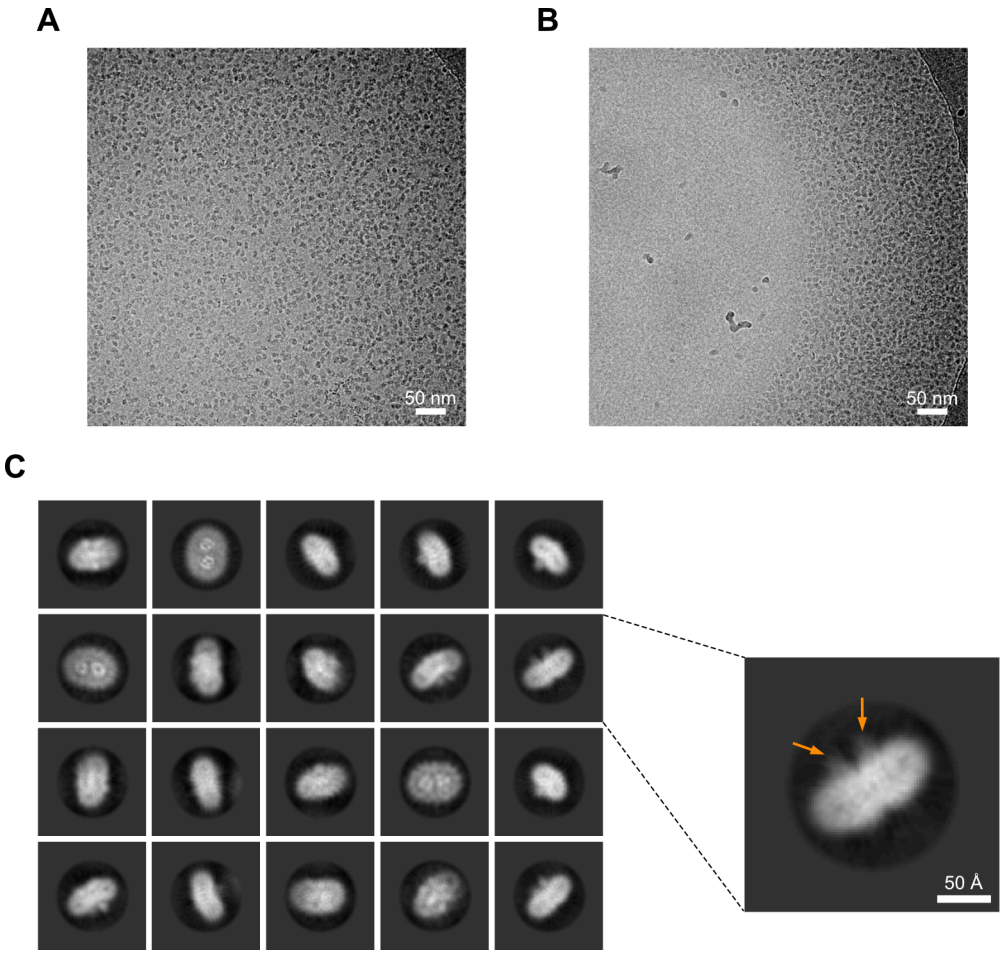


Figure 5.15. Initial CerS6 grid optimisation - protein purified using GDN. Representative Glacios micrographs acquired using a Falcon III direct electron detector (92,000 x magnification; approximately -2.4 µm defocus) show that **A.** good particle distribution throughout the holes can be obtained, and **B.** protein is excluded when ice is too thin. **C.** Representative 2D classes from a small Glacios linear collection (560 micrographs) show that protein in GDN displays good orientation distribution. A zoomed-in view is shown for a class corresponding to a side view of the protein, where the two homeodomains of the CerS6 dimer are labelled with orange arrows.

5.4.2 CerS6 Krios dataset

In order to evaluate the feasibility of this CerS6 dimer for cryo-EM structural determination, an initial dataset was acquired for the CerS6 GDN sample on a Titan Krios operating at 300 keV, at the electron Bio-Imaging Centre (eBIC, Didcot, UK). At this stage, it was not clear if the CerS6 dimer displayed real C2 symmetry, which could provide a significant advantage in downstream image processing (particularly given the small size of a CerS6 monomer).

Importantly, if that is the case, nanobody complexes could be pursued as a strategy towards obtaining high-resolution maps. In contrast, if these dimers display considerable heterogeneity away from the dimer interface, I could attempt to determine the structure of monomeric CerS6 instead. However, due to the smaller molecular weight of the CerS6-c025 monomer (42 kDa), this would likely require larger fiducial tools such as megabodies [295] to further increase particle size. Therefore, the central question that this experimental dataset aimed to answer was whether CerS6 formed a rigid C2-symmetrical dimer.

To this end, a dataset of 5,972 movies was collected. Full details on data collection and processing are described in Section 2.4.9, and statistics are displayed in Table 5.3. Processing of this dataset (Fig. 5.16A-D; workflow shown in Fig. 5.16A) revealed a mixed population of monomeric and dimeric CerS6 particles (approximately 50 % monomer, 50 % dimer) (Fig. 5.16A). However, the separation of monomer and dimer particles during processing and classification was not straightforward due to the lack of features protruding from the micelle. As expected, it became evident that this dataset would not yield high-resolution maps, with 3D refinement of dimer particles achieving a resolution of 6.9 Å (FSC threshold = 0.143) (Fig. 5.16A). Nonetheless, the CerS6 dimer appears to display C2 symmetry at this resolution. Also, maps obtained after masked local refinement of the protein region allowed the approximate placement of two copies of the CerS6 Alphafold2 model (Fig. 5.16D). This analysis encouraged follow-up work with CerS6-nanobody complexes in attempts to obtain high-resolution structures of this CerS6 dimer.

Table 5.3 Cryo-EM data collection and processing statistics

Data collection	
Microscope	Titan Krios (eBIC, UK)
Detector	K3
Voltage (keV)	300
Magnification	130,000
Collection mode	Counting (super-resolution, binned 2x)
Electron exposure (e- A^{-2})	50.9
No. frames	50
Pixel size	0.653 Å
Defocus range (μm; steps)	-0.8 to -2.4 (0.3)
No. movies	5,972
Data processing	
Initial number of particles	1,207,719
No. particles in 3D classification	208,113
No. particles in final reconstruction	78,859
Symmetry	C2
Map resolution (Å; FSC threshold = 0.143)	6.9
Map sharpening B-factor (A^2)	-493

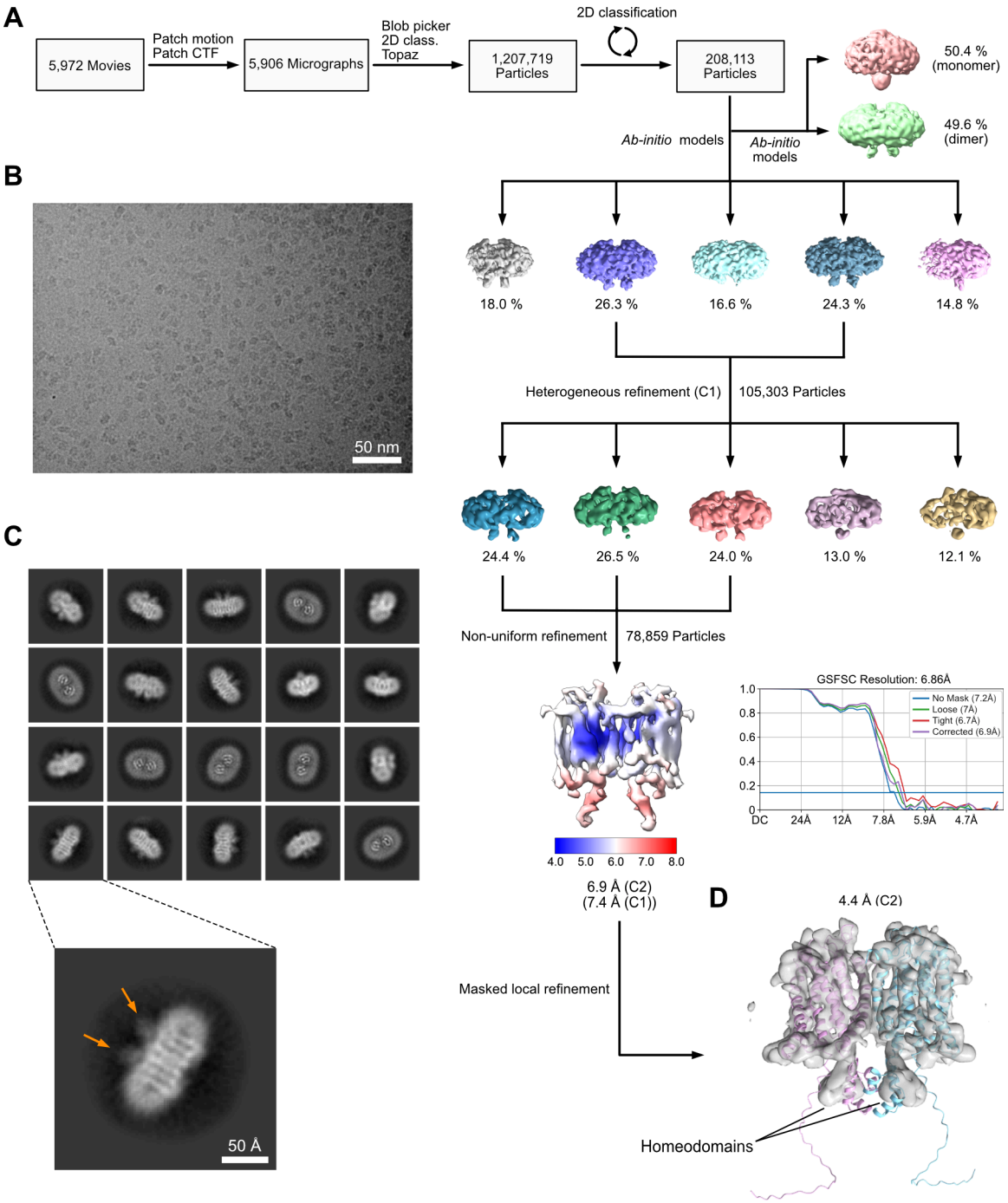


Figure 5.16. CerS6 cryo-EM dataset. A. EM processing workflow for CerS6 dataset. **B.** Representative micrographs acquired using a K3 direct electron detector (Gatan) (130,000 x magnification; approximately -2.0 µm defocus). **C.** Representative 2D classes. A zoomed-in view is shown for a class corresponding to a side view of the protein, where the two homeodomains of the CerS6 dimer are labelled with orange arrows. **D.** EM maps obtained after masked local refinement of the protein region, fitted with Alphafold2 models of two CerS6 monomers (pink / cyan).

5.4.3 Purification of CerS6-nanobody complexes

Next, nanobodies were employed as fiducial markers to overcome the small particle size and the lack of distinguishable features protruding from the micelle. Having established a priority list based on off-rate ranking of CerS6 nanobodies (Appendix 4), nanobodies with the slowest off-rates were selected for scaled-up production. Out of these, nanobodies D01 and E09 were initially selected to be used in cryo-EM efforts. These were purified from the periplasmic extract using Ni^{2+} Sepharose FF resin and buffer exchanged to remove the imidazole. Then, the nanobodies were concentrated and further purified by SEC. Peak fractions were pooled and concentrated to 20-25 mg mL^{-1}.

CerS6-nanobody complexes were reconstituted from recombinant protein components by adding the appropriate nanobody at a 2:1 molar ratio to the concentrated CerS6 sample in GDN and incubating the mixture on ice for 1h prior to further purification by SEC. The SEC elution profile again displays two partially overlapping CerS6 monomer and dimer peaks (Fig. 5.17A,B), with an additional peak at a higher elution volume representing uncomplexed nanobody. SDS-PAGE analysis of the dimer peaks confirmed the CerS6-nanobody complexes (Fig. 5.17C). As before, the CerS6 dimer peak was pooled and concentrated to 5 mg mL^{-1}.

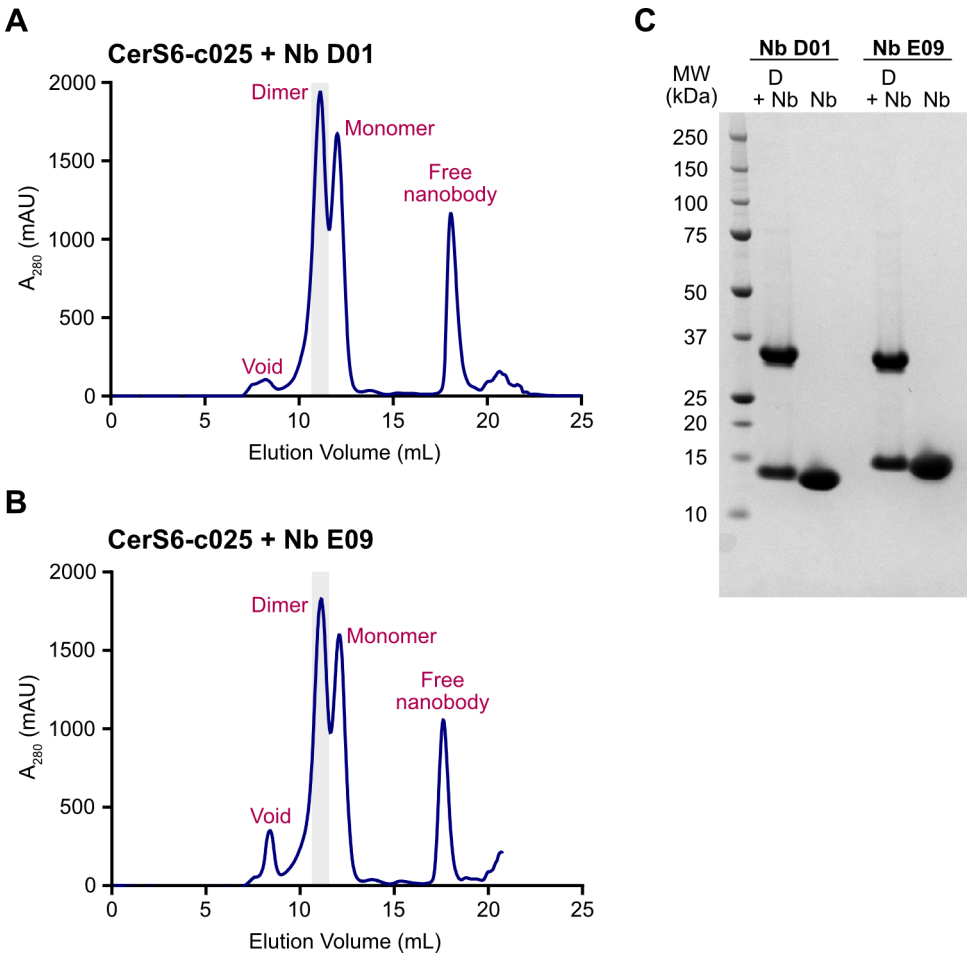


Figure 5.17. Purification of CerS6 – nanobody complexes. A,B. Elution profiles of (A) CerS6-c025 – nanobody D01 and (B) CerS6-c025 – nanobody E09 complexes from a Superdex 200 Increase 10/300 SEC column. Three included peaks are observed. Two of these peaks likely correspond to monomeric and dimeric CerS6 species. The third peak, which elutes at a higher volume, corresponds to free nanobody. The dimer peaks (shaded in grey) were pooled and concentrated to make grids. **C.** SDS-PAGE analysis of the final purified CerS6 dimer + nanobody (*D + Nb*) samples, revealing two bands: a higher molecular weight band corresponding to CerS6 and a lower molecular weight band corresponding to the nanobody. This confirmed the successful purification of the CerS6 – nanobody complexes. Nanobody-only samples (*Nb*) were included for comparison.

5.4.4 CerS6-nanobody complexes: grid screening

The two purified CerS6-nanobody complexes (CerS6-NbD01 and CerS6-NbE09) were applied onto 200 gold mesh Quantifoil holey carbon R 1.2/1.3 µm grids as described in Section 5.4.1 for the GDN sample, and blotting times between 5-8 s were sampled. The particle

distribution (Fig. 5.18A,B) again heavily varied with ice thickness, with particles being excluded from thinner ice. Small characterisation datasets were collected at OPIC (Oxford, UK) on a Glacios microscope operating at 200 keV, revealing that the CerS6 dimer configuration is preserved in both CerS6-NbD01 and CerS6-NbE09 complexes and that one nanobody molecule is bound to each CerS6 monomer of the dimer (Fig. 5.18C-F). In addition, both nanobodies (D01 and E09) bind to the ER lumen side of the CerS6 protein (Fig. 5.18C-F), as concluded on the observation that the cytoplasmic homeodomains are located on the opposite side of the protein. Reference-free 2D classification of extracted particles demonstrates that the bound nanobodies are clearly visible in the 2D classes (Fig. 5.18C,D), suggesting these could be useful for particle alignment. 3D reconstructions obtained from these small characterisation datasets are displayed in Fig. 5.18E,F, and a simplified EM processing workflow is shown in Fig.5.18G.

In summary, the preliminary cryo-EM characterisation revealed 2:2 CerS6:nanobody complexes, where the nanobodies bind on the ER lumen side of CerS6. In these complexes, the binders protrude away from the micelle, and thus could likely assist in particle alignment. These results warrant the collection of large datasets in the future, to assess whether high-resolution structures can be obtained.

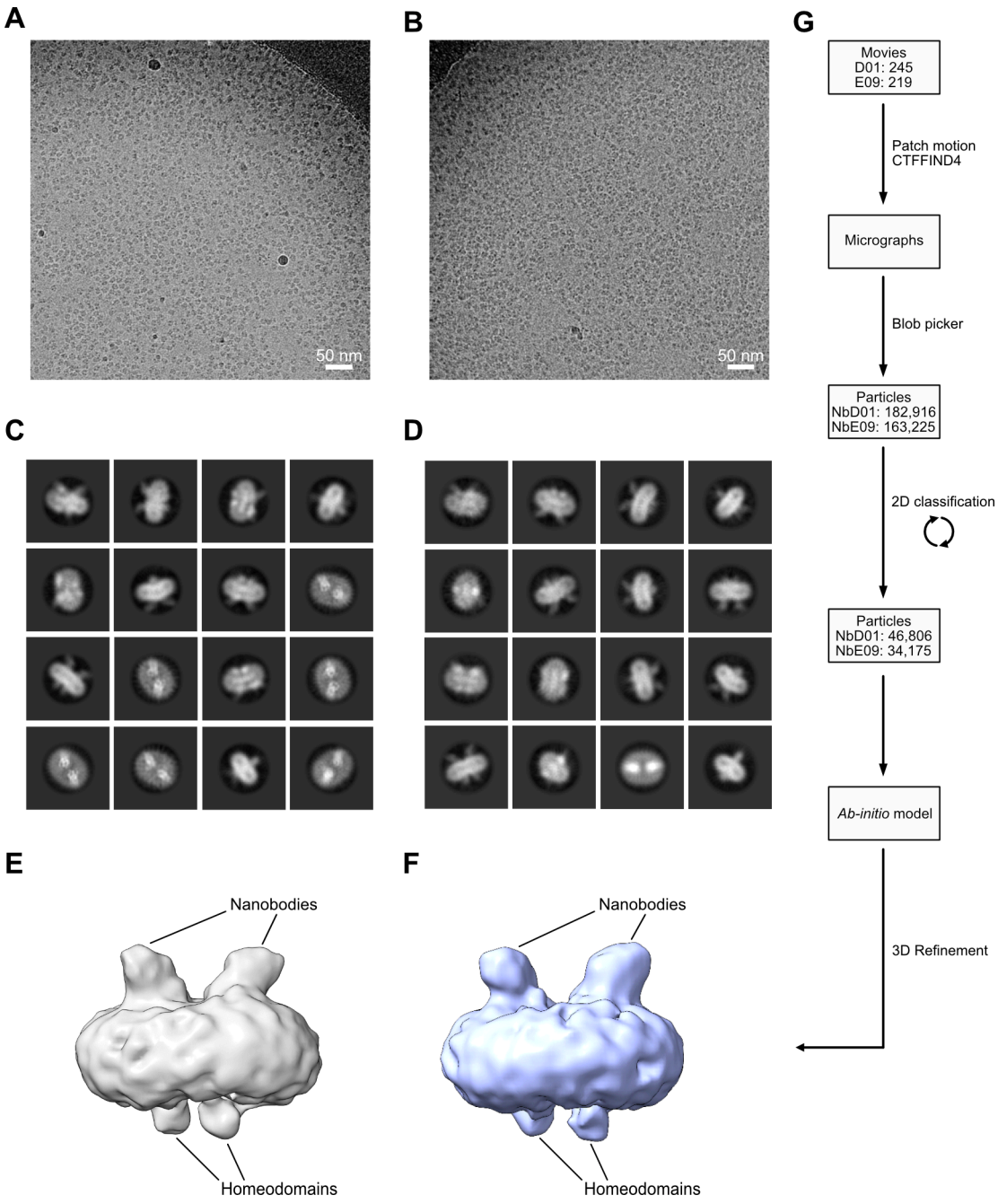


Figure 5.18. Initial evaluation of CerS6 – nanobody complexes by cryo-EM. A,B. Representative micrographs acquired using a Falcon III direct electron detector (92,000 x magnification; approximately -2.4 µm defocus) for (A) CerS6-NbD01 and (B) CerS6-NbE09. **C,D.** Representative 2D classes from small Glacios linear collection datasets. For both (C) CerS6-NbD01 and (D) CerS6-NbE09, two nanobodies can be observed bound to a CerS6 dimer. In both cases, the nanobodies bind on the ER lumen side. **E,F.** 3D reconstructions of the (E) CerS6-NbD01 and (F) CerS6-NbE09 complexes, where both the nanobodies and the homeodomains are clearly visible. **G.** Simplified EM processing workflow for the CerS6-nanobody complexes, leading to the 3D reconstructions shown in panels E and F.

5.5 Discussion

Human ceramide synthases, particularly CerS6 and potentially CerS5 and CerS1, are attractive targets for novel therapies for metabolic diseases [161]. As a result, there is great interest in obtaining structural insights into their catalytic mechanism, the basis of substrate specificity, and the binding mode of currently available non-selective inhibitors (described in Section 1.2.5). Here, I established the expression and purification of human CerS5 and CerS6, generated nanobody binders against these two proteins, and optimised conditions for the structural determination of CerS6-nanobody complexes by cryo-EM. Additionally, I identified well-expressing CerS1 constructs, which could be similarly pursued for structural studies.

While optimising purification conditions, I confirmed earlier reports that CerS5 and CerS6 are glycosylated in their N-terminus [139, 141] and generated glycosylation-deficient mutants to help crystallisation attempts. Crystallisation trials were carried out with EndoH-treated CerS5 and CerS6, as well as with the CerS6 glycosylation mutant, but no crystals could be obtained.

At this stage, I generated alpaca nanobodies against CerS5 and CerS6 through a collaboration with the Nanobody Service Facility (University of Zurich) and validated ELISA-positive clones using BLI. This strategy enabled prioritising successful binders for structural studies based on their off-rates. Overall, this work produced useful tools that could be used as crystallisation chaperones or as fiducial markers for cryo-EM.

I then focused on solving structures of CerS6 by cryo-EM, considering that none of my crystallisation attempts had progressed. However, the best CerS6 construct, CerS6-c025, has a monomeric molecular weight of ~42 kDa, which would pose a challenge for particle alignment. Therefore, I aimed to confirm the hypothesis from the literature that CerS6 can form

homodimers [140] and identified a dimeric population of CerS6 when expressed in Expi293F cells. Subsequently, I prepared grids with CerS6 purified in LMNG / CHS or GDN and found that, in both detergents, CerS6 dimers are present. Grids prepared with CerS6 in LMNG / CHS displayed severe particle orientation bias, and therefore I focused on using protein purified in GDN. I then collected a CerS6 dataset, to evaluate whether this CerS6 dimer was rigid and displayed real C2 symmetry and, therefore, whether I should continue to pursue this route in order to obtain high-resolution maps. This dataset clearly demonstrated that a mixed population of monomers and dimers is present, which is unsurprising given the partial overlap of monomer and dimer peaks in the SEC elution profile. Also, the lack of distinguishable protein structural features protruding from the micelle (except for the cytoplasmic homeodomain, connected by flexible linkers) made it challenging to efficiently separate monomer and dimer particles at the picking and classification steps. In addition, this lack of features likely creates difficulties during particle alignment, limiting the resulting map resolution to 6.9 Å. However, even at this resolution, the CerS6 dimer clearly displays the expected C2 symmetry. Therefore, the focus was shifted towards characterising the CerS6-nanobody complexes, as these could overcome the small size of the CerS6 particles. In particular, the use of nanobodies as fiducial markers would provide additional features outside the micelle, which could assist with particle alignment and improve monomer / dimer classification.

Two nanobodies with slow off-rates were initially selected, and small characterisation datasets collected for both CerS6-nanobody complexes revealed that two nanobodies bind per CerS6 dimer and that both these nanobodies bind on the ER side of the protein. This was somewhat unexpected given the high propensity of camelid nanobodies to recognise the active site of enzymes [308]. Importantly, the observation that the CerS6 nanobodies bind

on the ER lumen side means that these nanobodies do not directly block access to the substrate-binding site, located on the cytoplasmic side, and entertain the possibility that these nanobodies could also be used as tools to aid structure determination of substrate or inhibitor-bound complexes.

Now that the grid conditions for these two CerS6-nanobody complexes were optimised and testing datasets collected, the stage is now set for collecting large Krios datasets to determine a high-resolution structure. I have also prepared grids for the CerS6-NbE09 complex in the presence of the non-selective inhibitor Fumonisin B1, and will likewise collect a dataset for this sample in the future. As I found that this inhibitor stabilises the protein, it is possible that it could lock the protein in a particular conformation, thus reducing flexibility and potentially improving the chances of obtaining a higher resolution structure. Furthermore, there is great interest in obtaining structural insights into the binding mode of inhibitors in order to guide structure-based drug design efforts.

6 Conclusion

Obtaining mechanistic and three-dimensional structural information on target proteins at the molecular and atomic level is of the utmost significance to enable the rational design of novel therapeutics. The work presented here comprises a combination of X-ray crystallography, mass spectrometry, electron cryomicroscopy and activity assays, employed to study three human integral membrane proteins implicated in either disease or drug interactions. The first, MFSD10, is proposed to act as a multidrug efflux pump for NSAIDs and other unknown compounds. The other two proteins are central components of fatty acid and lipid metabolism and belong to two enzyme families that, together, determine the properties of newly synthesised sphingolipids. Importantly, members of both these enzyme families are considered to be attractive drug targets for the treatment of several metabolic disorders.

6.1 MFSD10

MFSD10 is an MFS transporter belonging to the drug:H^+ antiporter family 1 (DHA1) and was previously demonstrated to confer resistance to apoptosis mediated by NSAIDs [56]. I solved a crystal structure of human MFSD10 in an outward-facing partially occluded conformation with a bound citrate ion from the crystallisation condition. To the best of my knowledge, this was the first structure of a eukaryotic member of the MFS DHA1 family and the first structure of an atypical SLC. However, during the course of my thesis preparation, structures were published for two other atypical SLCs of the MFS, human UNC93B, which does not act as a transporter and instead regulates trafficking of toll-like receptors [309], and mouse and chicken MFSD2A, a sodium-dependent lysophosphatidylcholine symporter [310, 311]. Both these atypical SLCs are not related to MFSD10 [66].

Although the presence of the citrate ion prevented successful co-crystallisation with NSAID substrates, it revealed the presence of a polar carboxylate-binding pocket in the C-domain side of the central cavity. As known MFSD10 substrates (e.g. indomethacin and diclofenac) also contain a carboxylate group, I proposed that this polar pocket could likewise participate in recognition of the substrate carboxylates. In addition, TM4 and TM5 of MFSD10 form a long hydrophobic face which suggests it could enable non-specific hydrophobic interactions with the aromatic moieties of known NSAID substrates. Such non-specific hydrophobic interactions are a known hallmark of MFS-MDR transporters, as they must recognise multiple structurally diverse substrates [26].

Furthermore, I employed a pyranine-based fluorescence assay to track proton movement during the transport cycle of MFSD10 and observed that diclofenac transport resulted in proton antiport, thus confirming the proposed function of MFSD10 as an NSAID:H^+ antiporter. The structure also suggests putative proton-coupling roles for negatively charged residues on TM1 (Asp35) and TM7 (Glu290). Notably, acidic residues at either or both of these positions have been proposed to be involved in proton transport in other (but not all) DHA1 transporters [48, 50, 51, 246-250]. In the MFSD10 structure, Asp35 is not directly accessible from the central cavity in the outward-facing partially occluded conformation but could become accessible in other conformations to participate in proton translocation. Importantly, the structure suggests that Glu290 could be the site of initial protonation, as it is the only acidic residue accessible from the central cavity in this outward-facing conformation and is located closer to the extracellular entrance, at the bottom of a solvent-filled cavity. Glu290 forms a salt bridge with Arg408 (TM11), which participates in substrate recognition in the carboxylate pocket. Proton-coupled MFS antiporters couple substrate release with proton binding such that substrate binding typically prevents protonation of the

transporter, and the MFSD10 structure suggests that the interaction with Arg408 could modulate the pKa of Glu290, preventing its premature protonation prior to substrate release.

While the results obtained provide a useful starting point to study NSAID transport by MFSD10, the details of its transport mechanism and broader cellular function remain unknown. Although the crystal structure with bound citrate suggests how NSAIDs could be recognised in the substrate-binding site, their exact binding mode has not yet been determined, as co-crystallisation and soaking experiments were unsuccessful. Future work could focus on optimising the poorly diffracting crystals obtained in the absence of citrate, which could enable obtaining substrate co-structures. In addition, docking studies could also provide useful insights into the binding of NSAIDs to MFSD10.

Likewise, the proposed roles of Glu290 and likely Asp35 in the proton-coupling mechanism need to be evaluated experimentally and, to this end, I have designed Glu290Gln and Asp35Asn mutants, which could next be evaluated in the pyranine assay described in this work and in other functional assays currently being established through collaborations with partners within the ReSOLUTE consortium (see below). Similarly, residues in the substrate-binding site were also mutated (e.g. all the residues contacting the citrate ion in the structure and those forming the hydrophobic TM4/TM5 face of the substrate-binding site), and the resulting mutant proteins will be included in future studies to confirm their proposed role in substrate binding.

NSAID use is associated with gastrointestinal, cardiovascular and renal complications [62, 63, 65]. Available evidence suggests that MFSD10 could act as an NSAID efflux pump, similarly to multidrug resistance transporters in bacteria, which remove toxic antibiotics from the cell. Therefore, it could play a role in protecting against the adverse effects of this

class of drugs. However, the underlying mechanism of this transporter, as well as its subcellular localisation, are poorly understood. In addition, the endogenous substrates of MFSD10 are currently unknown. With this in mind, I solved the first crystal structure of human MFSD10 and developed additional tools and methodology for the expression, purification and reconstitution into proteoliposomes for the functional characterisation of this transporter. In line with the open-access mission of the SGC, I released the coordinates and structure factors onto the PDB in May 2020 and disseminated the structural and preliminary functional characterisation as a Target Enabling Package (TEP) in July 2021 [312]. TEPs are peer-reviewed open access tool kits (methods, structures, assays, chemical matter etc.) on potential drug targets [313] as a mandated deliverable of the SGC mission. These aim to provide reagents and knowledge to the community to accelerate chemical probe generation and drug discovery [313, 314].

In addition, MFSD10 has been nominated as a priority target in the ReSOLUTE consortium, an EU-funded IMI programme that aims to develop tools to enhance knowledge on the function of human SLCs, evaluate them as drug targets, and de-orphanise orphan transporters [315]. Within the ReSOLUTE remit, I have provided proteoliposomes for the functional characterisation of MFSD10 and have obtained and validated nanobody binders (in collaboration with Dr Saša Štefanić Nanobody Service Facility, University of Zurich). I cloned these nanobodies into an expression vector containing 3x FLAG tags (provided by Dr Eric Geertsma, Max Planck Institute of Molecular Cell Biology and Genetics) and supplied them to ReSOLUTE partners to enable studying the subcellular localisation of MFSD10 across different cell types. These efforts to determine the subcellular localisation of MFSD10 are underway and could provide crucial insights into its cellular function. Importantly, if a pH gradient exists across the identified membrane, the knowledge that MFSD10 acts as a proton-coupled antiporter can be used to understand the directionality of transport. Furthermore,

the nanobodies generated can also be used as crystallisation chaperones or as fiducial markers for cryo-EM. These approaches could enable obtaining structures of MFSD10 in complex with NSAID substrates (in the absence of citrate) and could perhaps lead to trapping of additional conformations of this transporter, which would greatly enhance our understanding of its transport mechanism. Lastly, as purified MFSD10 displays a dark-green colour (described in Section 3.2.1), a collaboration with Professor Douglas Kell's group (University of Liverpool) has been established to identify the presumed responsible compound using high-resolution ultra-high performance liquid chromatography tandem mass spectrometry (UHPLC-MS/MS). If successful, this ligand identification approach could suggest potential endogenous substrates of MFSD10, which would be invaluable in efforts towards its deorphanisation.

6.2 ELOVL7

ELO elongases (ELOVLs in mammals) catalyse the Claisen-like condensation of an acyl-CoA and a malonyl-CoA-derived two-carbon unit, yielding a 3-keto acyl-CoA with two additional carbons in the acyl chain. Despite their biological importance in determining the acyl chain specificity in the fatty acid elongation cycle, very little was known about their reaction mechanism.

In Chapter 4 of this thesis, I described the characterisation of the catalytic mechanism of human ELOVL7 using denaturing intact protein mass spectrometry as a follow-up study to the crystal structure previously obtained in the group by Dr Laiyin Nie and Dr Ashley Pike.

The crystal structure of ELOVL7 revealed the presence of a covalently bound 3-keto acyl-CoA product analogue. While it is difficult to rationalise exactly how this compound became

covalently attached to the enzyme to form what appears to be a dead-end complex, its presence provided useful insights into the architecture of the substrate-binding and active sites. Importantly, this 3-keto acyl-CoA is bound in a very narrow tunnel in the centre of the protein, which excluded a ternary complex mechanism for the reaction catalysed by ELOVL7, as there is not sufficient space for both substrates to bind simultaneously. Instead, the structure suggested a ping-pong type mechanism, where the substrates would react sequentially, with the first product leaving before the second substrate can bind.

In a ping-pong type mechanism, a covalent intermediate is expected, and therefore I developed a mass spectrometry-based assay to assess whether reaction with the first substrate (acyl-CoA) would lead to acyl-enzyme intermediate formation. Using this method, I demonstrated that reaction with acyl-CoA substrates leads to the covalent addition of the substrate acyl chain to the protein. Then, to confirm that the observed covalent addition of the substrate acyl chain corresponds to a real reaction intermediate, I optimised the experiment to enable the sequential reaction of ELOVL7 with an acyl-CoA, followed by malonyl-CoA. As expected, reaction of the acyl-enzyme intermediate with malonyl-CoA pushed the reaction to completion, leading to the loss of the covalent intermediate deconvoluted mass peak. In addition, the structure suggested that His150 of the conserved ^{147}HxxHH^{151} motif would be placed adjacent to the substrate acyl-CoA thioester. Strikingly, using the mass spectrometry approach described above, I observed that no covalent acyl-enzyme intermediate could be formed when the His150Ala mutant was incubated with an acyl-CoA substrate, indicating that His150 likely acts as the nucleophile during the first step of the reaction (transacylation). This was further supported by the finding that no co-purified covalent adducts were observed in the purified His150Ala mutant.

Taken together, my mass spectrometry results and the structural information obtained by Dr Laiyin Nie and Dr Ashley Pike revealed that catalysis by ELOVL7 proceeds via an unusual

acyl-imidazole intermediate formed at His150 of the ^{147}HxxHH^{151} motif. The proposed ping-pong mechanism is described in Section 4.3.1 of this work. We propose that ELO elongases use the conserved HxxHH motif to generate a His-His-Asp catalytic triad, where the second histidine of the HxxHH motif acts as the nucleophile. Notably, even though the histidines of the conserved HxxHH motif are known to be essential to the activity of ELO elongases [194, 195, 197], their precise role in the catalytic mechanism remained elusive until now. In addition, to the best of my knowledge, this is the first time a ping-pong mechanism has been proposed for ELOs.

While this work enabled expanding our understanding of the mechanism of ELO elongases, the details of the condensation step remain unclear. Despite modelling carried out by Dr Ashley Pike (described in Section 4.3.1) leading to the identification of a putative malonyl-CoA-binding pocket, we still do not know precisely how the malonyl moiety would be bound in this pocket. As a result, the interactions promoting decarboxylation and condensation remain unclear, and further work is required to attempt to obtain a malonyl-CoA-bound structure. While >90 % of the purified wild-type protein is covalently modified with the 3-keto acyl-CoA, the His150Ala mutant does not display any covalent modifications. His150 acts as the nucleophile during the first step of the reaction and therefore would not participate in malonyl-CoA binding, as it would be covalently linked to the acyl chain at the beginning of the condensation step. Therefore, I propose that co-crystallisation attempts using the His150Ala mutant and malonyl-CoA could potentially enable obtaining a malonyl-CoA-bound structure.

In addition, we do not know whether ELO elongases require prior acyl-enzyme intermediate formation to react with malonyl-CoA. Some KCSs, including FAE1-type KCSs [186, 282, 283], ensure that decarboxylation can only occur after the acyl-enzyme intermediate has been formed, thus preventing the unproductive depletion of malonyl-CoA. For ELOVL7,

this question could be answered in the future by detecting whether the incubation of malonyl-CoA with the wild-type enzyme in the absence of acyl-CoA would lead to CO_2 release. Alternatively, the His150Ala mutant could be used to fully exclude the possibility of prior reaction with any acyl-CoA substrates that could potentially be co-purified with the protein. CO_2 production could be specifically monitored by using [1–^{14}C]malonyl-CoA and carrying out the reaction in vacutainer tubes containing a filter paper saturated with an alkali metal hydroxide such as potassium hydroxide [316], which traps the $^{14}CO_2$ as potassium bicarbonate. Trapped radiolabelled bicarbonate can be measured by scintillation counting.

Furthermore, as described above, the formation of the dead-end complex observed in the crystal structure is hard to rationalise. Interestingly, this dead-end complex is also formed upon recombinant expression of ELOVL7 in Expi293F cells [213]. We speculated that, during recombinant overexpression, ELOVL7 could have elongated a 2,3-*trans*-enoyl-CoA, an intermediate metabolite of the fatty acid elongation cycle in the ER, and subsequently reacted with the α,β double bond through His150 (described in Section 4.3.2) [213]. It is possible that, during ELOVL7 overexpression, the fatty acid elongation cycle could become unbalanced, leading to the accumulation of this metabolite. However, it is unknown whether 2,3-*trans*-enoyl-CoAs are capable of acting as substrates of ELOVLs and whether this reaction could result in covalent attachment of the product. Therefore, the proposed mechanism for the formation of this dead-end complex remains speculative. To investigate this in more detail, one could determine whether 2,3-*trans*-enoyl-CoAs are elongated by ELOVL7 *in vitro*. This could be achieved by detection of the proposed elongated 3-keto-4,5-*trans*-enoyl-CoA products by mass spectrometry. Alternatively, radiolabelled malonyl-CoA could be used to enable product detection after the separation of the substrate and elongated species by thin layer chromatography. At the same time, the formation of any covalent adducts at

the end of this reaction could be evaluated using the denaturing intact protein mass spectrometry setup established in this work. Unfortunately, I could not commercially source 2,3-*trans*-enoyl-CoAs and therefore did not experimentally validate this hypothesis. However, even if this addition is possible, it is unlikely to be physiologically relevant in conditions where ELOVL7 is not overexpressed. Therefore, while the mechanism of formation of this adduct is intriguing, the main value of this covalently attached compound lies in stabilising the enzyme for crystallisation and revealing the architecture and interactions in the substrate-binding and active sites.

Lastly, although the histidine corresponding to His181 in *Dictyostelium discoideum* EloA is required for activity [194] and His181 of ELOVL7 is located in the active site and covalently attached to the product analogue in the structure, its precise role in catalysis remains unclear. His181 would be ideally placed to act as a proton donor to the leaving CoA during the trans-acylation step, but I found that the His181Ala mutation did not affect covalent intermediate formation. Moreover, His181 lines the side pocket proposed by Dr Ashley Pike to function as a malonyl-CoA binding / decarboxylation pocket. Therefore, it is possible that His181 could play a role in promoting malonyl-CoA decarboxylation by stabilising enolate formation. However, additional evidence is required to assign this role to His181. Obtaining a malonyl-CoA-bound structure, as described above, would be a valuable starting point.

ELOVL elongases have been proposed as potential targets for the development of novel therapies against X-linked adrenoleukodystrophy [211], cancer [209, 210], trypanosome infections [191] and type 2 diabetes [208]. Importantly, the acyl chain composition of sphingolipids plays a central role in metabolic diseases, including obesity-associated insulin resistance [164, 317]. Hence, the selective inhibition of ELOVL isoenzymes could enable regulating the availability of selected acyl-CoAs, thus affecting their incorporation into ceramides by ceramide synthases (the third topic of this thesis work). The unusual active

site architecture and catalytic mechanism of the ELO family of enzymes could therefore suggest novel strategies for the development of inhibitors.

6.3 Ceramide synthases

Ceramide synthases catalyse the N-acylation of sphingoid bases in the *de novo* and salvage pathways of ceramide biosynthesis, thus determining the acyl chain properties of synthesised ceramides and complex sphingolipids. Selective inhibition of the synthesis of C16 ceramides, particularly those produced by CerS6, has been proposed as a promising avenue to develop novel therapies for several metabolic disorders [88, 161, 164, 318] (described in Section 1.2.4). In addition, CerS5 and CerS1 have also been proposed as drug targets [161]. Therefore, understanding the mechanism of this family of enzymes, as well as obtaining structural information into the recognition of available non-selective inhibitors, is of significant interest, as it could enable the rational design of isoform-selective inhibitors.

To enable structural studies of human ceramide synthases, I initially established the large-scale recombinant expression and purification of human CerS5 and CerS6. To the best of my knowledge, the large-scale purification of human ceramide synthases has not been described in the literature. I then attempted to crystallise these proteins, but crystallisation was unsuccessful. Therefore, using the purified protein as antigen, I generated nanobody binders through a collaboration with the Nanobody Service Facility of the University of Zurich. These could serve a dual purpose as crystallisation chaperones and fiducial markers in electron cryomicroscopy. At present, I have not yet had the opportunity to attempt crystallisation with these nanobodies but have initiated efforts to solve the structures of CerS6-nanobody complexes by cryo-EM (discussed below).

The molecular weight of the monomeric 'best' CerS6 construct (CerS6-c025: Met1 to Asp350) is only 42 kDa, and therefore I initially validated literature reports that CerS6 could form homodimers [140]. Then, while nanobody validation was ongoing, I optimised grid conditions for CerS6 and collected a dataset on a Titan Krios microscope to evaluate the feasibility of pursuing structural determination of CerS6 by cryo-EM. Processing of this dataset revealed the presence of an approximately equal amount of monomer and dimer particles. These different particle classes could be somewhat separated during image processing, but the efficiency of this classification was perturbed by the lack of distinguishable protein features outside the detergent belt. As a result, I followed up on these observations by making grids of CerS6-nanobody complexes, which are now optimised and awaiting the collection of large datasets. Importantly, the use of nanobodies as fiducial markers could assist with particle alignment and monomer / dimer classification. Notably, nanobodies have enabled solving high-resolution cryo-EM structures of small membrane proteins, such as the 50 kDa proton-coupled folate transporter (PCFT) [319]. However, it is possible that the use of nanobodies may not provide sufficient features to enable accurate particle alignment and high-resolution reconstructions of CerS6. If this is the case, two additional approaches will be evaluated. Firstly, the nanobody binders validated and ranked in this study will be grafted onto a large rigid *E. coli* K12 glucosidase YgjK scaffold to produce ~100 kDa megabodies. This approach was recently developed in Jan Steyaert's group [295] and has enabled high-resolution cryo-EM reconstructions of small membrane proteins, such as the 58 kDa Hedgehog acyltransferase [320]. A second strategy that would enable increasing the size of the nanobody fiducial markers consists of using a nanobody-binding Fab fragment, which was recently made available by Kaspar Locher's group [296]. Overall, recent developments in the field have generated several useful tools to enhance the size of fiducial markers for cryo-

EM. These will be rapidly employed if high-resolution reconstructions cannot be obtained using nanobodies alone.

Furthermore, denaturing intact protein mass spectrometry analysis of purified CerS5 and CerS6 samples detected palmitoylation at one site in both proteins. While this may correspond to a post-translational modification of a surface-exposed cysteine, it must be noted that palmitoyl-CoA is the preferred substrate of both these enzymes. Therefore, if catalysis by ceramide synthases proceeds via a ping-pong type mechanism (as described for ELOVL7), it is possible that this mass addition could correspond to a trapped acyl-enzyme intermediate. Thus, in future studies to determine the site of palmitoyl addition, both these proteins will be enzymatically digested with trypsin and / or other proteases, and palmitoyl-containing peptides will be analysed by LC-MS/MS using the established facility in the CMD. In addition, purified CerS6 (but not CerS5) also contains another covalent adduct leading to a mass shift of +264 Da. Although the identity of this modification is unclear, this mass addition matches that expected for a C18:1 acyl chain. As a result, this adduct could also correspond to an unexpected covalent intermediate. Strikingly, this additional mass shift is only observed relative to the unmodified enzyme and not relative to the palmitoylated species. Thus, the mutual exclusivity between these two adducts suggests attachment at the same site. While CerS6 is not primarily responsible for C18 ceramide generation, there are reports in the literature stating that mouse CerS6 can react with C18 CoAs to generate C18 ceramides [139]. To clarify whether this covalent adduct corresponds to C18:1-CoA and determine where it is formed, the analysis of CerS6 peptides will also focus on this adduct. Overall, the identification of the site of these covalent attachments could either indicate that catalysis by CerS enzymes proceeds via a covalent acyl-enzyme intermediate or would otherwise likely identify the site of palmitoylation to be a surface-exposed cysteine. In the latter

case, this cysteine could then be mutated to remove the deconvoluted mass peak corresponding to palmitoylation. This, in turn, would enable the hypothesis of covalent intermediate formation to be investigated using the same experimental setup developed for ELOVL7, without the presence of contaminating signal from cysteine palmitoylation.

Lastly, as CerS1 has been proposed as a therapeutic target for the treatment of insulin resistance in skeletal muscle [159, 161, 168, 169], I designed constructs of human CerS1 (22 % sequence identity to CerS5; 25 % sequence identity to CerS6) and obtained initial data showing successful recombinant expression in mammalian cells. These results could constitute the starting point of future efforts to produce milligram amounts of CerS1 for structural or functional studies, as described in this work for CerS5 and CerS6.

6.4 Closing remarks

In this thesis, I used a combination of X-ray crystallography, cryo-EM, mass spectrometry and biochemical assays to study the mechanisms of three families of human membrane proteins.

Traditionally, X-ray crystallography constituted the main structural biology technique employed to obtain high-resolution three-dimensional information on proteins of interest, which is not only valuable for the rational design of novel therapeutics but also provides unique information that enables formulating hypotheses to describe the molecular mechanisms of those proteins. More recently, cryo-EM has emerged as a powerful alternative technique to obtain high-resolution structures and can now be routinely employed to solve structures of small membrane proteins, particularly where binders are available to act as fiducial

markers. Critically, this enables the structural determination of proteins that cannot be purified in sufficient amounts for crystallisation or that fail to crystallise (which is often the case for membrane proteins) and can reveal multiple conformations of dynamic proteins.

However, these structural techniques provide limited information on protein function, and additional data from biochemical, biophysical and computational approaches are required to complement the structural data and obtain a more complete understanding. As an example, in this work, I developed a denaturing intact protein mass spectrometry-based assay to characterise the mechanism of the human fatty acid elongase ELOVL7. Together with the structural data generated by others in the group, this enabled proposing a ping-pong mechanism for this enzyme and identified the nucleophile in the first step of the reaction. This work is a clear example of how different techniques can complement one another, leading to a more detailed understanding of protein function. I envisage that a similar approach combining cryo-EM and mass spectrometry could be employed to study the mechanism of human ceramide synthases. Likewise, other mass spectrometry-based methods can also be used in hybrid approaches to complement structural data. Native mass spectrometry enables identifying non-covalent interactions and can be used to study, for example, protein-lipid interactions, thus complementing other structural biology techniques. Importantly, lipid interactions are major but often underappreciated determinants of membrane protein function and oligomerisation. Moreover, when studying large multi-subunit complexes by cryo-EM, cross-linking mass spectrometry approaches enable identifying the relative positioning of the different subunits. Lastly, mass spectrometry can also provide useful quality control during the optimisation steps of structural biology projects, namely by identifying post-translational modifications, unexpected adducts and truncations, as evidenced in this thesis work.

Overall, as different structural and biophysical techniques become more accessible to biologists, such integrative cross-discipline approaches will likely become more popular, enabling faster project progress and a deeper understanding of protein function.

Appendices

Appendix 1 – Full-length protein sequences

Table A.1. Full-length amino acid sequences of the proteins studied in Chapters 3-5.

Protein	Uniprot ID	Full length sequence	Length (amino acids)
MFSD10	Q14728	MGWGGGGGCTPRPPIHQQPPERRVVTVVFLGLLLDLLAFTLLL PLLPGLLESHGRAHDPLYGSWQGGVDWFATAIGMPVEKRYNS VLFGGLIGSAFSVLQFLCAPLTGATSDCLGRRPVMLLCLMGVAT SYAVWATSRSFAAFLASRLIGGISKGNVSLSTAIVADLGSPLARS QGMAVIGVAFSLGFTLGPMLGASLPLEMAPWFALLFAASDLLFI FCFLPETLPLEKRAPSIALGFRDAADLLSPLALLRFSAVARGQDPP SGDRLSSLRRLGLVYFLYLFLFSGLEYTLSFLTHQRFQFSSLQQGK MFFLIGLTMATIQGAYARRIHPGGEVAAVKRALLLLVPAFLLIGW GRSLPVLGLGLLLYSFAAAVVVPCLSSVVAGYGSPGQKGTVMGT LRSLGALARAAGPLVAASVYWLAGAQACFTTWSGLFLLPFFLLQ KLSYPAQTLKAE	455
ELOVL7	A1L3X0	MAFSDLTSRTVHLYDNWIKDADPRVEDWLLMSSPLPQTILLGF YVYFVTSLGPKLMENRKPFELKKAMITYNFFIVLFSVYMCYEFVM SGWGIGYSFRCDIVDYSRSPTALRMARTCWLYYFSKFIELLDTIFF VLRKKNSQVTFLHVFHHTIMPWTWWFGVKFAAGGLGTFHALL NTAVHVVMYSYYGLSALGPAYQKYLWWKKYLTSLQLVQFVIVA IHISQFFFMEDCKYQFPVFACIIMSYSFMFLLLFLHFWYRAYTKG QRLPKTVKNGTCKNKDN	281
CerS1	P27544	MAAAGPAAGPTGPEPMPSYAQLVQRGWGSALAAARGCTDC GWGLARRGLAEHAHLAPPELLLLALGALGWTALRSAATARLFR PLAKRCCLQPRDAAKMPESAWKFLFYLGSWSYSAYLLFGTDYPF FHDPPSVFYDWTPGMAVPRDIAAAYLLQGSFYGHSIYATLYMD TWRKDSVVMLLHHVVTLILIVSSYAFRYHNVGILVLFLHDISDVQ LEFTKLNIYFKSRGGSYHRLHALAADLGCLSFGFSWFWFRLYWF PLKVLYATSHCSLRTVPDIPFYFFFNALLLLLTLMNLYWFLYIVAFA AKVLTGQVHELKDLREYDTAEAQSLKPSKAEKPLRNGLVKDKRF	350
CerS5	Q8N5B7	MATAAQGPLSLLWGWLWSERFWLPENVSWADLEGPADGYG YPRGRHILSVFPLAAGIFFVRLLFERFIAKPCALCIGIEDSGPYQAQ PNAILEKVFISITKYPDKKRLEGLSKQLDWNVRKIQCWFRHRRN QDKPPTLTKFCESMWRFTFYLCIFCYGIRFLWSSPWFWDIRQC WHNYPFQPLSSGLYHYYIMELAFYWSLMFSQFTDIKRKDFLIMF VHHLVTIGLISFSYINNMVRVGTLIMCLHDVSDFLLEAAKLANYA KYQRLCDTLFVIFSAVFMVTRLGIYPFWILNTTLFESWEIIGPYAS WWLLNGLLLTLQLLHVIWSYLIARIALKALIRGKVSKDDRSDVES SSEEEDVTTCTKSPCDSSSSNGANRVNGHMGGSYWAEE	392
CerS6	Q6ZMG9	MAGILAWFWNERFWLPHNVTWADLKNTEEATFPQAEDLYLA FPLAFCIFMVRLIFERFVAKPCAIALNIQANGPQIAPPNAILEKVF TAITKHPDEKRLEGLSKQLDWDVRSIQRWFRQRRNQEKPSTLT RFCESMWRFSFYLYVFTYGVRFLKKTPWLWNTRHCWYNYPYQ PLTTDLHYYYILELSFYWSLMFSQFTDIKRKDFGIMFLHHLVSIFLI TFSYVNNMARVGTLVLCLHDSADALLEAAKMANYAKFQKMCD LLFVMFAVVFITTRLGIFPLWVLNTTLFESWEIVGPYPSWWVFN LLLLLVQGLNCFWSYLIVKIACKAVSRGKAGKWNPLHVSKDDRS DIESSSDEEDSEPPGKNPHTATTTNGTSGTNGYLLTGSCSMDD	392

Appendix 2 – CerS1 construct screening

Table A.2. Expression levels of available CerS1 constructs. Constructs marked with * were designed in the SGC prior to my arrival. The expression level was judged by the intensity of the CerS1 SDS-PAGE bands obtained after a 3 mL purification using DDM / CHS as the extraction detergent (described in Section 5.2.2)

Construct ID	N-terminus	C-terminus	Expression system	Purification tags	Expression level
CerS1-c004*	T38	E337	*Sf*9	6x His	Low
CerS1-c005*	T38	L331	*Sf*9	6x His	Low
CerS1-c006*	T38	L320	*Sf*9	6x His	Low
CerS1-c010*	T38	E337	Mammalian	10x His - FLAG	Low
CerS1-c011*	T38	L331	Mammalian	10x His - FLAG	Low
CerS1-c012*	T38	L320	Mammalian	10x His - FLAG	None
CerS1-c013*	M16	L331	Mammalian	10x His - FLAG	None
CerS1-c014*	A30	L331	Mammalian	10x His - FLAG	None
CerS1-c016	A7	L320	Mammalian	10 x His - Twin Strep	Medium
CerS1-c017	A7	L331	Mammalian	10 x His - Twin Strep	Low
CerS1-c018	A7	E337	Mammalian	10 x His - Twin Strep	Low
CerS1-c019	A7	Q313	Mammalian	10 x His - Twin Strep	Medium
CerS1-c020	M16	L320	Mammalian	10 x His - Twin Strep	Medium
CerS1-c021	M16	L331	Mammalian	10 x His - Twin Strep	Medium
CerS1-c022	M16	E337	Mammalian	10 x His - Twin Strep	None
CerS1-c023	M16	Q313	Mammalian	10 x His - Twin Strep	Medium
CerS1-c024	T38	L320	Mammalian	10 x His - Twin Strep	Low
CerS1-c025	T38	L331	Mammalian	10 x His - Twin Strep	None
CerS1-c026	T38	E337	Mammalian	10 x His - Twin Strep	None
CerS1-c027	T38	Q313	Mammalian	10 x His - Twin Strep	Low

Appendix 3 – Nanobodies generated against CerS5 and CerS6

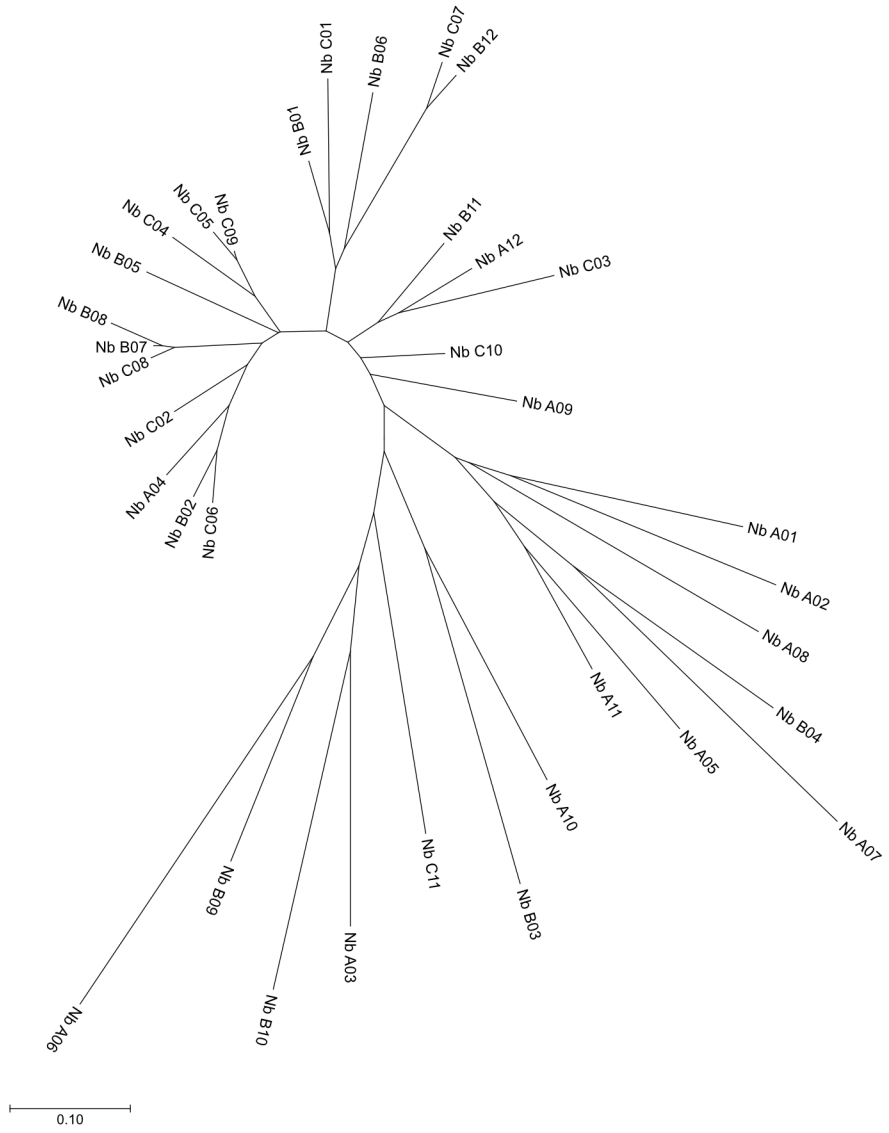


Figure A.1. CerS5 nanobodies. Radial phylogenetic tree of unique nanobodies raised against CerS5. Unique nanobodies were aligned using PROMALS3D [251] and the phylogenetic tree was generated in MEGA [321], using the maximum likelihood method.

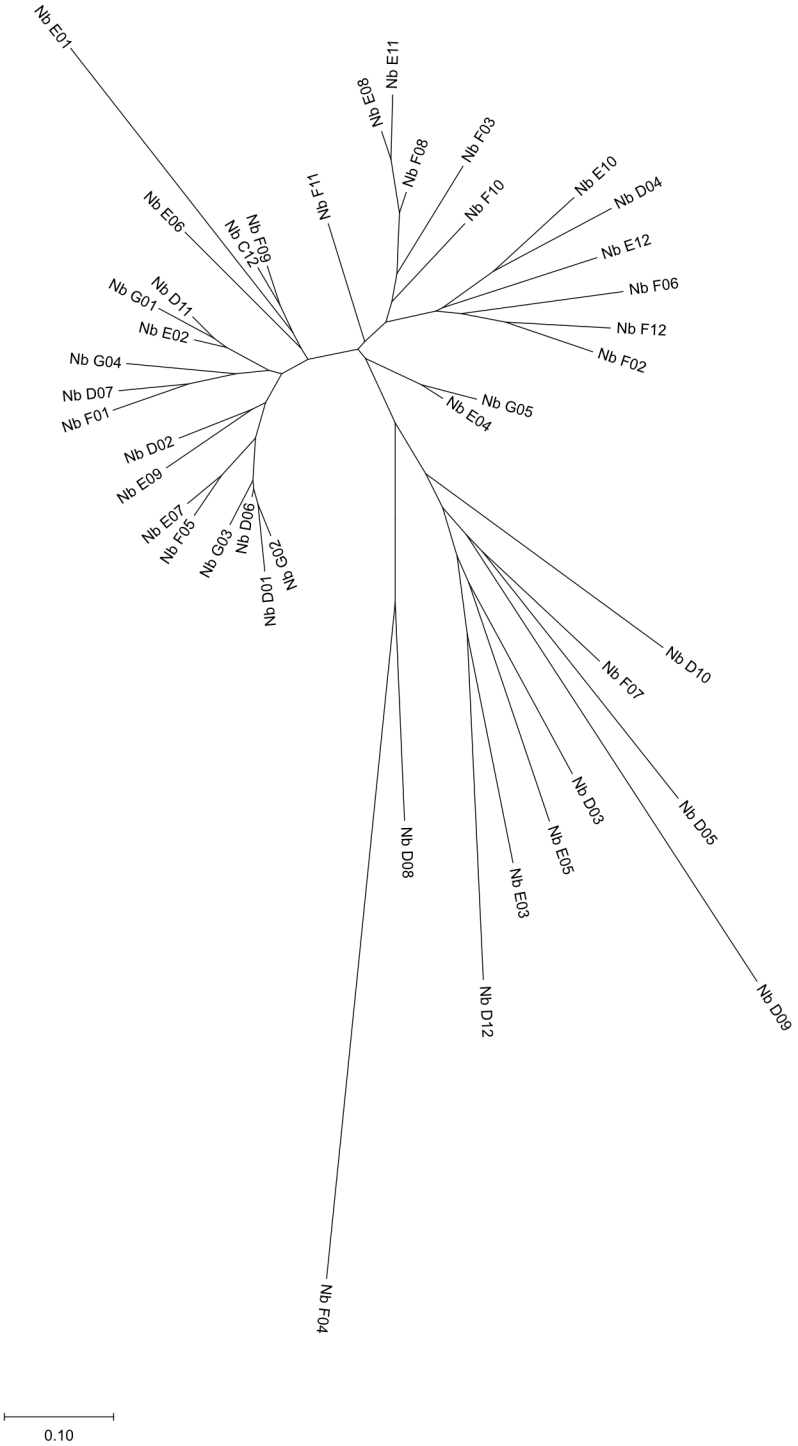


Figure A.2. CerS6 nanobodies. Radial phylogenetic tree of unique nanobodies raised against CerS6. Unique nanobodies were aligned using PROMALS3D [251] and the phylogenetic tree was generated in MEGA [321], using the maximum likelihood method.

Appendix 4 – Nanobody screening results

Table A.3 Calculated off rates of CerS5 nanobodies.

Nanobody Well ID	k_{off} (s^{-1})
B12	4.83×10^{-2}
B1	5.95×10^{-2}
A1	6.12×10^{-2}
B11	7.02×10^{-2}
C3	7.91×10^{-2}
A11	8.02×10^{-2}
A4	8.07×10^{-2}
B8	8.19×10^{-2}
A12	8.51×10^{-2}
C7	8.65×10^{-2}
C2	8.68×10^{-2}
C9	8.88×10^{-2}
C10	9.18×10^{-2}
B6	9.44×10^{-2}
C8	1.01×10^{-1}
C5	1.03×10^{-1}
C4	1.04×10^{-1}
B7	1.09×10^{-1}
C6	1.11×10^{-1}
B2	1.18×10^{-1}
B5	2.29×10^{-1}
A3	No binding detected
A5	No binding detected
A7	No binding detected
A9	No binding detected
A10	No binding detected
B3	No binding detected
B4	No binding detected
B9	No binding detected
B10	No binding detected
C1	No binding detected
C11	No binding detected
A2	No expression
A6	No expression
A8	No expression

Table A.4 Calculated off rates of CerS6 nanobodies.

Nanobody well ID	k_{off} (s^{-1})
D1	Positive slope during dissoc.*
E12	3.75×10^{-5}
E6	9.83×10^{-5}
F3	5.67×10^{-4}
G1	8.71×10^{-4}
D3	1.83×10^{-3}
E9	2.56×10^{-3}
F2	2.83×10^{-3}
E5	3.40×10^{-3}
G2	4.13×10^{-3}
F10	4.84×10^{-3}
D11	5.64×10^{-3}
E8	5.88×10^{-3}
G5	6.22×10^{-3}
F5	7.45×10^{-3}
G4	8.94×10^{-3}
E7	9.75×10^{-3}
F1	1.02×10^{-2}
D4	1.05×10^{-2}
F11	1.07×10^{-2}
D7	1.08×10^{-2}
F6	1.10×10^{-2}
G3	1.24×10^{-2}
E11	1.32×10^{-2}
E10	1.37×10^{-2}
C12	1.50×10^{-2}
F9	1.52×10^{-2}
E2	1.56×10^{-2}
E4	1.57×10^{-2}
D6	1.75×10^{-2}
D2	2.20×10^{-2}
F12	2.37×10^{-2}
D12	No binding detected
E1	No binding detected
E3	No binding detected
F7	No binding detected
D5	No expression
D8	No expression
D9	No expression
D10	No expression
F4	No expression
F8	No expression

* During the dissociation step, the sensogram obtained for nanobody D01 displayed a positive slope. Therefore, no k_{off} could be determined. However, this observation suggested that this nanobody displayed slow dissociation kinetics.

Appendix 5 – Vector Maps

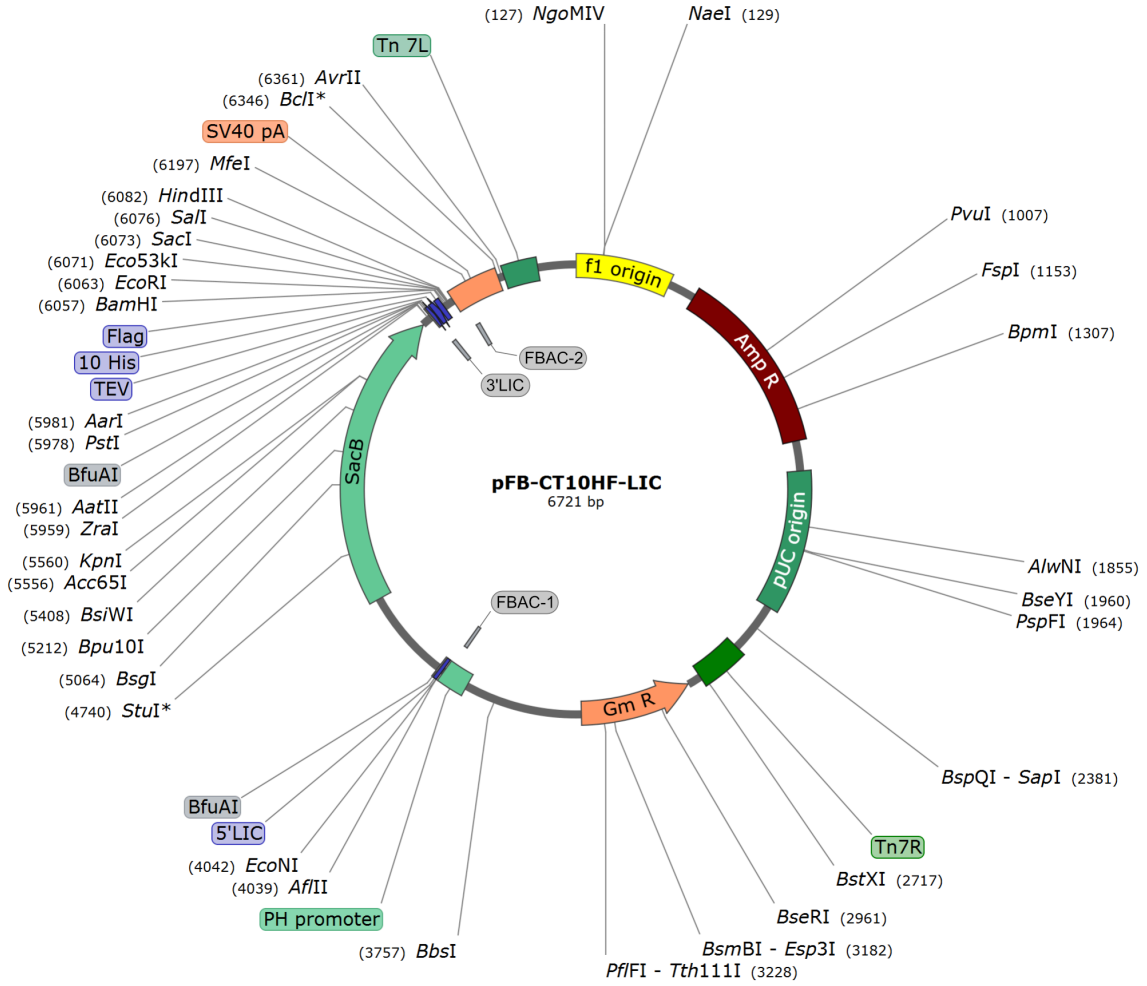


Figure A.3. Map of the pFB-CT10HF-LIC vector, used for the expression of MFSD10, ELOVL7 and CerS6-c003 in *Sf*9 insect cells.

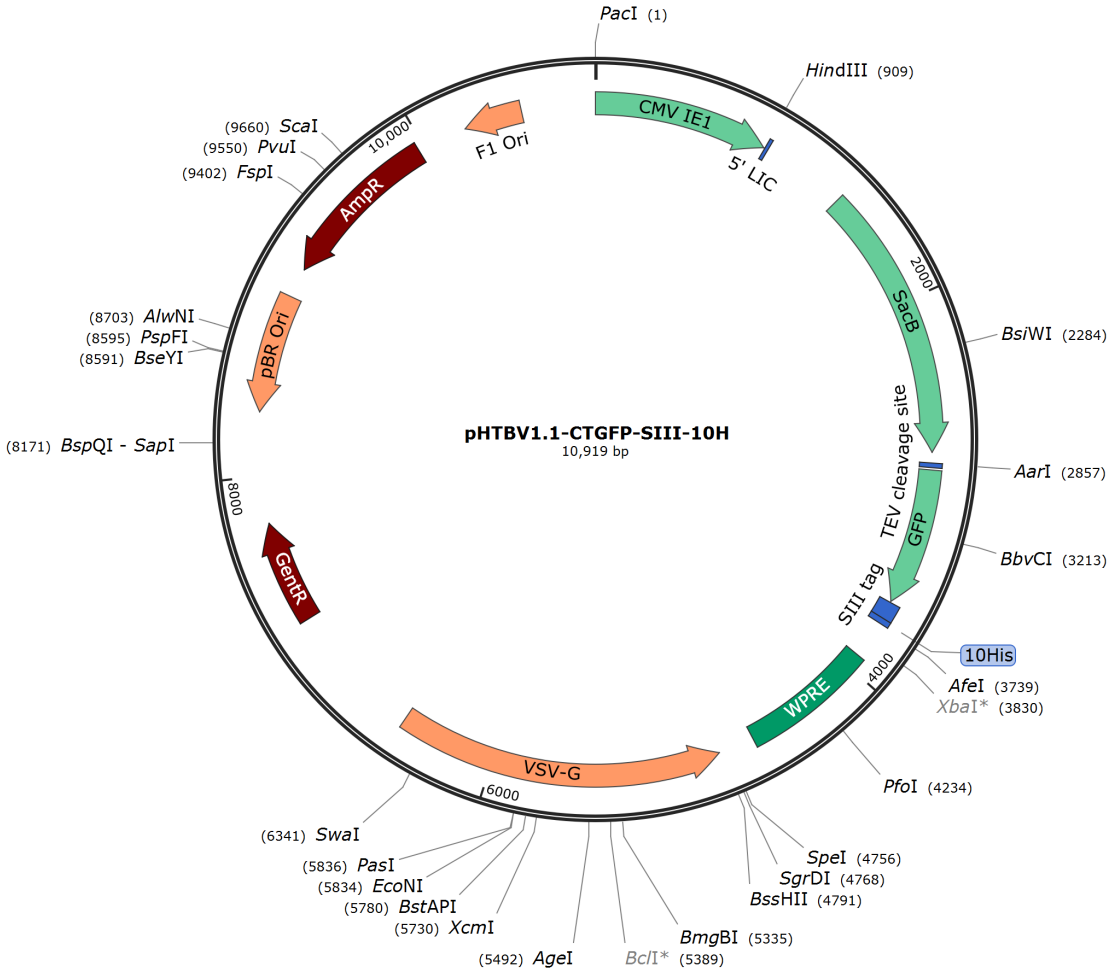


Figure A.4. Map of the pHTBV1.1-CTGFP-SIII-10H vector, used for the expression of MFSD10 in Expi293F cells.

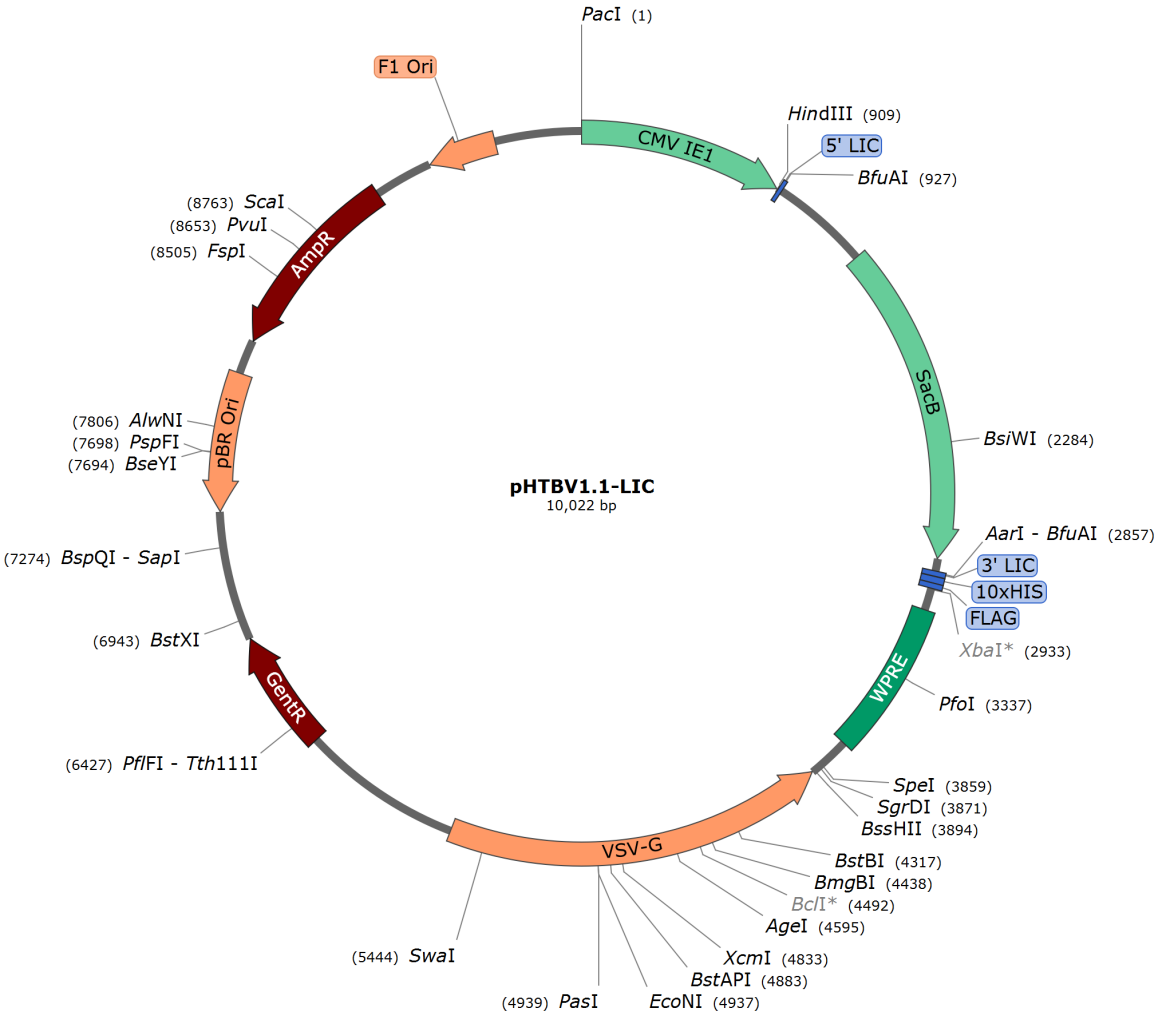


Figure A.5. Map of the pHTBV1.1-LIC vector, used for expression of CerS1, CerS5 and CerS6 constructs containing C-terminal 10x His and FLAG tags in Expi293F and Expi GnTI- cells.

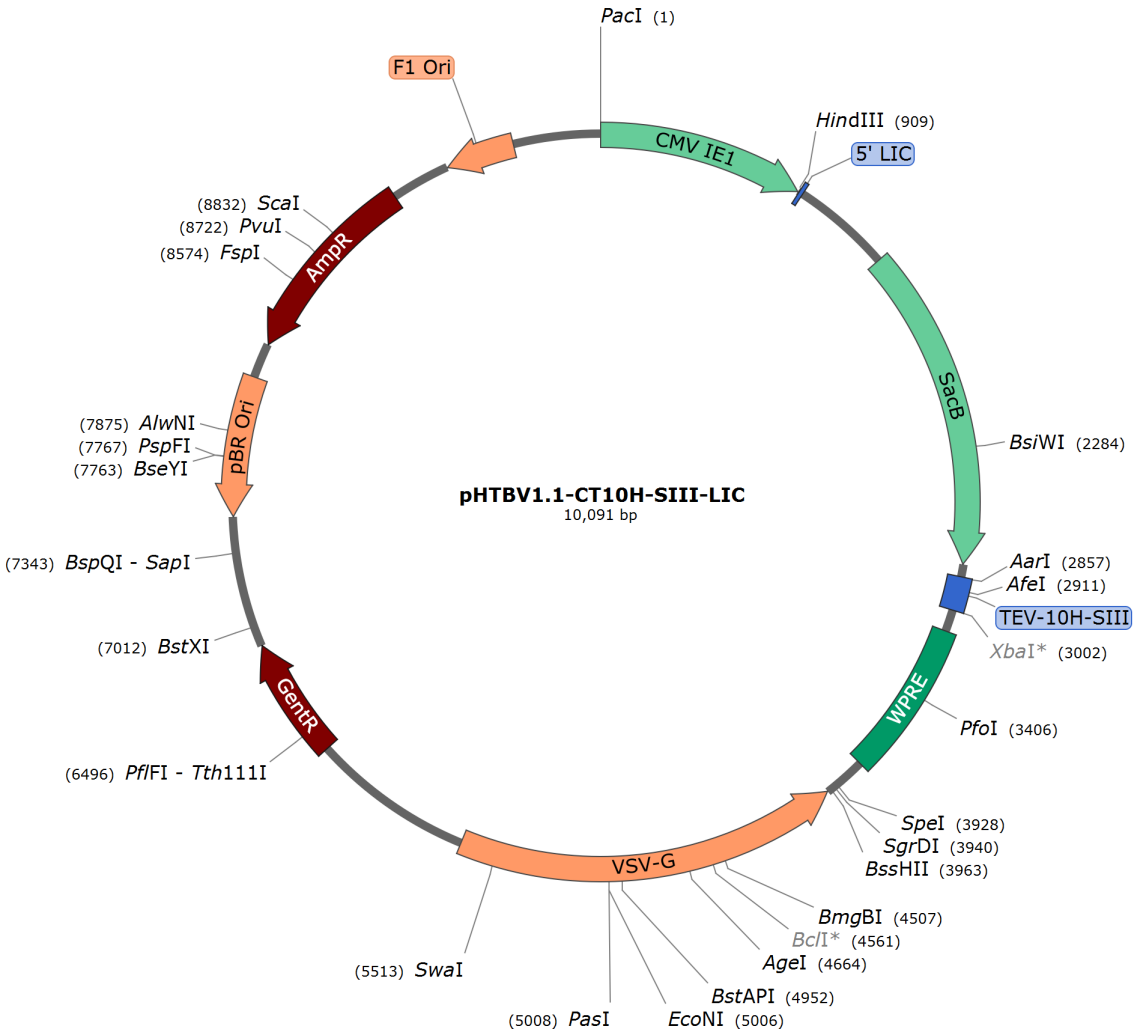


Figure A.6. Map of the pHTBV1.1-CT10H-SIII-LIC vector, used for expression of CerS5 and CerS6 constructs containing C-terminal 10x His and Twin-Strep tags in Expi293F and Expi GnTI- cells.

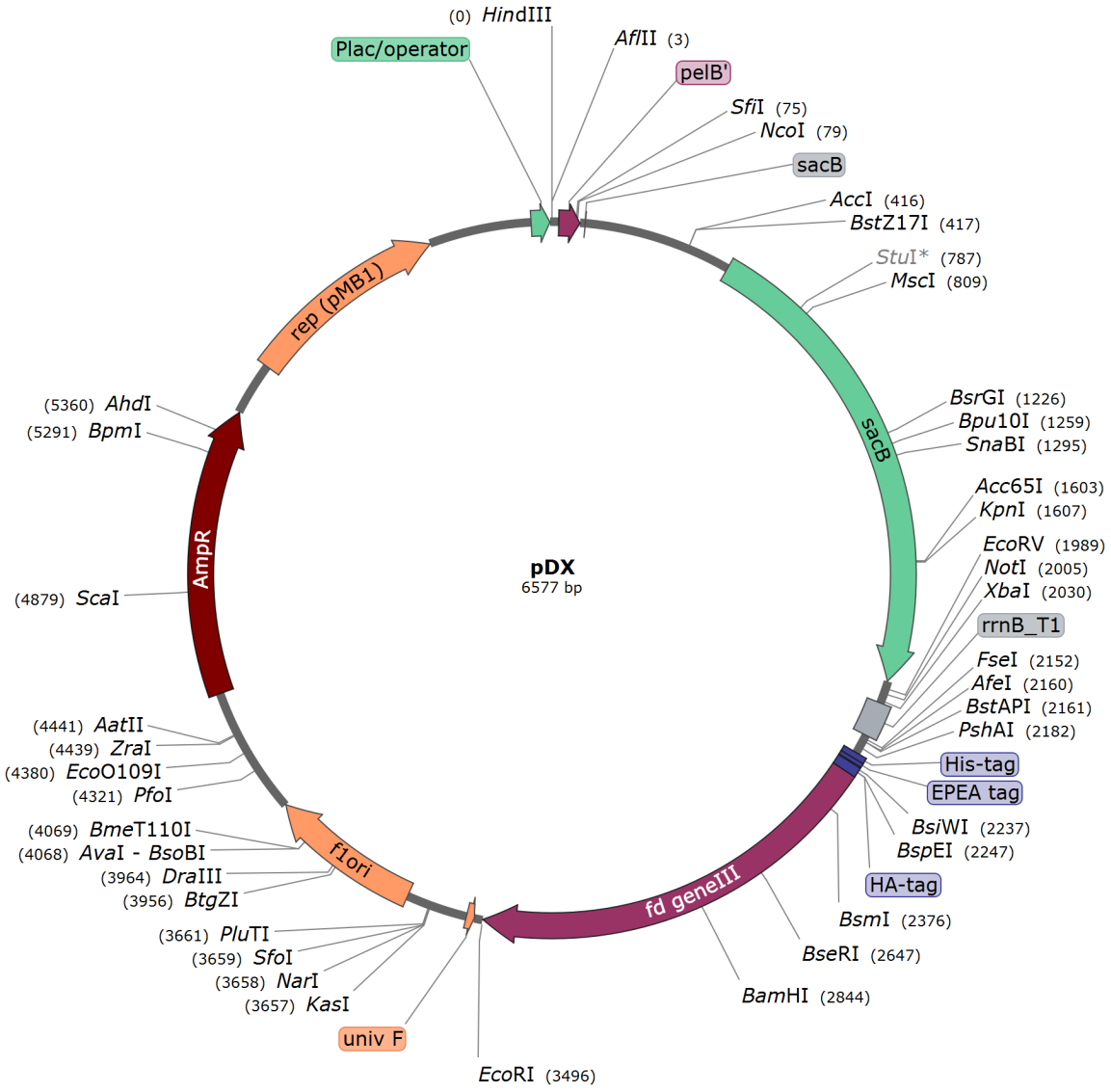


Figure A.7. Map of the pDX vector, used for expression of nanobodies in the *E. coli* strain WK6. This is also the original vector used for phage display and ELISA screening of CerS binders by the Nanobody Service Facility (University of Zurich).

Bibliography

1. Alberts B. Molecular biology of the cell. 6th ed. New York: Garland Science, Taylor and Francis Group; 2015.

2. Clausen MV, Hilbers F, Poulsen H. The Structure and Function of the Na,K-ATPase Isoforms in Health and Disease. Front Physiol. 2017;8:371.

3. Forgac M. Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology. Nat Rev Mol Cell Biol. 2007;8(11):917-29.

4. Mitchell P. Protonmotive redox mechanism of the cytochrome b-c1 complex in the respiratory chain: protonmotive ubiquinone cycle. FEBS Lett. 1975;56(1):1-6.

5. Wikstrom MKF. Proton pump coupled to cytochrome c oxidase in mitochondria. Nature. 1977;266:271-3.

6. Friedrich T, Steinmüller K, Weiss H. The proton-pumping respiratory complex I of bacteria and mitochondria and its homologue in chloroplasts. FEBS Lett. 1995;367(2):107-11.

7. Stuchebrukhov AA. Redox-Driven Proton Pumps of the Respiratory Chain. Biophys J. 2018;115(5):830-40.

8. Colas C, Ung PM-U, Schlessinger A. SLC transporters: structure, function, and drug discovery. MedChemComm. 2016;7(6):1069-81.

9. Jardetzky O. Simple allosteric model for membrane pumps. Nature. 1966;211:969-70.

10. Kazmier K, Claxton DP, Mchaourab HS. Alternating access mechanisms of LeuT-fold transporters: trailblazing towards the promised energy landscapes. Curr Opin Struct Biol. 2017;45:100-8.

11. Garaeva AA, Slotboom DJ. Elevator-type mechanisms of membrane transport. Biochem Soc Trans. 2020;48(3):1227-41.

12. Pao SS, Paulsen IT, Saier MH. Major Facilitator Superfamily. Microbiol Mol Biol Rev. 1998;62(1):1-34.

13. Law CJ, Maloney PC, Wang D-N. Ins and Outs of Major Facilitator Superfamily Antiporters. Annu Rev Microbiol. 2008;62(1):289-305.

14. Saier MH, Reddy VS, Tamang DG, Västermark Å. The Transporter Classification Database. Nucleic Acids Res. 2014;42(D1):D251-D8.

15. Wang SC, Davejan P, Hendargo KJ, Javadi-Razaz I, Chou A, Yee DC, et al. Expansion of the Major Facilitator Superfamily (MFS) to include novel transporters as well as transmembrane-acting enzymes. Biochim Biophys Acta. 2020;1862(9):183277.

16. Elbourne LDH, Tetu SG, Hassan KA, Paulsen IT. TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. Nucleic Acids Res. 2017;45(D1):D320-D4.

17. Cura AJ, Carruthers A. Role of Monosaccharide Transport Proteins in Carbohydrate Assimilation, Distribution, Metabolism, and Homeostasis. Compr Physiol. 2012;2(2):863-914.

18. Ishida N, Kawakita M. Molecular physiology and pathology of the nucleotide sugar transporter family (SLC35). Pflugers Arch. 2004;447(5):768-75.

19. Kawahara A, Nishi T, Hisano Y, Fukui H, Yamaguchi A, Mochizuki N. The sphingolipid transporter spns2 functions in migration of zebrafish myocardial precursors. Science. 2009;323(5913):524-7.

20. Nguyen LN, Ma D, Shui G, Wong P, Cazenave-Gassiot A, Zhang X, et al. Mfsd2a is a transporter for the essential omega-3 fatty acid docosahexaenoic acid. Nature. 2014;509:503-6.

21. Smith DE, Clémençon B, Hediger MA. Proton-coupled oligopeptide transporter family SLC15: Physiological, pharmacological and pathological implications. Mol Aspects Med. 2013;34(2-3):323-36.

22. Newstead S. Molecular insights into proton coupled peptide transport in the PTR family of oligopeptide transporters. Biochim Biophys Acta. 2015;1850(3):488-99.

23. Griffiths M, Beaumont N, Yao SYM, Sundaram M, Boumah CE, Davies A, et al. Cloning of a human nucleoside transporter implicated in the Cellular uptake of adenosine and chemotherapeutic drugs. Nat Med. 1997;3(1):89-93.

24. Roth M, Obaidat A, Hagenbuch B. OATPs, OATs and OCTs: the organic anion and cation transporters of the SLCO and SLC22A gene superfamilies. Br J Pharmacol. 2012;165(5):1260-87.

25. Paulsen IT, Brown MH, Skurray RA. Proton-dependent multidrug efflux systems. Microbiol Rev. 1996;60(4):575-608.

26. Fluman N, Bibi E. Bacterial multidrug transport through the lens of the major facilitator superfamily. Biochim Biophys Acta. 2009;1794(5):738-47.

27. Edgar R, Bibi E. MdfA, an Escherichia coli multidrug resistance protein with an extraordinarily broad spectrum of drug recognition. J Bacteriol. 1997;179(7):2274-80.

28. Floyd JL, Smith KP, Kumar SH, Floyd JT, Varela MF. LmrS Is a Multidrug Efflux Pump of the Major Facilitator Superfamily from Staphylococcus aureus. Antimicrob Agents Chemoter. 2010;54(12):5406-12.

29. Reddy VS, Shlykov MA, Castillo R, Sun EI, Saier MH. The major facilitator superfamily (MFS) revisited. FEBS J. 2012;279(11):2022-35.

30. Fowler W, Philip, Orwick-Rydmark M, Radestock S, Solcan N, Dijkman M, Patricia, Lyons A, Joseph, et al. Gating Topology of the Proton-Coupled Oligopeptide Symporters. Structure. 2015;23(2):290-301.

31. Wright NJ, Lee S-Y. Structures of human ENT1 in complex with adenosine reuptake inhibitors. Nat Struct Mol Biol. 2019;26(7):599-606.

32. Quistgaard EM, Löw C, Guettou F, Nordlund P. Understanding transport by the major facilitator superfamily (MFS): structures pave the way. Nat Rev Mol Cell Biol. 2016;17(2):123-32.

33. Drew D, North RA, Nagarathinam K, Tanabe M. Structures and General Transport Mechanisms by the Major Facilitator Superfamily (MFS). Chem Rev. 2021;121(9):5289-335.

34. Yan N. Structural advances for the major facilitator superfamily (MFS) transporters. Trends Biochem Sci. 2013;38(3):151-9.

35. Huang Y, Lemieux MJ, Song J, Auer M, Wang DN. Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. Science. 2003;301(5633):616-20.

36. Abramson J, Smirnova I, Kasho V, Verner G, Kaback HR, Iwata S. Structure and mechanism of the lactose permease of Escherichia coli. Science. 2003;301(5633):610-5.

37. Lemieux MJ, Huang Y, Wang D-N. The structural basis of substrate translocation by the Escherichia coli glycerol-3-phosphate transporter: a member of the major facilitator superfamily. Curr Opin Struct Biol. 2004;14(4):405-12.

38. Law CJ, Almqvist J, Bernstein A, Goetz RM, Huang Y, Soudant C, et al. Salt-bridge Dynamics Control Substrate-induced Conformational Change in the Membrane Transporter GlpT. J Mol Biol. 2008;378(4):828-39.

39. Jiang D, Zhao Y, Wang X, Fan J, Heng J, Liu X, et al. Structure of the YajR transporter suggests a transport mechanism based on the conserved motif A. Proc Natl Acad Sci U S A. 2013;110(36):14664-9.

40. Quistgaard EM, Löw C, Moberg P, Trésaugues L, Nordlund P. Structural basis for substrate transport in the GLUT-homology family of monosaccharide transporters. Nat Struct Mol Biol. 2013;20(6):766-8.

41. Forrest LR, Krämer R, Ziegler C. The structural basis of secondary active transport mechanisms. Biochim Biophys Acta. 2011;1807(2):167-88.

42. Newstead S, Drew D, Cameron AD, Postis VLG, Xia X, Fowler PW, et al. Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. EMBO J. 2011;30(2):417-26.

43. Zhang XC, Zhao Y, Heng J, Jiang D. Energy coupling mechanisms of MFS transporters. Protein Sci. 2015;24(10):1560-79.

44. Sahin-Tóth M, Karlin A, Kaback HR. Unraveling the mechanism of the lactose permease of Escherichia coli. Proc Natl Acad Sci U S A. 2000;97(20):10729-32.

45. Solcan N, Kwok J, Fowler PW, Cameron AD, Drew D, Iwata S, et al. Alternating access mechanism in the POT family of oligopeptide transporters. EMBO J. 2012;31(16):3411-21.

46. Doki S, Kato HE, Solcan N, Iwaki M, Koyama M, Hattori M, et al. Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. Proc Natl Acad Sci U S A. 2013;110(28):11343-8.

47. Li C, Yue Z, Newstead S, Voth GA. Proton Coupling and the Multiscale Kinetic Mechanism of a Peptide Transporter. bioRxiv. 2021. doi: 10.1101/2021.09.10.459748

48. Fluman N, Ryan M, Christopher, Whitelegge P, Julian, Bibi E. Dissection of Mechanistic Principles of a Secondary Multidrug Efflux Protein. Mol Cell. 2012;47(5):777-87.

49. Schuldiner S. Competition as a way of life for H^+-coupled antiporters. J Mol Biol. 2014;426(14):2539-46.

50. Law CJ. Competition between protons and substrate for binding to the major facilitator superfamily multidrug/H^+ antiporter MdtM. Experimental Results. 2021;2(e34):1-9.

51. Brawley DN, Sauer DB, Li J, Zheng X, Koide A, Jedhe GS, et al. Structural basis for inhibition of the drug efflux pump NorA from Staphylococcus aureus. Nat Chem Biol. 2022.

52. Kumar S, Mukherjee MM, Varela MF. Modulation of Bacterial Multidrug Resistance Efflux Pumps of the Major Facilitator Superfamily. International Journal of Bacteriology. 2013;2013:1-15.

53. Costa C, Dias PJ, Sa-Correia I, Teixeira MC. MFS multidrug transporters in pathogenic fungi: do they have real clinical impact? Front Physiol. 2014;5:197.

54. Choi YH, Yu AM. ABC transporters in multidrug resistance and pharmacokinetics, and strategies for drug development. Curr Pharm Des. 2014;20(5):793-807.

55. Robey RW, Pluchino KM, Hall MD, Fojo AT, Bates SE, Gottesman MM. Revisiting the role of ABC transporters in multidrug-resistant cancer. Nat Rev Cancer. 2018;18(7):452-64.

56. Mima S, Ushijima H, Hwang H-J, Tsutsumi S, Makise M, Yamaguchi Y, et al. Identification of the TPO1 gene in yeast, and its human orthologue TETRAN, which cause resistance to NSAIDs. FEBS Lett. 2007;581(7):1457-63.

57. Reece M, Prawitt D, Landers J, Kast C, Gros P, Housman D, et al. Functional characterization of ORCTL2 - an organic cation transporter expressed in the renal proximal tubules. FEBS Lett. 1998;433(3):245-50.

58. Duyao MP, Taylor SAM, Buckler AJ, Ambrose CM, Lin C, Groot N, et al. A gene from chromosome 4p 16.3 with similarity to a superfamily of transporter proteins. Hum Mol Genet. 1993;2(6):673-6.

59. Cooper PR, Smilinich NJ, Day CD, Nowak NJ, Reid LH, Pearsall RS, et al. Divergently Transcribed Overlapping Genes Expressed in Liver and Kidney and Located in the 11p15.5 Imprinted Domain. Genomics. 1998;49(1):38-51.

60. Ushijima H, Hiasa M, Namba T, Hwang HJ, Hoshino T, Mima S, et al. Expression and function of TETRAN, a new type of membrane transporter. Biochem Biophys Res Commun. 2008;374(2):325-30.

61. Smyth EM, Grosser T, Wang M, Yu Y, FitzGerald GA. Prostanoids in health and disease. J Lipid Res. 2009;50 Suppl(Suppl):S423-S8.

62. Harirforoosh S, Asghar W, Jamali F. Adverse effects of nonsteroidal antiinflammatory drugs: An update of gastrointestinal, cardiovascular and renal complications. J Pharm Pharm Sci. 2013;16(5):821-47.

63. Schjerning A-M, McGettigan P, Gislason G. Cardiovascular effects and safety of (non-aspirin) NSAIDs. Nat Rev Cardiol. 2020;17(9):574-84.

64. McGettigan P, Henry D. Use of non-steroidal anti-inflammatory drugs that elevate cardiovascular risk: an examination of sales and essential medicines lists in low-, middle-, and high-income countries. PLoS Med. 2013;10(2):e1001388-e.

65. Fosbøl EL, Folke F, Jacobsen S, Rasmussen JN, Sørensen R, Schramm TK, et al. Cause-Specific Cardiovascular Risk Associated With Nonsteroidal Antiinflammatory Drugs Among Healthy Individuals. Circ Cardiovasc Qual Outcomes. 2010;3(4):395-405.

66. Perland E, Bagchi S, Klaesson A, Fredriksson R. Characteristics of 29 novel atypical solute carriers of major facilitator superfamily type: evolutionary conservation, predicted structure and neuronal co-expression. Open Biol. 2017;7(9):170142.

67. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580-5.

68. Samaras P, Schmidt T, Frejno M, Gessulat S, Reinecke M, Jarzab A, et al. ProteomicsDB: a multi-omics and multi-organism resource for life science research. Nucleic Acids Res. 2019;48(D1):D1153–D63.

69. Motoi Y, Saeki M, Nishimura T, Katayama K, Kitamura N, Ichikawa H, et al. Establishment of monoclonal antibodies against a novel eosinophil-specific cell surface molecule, major facilitator super family domain containing 10. Immunol Lett. 2012;147(1-2):80-4.

70. Nishimura T, Saeki M, Motoi Y, Kitamura N, Mori A, Kaminuma O, et al. Selective Suppression of Th2 Cell-Mediated Lung Eosinophilic Inflammation by Anti-Major Facilitator Super Family Domain Containing 10 Monoclonal Antibody. Allergol Int. 2014;63:29-35.

71. Cheng L-C, Baboo S, Lindsay C, Brusman L, Martinez-Bartolomé S, Tapia O, et al. Identification of new transmembrane proteins concentrated at the nuclear envelope using organellar proteomics of mesenchymal cells. Nucleus. 2019;10(1):126-43.

72. Go CD, Knight JDR, Rajasekharan A, Rathod B, Hesketh GG, Abe KT, et al. A proximity-dependent biotinylation map of a human cell. Nature. 2021;595:120-4.

73. Varela MF, Sansom CE, Griffith JK. Mutational analysis and molecular modelling of an amino acid sequence motif conserved in antiporters but not symporters in a transporter superfamily. Mol Membr Biol. 1995;12(4):313-9.

74. Nagarathinam K, Nakada-Nakura Y, Parthier C, Terada T, Juge N, Jaenecke F, et al. Outward open conformation of a Major Facilitator Superfamily multidrug/H+ antiporter provides insights into switching mechanism. Nat Commun. 2018;9:4005.

75. Saier J, Milton H, Paulsen IT. Phylogeny of multidrug transporters. Semin Cell Dev Biol. 2001;12(3):205-13.

76. Reimer RJ, Fon EA, Edwards RH. Vesicular neurotransmitter transport and the presynaptic regulation of quantal size. Curr Opin Neurobiol. 1998;8(3):405-12.

77. Wimalasena K. Vesicular monoamine transporters: structure-function, pharmacology, and medicinal chemistry. Med Res Rev. 2011;31(4):483-519.

78. Lawal HO, Krantz DE. SLC18: Vesicular neurotransmitter transporters for monoamines and acetylcholine. Mol Aspects Med. 2013;34(2-3):360-72.

79. Hannun YA, Obeid LM. Principles of bioactive lipid signalling: lessons from sphingolipids. Nat Rev Mol Cell Biol. 2008;9(2):139-50.

80. Merrill AH, Jr. Sphingolipid and glycosphingolipid metabolic pathways in the era of sphingolipidomics. Chem Rev. 2011;111(10):6387-422.

81. Mullen D, Thomas, Hannun A, Yusuf, Obeid M, Lina. Ceramide synthases at the centre of sphingolipid metabolism and biology. Biochem J. 2012;441(3):789-802.

82. Hannun YA, Obeid LM. Sphingolipids and their metabolism in physiology and disease. Nat Rev Mol Cell Biol. 2018;19(3):175-91.

83. Summers SA, Chaurasia B, Holland WL. Metabolic Messengers: ceramides. Nat Metab. 2019;1(11):1051-8.

84. Summers SA, Garza LA, Zhou H, Birnbaum MJ. Regulation of Insulin-Stimulated Glucose Transporter GLUT4 Translocation and Akt Kinase Activity by Ceramide. Mol Cell Biol. 1998;18(9):5457-64.

85. Raichur S, Wang T, Siew, Chan W, Puck, Li Y, Ching J, Chaurasia B, et al. CerS2 Haploinsufficiency Inhibits β-Oxidation and Confers Susceptibility to Diet-Induced Steatohepatitis and Insulin Resistance. Cell Metab. 2014;20(4):687-95.

86. García-Ruiz C, Colell A, Marí M, Morales A, Fernández-Checa JC. Direct Effect of Ceramide on the Mitochondrial Electron Transport Chain Leads to Generation of Reactive Oxygen Species. J Biol Chem. 1997;272(17):11369-77.

87. Smith E, Melissa, Tippetts S, Trevor, Brassfield S, Eric, Tucker J, Braden, Ockey A, Swensen C, Adam, et al. Mitochondrial fission mediates ceramide-induced metabolic disruption in skeletal muscle. Biochem J. 2013;456(3):427-39.

88. Hammerschmidt P, Ostkotte D, Nolte H, Gerl MJ, Jais A, Brunner HL, et al. CerS6-Derived Sphingolipids Interact with Mff and Promote Mitochondrial Fragmentation in Obesity. Cell. 2019;177(6):1536-52.e23.

89. Chaurasia B, Talbot CL, Summers SA. Adipocyte Ceramides-The Nexus of Inflammation and Metabolic Disease. Front Immunol. 2020;11:576347.

90. Chaurasia B, Summers SA. Ceramides in Metabolism: Key Lipotoxic Players. Annu Rev Physiol. 2021;83:303-30.

91. Dressler KA, Kan CC, Kolesnick RN. Sphingomyelin synthesis is involved in adherence during macrophage differentiation of HL-60 cells. J Biol Chem. 1991;266(18):11522-7.

92. Hanada K, Nishijima M, Kiso M, Hasegawa A, Fujita S, Ogawa T, et al. Sphingolipids are essential for the growth of Chinese hamster ovary cells. Restoration of the growth of a mutant defective in sphingoid base biosynthesis by exogenous sphingolipids. J Biol Chem. 1992;267(33):23527-33.

93. Tafesse FG, Huitema K, Hermansson M, van der Poel S, van den Dikkenberg J, Uphoff A, et al. Both sphingomyelin synthases SMS1 and SMS2 are required for

sphingomyelin homeostasis and growth in human HeLa cells. J Biol Chem. 2007;282(24):17537-47.

94. Contreras F-X, Ernst AM, Haberkant P, Björkholm P, Lindahl E, Gönen B, et al. Molecular recognition of a single sphingolipid species by a protein's transmembrane domain. Nature. 2012;481:525-9.

95. Ramírez-Camacho I, Bautista-Pérez R, Correa F, Buelna-Chontal M, Román-Anguiano NG, Medel-Franco M, et al. Role of sphingomyelinase in mitochondrial ceramide accumulation during reperfusion. Biochim Biophys Acta. 2016;1862(10):1955-63.

96. Iwabuchi K, Nakayama H, Iwahara C, Takamori K. Significance of glycosphingolipid fatty acid chain length on membrane microdomain-mediated signal transduction. FEBS Lett. 2010;584(9):1642-52.

97. D'Angelo G, Capasso S, Sticco L, Russo D. Glycosphingolipids: synthesis and functions. FEBS J. 2013;280(24):6338-53.

98. Russo D, Parashuraman S, D'Angelo G. Glycosphingolipid-Protein Interaction in Signal Transduction. Int J Mol Sci. 2016;17(10).

99. Hakomori S-I, Handa K, Iwabuchi K, Yamamura S, Prinetti A. New insights in glycosphingolipid function: "glycosignaling domain," a cell surface assembly of glycosphingolipids with signal transducer molecules, involved in cell adhesion coupled with signaling. Glycobiology. 1998;8(10):xi-xviii.

100. Todeschini AR, Hakomori S-I. Functional role of glycosphingolipids and gangliosides in control of cell adhesion, motility, and growth, through glycosynaptic microdomains. Biochim Biophys Acta. 2008;1780(3):421-33.

101. Boggs JM, Gao W, Hirahara Y. Myelin glycosphingolipids, galactosylceramide and sulfatide, participate in carbohydrate-carbohydrate interactions between apposed membranes and may form glycosynapses between oligodendrocyte and/or myelin membranes. Biochim Biophys Acta. 2008;1780(3):445-55.

102. Rani CS, Abe A, Chang Y, Rosenzweig N, Saltiel AR, Radin NS, et al. Cell cycle arrest induced by an inhibitor of glucosylceramide synthase. Correlation with cyclin-dependent kinases. J Biol Chem. 1995;270(6):2859-67.

103. Li JF, Zheng SJ, Wang LL, Liu S, Ren F, Chen Y, et al. Glucosylceramide synthase regulates the proliferation and apoptosis of liver cells in vitro by Bcl-2/Bax pathway. Mol Med Rep. 2017;16(5):7355-60.

104. Reza S, Ugorski M, Suchański J. Glucosylceramide and galactosylceramide, small glycosphingolipids with significant impact on health and disease. Glycobiology. 2021;31(11):1416-34.

105. Lingwood D, Simons K. Lipid rafts as a membrane-organizing principle. Science. 2010;327(5961):46-50.

106. Stancevic B, Kolesnick R. Ceramide-rich platforms in transmembrane signaling. FEBS Lett. 2010;584(9):1728-40.

107. Burgert A, Schlegel J, Bécam J, Doose S, Bieberich E, Schubert-Unkmeir A, et al. Characterization of Plasma Membrane Ceramides by Super-Resolution Microscopy. Angew Chem Int Ed Engl. 2017;56(22):6131-5.

108. Courtney KC, Pezeshkian W, Raghupathy R, Zhang C, Darbyson A, Ipsen JH, et al. C24 Sphingolipids Govern the Transbilayer Asymmetry of Cholesterol and Lateral Organization of Model and Live-Cell Plasma Membranes. Cell Rep. 2018;24(4):1037-49.

109. Imgrund S, Hartmann D, Farwanah H, Eckhardt M, Sandhoff R, Degen J, et al. Adult ceramide synthase 2 (CERS2)-deficient mice exhibit myelin sheath defects, cerebellar degeneration, and hepatocarcinomas. J Biol Chem. 2009;284(48):33549-60.

110. Isokawa M, Sassa T, Hattori S, Miyakawa T, Kihara A. Reduced chain length in myelin sphingolipids and poorer motor coordination in mice deficient in the fatty acid elongase Elovl1. FASEB bioAdvances. 2019;1(12):747-59.

111. Li W, Sandhoff R, Kono M, Zerfas P, Hoffmann V, Ding BC, et al. Depletion of ceramides with very long chain fatty acids causes defective skin permeability barrier function, and neonatal lethality in ELOVL4 deficient mice. Int J Biol Sci. 2007;3(2):120-8.

112. Cameron DJ, Tong Z, Yang Z, Kaminoh J, Kamiyah S, Chen H, et al. Essential role of Elovl4 in very long chain fatty acid synthesis, skin permeability barrier function, and neonatal survival. Int J Biol Sci. 2007;3(2):111-9.

113. Sassa T, Ohno Y, Suzuki S, Nomura T, Nishioka C, Kashiwagi T, et al. Impaired epidermal permeability barrier in mice lacking elovl1, the gene responsible for very-long-chain fatty acid production. Mol Cell Biol. 2013;33(14):2787-96.

114. Rabionet M, Bayerle A, Jennemann R, Heid H, Fuchser J, Marsching C, et al. Male meiotic cytokinesis requires ceramide synthase 3-dependent sphingolipids with unique membrane anchors. Hum Mol Genet. 2015;24(17):4792-808.

115. Harkewicz R, Du H, Tong Z, Alkuraya H, Bedell M, Sun W, et al. Essential role of ELOVL4 protein in very long chain fatty acid synthesis and retinal function. J Biol Chem. 2012;287(14):11469-80.

116. Pewzner-Jung Y, Brenner O, Braun S, Laviad EL, Ben-Dor S, Feldmesser E, et al. A critical role for ceramide synthase 2 in liver homeostasis: II. insights into molecular changes leading to hepatopathy. J Biol Chem. 2010;285(14):10911-23.

117. Braun PE, Snell EE. Biosynthesis of sphingolipid bases. II. Keto intermediates in synthesis of sphingosine and dihydrosphingosine by cell-free extracts of Hansenula ciferri. J Biol Chem. 1968;243(14):3775-83.

118. Hanada K. Serine palmitoyltransferase, a key enzyme of sphingolipid metabolism. Biochim Biophys Acta. 2003;1632(1-3):16-30.

119. Beeler T, Bacikova D, Gable K, Hopkins L, Johnson C, Slife H, et al. The Saccharomyces cerevisiae TSC10/YBR265w gene encoding 3-ketosphinganine reductase is identified in a screen for temperature-sensitive suppressors of the Ca2+-sensitive csg2Delta mutant. J Biol Chem. 1998;273(46):30688-94.

120. Kihara A, Igarashi Y. FVT-1 is a mammalian 3-ketodihydrosphingosine reductase with an active site that faces the cytosolic side of the endoplasmic reticulum membrane. J Biol Chem. 2004;279(47):49243-50.

121. Ternes P, Franke S, Zähringer U, Sperling P, Heinz E. Identification and characterization of a sphingolipid delta 4-desaturase family. J Biol Chem. 2002;277(28):25512-8.

122. Omae F, Miyazaki M, Enomoto A, Suzuki M, Suzuki Y, Suzuki A. DES2 protein is responsible for phytoceramide biosynthesis in the mouse small intestine. Biochem J. 2004;379(3):687-95.

123. Riboni L, Prinetti A, Bassi R, Tettamanti G. Formation of bioactive sphingoid molecules from exogenous sphingomyelin in primary cultures of neurons and astrocytes. FEBS Lett. 1994;352(3):323-6.

124. Kitatani K, Sheldon K, Rajagopalan V, Anelli V, Jenkins RW, Sun Y, et al. Involvement of acid beta-glucosidase 1 in the salvage pathway of ceramide formation. J Biol Chem. 2009;284(19):12972-8.

125. Chatelut M, Leruth M, Harzer K, Dagan A, Marchesini S, Gatt S, et al. Natural ceramide is unable to escape the lysosome, in contrast to a fluorescent analogue. FEBS Lett. 1998;426(1):102-6.

126. Bernardo K, Hurwitz R, Zenk T, Desnick RJ, Ferlinz K, Schuchman EH, et al. Purification, characterization, and biosynthesis of human acid ceramidase. J Biol Chem. 1995;270(19):11098-102.

127. Gebai A, Gorelik A, Li Z, Illes K, Nagar B. Structural basis for the activation of acid ceramidase. Nat Commun. 2018;9:1621.

128. Riboni L, Bassi R, Caminiti A, Prinetti A, Viani P, Tettamanti G. Metabolic fate of exogenous sphingosine in neuroblastoma neuro2A cells. Dose-dependence and biological effects. Ann N Y Acad Sci. 1998;845:46-56.

129. Smith ER, Merrill AH, Jr. Differential roles of de novo sphingolipid biosynthesis and turnover in the "burst" of free sphingosine and sphinganine, and their 1-phosphates and N-acyl-derivatives, that occurs upon changing the medium of cells in culture. J Biol Chem. 1995;270(32):18749-58.

130. Ogretmen B, Pettus BJ, Rossi MJ, Wood R, Usta J, Szulc Z, et al. Biochemical mechanisms of the generation of endogenous long chain ceramide in response to exogenous short chain ceramide in the A549 human lung adenocarcinoma cell line. Role for endogenous ceramide in mediating the action of exogenous ceramide. J Biol Chem. 2002;277(15):12960-9.

131. Becker KP, Kitatani K, Idkowiak-Baldys J, Bielawski J, Hannun YA. Selective inhibition of juxtanuclear translocation of protein kinase C betaII by a negative feedback mechanism involving ceramide formed from the salvage pathway. J Biol Chem. 2005;280(4):2606-12.

132. Strub GM, Maceyka M, Hait NC, Milstien S, Spiegel S. Extracellular and intracellular actions of sphingosine-1-phosphate. Adv Exp Med Biol. 2010;688:141-55.

133. Kitatani K, Idkowiak-Baldys J, Hannun YA. The sphingolipid salvage pathway in ceramide metabolism and signaling. Cell Signal. 2008;20(6):1010-8.

134. Kolesnick R. Ceramide: a novel second messenger. Trends Cell Biol. 1992;2(8):232-6.

135. Boulgaropoulos B, Amenitsch H, Laggner P, Pabst G. Implication of sphingomyelin/ceramide molar ratio on the biological activity of sphingomyelinase. Biophys J. 2010;99(2):499-506.

136. Holmes RS, Barron KA, Krupenko NI. Ceramide Synthase 6: Comparative Analysis, Phylogeny and Evolution. Biomolecules. 2018;8(4):111.

137. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583-9.

138. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science. 2021;373(6557):871-6.

139. Mizutani Y, Kihara A, Igarashi Y. Mammalian Lass6 and its related family members regulate synthesis of specific ceramides. Biochem J. 2005;390(1):263-71.

140. Laviad EL, Kelly S, Merrill AH, Jr., Futerman AH. Modulation of ceramide synthase activity via dimerization. J Biol Chem. 2012;287(25):21025-33.

141. Tidhar R, Zelnik ID, Volpert G, Ben-Dor S, Kelly S, Merrill AH, Jr., et al. Eleven residues determine the acyl chain specificity of ceramide synthases. J Biol Chem. 2018;293(25):9912-21.

142. Tidhar R, Futerman AH. The complexity of sphingolipid biosynthesis in the endoplasmic reticulum. Biochim Biophys Acta. 2013;1833(11):2511-8.

143. Winter E, Ponting CP. TRAM, LAG1 and CLN8: members of a novel family of lipid-sensing domains? Trends Biochem Sci. 2002;27(8):381-3.

144. Spassieva S, Seo JG, Jiang JC, Bielawski J, Alvarez-Vasquez F, Jazwinski SM, et al. Necessary role for the Lag1p motif in (dihydro)ceramide synthase activity. J Biol Chem. 2006;281(45):33931-8.

145. Tidhar R, Ben-Dor S, Wang E, Kelly S, Merrill AH, Jr., Futerman AH. Acyl chain specificity of ceramide synthases is determined within a region of 150 residues in the Tram-Lag-CLN8 (TLC) domain. J Biol Chem. 2012;287(5):3197-206.

146. Mesika A, Ben-Dor S, Laviad EL, Futerman AH. A new functional motif in Hox domain-containing ceramide synthases: identification of a novel region flanking the Hox and TLC domains essential for activity. J Biol Chem. 2007;282(37):27366-73.

147. Voelzmann A, Wulf AL, Eckardt F, Thielisch M, Brondolin M, Pesch YY, et al. Nuclear Drosophila CerS Schlank regulates lipid homeostasis via the homeodomain, independent of the lag1p motif. FEBS Lett. 2016;590(7):971-81.

148. Sociale M, Wulf AL, Breiden B, Klee K, Thielisch M, Eckardt F, et al. Ceramide Synthase Schlank Is a Transcriptional Regulator Adapting Gene Expression to Energy Requirements. Cell Rep. 2018;22(4):967-78.

149. Mendelson K, Pandey S, Hisano Y, Carellini F, Das BC, Hla T, et al. The ceramide synthase 2b gene mediates genomic sensing and regulation of sphingosine levels during zebrafish embryogenesis. eLife. 2017;6:e21992.

150. Sassa T, Hirayama T, Kihara A. Enzyme Activities of the Ceramide Synthases CERS2-6 Are Regulated by Phosphorylation in the C-terminal Region. J Biol Chem. 2016;291(14):7477-87.

151. Mesicek J, Lee H, Feldman T, Jiang X, Skobeleva A, Berdyshev EV, et al. Ceramide synthases 2, 5, and 6 confer distinct roles in radiation-induced apoptosis in HeLa cells. Cell Signal. 2010;22(9):1300-7.

152. Jensen SA, Calvert AE, Volpert G, Kouri FM, Hurley LA, Luciano JP, et al. Bcl2L13 is a ceramide synthase inhibitor in glioblastoma. Proc Natl Acad Sci U S A. 2014;111(15):5682-7.

153. Ohno Y, Suto S, Yamanaka M, Mizutani Y, Mitsutake S, Igarashi Y, et al. ELOVL1 production of C24 acyl-CoAs is linked to C24 sphingolipid synthesis. Proc Natl Acad Sci U S A. 2010;107(43):18439-44.

154. Kim JL, Mestre B, Malitsky S, Itkin M, Kupervaser M, Futerman AH. Fatty acid transport protein 2 interacts with ceramide synthase 2 to promote ceramide synthesis. J Biol Chem. 2022;298(4):101735.

155. Haus JM, Kashyap SR, Kasumov T, Zhang R, Kelly KR, Defronzo RA, et al. Plasma ceramides are elevated in obese subjects with type 2 diabetes and correlate with the severity of insulin resistance. Diabetes. 2009;58(2):337-43.

156. Hajduch E, Lachkar F, Ferre P, Foufelle F. Roles of Ceramides in Non-Alcoholic Fatty Liver Disease. J Clin Med. 2021;10(4):792.

157. Sheka AC, Adeyi O, Thompson J, Hameed B, Crawford PA, Ikramuddin S. Nonalcoholic Steatohepatitis: A Review. JAMA. 2020;323(12):1175-83.

158. Koo SH. Nonalcoholic fatty liver disease: molecular mechanisms for the hepatic steatosis. Clin Mol Hepatol. 2013;19(3):210-5.

159. Bergman BC, Brozinick JT, Strauss A, Bacon S, Kerege A, Bui HH, et al. Muscle sphingolipids during rest and exercise: a C18:0 signature for insulin resistance in humans. Diabetologia. 2016;59(4):785-98.

160. Turpin SM, Nicholls HT, Willmes DM, Mourier A, Brodesser S, Wunderlich CM, et al. Obesity-induced CerS6-dependent C16:0 ceramide production promotes weight gain and glucose intolerance. Cell Metab. 2014;20(4):678-86.

161. Raichur S. Ceramide Synthases Are Attractive Drug Targets for Treating Metabolic Diseases. Front Endocrinol. 2020;11:483.

162. Kim YR, Lee EJ, Shin KO, Kim MH, Pewzner-Jung Y, Lee YM, et al. Hepatic triglyceride accumulation via endoplasmic reticulum stress-induced SREBP-1 activation is regulated by ceramide synthases. Exp Mol Med. 2019;51(11):1-16.

163. Gosejacob D, Jäger PS, Vom Dorp K, Frejno M, Carstensen AC, Köhnke M, et al. Ceramide Synthase 5 Is Essential to Maintain C16:0-Ceramide Pools and Contributes to the Development of Diet-induced Obesity. J Biol Chem. 2016;291(13):6989-7003.

164. Raichur S, Brunner B, Bielohuby M, Hansen G, Pfenninger A, Wang B, et al. The role of C16:0 ceramide in the development of obesity and type 2 diabetes: CerS6 inhibition as a novel therapeutic approach. Mol Metab. 2019;21:36-50.

165. Montgomery MK, Brown SH, Lim XY, Fiveash CE, Osborne B, Bentley NL, et al. Regulation of glucose homeostasis and insulin action by ceramide acyl-chain length: A beneficial role for very long-chain sphingolipid species. Biochim Biophys Acta. 2016;1861(11):1828-39.

166. Park JW, Park WJ, Kuperman Y, Boura-Halfon S, Pewzner-Jung Y, Futerman AH. Ablation of very long acyl chain sphingolipids causes hepatic insulin resistance in mice due to altered detergent-resistant membranes. Hepatology. 2013;57(2):525-32.

167. Pewzner-Jung Y, Park H, Laviad EL, Silva LC, Lahiri S, Stiban J, et al. A critical role for ceramide synthase 2 in liver homeostasis: I. alterations in lipid metabolic pathways. J Biol Chem. 2010;285(14):10902-10.

168. Turpin-Nolan SM, Hammerschmidt P, Chen W, Jais A, Timper K, Awazawa M, et al. CerS1-Derived C(18:0) Ceramide in Skeletal Muscle Promotes Obesity-Induced Insulin Resistance. Cell Rep. 2019;26(1):1-10.e7.

169. Turner N, Lim XY, Toop HD, Osborne B, Brandon AE, Taylor EN, et al. A selective inhibitor of ceramide synthase 1 reveals a novel role in fat metabolism. Nat Commun. 2018;9:3165.

170. Riley RT, Merrill AH, Jr. Ceramide synthase inhibition by fumonisins: a perfect storm of perturbed sphingolipid metabolism, signaling, and disease. J Lipid Res. 2019;60(7):1183-9.

171. Marasas WF. Discovery and occurrence of the fumonisins: a historical perspective. Environ Health Perspect. 2001;109 Suppl 2(Suppl 2):239-43.

172. Marasas WF, Riley RT, Hendricks KA, Stevens VL, Sadler TW, Gelineau-van Waes J, et al. Fumonisins disrupt sphingolipid metabolism, folate transport, and neural tube development in embryo culture and in vivo: a potential risk factor for human neural tube defects among populations consuming fumonisin-contaminated maize. J Nutr. 2004;134(4):711-6.

173. Merrill AH, Jr., van Echten G, Wang E, Sandhoff K. Fumonisin B1 inhibits sphingosine (sphinganine) N-acyltransferase and de novo sphingolipid biosynthesis in cultured neurons in situ. J Biol Chem. 1993;268(36):27299-306.

174. Berdyshev EV, Gorshkova I, Skobeleva A, Bittman R, Lu X, Dudek SM, et al. FTY720 Inhibits Ceramide Synthases and Up-regulates Dihydrosphingosine 1-Phosphate Formation in Human Lung Endothelial Cells. J Biol Chem. 2009;284(9):5467-77.

175. Schiffmann S, Hartmann D, Fuchs S, Birod K, Ferreiròs N, Schreiber Y, et al. Inhibitors of specific ceramide synthases. Biochimie. 2012;94(2):558-65.

176. Toop HD, Don AS, Morris JC. Synthesis and biological evaluation of analogs of AAL(S) for use as ceramide synthase 1 inhibitors. Org Biomol Chem. 2015;13(48):11593-6.

177. Kihara A. Very long-chain fatty acids: elongation, physiology and related disorders. J Biochem. 2012;152(5):387-95.

178. Furland NE, Zanetti SR, Oresti GM, Maldonado EN, Aveldaño MI. Ceramides and Sphingomyelins with High Proportions of Very Long-chain Polyunsaturated Fatty acids in Mammalian Germ Cells. J Biol Chem. 2007;282(25):18141-50.

179. Hajeyah AA, Griffiths WJ, Wang Y, Finch AJ, O'Donnell VB. The Biosynthesis of Enzymatically Oxidized Lipids. Front Endocrinol. 2020;11:591819.

180. Soupene E, Kuypers FA. Mammalian long-chain acyl-CoA synthetases. Exp Biol Med. 2008;233(5):507-21.

181. Paiva P, Medina FE, Viegas M, Ferreira P, Neves RPP, Sousa JPM, et al. Animal Fatty Acid Synthase: A Chemical Nanofactory. Chem Rev. 2021;121(15):9502-53.

182. Moon Y-A, Horton JD. Identification of Two Mammalian Reductases Involved in the Two-carbon Fatty Acyl Elongation Cascade. J Biol Chem. 2003;278(9):7335-43.

183. Ikeda M, Kanao Y, Yamanaka M, Sakuraba H, Mizutani Y, Igarashi Y, et al. Characterization of four mammalian 3-hydroxyacyl-CoA dehydratases involved in very long-chain fatty acid synthesis. FEBS Lett. 2008;582(16):2435-40.

184. Leonard AE, Pereira SL, Sprecher H, Huang Y-S. Elongation of long-chain fatty acids. Prog Lipid Res. 2004;43(1):36-54.

185. Venegas-Calerón M, Beaudoin F, Sayanova O, Napier JA. Co-transcribed Genes for Long Chain Polyunsaturated Fatty Acid Biosynthesis in the Protozoon Perkinsus marinus Include a Plant-like FAE1 3-Ketoacyl Coenzyme A Synthase. J Biol Chem. 2007;282(5):2996-3003.

186. Haslam TM, Kunst L. Extending the story of very-long-chain fatty acid elongation. Plant Sci. 2013;210:93-107.

187. Jakobsson A, Westerberg R, Jacobsson A. Fatty acid elongases in mammals: Their regulation and roles in metabolism. Prog Lipid Res. 2006;45(3):237-49.

188. Toke DA, Martin CE. Isolation and Characterization of a Gene Affecting Fatty Acid Elongation in Saccharomyces cerevisiae. J Biol Chem. 1996;271(31):18413-22.

189. Oh C-S, Toke DA, Mandala S, Martin CE. ELO2 and ELO3, Homologues of the Saccharomyces cerevisiae ELO1 Gene, Function in Fatty Acid Elongation and Are Required for Sphingolipid Formation. J Biol Chem. 1997;272(28):17376-84.

190. Lee SH, Stephens JL, Paul KS, Englund PT. Fatty Acid Synthesis by Elongases in Trypanosomes. Cell. 2006;126(4):691-9.

191. Lee SH, Stephens JL, Englund PT. A fatty-acid synthesis mechanism specialized for parasitism. Nat Rev Microbiol. 2007;5(4):287-97.

192. Ramakrishnan S, Docampo MD, Macrae JI, Pujol FM, Brooks CF, Van Dooren GG, et al. Apicoplast and Endoplasmic Reticulum Cooperate in Fatty Acid Biosynthesis in Apicomplexan Parasite Toxoplasma gondii*. J Biol Chem. 2012;287(7):4957-71.

193. Heath RJ, Rock CO. The Claisen condensation in biology. Nat Prod Rep. 2002;19(5):581-96.

194. Hernandez-Buquer S, Blacklock BJ. Site-directed mutagenesis of a fatty acid elongase ELO-like condensing enzyme. FEBS Lett. 2013;587(23):3837-42.

195. Denic V, Weissman JS. A molecular caliper mechanism for determining very long-chain fatty acid length. Cell. 2007;130(4):663-77.

196. Tamura K, Makino A, Hullin-Matsuda F, Kobayashi T, Furihata M, Chung S, et al. Novel lipogenic enzyme ELOVL7 is involved in prostate cancer growth through saturated long-chain fatty acid metabolism. Cancer Res. 2009;69(20):8133-40.

197. Naganuma T, Sato Y, Sassa T, Ohno Y, Kihara A. Biochemical characterization of the very long-chain fatty acid elongase ELOVL7. FEBS Lett. 2011;585(20):3337-41.

198. Zhang K, Kniazeva M, Han M, Li W, Yu Z, Yang Z, et al. A 5-bp deletion in ELOVL4 is associated with two related forms of autosomal dominant macular dystrophy. Nat Genet. 2001;27:89-93.

199. Cadieux-Dion M, Turcotte-Gauthier M, Noreau A, Martin C, Meloche C, Gravel M, et al. Expanding the clinical phenotype associated with ELOVL4 mutation: study of a large French-Canadian family with autosomal dominant spinocerebellar ataxia and erythrokeratodermia. JAMA Neurol. 2014;71(4):470-5.

200. Di Gregorio E, Borroni B, Giorgio E, Lacerenza D, Ferrero M, Lo Buono N, et al. ELOVL5 mutations cause spinocerebellar ataxia 38. Am J Hum Genet. 2014;95(2):209-17.

201. Aldahmesh MA, Mohamed JY, Alkuraya HS, Verma IC, Puri RD, Alaiya AA, et al. Recessive mutations in ELOVL4 cause ichthyosis, intellectual disability, and spastic quadriplegia. Am J Hum Genet. 2011;89(6):745-50.

202. Mir H, Raza SI, Touseef M, Memon MM, Khan MN, Jaffar S, et al. A novel recessive mutation in the gene ELOVL4 causes a neuro-ichthyotic disorder with variable expressivity. BMC Med Genet. 2014;15:25.

203. Kutkowska-Kazmierczak A, Rydzanicz M, Chlebowski A, Klosowska-Kosicka K, Mika A, Gruchota J, et al. Dominant ELOVL1 mutation causes neurological disorder with ichthyotic keratoderma, spasticity, hypomyelination and dysmorphic features. J Med Genet. 2018;55(6):408-14.

204. Mueller N, Sassa T, Morales-Gonzalez S, Schneider J, Salchow DJ, Seelow D, et al. De novo mutation in ELOVL1 causes ichthyosis, acanthosis nigricans, hypomyelination, spastic paraplegia, high frequency deafness and optic atrophy. J Med Genet. 2019;56(3):164-75.

205. Ofman R, Dijkstra IME, Van Roermund CWT, Burger N, Turkenburg M, Van Cruchten A, et al. The role of ELOVL1 in very long-chain fatty acid homeostasis and X-linked adrenoleukodystrophy. EMBO Mol Med. 2010;2(3):90-7.

206. Westerberg R, Tvrdik P, Unden AB, Mansson JE, Norlen L, Jakobsson A, et al. Role for ELOVL3 and fatty acid chain length in development of hair and skin function. J Biol Chem. 2004;279(7):5621-9.

207. Moon YA, Hammer RE, Horton JD. Deletion of ELOVL5 leads to fatty liver through activation of SREBP-1c in mice. J Lipid Res. 2009;50(3):412-23.

208. Matsuzaka T, Shimano H, Yahagi N, Kato T, Atsumi A, Yamamoto T, et al. Crucial role of a long-chain fatty acid elongase, Elovl6, in obesity-induced insulin resistance. Nat Med. 2007;13(10):1193-202.

209. Marien E, Meister M, Muley T, Del Pulgar TG, Derua R, Spraggins JM, et al. Phospholipid profiling identifies acyl chain elongation as a ubiquitous trait and potential target for the treatment of lung squamous cell carcinoma. Oncotarget. 2016;7(11):12582-97.

210. Centenera MM, Scott JS, Machiels J, Nassar ZD, Miller DC, Zinonos I, et al. ELOVL5 Is a Critical and Targetable Fatty Acid Elongase in Prostate Cancer. Cancer Res. 2021;81(7):1704-18.

211. Come JH, Senter TJ, Clark MP, Court JJ, Gale-Day Z, Gu W, et al. Discovery and Optimization of Pyrazole Amides as Inhibitors of ELOVL1. J Med Chem. 2021;64(24):17753-76.

212. Uttaro AD. Acquisition and biosynthesis of saturated and unsaturated fatty acids by trypanosomatids. Mol Biochem Parasitol. 2014;196(1):61-70.

213. Nie L, Pascoa TC, Pike ACW, Bushell SR, Quigley A, Ruda GF, et al. The structural basis of fatty acid elongation by the ELOVL elongases. Nat Struct Mol Biol. 2021;28(6):512-20.

214. Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010;66(2):125-32.

215. Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr. 2013;69(7):1204-14.

216. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007;40(4):658-74.

217. Stein N. CHAINSAW: A program for mutating pdb files used as templates in molecular replacement. J Appl Crystallogr. 2008;41(3):641-3.

218. Cowtan K. Recent developments in classical density modification. Acta Crystallogr D Biol Crystallogr. 2010;66(4):470-8.

219. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr. 2006;62(9):1002-11.

220. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 4):486-501.

221. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 2011;67(4):355-67.

222. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr. 2006;62(4):439-50.

223. Bepler T, Morin A, Rapp M, Brasch J, Shapiro L, Noble AJ, et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. Nat Methods. 2019;16(11):1153-60.

224. Punjani A, Zhang H, Fleet DJ. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. Nat Methods. 2020;17(12):1214-21.

225. Ghosh S, Yu CL, Ferraro DJ, Sudha S, Pal SK, Schaefer WF, et al. Blue protein with red fluorescence. Proc Natl Acad Sci U S A. 2016;113(41):11513-8.

226. Caffrey M. Crystallizing membrane proteins for structure determination: use of lipidic mesophases. Annu Rev Biophys. 2009;38:29-51.

227. Caffrey M. A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. Acta Crystallogr F Struct Biol Commun. 2015;71(Pt 1):3-18.

228. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr D Biol Crystallogr. 2011;67(Pt 4):271-81.

229. Tickle IJ, Flensburg C, Keller P, Paciorek W, Sharff A, Vonrhein C, et al. STARANISO. Cambridge, UK: Global Phasing Ltd; 2018.

230. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: All-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010;66(1):12-21.

231. Parker JL, Newstead S. Membrane Protein Crystallisation: Current Trends and Future Perspectives. Adv Exp Med Biol. 2016;922:61-72.

232. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 2021;30(1):70-82.

233. Sun L, Zeng X, Yan C, Sun X, Gong X, Rao Y, et al. Crystal structure of a bacterial homologue of glucose transporters GLUT1–4. Nature. 2012;490:361-6.

234. Olsson MH, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. J Chem Theory Comput. 2011;7(2):525-37.

235. Deng D, Xu C, Sun P, Wu J, Yan C, Hu M, et al. Crystal structure of the human glucose transporter GLUT1. Nature. 2014;510:121-5.

236. Nomura N, Verdon G, Kang HJ, Shimamura T, Nomura Y, Sonoda Y, et al. Structure and mechanism of the mammalian fructose transporter GLUT5. Nature. 2015;526:397-401.

237. Qureshi AA, Suades A, Matsuoka R, Brock J, McComas SE, Nji E, et al. The molecular basis for sugar import in malaria parasites. Nature. 2020;578:321-5.

238. Newby ZE, O'Connell JD, 3rd, Gruswitz F, Hays FA, Harries WE, Harwood IM, et al. A general protocol for the crystallization of membrane proteins for X-ray structural investigation. Nat Protoc. 2009;4(5):619-37.

239. Caffrey M, Li D, Dukkipati A. Membrane protein structure determination using crystallography and lipidic mesophases: recent advances and successes. Biochemistry. 2012;51(32):6266-88.

240. Jungnickel KEJ, Parker JL, Newstead S. Structural basis for amino acid transport by the CAT family of SLC7 transporters. Nat Commun. 2018;9:550.

241. Thomas JA, Tate CG. Quality control in eukaryotic membrane protein overproduction. J Mol Biol. 2014;426(24):4139-54.

242. Birch J, Cheruvara H, Gamage N, Harrison PJ, Lithgo R, Quigley A. Changes in Membrane Protein Structural Biology. Biology (Basel). 2020;9(11):401.

243. Parker JL, Mindell JA, Newstead S. Thermodynamic evidence for a dual transport mechanism in a POT peptide transporter. eLife. 2014;3:e04273.

244. Uzdavinys P, Coinçon M, Nji E, Ndi M, Winkelmann I, Von Ballmoos C, et al. Dissecting the proton transport pathway in electrogenic Na+/H+ antiporters. Proc Natl Acad Sci U S A. 2017;114(7):E1101-E10.

245. Heng J, Zhao Y, Liu M, Liu Y, Fan J, Wang X, et al. Substrate-bound structure of the E. coli multidrug resistance transporter MdfA. Cell Res. 2015;25(9):1060-73.

246. Yin Y, He X, Szewczyk P, Nguyen T, Chang G. Structure of the Multidrug Transporter EmrD from Escherichia coli. Science. 2006;312(5774):741-4.

247. Schaedler TA, Van Veen HW. A flexible cation binding site in the multidrug major facilitator superfamily transporter LmrP is associated with variable proton coupling. FASEB J. 2010;24(10):3653-61.

248. Baker J, Wright SH, Tama F. Simulations of substrate transport in the multidrug transporter EmrD. Proteins. 2012;80(6):1620-32.

249. Yaffe D, Vergara-Jaque A, Forrest LR, Schuldiner S. Emulating proton-induced conformational changes in the vesicular monoamine transporter VMAT2 by mutagenesis. Proc Natl Acad Sci U S A. 2016;113(47):E7390-e8.

250. Yaffe D, Forrest LR, Schuldiner S. The ins and outs of vesicular monoamine transporters. J Gen Physiol. 2018;150(5):671-82.

251. Pei J, Grishin NV. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. Methods Mol Biol. 2014;1079:263-71.

252. Lee C, Kang HJ, Von Ballmoos C, Newstead S, Uzdavinys P, Dotson DL, et al. A two-domain elevator mechanism for sodium/proton antiport. Nature. 2013;501:573-7.

253. Gottlieb Y, Truman M, Cohen LA, Leichtmann-Bardoogo Y, Meyron-Holtz EG. Endoplasmic reticulum anchored heme-oxygenase 1 faces the cytosol. Haematologica. 2012;97(10):1489-93.

254. Shum M, Shintre CA, Althoff T, Gutierrez V, Segawa M, Saxberg AD, et al. ABCB10 exports mitochondrial biliverdin, driving metabolic maladaptation in obesity. Sci Transl Med. 2021;13(594):eabd1869.

255. Baranano DE, Rao M, Ferris CD, Snyder SH. Biliverdin reductase: a major physiologic cytoprotectant. Proc Natl Acad Sci U S A. 2002;99(25):16093-8.

256. Nytofte NS, Serrano MA, Monte MJ, Gonzalez-Sanchez E, Tumer Z, Ladefoged K, et al. A homozygous nonsense mutation (c.214C->A) in the biliverdin reductase alpha gene (BLVRA) results in accumulation of biliverdin during episodes of cholestasis. J Med Genet. 2011;48(4):219-25.

257. Parisi LR, Sowlati-Hashjin S, Berhane IA, Galster SL, Carter KA, Lovell JF, et al. Membrane Disruption by Very Long Chain Fatty Acids during Necroptosis. ACS Chem Biol. 2019;14(10):2286-94.

258. Tolkach Y, Merseburger A, Herrmann T, Kuczyk M, Serth J, Imkamp F. Signatures of Adverse Pathological Features, Androgen Insensitivity and Metastatic Potential in Prostate Cancer. Anticancer Res. 2015;35(10):5443-51.

259. Han W, Gao S, Barrett D, Ahmed M, Han D, Macoska JA, et al. Reactivation of androgen receptor-regulated lipid biosynthesis drives the progression of castration-resistant prostate cancer. Oncogene. 2018;37(6):710-21.

260. Zhang X, Wang Y. Identification of hub genes and key pathways associated with the progression of gynecological cancer. Oncol Lett. 2019;18(6):6516-24.

261. Li G, Cui S, Du J, Liu J, Zhang P, Fu Y, et al. Association of GALC, ZNF184, IL1R2 and ELOVL7 With Parkinson's Disease in Southern Chinese. Front Aging Neurosci. 2018;10:402.

262. Shanklin J, Whittle E, Fox BG. Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, stearoyl-CoA desaturase, and are conserved in alkane hydroxylase and xylene monooxygenase. Biochemistry. 1994;33(43):12787-94.

263. Ternes P, Franke S, Zähringer U, Sperling P, Heinz E. Identification and Characterization of a Sphingolipid Δ4-Desaturase Family. Journal of Biological Chemistry. 2002;277(28):25512-8.

264. Wang H, Klein MG, Zou H, Lane W, Snell G, Levin I, et al. Crystal structure of human stearoyl-coenzyme A desaturase in complex with substrate. Nat Struct Mol Biol. 2015;22(7):581-5.

265. Wyatt PJ. Light scattering and the absolute characterization of macromolecules. Anal Chim Acta. 1993;272(1):1-40.

266. Slotboom DJ, Duurkens RH, Olieman K, Erkens GB. Static light scattering to characterize membrane proteins in detergent solution. Methods. 2008;46(2):73-82.

267. Miercke LJW, Robbins RA, Stroud RM. Tetra Detector Analysis of Membrane Proteins. Curr Protoc Protein Sci. 2014;77:29.10.1-29.10.30.

268. Röttig A, Steinbüchel A. Acyltransferases in bacteria. Microbiol Mol Biol Rev. 2013;77(2):277-321.

269. Paiva P, Sousa SF, Ramos MJ, Fernandes PA. Understanding the Catalytic Machinery and the Reaction Pathway of the Malonyl-Acetyl Transferase Domain of Human Fatty Acid Synthase. ACS Catalysis. 2018;8(6):4860-72.

270. Jez JM, Noel JP. Mechanism of Chalcone Synthase. J Biol Chem. 2000;275(50):39640-6.

271. Wedekind JE, Frey PA, Rayment I. The structure of nucleotidylated histidine-166 of galactose-1-phosphate uridylyltransferase provides insight into phosphoryl group transfer. Biochemistry. 1996;35(36):11560-9.

272. Zhou X, Chou T-F, Aubol BE, Park CJ, Wolfenden R, Adams J, et al. Kinetic Mechanism of Human Histidine Triad Nucleotide Binding Protein 1. Biochemistry. 2013;52(20):3588-600.

273. Veeramani S, Lee MS, Lin MF. Revisiting histidine-dependent acid phosphatases: a distinct group of tyrosine phosphatases. Trends Biochem Sci. 2009;34(6):273-8.

274. Gottlin EB, Rudolph AE, Zhao Y, Matthews HR, Dixon JE. Catalytic mechanism of the phospholipase D superfamily proceeds via a covalent phosphohistidine intermediate. Proc Natl Acad Sci U S A. 1998;95(16):9202-7.

275. Leiros I, Secundo F, Zambonelli C, Servi S, Hough E. The first crystal structure of a phospholipase D. Structure. 2000;8(6):655-67.

276. Trent MS, Worsham LMS, Ernst-Fonberg ML. HlyC, the Internal Protein Acyltransferase That Activates Hemolysin Toxin: Roles of Various Conserved Residues in Enzymatic Activity As Probed by Site-Directed Mutagenesis. Biochemistry. 1999;38(29):9541-8.

277. Dodds AW, Ren XD, Willis AC, Law SK. The reaction mechanism of the internal thioester in the human complement component C4. Nature. 1996;379:177-9.

278. Gadjeva M, Dodds AW, Taniguchi-Sidle A, Willis AC, Isenman DE, Law SK. The covalent binding reaction of complement component C3. J Immunol. 1998;161(2):985-90.

279. Elliott RJ, Bennet AJ, Braun CA, MacLeod AM, Borgford TJ. Active-site variants of Streptomyces griseus protease B with peptide-ligation activity. Chem Biol. 2000;7(3):163-71.

280. Bolon DN, Mayo SL. Enzyme-like proteins by computational design. Proc Natl Acad Sci U S A. 2001;98(25):14274-9.

281. Burke AJ, Lovelock SL, Frese A, Crawshaw R, Ortmayer M, Dunstan M, et al. Design and evolution of an enzyme with a non-canonical organocatalytic mechanism. Nature. 2019;570:219-23.

282. Witkowski A, Joshi AK, Lindqvist Y, Smith S. Conversion of a beta-ketoacyl synthase to a malonyl decarboxylase by replacement of the active-site cysteine with glutamine. Biochemistry. 1999;38(36):11643-50.

283. Ghanevati M, Jaworski JG. Engineering and mechanistic studies of the Arabidopsis FAE1 β-ketoacyl-CoA synthase, FAE1 KCS. Eur J Biochem. 2002;269(14):3531-9.

284. Naganuma T, Kihara A. Two modes of regulation of the fatty acid elongase ELOVL6 by the 3-ketoacyl-CoA reductase KAR in the fatty acid elongation cycle. PLoS One. 2014;9(7):e101823.

285. Ghanevati M, Jaworski JG. Active-site residues of a plant membrane-bound fatty acid elongase beta-ketoacyl-CoA synthase, FAE1 KCS. Biochim Biophys Acta. 2001;1530(1):77-85.

286. Aslanidis C, de Jong PJ. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. 1990;18(20):6069-74.

287. Sonnhammer EL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol. 1998;6:175-82.

288. Ishida T, Kinoshita K. PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res. 2007;35(Web Server issue):W460-W4.

289. Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res. 2003;31(13):3701-8.

290. Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. 2013;41(Web Server issue):W349-57.

291. Kotov V, Bartels K, Veith K, Josts I, Subhramanyam UKT, Günther C, et al. High-throughput stability screening for detergent-solubilized membrane proteins. Sci Rep. 2019;9(1):10379.

292. Drew D, Newstead S, Sonoda Y, Kim H, von Heijne G, Iwata S. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae. Nat Protoc. 2008;3(5):784-98.

293. Gao K, Oerlemans R, Groves MR. Theory and applications of differential scanning fluorimetry in early-stage drug discovery. Biophys Rev. 2020;12(1):85-104.

294. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. Pac Symp Biocomput. 2002;7:310-22.

295. Uchański T, Masiulis S, Fischer B, Kalichuk V, López-Sánchez U, Zarkadas E, et al. Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. Nat Methods. 2021;18:60-8.

296. Bloch JS, Mukherjee S, Kowal J, Filippova EV, Niederer M, Pardon E, et al. Development of a universal nanobody-binding Fab module for fiducial-assisted cryo-EM studies of membrane proteins. Proc Natl Acad Sci U S A. 2021;118(47).

297. Wu X, Rapoport TA. Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies). Proc Natl Acad Sci U S A. 2021;118(41).

298. Pardon E, Laeremans T, Triest S, Rasmussen SGF, Wohlkönig A, Ruf A, et al. A general protocol for the generation of Nanobodies for structural biology. Nat Protoc. 2014;9(3):674-93.

299. Zimmermann I, Egloff P, Hutter CA, Arnold FM, Stohler P, Bocquet N, et al. Synthetic single domain antibodies for the conformational trapping of membrane proteins. elife. 2018;7:e34317.

300. Stengele I, Bross P, Garcés X, Giray J, Rasched I. Dissection of functional domains in phage fd adsorption protein. Discrimination between attachment and penetration sites. J Mol Biol. 1990;212(1):143-9.

301. McCafferty J, Griffiths AD, Winter G, Chiswell DJ. Phage antibodies: filamentous phage displaying antibody variable domains. Nature. 1990;348:552-4.

302. Hoogenboom HR, Griffiths AD, Johnson KS, Chiswell DJ, Hudson P, Winter G. Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. Nucleic Acids Res. 1991;19(15):4133-7.

303. Hoogenboom HR. Overview of antibody phage-display technology and its applications. Methods Mol Biol. 2002;178:1-37.

304. Zimmermann I, Egloff P, Hutter CAJ, Kuhn BT, Bräuer P, Newstead S, et al. Generation of synthetic nanobodies against delicate proteins. Nat Protoc. 2020;15(5):1707-41.

305. Le Bon C, Michon B, Popot JL, Zoonens M. Amphipathic environments for determining the structure of membrane proteins by single-particle electron cryo-microscopy. Q Rev Biophys. 2021;54:e6.

306. Chae PS, Rasmussen SG, Rana RR, Gotfryd K, Kruse AC, Manglik A, et al. A new class of amphiphiles bearing rigid hydrophobic groups for solubilization and stabilization of membrane proteins. Chemistry. 2012;18(31):9485-90.

307. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods. 2017;14(3):290-6.

308. De Genst E, Silence K, Decanniere K, Conrath K, Loris R, Kinne J, et al. Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. Proc Natl Acad Sci U S A. 2006;103(12):4586-91.

309. Ishida H, Asami J, Zhang Z, Nishizawa T, Shigematsu H, Ohto U, et al. Cryo-EM structures of Toll-like receptors in complex with UNC93B1. Nat Struct Mol Biol. 2021;28(2):173-80.

310. Cater RJ, Chua GL, Erramilli SK, Keener JE, Choy BC, Tokarz P, et al. Structural basis of omega-3 fatty acid transport across the blood–brain barrier. Nature. 2021;595:315-9.

311. Wood CAP, Zhang J, Aydin D, Xu Y, Andreone BJ, Langen UH, et al. Structure and mechanism of blood–brain-barrier lipid transporter MFSD2A. Nature. 2021;596:444-8.

312. Pascoa TC, Pike ACW, Bushell SR, Quigley A, Mukhopadhyay SMM, Burgess-Brown NA, et al. Major facilitator superfamily domain-containing protein 10 (MFSD10) A Target Enabling Package (TEP) (Version 1) [Data set]. Zenodo. 2021. doi: 10.5281/ZENODO.5126807

313. Lee WH. Open Access Target Validation Is a More Efficient Way to Accelerate Drug Discovery. PLoS Biol. 2015;13(6):e1002164.

314. Bradley R, Anthony, Echalier A, Fairhead M, Strain-Damerell C, Brennan P, Bullock N, Alex, et al. The SGC beyond structural genomics: redefining the role of 3D structures by coupling genomic stratification with fragment-based discovery. Essays Biochem. 2017;61(5):495-503.

315. Superti-Furga G, Lackner D, Wiedmer T, Ingles-Prieto A, Barbosa B, Girardi E, et al. The RESOLUTE consortium: unlocking SLC transporters for drug discovery. Nat Rev Drug Discov. 2020;19(7):429-30.

316. Bland ML. Measurement of Carbon Dioxide Production from Radiolabeled Substrates in Drosophila melanogaster. J Vis Exp. 2016(112):54045.

317. Turpin-Nolan SM, Brüning JC. The role of ceramides in metabolic disorders: when size and localization matters. Nat Rev Endocrinol. 2020;16(4):224-33.

318. Park JW, Park WJ, Futerman AH. Ceramide synthases as potential targets for therapeutic intervention in human diseases. Biochim Biophys Acta. 2014;1841(5):671-81.

319. Parker JL, Deme JC, Kuteyi G, Wu Z, Huo J, Goldman ID, et al. Structural basis of antifolate recognition and transport by PCFT. Nature. 2021;595:130-4.

320. Coupland CE, Andrei SA, Ansell TB, Carrique L, Kumar P, Sefer L, et al. Structure, mechanism, and inhibition of Hedgehog acyltransferase. Mol Cell. 2021;81(24):5025-38.e10.

321. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. 2021;38(7):3022-7.