Structural & Functional Studies on Cystinosin, the Proton-Coupled Cystine Transporter

Mark Löbel
Department of Biochemistry
Christ Church

Supervisors:
Prof. Simon Newstead
Prof. Francis Barr

A thesis submitted for the degree of
Doctor of Philosophy

2024

''I should not like to leave an impression that all structural problems can be settled by X-ray analysis or that all crystal structures are easy to solve. I seem to have spent much more of my life not solving structures than solving them.''

– Dorothy Crowfoot Hodgkin (1965)

Acknowledgements

This thesis simply would not have come together without the support and encouragement of many incredible people. I am truly grateful to each and every one of you.

First and foremost, I would like to express my deepest gratitude to Prof. Simon Newstead and Dr. Joanne Parker, for your mentorship and invaluable guidance throughout my doctoral journey. Your patient encouragement, insightful feedback, and dedication to scientific excellence have shaped not only the course of this research but also my growth as a scientist. Simon and Jo, your euphoria about science and your remarkable efficiency in the lab remain unmatched and have been a constant source of motivation for me.

I would also like to thank the colleagues who contributed their expertise and time to various parts of this work. I appreciate your collaboration and support. In no particular order:

- Prof. Simon Newstead – for crystal harvesting and initial model building.
- Dr. Joanne Parker – for your help with the cystine-bound structure and for providing the GST-BirA and PNGase F enzymes.
- Sacha Salphati & Prof. Stephen Tucker – for conducting the electrophysiological analyses.
- Dr. Kamel El Omari & Dr. Armin Wagner – for carrying out the S-SAD experiments.
- Dr. Takafumi Kato – for your help with the structural studies using cryo-EM.
- Dr. Jiandong Huo & Dr. Raymond Owens – for carrying out the llama immunization to generate nanobodies.
- Dr. Edward Lowe & Dr. Alicia Churchill-Angus – for your help with crystal data analysis.
- Dr. Feng Qu – for providing PepT*So*-Avi-biotin.
- Dr. Abraham Oluwole – for your native mass spectrometry expertise.
- Dr. Gabriel Kuteyi, Dr. Oliver Adams, Sigurbjörn Markússon & Abby Lin Yifei – for your support with mammalian cell culture.

Furthermore, I am grateful to Christ Church, the Clarendon Fund, and the Oxford Medical Research Council for kindly funding my DPhil.

Next, I want to thank my friends, especially Max, Siggi, Feng, Oli, Dimitris, Bjarne, Daniel, and Owen, for making my time in Oxford truly special. Thank you for the camaraderie, for all the

fruitful – and not-so-fruitful – discussions, and for the constant encouragement along the way. Thank you for all the adventures we shared: the countless tennis sessions with Max, winning the English Handball University Championships and Varsity Match with Siggi, regular squash battles with Owen, and the many trips on the water – whether in punts or kayaks – especially with Bjarne. I will always remember the exquisite formal dinners, pub trips, and board game nights, along with countless other moments outside the lab that offered a welcome change of pace from the daily grind of pursuing a doctoral degree. Also, a special shout-out to the lunch team.

Finally, I owe a huge debt of gratitude to my family and friends back home. Your unwavering support, encouragement, and sacrifices over the years have been my foundation. Knowing you were behind me every step of the way gave me strength, especially during the tough times. This achievement is as much yours as it is mine.

Abstract

The lysosome is a major signalling centre within the cell and an important regulator of metabolism through its regulation of mTORC1 activity. A key function of the lysosome is to degrade proteins to their constituent amino acids and export these to the cytoplasm for metabolic recycling or sensing. In comparison to amino acid transport across the plasma membrane, transport across organellar membranes is currently poorly understood at both the cellular and molecular levels. Several physiological disorders are linked to mutations in lysosomal transporters, sparking a clear clinical and biochemical drive to understand their structure and mechanism of action. My thesis presents a series of structures of the lysosomal cystine transporter cystinosin (CTNS), captured in the ligand-bound inward-open and outward-open states. CTNS (SLC66A4), a member of the PQ-loop family, is responsible for exporting cystine from the lysosome using the pre-established proton gradient. Mutations in the *CTNS* gene lead to the accumulation of cystine in the lysosome, resulting in impaired lysosomal function, which leads to cystinosis, a rare disease that causes progressive glomerular damage and end-stage renal failure during mid-childhood. Until now, the molecular basis for proton-coupled cystine transport *via* CTNS remained elusive, hampering efforts to develop novel therapeutics for cystinosis and understand the role of CTNS in lysosomal homeostasis. The structural analysis of the transporter undertaken in this work reveals novel features in CTNS that are not observed in other PQ-loop proteins, such as the presence of two salt bridges that couple cystine transport to the translocation of protons. Using a combination of *in vitro* liposome assays targeting the *Arabidopsis thaliana* homologue of the transporter and *in vivo* experiments on human CTNS, I propose a mechanism for specific cystine recognition and proton-coupled transport. In contrast to current models regarding PQ-loop protein-mediated transport, the data indicate a more asymmetrical mode of action, exposing an interface which may allow CTNS to form functional dimers. Finally, the presented structures, coupled with the functional analysis of CTNS, provide a framework for understanding how specific hereditary defects cause cystinosis.

Table of Contents

Acknowledgements iii

Abstract v

Table of Contents vii

List of Figures xiii

List of Tables xvii

List of Abbreviations & Acronyms xix

1 Introduction 1

1.1 Lysosomal Storage Diseases . . . 1
1.2 Cystinosis . . . 2
1.3 Cystinosin (CTNS) . . . 4
1.4 Related Transporters & Receptors . . . 7
1.5 Nanobodies & Sybodies as Crystallization Chaperones . . . 9
1.6 Aims & Objectives . . . 11

2 Materials & Methods 13

2.1 Construct Design . . . 13
2.1.1 Expression in Yeast Cells . . . 13
2.1.2 Expression in Mammalian Cells . . . 14
2.1.3 Expression in Bacterial Cells . . . 14
2.2 Recombinant Protein Production . . . 18
2.2.1 Yeast Transformation . . . 18
2.2.2 Expression Trials in Yeast Cells . . . 18
2.2.3 Small-Scale Membrane Fraction Preparation . . . 18
2.2.4 Fluorescent Size Exclusion Chromatography . . . 19
2.2.5 Expression in Yeast Cells . . . 19
2.2.6 Membrane Fraction Preparation . . . 19
2.2.7 Membrane Protein Purification . . . 20

2.2.8 Expression in Bacterial Cells . . . 20
2.2.9 Purification of Sybodies & Nanobodies . . . 21
2.3 Membrane Protein Characterization . . . 22
2.3.1 Heat-Fluorescent Size Exclusion Chromatography . . . 22
2.3.2 Polyacrylamide Gel Electrophoresis & Western Blotting . . . 22
2.3.3 Circular Dichroism Spectroscopy . . . 22
2.3.4 Nano Differential Scanning Fluorimetry . . . 23
2.4 Generation of Sybodies & Nanobodies . . . 23
2.4.1 Target Protein Biotinylation . . . 23
2.4.2 Sybody Screening . . . 24
2.4.3 Nanobody Screening . . . 28
2.4.4 Size Exclusion Chromatography Co-Elutions . . . 31
2.4.5 Pulldowns Using His-Tagged Binders . . . 31
2.4.6 Bio-Layer Interferometry . . . 31
2.5 Structural Analysis of CTNS . . . 32
2.5.1 Crystallization . . . 32
2.5.2 Heavy Atom Soaking . . . 32
2.5.3 X-Ray Diffraction Data Collection . . . 33
2.5.4 X-Ray Diffraction Data Processing . . . 33
2.5.5 Crystal Structure Determination . . . 33
2.5.6 Cryo-EM Grids Preparation & Screening . . . 34
2.5.7 Cryo-EM Data Collection, Processing & Model Building . . . 34
2.6 Functional Analysis of CTNS . . . 35
2.6.1 Cystine Binding Assay . . . 35
2.6.2 Proteoliposome Reconstitution . . . 36
2.6.3 Transport Assay Using Pyranine . . . 36
2.6.4 Radioactive Transport Assay . . . 37

3 Screening & Characterization of CTNS 39
3.1 Homologue Screening . . . 39
3.1.1 Construct Comparison . . . 39
3.1.2 Expression Trials . . . 41
3.2 Preparation of *At*CTNS for Structural & Functional Studies . . . 44
3.2.1 Recombinant Protein Production . . . 44
3.2.2 Proteoliposome Reconstitution . . . 46
3.3 Biophysical Characterization of *At*CTNS . . . 47
3.3.1 Thermal Stability Determination . . . 47
3.3.2 Ligand Binding Analysis . . . 49

4 Crystal Trials Using *At*CTNS **53**
4.1 Choice of Cubic Phase Host Lipid . . . 53
4.2 Crystallization . . . 54
4.3 Data Collection & Processing . . . 54
4.4 Data Quality Analysis . . . 57
4.5 Structure Solution Attempts Using Homology Models . . . 59

5 Generation of Sybodies & Nanobodies against *At*CTNS **65**
5.1 Purification & Biotinylation of *At*CTNS-Avi . . . 65
5.2 Sybodies . . . 67
5.2.1 Screening . . . 67
5.2.2 Biochemical & Biophysical Analyses of *At*CTNS-Sybody Interactions . 71
5.2.3 Repeated Screening at pH 5.5 . . . 75
5.3 Llama Derived Nanobodies . . . 76
5.3.1 Screening . . . 76
5.3.2 Biochemical & Biophysical Analyses of *At*CTNS-Nanobody Interactions 78

6 Crystal Optimization Using Sybody- & Nanobody-Bound *At*CTNS Complexes **81**
6.1 Co-Crystal Trials Using Sybodies . . . 81
6.1.1 Purification of Tag-Free Sybodies . . . 81
6.1.2 Complex Formation & Co-Crystallization . . . 83
6.1.3 Data Collection, Processing & Analysis . . . 84
6.1.4 Structure Solution Attempt by Molecular Replacement . . . 86
6.2 Co-Crystal Trials Using Nanobodies . . . 88
6.2.1 Nanobody Production & Complex Formation . . . 88
6.2.2 Co-Crystallization & Data Analysis . . . 89
6.2.3 Structure Solution Attempt by Molecular Replacement . . . 91
6.3 Co-Crystal Trials Using Multiple Binders . . . 91
6.3.1 Complex Formation & Co-Crystallization . . . 91
6.3.2 Data Collection, Processing & Analysis . . . 93
6.3.3 Structure Solution Attempt by Molecular Replacement . . . 96

7 *At*CTNS Structure Solution & Co-Crystal Structures **99**
7.1 Overview . . . 99
7.2 Structural Solution of *At*CTNS . . . 101
7.2.1 Anomalous Data Collection, Processing & Analysis . . . 101
7.2.2 Initial Phase Determination & Model Building . . . 103
7.2.3 Structure Refinement . . . 104
7.2.4 Structure Validation . . . 108
7.2.5 Structure of *At*CTNS in the Apo & Inward-Open States . . . 109

7.3 *At*CTNS Structures from Unsuccessful Co-Crystal Trials Using Ligand 111
7.3.1 Higher-Resolution Structure of *At*CTNS Confirms Previous Model . . 113
7.3.2 High-Resolution Structure of *At*CTNS Reveals Dimeric Arrangement . 118
7.3.3 High-Resolution S-SAD Data Validate *At*CTNS Model 122
7.4 Lower-Resolution Ligand-Bound Structure of *At*CTNS 124

8 Structural Studies on Human CTNS 129
8.1 *At*CTNS Dimer as a Potential Template for *Hs*CTNS Dimerization 129
8.2 Screening & Characterization of *Hs*CTNS Constructs 131
8.2.1 Expression Trials 131
8.2.2 Purification Trials & Biophysical Characterization 133
8.3 Generation of Sybodies against *Hs*CTNS(Δ2–115) 137
8.3.1 Purification & Biotinylation of *Hs*CTNS(Δ2–115)-Avi 137
8.3.2 Screening 138
8.3.3 Biochem. & Biophys. Analyses of *Hs*CTNS(Δ2–115)-Sybody Interactions 139
8.4 Structural Trials Using Cryo-EM 143
8.4.1 Sample Preparation 144
8.4.2 Grid Preparation, Data Collection & Processing 145
8.4.3 Model Building, Refinement & Validation 148
8.4.4 *Hs*CTNS(Δ2–115) Dimer Structure 149
8.4.5 Potential Full-Length *Hs*CTNS Dimer 151

9 Structural & Functional Analysis of CTNS 153
9.1 Structure of CTNS 153
9.1.1 Architectural Comparison with Related Proteins 155
9.1.2 Structural Characteristics of CTNS 159
9.2 Assay Development to Address CTNS Function 163
9.2.1 Cystine Binding Assay 163
9.2.2 Transport Assay Using Pyranine 163
9.2.3 Radioactive Transport Assay 165
9.2.4 Electrophysiological TEVC Assay 166
9.3 Functional Characterization of CTNS 166
9.3.1 Molecular Basis for Cystine Recognition 168
9.3.2 Molecular Basis for Cystine Transport 171
9.3.3 Mechanism of Cystine Transport 174
9.3.4 Disease-Causing Mutation Mapping 178

10 Conclusion 181

Appendix **183**
Construct Design . . . 183
Generation of *At*CTNS-Specific Sybodies . . . 191
Generation of *At*CTNS-Specific Nanobodies . . . 195
Crystallography Using Sybody- & Nanobody-Bound *At*CTNS Complexes . . . 196
Generation of *Hs*CTNS(Δ2–115)-Specific Sybodies . . . 205
Functional Analysis of CTNS . . . 213
Publications . . . 214

Bibliography **235**

List of Figures

1.1 Cystine Structure & Crystals . . . 2
1.2 Metabolic Cystine Pathway . . . 3
1.3 CTNS Isoforms . . . 4
1.4 CTNS-Mediated Cell Biological Effects . . . 7
1.5 Alternating Access Glucose Transport *via Lb*SemiSWEET . . . 8
1.6 Nanobodies as Crystallization Chaperones . . . 9
1.7 Sybodies as Crystallization Chaperones. . . . 10

3.1 Sequence Alignment of CTNS Homologues . . . 40
3.2 Expression Trials of CTNS Homologues . . . 42
3.3 *At*CTNS Purification . . . 45
3.4 *At*CTNS Reconstitution into Liposomes . . . 46
3.5 Thermal Stability Analyses of *At*CTNS . . . 48
3.6 *At*CTNS Ligand Binding Analysis Using NanoDSF . . . 50

4.1 LCP Crystals of *At*CTNS . . . 54
4.2 Best Diffraction Patterns of *At*CTNS . . . 57
4.3 Homology Models for Molecular Replacement . . . 61
4.4 Exemplary Electron Density Maps After MR . . . 62

5.1 Purification & Biotinylation of *At*CTNS-Avi . . . 66
5.2 ELISA of *At*CTNS-Specific Sybodies . . . 68
5.3 Phylogenetic Analysis of *At*CTNS-Specific Concave (G) Sybodies . . . 69
5.4 Phylogenetic Analysis of *At*CTNS-Specific Loop (B) Sybodies . . . 70
5.5 α-*At*CTNS Sybody CDR Sequence Comparison . . . 71
5.6 SEC Co-Elutions of *At*CTNS & Sybodies . . . 72
5.7 Pulldowns of *At*CTNS Using His-Tagged Sybodies . . . 73
5.8 *At*CTNS-Sybody Binding Analysis Using BLI . . . 74
5.9 K_D Analysis of Important Sybodies Using BLI . . . 75
5.10 ELISA & Phylogenetic Analysis of *At*CTNS-Specific Nanobodies . . . 77
5.11 α-*At*CTNS Nanobody CDR Sequence Comparison . . . 78
5.12 Purification of His-Tagged Nanobodies & *At*CTNS Pulldowns . . . 79

5.13 *At*CTNS-Nanobody Binding Analysis Using BLI 79
5.14 K_D Analysis of Important Nanobodies Using BLI 80

6.1 *At*CTNS-Sybody Complex Formation for Co-Crystallization 82
6.2 LCP Crystals of *At*CTNS in Complex with Syb39 83
6.3 Typical Diffraction Patterns of *At*CTNS-Syb39 85
6.4 Twinning Detection Using the *L*-Test 86
6.5 Electron Density After MR Using a Sybody Frame 87
6.6 *At*CTNS-Nanobody Complex Formation for Co-Crystallization 88
6.7 LCP Crystals of *At*CTNS in Complex with Nb4 89
6.8 Best Diffraction Patterns of *At*CTNS-Nb4 91
6.9 Complex Formation & Co-Crystallization of *At*CTNS-Syb39-Nb4 92
6.10 Typical Diffraction Patterns of *At*CTNS-Syb39-Nb4 93
6.11 Electron Density After MR Using Sybody & Nanobody Frames 96

7.1 Energy Scans of Anomalous Scatterers S & Hg 101
7.2 Placement of the First *At*CTNS Model 104
7.3 *At*CTNS Model & Map Refinements 105
7.4 Essential Electron Density Maps of *At*CTNS-Syb39-Nb4 at 2.65 Å 106
7.5 Anomalous Difference Maps of *At*CTNS-Syb39-Nb4 107
7.6 2.65 Å Co-Crystal Structure of *At*CTNS-Syb39-Nb4 110
7.7 Complex Formation & Co-Crystallization Using Ligand-Supplemented Complexes 112
7.8 Typical Diffraction Patterns of *At*CTNS-Nb1-Nb4 113
7.9 Essential Electron Density Maps of *At*CTNS-Nb1-Nb4 at 2.33 Å 115
7.10 2.33 Å Co-Crystal Structure of *At*CTNS-Nb1-Nb4 116
7.11 Typical Diffraction Pattern of *At*CTNS-Syb1-Nb4 117
7.12 Essential Electron Density Maps of *At*CTNS-Syb1-Nb4 at 1.61 Å 119
7.13 1.61 Å Co-Crystal Structure of *At*CTNS-Syb1-Nb4 121
7.14 Anomalous Difference Map of *At*CTNS-Syb1-Nb4 124
7.15 Essential Electron Density Maps of *At*CTNS-Syb39-Cystine at 3.47 Å 125
7.16 3.47 Å Co-Crystal Structure of Ligand-Bound *At*CTNS in Complex with Syb39 . 126

8.1 *At*CTNS Crystal Dimer Interface & Comparison with *Hs*CTNS 130
8.2 Oligomeric State of *At*CTNS in Solution 131
8.3 Expression Trials of Human CTNS Constructs in Yeast Cells 132
8.4 Expression Trials of Human CTNS Constructs in Mammalian Cells 132
8.5 Purification & Thermal Stability Analysis of *Hs*CTNS Expressed in Yeast Cells . 133
8.6 *Hs*CTNS(Δ2–115) Purification Detergent Screening 134
8.7 Thermal Stability Analyses of *Hs*CTNS(Δ2–115) 135
8.8 *Hs*CTNS(Δ2–115) Split Species Purification & Oligomeric State Analyses 136

8.9 Purification & Biotinylation of *Hs*CTNS(Δ2–115)-Avi . . . 137
8.10 ELISA of *Hs*CTNS(Δ2–115)-Specific Sybodies . . . 139
8.11 Phylogenetic Analysis of *Hs*CTNS(Δ2–115)-Specific Sybodies . . . 140
8.12 α-*Hs*CTNS(Δ2–115) Sybody CDR Sequence Comparison . . . 141
8.13 Purification & Biochemical Analyses of Tagged α-*Hs*CTNS(Δ2–115) Sybodies . . . 142
8.14 *Hs*CTNS(Δ2–115) Dimer-Tagged Sybody Binding Analysis Using BLI . . . 144
8.15 *Hs*CTNS(Δ2–115)-Sybody-Myc-His Complex Formation for Cryo-EM Trials . . . 145
8.16 Cryo-EM Grid Screening . . . 146
8.17 Cryo-EM Data Processing Workflow for *Hs*CTNS(Δ2–115)-Syb30-Myc-His . . . 147
8.18 Essential Cryo-EM Maps of *Hs*CTNS(Δ2–115)-Syb30-Myc-His at 2.45 Å . . . 148
8.19 *Hs*CTNS(Δ2–115) Dimer Structure & Interface . . . 150
8.20 Initial Dimerization Analysis of Full-Length *Hs*CTNS . . . 151

9.1 *At*CTNS Structure in the Inward-Open Conformation . . . 154
9.2 *Hs*CTNS(Δ2–115) Structure in the Outward-Open Conformation . . . 155
9.3 Structural Comparison Using *At*CTNS in the Inward-Open State . . . 156
9.4 Structural Comparison Using *Hs*CTNS(Δ2–115) in the Outward-Open State . . . 157
9.5 PQ-Motif Prolines as Molecular Hinges . . . 159
9.6 PQ-Motif Glutamines Seal CTNS on the Cytoplasmic Side . . . 160
9.7 Unique Salt Bridge Seals CTNS on the Luminal Side . . . 161
9.8 Charged Binding Site Residues of CTNS . . . 162
9.9 Assays Used to Study CTNS Function . . . 164
9.10 Purification & Reconstitution of *At*CTNS Variants . . . 165
9.11 Annotated Sequence Alignment of CTNS Homologues . . . 167
9.12 Functional Characterization of CTNS . . . 168
9.13 Functional Characterization of Binding Site Variants . . . 170
9.14 Functional Characterization of PQ-Motif Proline Variants . . . 172
9.15 Functional Characterization of Salt Bridge Residue Variants . . . 173
9.16 Alternating Access Cystine Transport *via* CTNS . . . 174
9.17 Proposed CTNS Transport Mechanism . . . 175
9.18 Water Networks in the 1.61 Å *At*CTNS-Syb1-Nb4 Co-Crystal Structure . . . 176
9.19 Model for the Asymmetric Transition Between Conformational States . . . 177
9.20 Structural Context of Cystinosis-Causing Mutations . . . 178

A.1 SEC Traces of Myc-His-Tagged α-*At*CTNS Sybodies . . . 193
A.2 SEC Co-Elutions of *At*CTNS & Myc-His-Tagged Sybodies . . . 194
A.3 SEC Traces of His-Myc-Tagged α-*At*CTNS Nanobodies . . . 195
A.4 Best Diffraction Patterns of *At*CTNS-Syb39-Nb4 (Hg-SAD) . . . 198
A.5 Ramachandran Plots: *At*CTNS-Syb39-Nb4 Model . . . 201
A.6 Ramachandran Plots: *At*CTNS-Nb1-Nb4 Model . . . 202

A.7 Ramachandran Plots: *At*CTNS-Syb1-Nb4 Model 203
A.8 Ramachandran Plots: *At*CTNS-Syb39-Cystine Model 204
A.9 SEC Traces of Myc-His-Tagged α-*Hs*CTNS(Δ2–115) Sybodies 209
A.10 Ramachandran Plots: *Hs*CTNS(Δ2–115)-Syb30-Myc-His Model 212
A.11 *At*CTNS Mutants Cystine Binding NanoDSF Assay Data 213

List of Tables

1.1 List of Cystinosis-Causing Missense Mutations . . . 5

2.1 Constructs for Yeast Cell Culture . . . 15
2.2 Constructs for Mammalian Cell Culture . . . 16
2.3 Constructs for Bacterial Cell Culture . . . 17

3.1 Main Detergents Used in This Research . . . 43
3.2 Main Lipids Used in This Research . . . 47

4.1 X-Ray Data Collection & Processing Statistics of *At*CTNS . . . 56

6.1 X-Ray Data Collection & Processing Statistics of *At*CTNS-Syb39 . . . 84
6.2 X-Ray Data Collection & Processing Statistics of *At*CTNS-Nb4 . . . 90
6.3 X-Ray Data Collection & Processing Statistics of *At*CTNS-Syb39-Nb4 . . . 94

7.1 Overview of *At*CTNS Co-Crystal Structures . . . 100
7.2 Anomalous Data Statistics of *At*CTNS-Syb39-Nb4 . . . 102
7.3 Anomalous Peaks of *At*CTNS-Syb39-Nb4 (S-SAD & Hg-SAD) . . . 108
7.4 X-Ray Data Collection & Processing Statistics of *At*CTNS-Nb1-Nb4 . . . 114
7.5 X-Ray Data Collection & Processing Statistics of *At*CTNS-Syb1-Nb4 . . . 118
7.6 Anomalous Data of *At*CTNS-Syb1-Nb4 (S-SAD) . . . 123
7.7 Crystal Structure Refinement & Validation Statistics . . . 127

8.1 Phage Display Enrichment Factors for *Hs*CTNS(Δ2–115) Sybody Screening . . . 138
8.2 Cryo-EM Structure Refinement & Validation Statistics . . . 149

9.1 Sequence Identities among PQ-Loop & Related Proteins . . . 158
9.2 Annotated List of Cystinosis-Causing Missense Mutations . . . 179

A.1 Gene Strings . . . 183
A.2 Primers Used for Sequencing & Cloning . . . 184
A.3 Construct Sequences . . . 187
A.4 Additional Constructs for Bacterial Cell Culture . . . 189
A.5 ELISA Results & Analysis of *At*CTNS-Specific Sybodies . . . 191

A.6 ELISA Results & Analysis of *At*CTNS-Specific Nanobodies 195
A.7 X-Ray Data Collection & Processing Statistics of *At*CTNS-Syb39-Nb4 (S-SAD) . 196
A.8 X-Ray Data Collection & Processing Statistics of *At*CTNS-Syb39-Nb4 (Hg-SAD) 197
A.9 X-Ray Data Collection & Processing Statistics of *At*CTNS-Syb1-Nb4 (S-SAD) . . 199
A.10 X-Ray Data Collection & Processing Statistics of *At*CTNS-Syb39-Cystine 200
A.11 ELISA Results & Analysis of *Hs*CTNS(Δ2–115)-Specific Sybodies 205
A.12 Further ELISA Results of Untested *Hs*CTNS(Δ2–115)-Specific Sybodies 207
A.13 *Hs*CTNS(Δ2–115)-Sybody Binding Analysis Overview 211

List of Abbreviations & Acronyms

3C	human rhinovirus 3C protease
7NA	mutations N36A, N41A, N51A, N66A, N84A, N104A, and N107A
AA	amino acid
AC	affinity chromatography
AEBSF	4-(2-aminoethyl)benzenesulfonyl fluoride
Amp	ampicillin
At	*Arabidopsis thaliana*
AU	absorbance units
A_λ	absorbance at wavelength λ in nm
BF	bright-field
BLI	bio-layer interferometry
BN	blue native
BSA	bovine serum albumin
CAM	chloramphenicol
CD	circular dichroism
CDR	complementarity-determining region
CFU	colony-forming units
CHL	cholesterol
CHS	cholesteryl hemisuccinate
CMA	chaperone-mediated autophagy
CMC	critical micelle concentration
cryo-EM	cryogenic electron microscopy
cryo-TEM	cryogenic transmission electron microscope
CTF	contrast transfer function
CTNS	cystinosin
CV	column volumes
DDM	*n*-dodecyl-β-D-maltopyranoside
DLS	Diamond Light Source
DM	*n*-decyl-β-D-maltopyranoside
DME	dimethyl ether
DMSO	dimethyl sulfoxide
Dr	*Danio rerio*
E	elution

Ec	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
ER	endoplasmic reticulum
Fab	antigen-binding fragment
Fc	crystallizable fragment
FSC	Fourier shell correlation
FSEC	fluorescent size exclusion chromatography
FT	flow-through
FX	fragment exchange
F_X	fluorescence of molecule X
GFP	green fluorescent protein
Gg	*Gallus gallus*
GPCR	G protein-coupled receptor
GSH	glutathione
GSSG	glutathione disulfide
GST	glutathione-*S*-transferase
HBS	HEPES-buffered saline
HCAb	heavy chain antibody
HEK	human embryonic kidney
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
His	polyhistidine
HR	homologous recombination
HRP	horseradish peroxidase
Hs	*Homo sapiens*
Ig	immunoglobulin
I_m	membrane current
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
ISPyB	Information System for Protein Crystallography Beamline
Kan	kanamycin
K_D	equilibrium dissociation constant
kDa	kilodalton
KDELR	KDEL receptor
K_M	Michaelis-Menten constant
k_{off}	dissociation rate constant
k_{on}	association rate constant
LAMP2A	lysosome-associated membrane protein 2A
LB	Lennox Broth
Lb	*Leptospira biflexa*
LC	loading control
LCP	lipidic cubic phase

LLG	log-likelihood gain
LMNG	lauryl maltose neopentyl glycol
LSD	lysosomal storage disease
LTM	lysosomal targeting motif
MAG	monoacylglycerol
MBP	maltose-binding protein
MBS	MES-buffered saline
MES	2-(*N*-morpholino)ethanesulfonic acid
Mm	*Mus musculus*
MME	monomethyl ether
MO	monoolein
mQ-H_2O	Milli-Q water
MR	molecular replacement
mTORC1	mechanistic target of rapamycin complex 1
MW	molecular weight
MWCO	molecular weight cut-off
N/A	not available
nanoDSF	nano differential scanning fluorimetry
Nb	nanobody
NC	negative control
NG	*n*-nonyl-β-D-glucopyranoside
Nm	*Neisseria mucosa*
NMR	nuclear magnetic resonance
nMS	native mass spectrometry
NTA	nitrilotriacetic acid
NTD	N-terminal domain
NU	non-uniform
O/N	overnight
OD_{600}	optical density at 600 nm wavelength
Os	*Oryza sativa*
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PD	phage display
PDB	protein data bank
PEG	polyethylene glycol
pelB	pectate lyase B
$PepT_{So}$	peptide transporter from *Shewanella oneidensis*
PFU	plaque-forming units
PNGase F	peptide:N-glycosidase F
POPC	1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine

POPG	1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol)
qPCR	quantitative polymerase chain reaction
RFU	relative fluorescence units
RMSD	root mean square deviation
ROS	reactive oxygen species
RPM	revolutions per minute
RT	room temperature
SAD	single-wavelength anomalous dispersion
scFv	single-chain variable fragment
SD	standard deviation
SDS	sodium dodecyl sulfate
SEC	size exclusion chromatography
SEM	standard error of the mean
SLC	solute carrier
SN	supernatant
SWEET	sugars will eventually be exported transporter
Syb	sybody
TB	Terrific Broth
TBS	Tris-buffered saline
TEV	tobacco etch virus protease
TEVC	two-electrode voltage clamp
TFEB	transcription factor EB
TFZ	translation-function *Z*-score
THB	three-helix bundle
TLS	translation, libration, and screw
TM	transmembrane
T_m	melting temperature
TMB	3,3',5,5'-tetramethylbenzidine
TMD	transmembrane domain
TOG	transporter–opsin–GPCR
UV	ultraviolet
v	valinomycin
V_c	command voltage
VD	vapour diffusion
V_M	Matthews coefficient
V_m	membrane potential
W	wash
WT	wild type
XRC	X-ray crystallography
β_2-AR	β_2-adrenergic receptor

CHAPTER 1

Introduction

Lysosomes were discovered in the early to mid-1950s after a number of hydrolytic enzymes were found in a previously undefined membrane-enclosed cell compartment (de Duve *et al.*, 1955). They were soon established as the ''digestive tract of the cell'' (de Duve, 1963), responsible for the degradation and recycling of unneeded macromolecules stemming from both inside and outside the cell – this includes the breakdown of proteins to their constituent amino acids and their subsequent export from the lysosomal lumen back to the cytosol using dedicated transporters. Whilst the degradative function was known for a long time, more recent discoveries have shown that lysosomes hold additional functions such as nutrient sensing, complex cell signalling, and gene regulation, among others, and have advanced our view of the lysosome to being a key regulator of metabolism in response to environmental cues (Ballabio & Bonifacino, 2020; Lawrence & Zoncu, 2019). In particular, lysosomes dynamically recruit the mechanistic target of rapamycin complex 1 (mTORC1) kinase to their surfaces and help activate it through the supply of amino acids, initiating a signalling cascade that promotes cell growth (Saxton & Sabatini, 2017). In the absence of amino acids, mTORC1 signalling halts, allowing transcription factor EB (TFEB) to translocate to the nucleus, where it promotes lysosomal biogenesis and autophagy, thus promoting the recycling of cellular components to meet energy needs (Napolitano & Ballabio, 2016). Abnormalities within this network obstruct cellular and organismal homeostasis and therefore result in an ever increasing number of medical conditions including lysosomal storage diseases (LSDs), metabolic and neurodegenerative abnormalities, and cancer (Ballabio & Bonifacino, 2020).

1.1 Lysosomal Storage Diseases

LSDs are defined by the gradual accumulation of diverse substrates inside lysosomes, ultimately leading to their dysregulation and malfunction (Platt *et al.*, 2018; Ballabio & Bonifacino, 2020). As rare inherited metabolic recessive diseases, LSDs are caused by mutation in lysosomal proteins, including resident proteases and glycosidases, as well as integral membrane proteins, such as transporters. Depending on the specific mutation, LSDs may present earlier or later in life – most are classified in different phenotypes, such as infantile and juvenile, based on the first display of symptoms; the disease severity generally worsens progressively. The specific cell biological effects caused by the substrate accumulation within lysosomes differ among

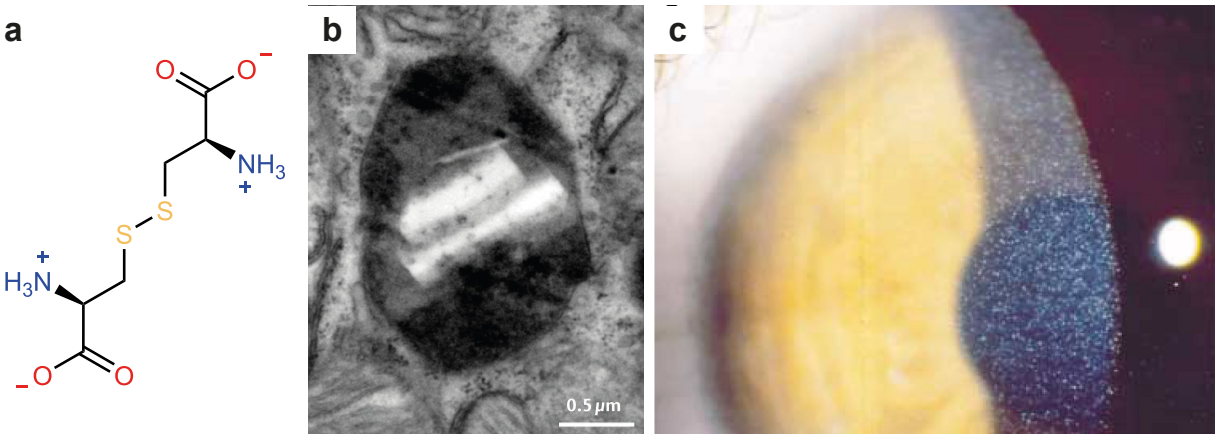


Fig. 1.1 – Cystine Structure & Crystals.
a, Chemical structure of L-cystine. **b**, Electron micrograph of a kidney proximal tubule cell of an 8-month-old *Ctns*-knockout mouse, showing cystine crystals inside an enlarged lysosome. Figure taken from Cherqui & Courtoy (2017). **c**, Corneal cystine crystals in the eyes of a 32-month-old boy. Figure taken from Gahl *et al.* (2002).

individual LSDs but often include impaired cellular transport and degradation, oxidative stress, dysregulation of immune responses, and cell death pathways (Platt *et al.*, 2018). Lysosomes are common in cells of all tissues, however, the impact of individual LSDs on the different organs varies. Most affect multiple organs but are defined by one specifically; the central nervous system is damaged in the majority of LSDs (Platt *et al.*, 2018).

1.2 Cystinosis

Cystinosis is an inherited LSD that occurs in approximately 1 in 100,000 to 200,000 live births (Gahl *et al.*, 2002). It is caused by defects within cystinosin (CTNS) (Town *et al.*, 1998), an integral membrane protein responsible for the transport of cystine, the oxidized dimer of cysteine (**Fig. 1.1 a**), out of lysosomes (Kalatzis *et al.*, 2001; see **Fig. 1.2** and **Sec. 1.3**). Cystinosis is characterized by the accumulation of cystine (Abderhalden, 1903) in lysosomes (Schulman *et al.*, 1969), which is caused by defective transport (Gahl *et al.*, 1982), and results in crystal formation (**Fig. 1.1 b**) and lysosomal dysfunction (Cherqui & Courtoy, 2017). Like most LSDs, cystinosis is a multi-systemic disorder affecting all tissues and organs, however, the first manifestations of cystinosis are usually made through renal damage (Gahl *et al.*, 2002). Typical signs include growth failure, polyuria, polydipsia, electrolyte imbalance, and dehydration, as well as the development of rickets, which are usually associated with vitamin D deficiency. In most cases, the so-called Fanconi syndrome, summarizing these symptoms, is developed within the first 6–12 months of life; only a small number of cases emerge at a juvenile age. Because renal insufficiency is frequently reached within 10 years of disease progression, kidney transplantations play an integral role in the survival of patients with nephropathic infantile cystinosis (Malekzadeh *et al.*, 1977; Gahl *et al.*, 2002). Deposition of cystine crystals in the

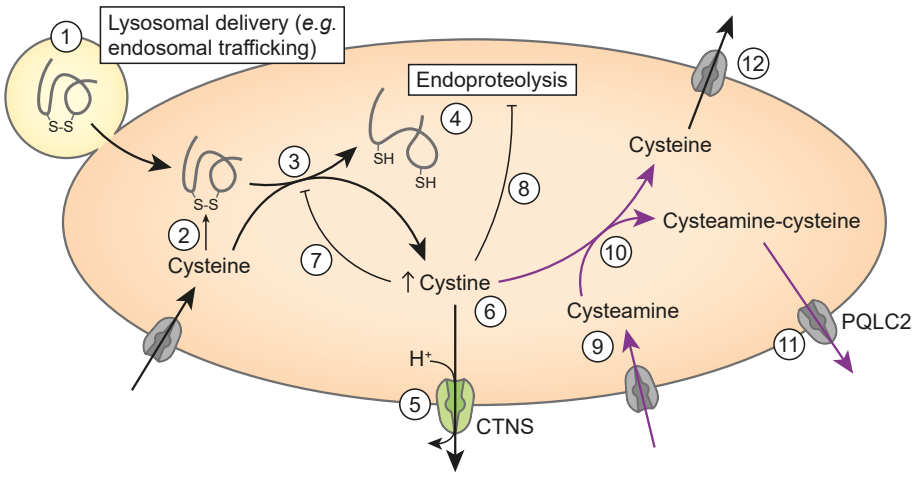


Fig. 1.2 – Metabolic Cystine Pathway.
Cystine originates from disulfide-bearing proteins (1) and is released after their degradation in the lysosome. This process depends on cysteine to reduce the disulfide bridges (2), which is necessary to unfold the protein (3), exposing the core for cleavage by cysteine proteases like cathepsin B (4). Under normal conditions (black pathway), cystine is readily exported by CTNS out of the lysosome (5), however, in cystinotic cells, CTNS function is abolished, and cystine accumulates within the lysosome (6). Cystine accumulation is thought to disturb cysteine-mediated protein reduction (7) as well as the function of relevant proteases, impairing endoproteolysis altogether (8). Cystinosis is usually treated with cysteamine (purple pathway), which enters the lysosome (9) and cleaves cystine, forming cysteamine-cysteine mixed disulfide and cysteine (10), both of which use distinct transporters to exit the lysosome (11 and 12). Figure and legend adapted from Cherqui & Courtoy (2017).

cornea (**Fig. 1.1 c**) also occurs early during the disease, causing photophobia and painful corneal erosions, potentially leading to retinal blindness (Gahl *et al.*, 2002). At a later stage, patients can develop hypothyroidism, hypogonadism, diabetes, myopathy, and deterioration of the central nervous system. Patients with nephropathic juvenile cystinosis have the same features, but disease progression is usually slowed; ocular cystinosis only affects the eyes without renal manifestations (Gahl *et al.*, 2002).

Early treatment of cystinosis is essential and is carried out through strictly clocked application of high doses of cysteamine, which is effective in depleting cells of cystine (Gahl *et al.*, 1987; mechanism described in **Fig. 1.2**). However, the drug is not always tolerated well by patients. The main problems are caused by the unpleasant taste and smell of cysteamine capsules, as well as their low bioavailability, which often cause nausea and vomiting, and a pungent body odour including bad breath, respectively (Gahl *et al.*, 2002). To improve tolerance, the development of a prodrug is the subject of current research (Frost *et al.*, 2016; current code name: CF10). Although early and extensive administration of cysteamine is effective at reducing the onset of cystinosis symptoms, it does not affect established renal Fanconi syndrome (Cherqui & Courtoy, 2017). This indicates that the lysosomal cystine transporter CTNS holds functions beyond cystine transport, which contribute to its pathogenesis (see **Sec. 1.3**). Cysteamine is also used in eye drops to dissolve corneal cystine crystals (Kaiser-Kupfer *et al.*, 1987).

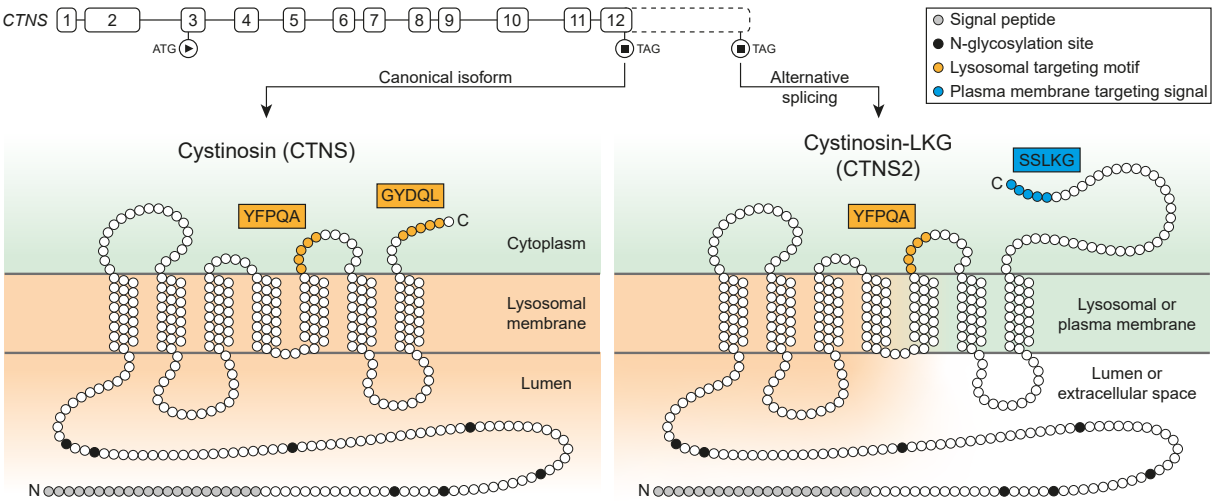


Fig. 1.3 – CTNS Isoforms.
Topology diagrams of the two human CTNS isoforms are shown, and their main features are defined. Figure adapted from Cherqui & Courtoy (2017).

1.3 Cystinosin (CTNS)

Cystinosis is caused by defects in the *CTNS* gene, which is located on the short arm of chromosome 17 (McDowell *et al.*, 1995). It codes for the 367-amino acid lysosomal cystine transporter cystinosin (CTNS) (Kalatzis *et al.*, 2001; **Fig. 1.3**), an integral membrane protein that is ubiquitously expressed in cells of all tissues (Cherqui & Courtoy, 2017). With no structural information available, CTNS was predicted to contain a long N-terminal segment, followed by seven transmembrane (TM) helices, and a short C-terminal tail (Cherqui *et al.*, 2001). The N-terminal segment includes seven asparagine-linked glycosylation sites, as well as a signal peptide, which ensures the correct orientation of the protein when it is inserted into the membrane during translation (von Heijne, 1990). Despite earlier reports that the signal peptide is uncleavable (Cherqui *et al.*, 2001), it was later shown that it is removed post-translationally, generating a new N-terminal end for CTNS starting with a serine at residue number 23 (Nevo *et al.*, 2017). The C-terminal end of CTNS bears a tyrosine-based lysosomal targeting motif (LTM) with sequence GYDQL (Cherqui *et al.*, 2001). Deletion of the entire LTM or mutation of either the tyrosine or leucine residue to alanine causes partial mislocalization to the plasma membrane, however, complete lysosomal targeting is only abolished after additional removal of another sequence, YFPQA, located within the third cytoplasmic loop (Cherqui *et al.*, 2001). YFPQA was proposed to act as a conformational signal, but the exact mechanism by which it mediates lysosomal targeting remains unknown (Cherqui *et al.*, 2001; Kalatzis *et al.*, 2004).

Alternative splicing of exon 12 of the *CTNS* gene yields a second isoform with an elongated C-terminal tail ending in SSLKG, replacing the LTM (Taranta *et al.*, 2008; **Fig. 1.3**). This isoform, termed CTNS-LKG or CTNS2, accounts for 5–20% of CTNS transcripts in most tissues (Taranta

Table 1.1 – List of Cystinosis-Causing Missense Mutations.
Adapted from David *et al.* (2019).

Mutation	Clinical phenotype	Reference(s)
V42I	Atypical/juvenile	Kalatzis *et al.*, 2002
G110V	Atypical	Kalatzis *et al.*, 2002
I133F	Infantile	McGowan-Jordan *et al.*, 1999
S139F	Atypical/juvenile	Attard *et al.*, 1999
S141F	Infantile	Kalatzis *et al.*, 2004
R151G	Infantile	Topaloglu *et al.*, 2012
G157D	Infantile	Topaloglu *et al.*, 2012
L158P	Infantile	McGowan-Jordan *et al.*, 1999
G169D	Infantile	Shotelersuk *et al.*, 1998
Y173H/C	Infantile	Topaloglu *et al.*, 2012; Sadeghipour *et al.*, 2017
N177S/T	Infantile/juvenile	Kalatzis *et al.*, 2002; Aldahmesh *et al.*, 2009
W182R	Infantile	Shotelersuk *et al.*, 1998
K187R	Infantile	Soliman *et al.*, 2014
G197R	Ocular	Anikster *et al.*, 1999
P200L	Juvenile	Kalatzis *et al.*, 2002
D205N	Infantile	Shotelersuk *et al.*, 1998
T216R	Infantile	Alcantara-Ortigoza *et al.*, 2008
Q222R	Infantile	Kalatzis *et al.*, 2002
S270F/Y	Infantile	Macías-Vidal *et al.*, 2009
K280R	Juvenile	Thoene *et al.*, 1999
M287I	Infantile	Mason *et al.*, 2003
N288K	Infantile	Kalatzis *et al.*, 2002
S293I	Infantile	Topaloglu *et al.*, 2017
S298N	Infantile	Shotelersuk *et al.*, 1998
D305G/Y	Infantile	Shotelersuk *et al.*, 1998; Attard *et al.*, 1999
G308R/E/V	Infantile	Shotelersuk *et al.*, 1998; Kiehntopf *et al.*, 2002; Shahkarami *et al.*, 2013
G309D/V	Infantile	Macías-Vidal *et al.*, 2009; Yeetong *et al.*, 2012
N323K	Juvenile	Thoene *et al.*, 1999
T334N	Juvenile	Servais *et al.*, 2008
G337R	Infantile	Mason *et al.*, 2003
L338R/P	Infantile	Attard *et al.*, 1999; Aldahmesh *et al.*, 2009
G339R	Infantile	Shotelersuk *et al.*, 1998
D346N	Atypical/juvenile	Attard *et al.*, 1999
G362R	Infantile	Soliman *et al.*, 2014

et al., 2012), with the higher distributions found in epithelial cells (Taranta *et al.*, 2017). The SSLKG motif acts as a targeting motif and redirects CTNS-LKG to the plasma membrane (Bellomo *et al.*, 2016). However, due to clathrin-mediated endocytosis, plasma membrane-resident CTNS-LKG is also trafficked to the lysosome, as well as other cytosolic structures (Taranta *et al.*, 2008; Bellomo *et al.*, 2016).

CTNS has been described as the lysosomal L-cystine transporter (**Fig. 1.2** step 5) using the proton potential established between the cytoplasm and the lysosome to facilitate cystine and proton symport (Kalatzis *et al.*, 2001) at a 1:1 stoichiometry (Ruivo *et al.*, 2012). CTNS-LKG also

functions as a cystine transporter (Taranta *et al.*, 2008), even at the plasma membrane, however, only in the presence of a steep proton gradient (Bellomo *et al.*, 2016; Taranta *et al.*, 2017). A multitude of deletions, insertions, and mutations within the *CTNS* gene have been linked to the specific phenotypes of cystinosis. The most common cause is a 57 kb deletion, which cuts exon 10 and abolishes CTNS expression altogether (Anikster *et al.*, 1999). However, shorter deletions as well as insertions also occur frequently (Town *et al.*, 1998; Kalatzis *et al.*, 2004). To date, 42 missense mutations have been identified to cause cystinosis (David *et al.*, 2019; **Table 1.1**) due to their inhibitory effect on cystine transport (Kalatzis *et al.*, 2004). The level of transport inhibition was shown to correlate directly with the severity of disease for most mutants tested (Kalatzis *et al.*, 2004). Interestingly, exceptions were found: S139F, K280R, and N323K, which are only associated with nephropathic juvenile cystinosis, abolish cystine transport completely, whilst W182R and S298N – both linked to nephropathic infantile cystinosis – have only little impact on transport (Kalatzis *et al.*, 2004). This indicates that some CTNS mutants may interfere with potential additional functions of CTNS besides cystine transport, whilst others may still facilitate such a function due to their lysosomal localization, even when trapped in a transport-deficient state (Cherqui & Courtoy, 2017).

Recent discoveries have indicated that CTNS interacts physically with other lysosomal structures and, generally, may be vital for lysosomal function. This would transform CTNS into an integral contributor to the maintenance of lysosomal homeostasis. Specifically, CTNS was shown to interact with components of the mTORC1 kinase, including the Ragulator-RagA/C complex, as well as the proton pump v-ATPase, which functions to acidify the lysosomal lumen (Andrzejewska *et al.*, 2016). Interestingly, this interaction seems to be conformation specific since CTNS knockdown, as well as the deletion of only the YFPQA motif (see above), result in the same reduction of lysosomal recruitment and activation of mTORC1 (Andrzejewska *et al.*, 2016; Guo *et al.*, 2022). It was also proposed that CTNS inherits a role in trafficking reception, where it interacts with the lysosome-associated membrane protein 2A (LAMP2A) and regulates its trafficking to lysosomes (Zhang *et al.*, 2017). It was shown that CTNS deficiency in cystinotic cells causes LAMP2A mislocalization and thus down-regulates chaperone-mediated autophagy (CMA), which is facilitated by LAMP2A multimerization at the lysosomal membrane (Cherqui & Courtoy, 2017; **Fig. 1.4**). Whilst CMA is inactivated, CTNS loss actually promotes autophagic processes and lysosomal biogenesis *via* the TFEB master transcription factor, which translocates into the nucleus when it is not phosphorylated by the mTORC1 kinase (Cherqui & Courtoy, 2017; **Fig. 1.4**).

Another substantial cell biological effect caused by CTNS loss is the decrease in cytosolic cysteine levels due to the accumulation of cystine within lysosomes (Cherqui & Courtoy, 2017). Lysosomal cystine is the primary source for the production of cytosolic cysteine, although others exist, and cysteine is used to synthesize glutathione (GSH), which protects cells from reactive oxygen species (ROS), frequently formed during mitochondrial energy production

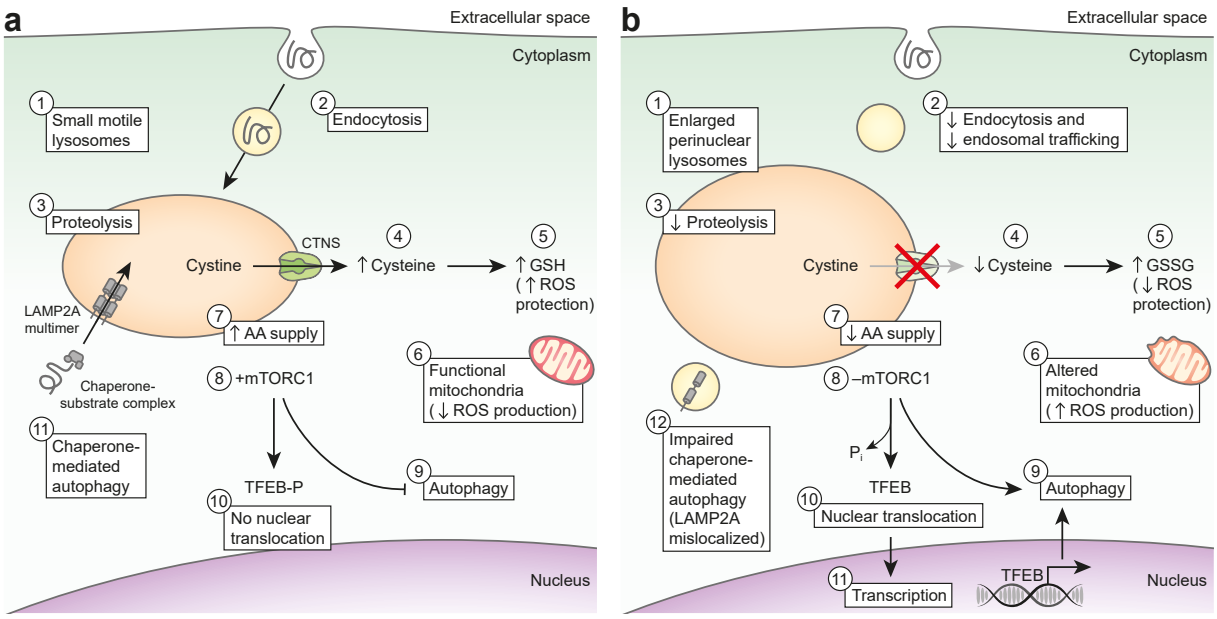


Fig. 1.4 – CTNS-Mediated Cell Biological Effects.
a, In healthy cells, small and motile lysosomes fuse frequently with late endosomes (1 and 2). Proteolysis (3) of its contents creates cystine (4), which is exported by CTNS and subsequently cleaved into cysteine, favouring the production of reduced GSH to protect from ROS (5). The amino acid supply also keeps mTORC1 in its active state, where it inactivates autophagy processes and phosphorylates TFEB, thereby abolishing nuclear translocation and transcriptional effects (7–10). LAMP2A is successfully trafficked to the lysosomal membrane, enabling CMA (11). **b**, In cystinotic cells, the accumulation of cystine causes enlarged and less functional lysosomes (1–3). Consequently, cysteine production from cystine (4) and GSH production are down-regulated, leading to decreased protection from ROS (5). The shortened amino acid supply also inactivates mTORC1, allowing TFEB to hydrolyse to its active unphosphorylated form. Both events result in the activation of autophagic processes (8–11). CTNS loss also causes LAMP2A mislocalization (12), abolishing the clearance of altered cytosolic proteins by CMA. Figure and legend adapted from Cherqui & Courtoy (2017).

(Cherqui & Courtoy, 2017; **Fig. 1.4**). With cytosolic cysteine levels dropping, the equilibrium between GSH and glutathione disulfide (GSSG) shifts towards GSSG, leading to a decreased protection from ROS and an increased chance of the initiation of programmed cell death by apoptosis (Cherqui & Courtoy, 2017). The stimulation of lysosomal and autophagic processes resulting from CTNS loss (as described above) may in fact increase oxidative stress levels. This would explain why proximal tubular cells of the kidney, which rely on mTORC1 signalling (Fantus *et al.*, 2016) and are prone to apoptosis (Park *et al.*, 2002), are most vulnerable to defects within CTNS (Cherqui & Courtoy, 2017; Jamalpoor *et al.*, 2021).

1.4 Related Transporters & Receptors

CTNS belongs to the family of PQ-loop proteins, a diverse group of transmembrane proteins featuring the topology of seven helices and a frequently occurring proline/glutamine (PQ)-motif (Saudek, 2012). Transporters of this group use an alternating access mechanism to move their

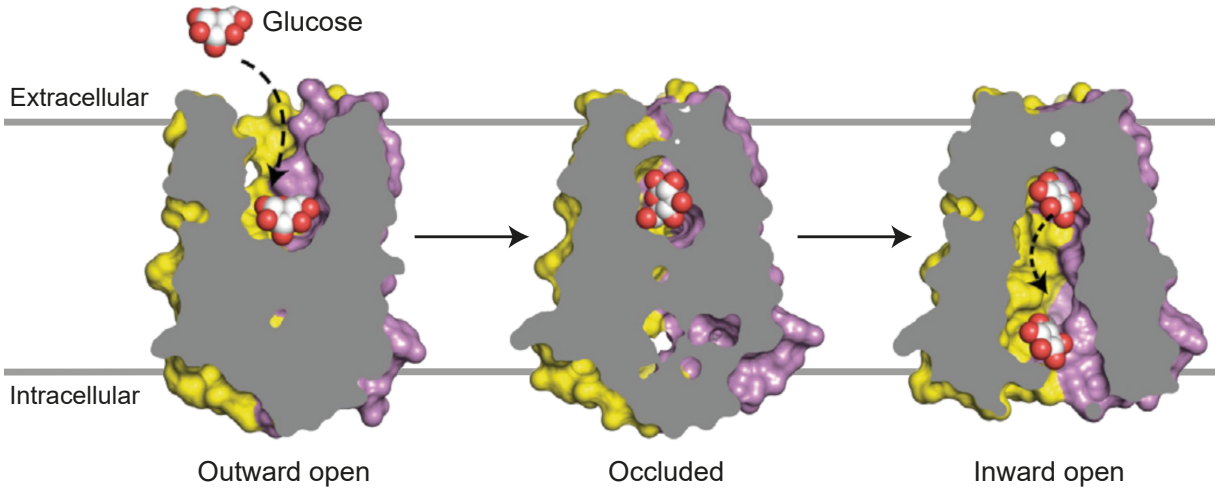


Fig. 1.5 – Alternating Access Glucose Transport *via Lb*SemiSWEET.
Slab representation of the major conformational states obtained using XRC and molecular dynamics simulations. Figure adapted from Latorraca *et al.* (2017).

substrates across the membrane (Jardetzky, 1966). However, to date, a complete transport cycle has only been demonstrated for a bacterial homologue of the SWEET transporter (Latorraca *et al.*, 2017; **Fig. 1.5**), but not for any eukaryotic relatives. Although the bacterial orthologs called SemiSWEETs do not contain seven TM helices and are instead constructed from two identical three-helix bundles (THBs), they do preserve a PQ-motif in the first helix of each THB, highlighting its functional importance in mediating the conformational changes required for transport (Latorraca *et al.*, 2017). Only due to the ''helix-breaking'' effect of prolines (Nilsson *et al.*, 1998) are these proteins able to move around a central ligand binding site to allow for alternating access from one side of the membrane, whilst gating the other (Newstead & Barr, 2020). Interestingly, the gating mechanism on the cytoplasmic-facing side is, in part, mediated by the PQ-motif glutamine through hydrogen bonding with loop residues of the opposite THB (Xu *et al.*, 2014). Due to their architecture, SemiSWEETs move symmetrically around the ligand to facilitate transport (Lee *et al.*, 2015; Latorraca *et al.*, 2017). Their transport mechanism was established as the simplest version of alternating access transport and named the rocker-switch model (Drew & Boudker, 2016).

The only structures of eukaryotic PQ-loop proteins published to date show two members of the SWEET transporter (Tao *et al.*, 2015; Han *et al.*, 2017) and the KDEL receptor (Bräuer *et al.*, 2019), each only in one conformational state. The SWEETs lack the glutamines in the PQ-motifs in TMs 1 and 5, and consequently use a different gating mechanism that is yet to be described. Despite that, the mechanism by which they transport sugars remains symmetric (Newstead & Barr, 2020), following the rocker-switch model, owing to the fact that mutagenesis of either PQ-motif proline causes complete loss of function (Tao *et al.*, 2015). The KDELR is a trafficking receptor that specifically binds the C-terminus of a peptide ending in KDEL (Bräuer *et al.*, 2019). It shares a similar architecture with other PQ-loop transporters but only retains one of the

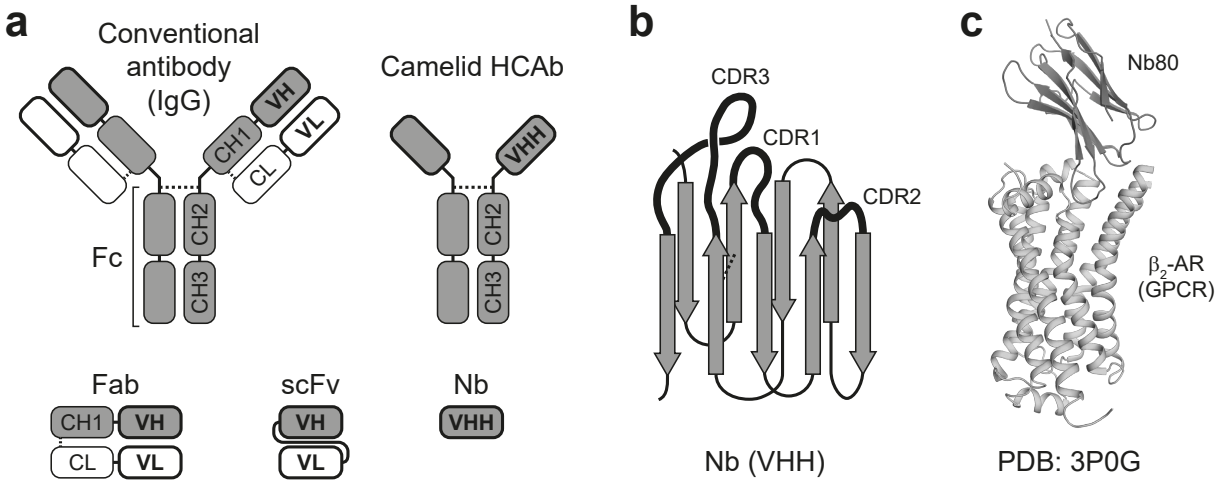


Fig. 1.6 – Nanobodies as Crystallization Chaperones.
a, Schematic representation of naturally occurring antibodies: conventional heterotetrameric antibodies present in mammals and camelids, and heavy chain antibodies (HCAbs), only present in camelids. Schematic adapted from Muyldermans (2013). Antibody-derived fragments used in X-ray crystallography are shown. Constant (C, regular font) and variable (V, bold) regions of heavy (H, grey) and light (L, white) chains are highlighted. Inter-chain disulfides are shown using dashed lines. **b**, Typical Nb fold, highlighting complementarity-determining regions (CDRs) 1, 2, and 3. Schematic adapted from Manglik *et al.* (2017). **c**, Cartoon representation of a Nb bound to a GPCR (β_2-adrenergic receptor, β_2-AR).

PQ-motifs – that in TM5. As a result, KDEL binding causes an asymmetric movement within the receptor and exposes a retrieval motif positioned in the end of TM7 (Newstead & Barr, 2020).

1.5 Nanobodies & Sybodies as Crystallization Chaperones

The use of crystallization chaperones can have a substantial impact on the success of a challenging membrane protein structural biology project. This can be demonstrated well using G protein-coupled receptors (GPCRs) as an example: The first structure of a GPCR (bovine rhodopsin) was solved in 2000 (Palczewski *et al.*, 2000), but despite many attempts, other (less stable) members of this large family of receptors could not be used to obtain well-diffracting crystals. This changed when a GPCR (β_2-adrenergic receptor) was fused to lysozyme from bacteriophage T4 (Rosenbaum *et al.*, 2007), which contributed vitally to the formation of crystal contacts, enabling crystallization and structure determination (Cherezov *et al.*, 2007). Such major covalent modifications come with risks, though, as the integrity and function of the membrane protein may be disturbed (Rosenbaum *et al.*, 2009). To circumvent these caveats, antibody-based crystallization chaperones such as antigen-binding fragments (Fabs) or nanobodies (Nbs) can be used (**Fig. 1.6**). Specific binding stabilizes the target protein and, depending on the epitope, may lock it in a specific conformation, enhancing the formation of homogeneous crystals as well as downstream analysis of structure-function correlations

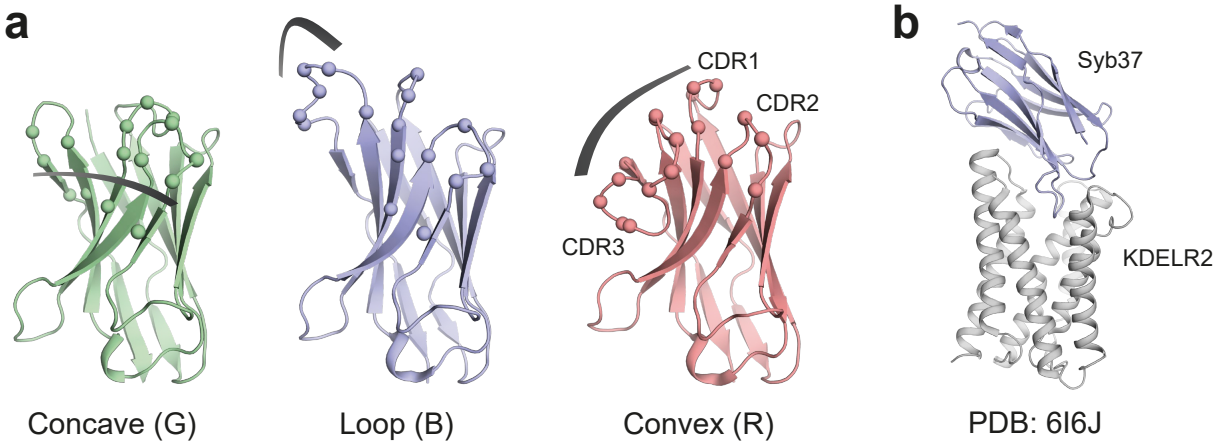


Fig. 1.7 – Sybodies as Crystallization Chaperones.
a, Library-encoding synthetic nanobodies (Sybs) as published by Zimmermann *et al.* (2020). Spheres indicate the position of hypermutated residues. **b**, Cartoon representation of a loop Syb bound to KDELR2.

(Manglik *et al.*, 2017; Uchański *et al.*, 2020). The application of these crystallization chaperones, in combination with other technological breakthroughs in the field, as well as the resolution revolution in cryo-EM, have resulted in structures of more than 80 members of the GPCR family (Yang *et al.*, 2021).

The generation of synthetic nanobodies, known as sybodies (Sybs), is based on three mRNA libraries that are hypermutated in the CDRs (Zimmermann *et al.*, 2020). They code for three different types of Sybs, which mainly differ in the length of their third CDR loop (**Fig. 1.7 a**, sequences shown in **Table A.3**). Sybs with a short CDR3 expose a concave binding site. Sybs with longer CDR3s are termed loop Sybs. These have been found useful in forming interactions with the ligand binding cavity of small membrane proteins, such as the KDEL receptor (Bräuer *et al.*, 2019; **Fig. 1.7 b**). Further CDR3 elongation may cause the loop to collapse towards the concave surface of the Syb, forming a convex binding interface. The concave, loop, and convex libraries are colour-coded green (G), blue (B), and red (R), respectively. The screening procedure consists of consecutive rounds of ribosome and phage displays, utilizing a biotinylated version of the target protein to facilitate specific immobilization *via* streptavidin and neutravidin. The avidin-biotin (vitamin B_7) interaction is known to form one of the strongest non-covalent bonds in biochemistry, featuring an equilibrium dissociation constant (K_D) of ~1×10^{-15} M, or one femtomolar (Green, 1963), and is therefore well suited for biotechnological applications like ribosome and phage displays. Using this Syb screening protocol, specific binders against any purified and biotinylated protein can be generated quickly and purely through *in vitro* experiments without the use of lab animals (Zimmermann *et al.*, 2020).

1.6 Aims & Objectives

No molecular studies have been undertaken at a structural or biochemical level on CTNS, hampering efforts to link pathologies associated with defects in the *CTNS* gene to the reported cellular defects. At the start of my DPhil, the following were some of the main questions that needed to be answered to further our understanding of lysosomal biology and the role of CTNS in human health and disease:

- How does CTNS fit structurally in the diverse PQ-loop family? What characteristics are shared, and are there any novel features?
- What is the underlying mechanism for cystine transport by CTNS? How does CTNS couple cystine movement to the translocation of protons?
- Do the reported missense mutations affect the transport function of CTNS? Are these effects sufficient to explain the progression of cystinosis?
- How does CTNS interact with and regulate other lysosomal structures?

To answer these questions as precisely as possible, a combination of structural biology, biophysical, and biochemical techniques was employed. A substantial part of my DPhil work was dedicated to the structural solution and analysis of the CTNS protein, followed by the characterization of its transport function. The following objectives were pursued, reached, and included within the scope of this thesis:

- Identification of CTNS constructs most suitable for recombinant production in yeast and mammalian systems (**Chs. 3** and **8**).
- Biophysical characterization to assess their suitability for structural and functional studies (**Chs. 3** and **8**).
- Generation of CTNS-specific nanobody and sybody binders (**Chs. 5** and **8**).
- Crystallization in the lipidic cubic phase and structure determination using X-ray crystallography (**Chs. 4**, **6**, and **7**).
- Sample preparation and structure solution using cryogenic electron microscopy (**Ch. 8**).
- Development of a robust CTNS ligand binding assay (**Chs. 3** and **9**).
- Establishment of robust transport assays using reconstituted proteoliposomes (**Ch. 9**).
- Detailed structural and functional analysis of CTNS, contextualized using the existing literature (**Ch. 9**).

CHAPTER 2

Materials & Methods

2.1 Construct Design

2.1.1 Expression in Yeast Cells

The genes encoding CTNS of various organisms were codon optimized for expression in *Saccharomyces cerevisiae*, flanked with sites homologous to the insertion site of the pDDGFP-Leu2d expression vector (Parker & Newstead, 2014; Addgene #58352), and synthesized as strings from DNA fragments (GeneArt, Invitrogen; **Table A.1**). Constructs were amplified through standard polymerase chain reaction (PCR) using appropriate primers (**Table A.2**, synthesized by Sigma-Aldrich) and the KOD Hot Start DNA Polymerase Kit (Sigma-Aldrich) according to the manufacturer's instructions. Amplified DNA fragments were subsequently inserted into the vector by homologous recombination (HR). To do so, 2–4 µg of a construct were transformed along with 200–400 ng of SmaI-linearized pDDGFP-Leu2d (FastDigest Restriction Enzymes, Thermo Scientific, if not stated otherwise) into yeast strain BJ5460 (see **Sec. 2.2.1**). After a 3-day selection at 30 °C on a –Ura plate (synthetic complete medium minus uracil, 2% (w/v) agar, 2% (w/v) D-glucose), 8–10 colonies were picked and used to inoculate a 10 mL –Ura culture (as before without agar), which was grown overnight (O/N) at 30 °C and 220–250 revolutions per minute (RPM). Yeast cells were harvested by centrifugation at 4,000 × *g* for 5 min, resuspended in 200 µL phosphate-buffered saline (PBS), and added to ~200 µL of silica beads. The mixture was vortexed for 15 min at 4 °C, and after a 2 s centrifugation step at 21,000 × *g*, the supernatant (SN) was taken and used for plasmid DNA recovery (GeneJET Plasmid Miniprep Kit, Thermo Scientific). For selection and amplification, the eluted DNA was transformed into *Escherichia coli* OmniMAX cells (Invitrogen) using the heat-shock method (incubation on ice for 15 min, followed by 45 s at 42 °C, and 3 min on ice) and plated on Lennox Broth (LB)-agar supplemented with 100 µg mL^{-1} ampicillin (Amp). Single colonies yielded after O/N incubation at 37 °C were used to inoculate 3–4 mL LB-Amp cultures. Those were grown O/N at 37 °C and 180–200 RPM, and the plasmid DNA again purified. The integrity of the constructs was verified by diagnostic restriction digest using enzymes BamHI and XhoI, as well as Sanger sequencing (Source BioScience).

In pDDGFP-Leu2d, CTNS-encoding sequences (amino acid sequences shown in **Table A.3**) are followed by a tobacco etch virus protease (TEV) recognition site (GENLYFQ*GQF; *indicates cleavage site), a copy of yeast-enhanced green fluorescent protein (GFP) (Cormack *et al.*, 1997), and a polyhistidine (His)-tag (Hochuli *et al.*, 1987; 8 × H). Avi-tags (Beckett *et al.*, 1999; GLNDIFEAQKIEWHEVN) were added to the N- (or C-) terminal end of desired constructs, and deletions introduced, by PCR amplification using complementary overhang primers (**Table A.2**). Variant forms of CTNS were created by site-directed mutagenesis using overlap-extension PCR. The 7NA alteration of human CTNS, indicating the mutation of asparagines 36, 41, 51, 66, 84, 104, and 107 to alanines, was generated using a specifically designed gene string that ranges from the 5' HR site to the bases encoding residue 121 (**Table A.1**). To exchange the 5' DNA fragment, the *Hs*CTNS construct was cut *via* the intrinsic BglII site present at the bases encoding residues 119–120. The 3' end was purified by gel extraction and subsequently fused with the new 5' end using the Quick Ligation Kit (New England Biolabs) according to the manufacturer's instructions. The second isoform of human CTNS (*Hs*CTNS2) was created by extending the construct of isoform 1 using consecutive primers (**Table A.2**). In all these cases, DNA fragments were purified and ligated back into the pDDGFP-Leu2d vector by HR as described before. All ensuing constructs are listed in **Table 2.1**.

2.1.2 Expression in Mammalian Cells

Several CTNS constructs from **Sec. 2.1.1** were subcloned into the pLEXm plasmid (Aricescu *et al.*, 2006; Addgene #99844) using the SacI and XhoI restriction sites to enable expression in mammalian human embryonic kidney (HEK) 293F cells. Constructs were designed to bear different C-terminal tags, including a TEV-cleavable GFP-His-tag, as described before, as well as uncleavable Flag- (Hopp *et al.*, 1988; LEDYKDDDDK) and StrepII-tags (Schmidt *et al.*, 1996; SAWSHPQFEK). They are listed in **Table 2.2** but not further described in this chapter as they were only used in preliminary experiments.

2.1.3 Expression in Bacterial Cells

The generation of sybodies (Sybs) and nanobodies (Nbs) is described in **Sec. 2.4**. Unique constructs were isolated after selection by enzyme-linked immunosorbent assay (ELISA). At this time, individual Sybs and Nbs were contained in the vectors pSb_init (Addgene #110100) and pADL-23c (Antibody Design Laboratories), respectively. In these vectors, the binder-encoding sequences (amino acid sequences shown in **Table A.3**) are preceded by a pectate lyase B (pelB) leader (MSKYLLPTAAAGLLLLAAQPAMA*GSSS in pSb_init and MKYLLPTAAAGLLLLAAQPAMA* in pADL-23c; *indicates natural cleavage site) and followed by a linker, c-Myc-, and His-tag (AGRAGEQKLISEEDLNSAVDHHHHHH in pSb_init and SEPKTPKPQPAAGPGGQHHHHHHGAEQKLISEEDLS in pADL-23c; tags are underlined). The linkers located directly before and after the construct in the pSb_init vector relate to BspQI

Table 2.1 – Constructs for Yeast Cell Culture.
The listed constructs were cloned into the vector pDDGFP-Leu2d, allowing for dual selection using either the uracil or leucine marker. The additional Amp resistance is used for cloning and plasmid amplification purposes. Expression is induced using D-galactose and yields the constructs C-terminally fused to a TEV-cleavable GFP-His-tag. Newstead lab plasmid collection identifiers are annotated for each construct.

Identifier	Construct
1669	*At*CTNS
1766	*At*CTNS-Avi
1915	Avi-*At*CTNS
2286	*At*CTNS(W24A)
2287	*At*CTNS(W24F)
2288	*At*CTNS(S27A)
2289	*At*CTNS(S27C)
2475	*At*CTNS(P30A)
2476	*At*CTNS(P30G)
2290	*At*CTNS(D47A)
2291	*At*CTNS(D47N)
2292	*At*CTNS(N52A)
2293	*At*CTNS(N52Q)
2294	*At*CTNS(K55A)
2482	*At*CTNS(K55A/K166A)
2295	*At*CTNS(K55Q)
2296	*At*CTNS(K55R)
2297	*At*CTNS(H56A)
2298	*At*CTNS(H56F)
2299	*At*CTNS(P87A)
2300	*At*CTNS(P87G)
2301	*At*CTNS(N91A)
2302	*At*CTNS(N91Q)
2303	*At*CTNS(D92A)
2304	*At*CTNS(D92N)
2305	*At*CTNS(K166A)
2306	*At*CTNS(K166Q)
2307	*At*CTNS(K166R)
2308	*At*CTNS(Y167A)
2309	*At*CTNS(Y167F)
2477	*At*CTNS(P169A)
2478	*At*CTNS(P169G)
2310	*At*CTNS(D191A)
2479	*At*CTNS(D191G)
2311	*At*CTNS(D191N)
2480	*At*CTNS(D191Y)
2312	*At*CTNS(Q201A)
2313	*At*CTNS(Q201N)
2314	*At*CTNS(K221A)

Identifier	Construct
2315	*At*CTNS(K221Q)
2316	*At*CTNS(K221R)
2481	*At*CTNS(S228A)
2317	*At*CTNS(D232A)
2318	*At*CTNS(D232N)
2268	*At*CTNS(Δ246–270)
2269	*At*CTNS(D232N/Δ246–270)
1674	*Dr*CTNS
2213	*Dr*CTNS(Δ2–120)
1670	*Gg*CTNS
1767	*Gg*CTNS-Avi
2214	*Gg*CTNS(Δ2–114)
1671	*Hs*CTNS
2485	*Hs*CTNS(D346N)
2483	*Hs*CTNS(7NA)
2484	*Hs*CTNS(7NA)-Avi
2489	*Hs*CTNS(7NA/D346N)
2731	*Hs*CTNS(Δ2–23)
2735	*Hs*CTNS(Δ2–23/Δ360–367)
2736	*Hs*CTNS(Δ2–23/7NA/Δ360–367)
2732	*Hs*CTNS(Δ2–23/7NA)
1914	*Hs*CTNS(Δ2–115)
2215	*Hs*CTNS(Δ2–115)-Avi
2486	*Hs*CTNS(Δ2–115/D346N)
2487	*Hs*CTNS(Δ2–115/D346N)-Avi
2737	*Hs*CTNS(Δ2–115/Δ360–367)
2270	*Hs*CTNS(Δ2–118)
2271	*Hs*CTNS(Δ2–118/Δ360–367)
2488	*Hs*CTNS(Δ2–118/D346N/Δ360–367)
2733	*Hs*CTNS(Δ360–367)
2734	*Hs*CTNS(7NA/Δ360–367)
2210	*Hs*CTNS2
2738	*Hs*CTNS2(7NA)
2739	*Hs*CTNS2(Δ2–23)
2740	*Hs*CTNS2(Δ2–23/7NA)
2209	*Hs*CTNS2(Δ2–115)
2741	*Hs*CTNS2(Δ2–118)
1672	*Mm*CTNS
2212	*Mm*CTNS(Δ2–115)

Table 2.2 – Constructs for Mammalian Cell Culture.
Newstead lab plasmid collection identifiers are annotated for each construct.

a, Constructs bearing a C-terminal, TEV-cleavable GFP-His-tag.

Identifier	Construct	Plasmid	Selection
2211	*Hs*CTNS	pLEXm	Amp
2813	*Hs*CTNS(7NA)	pLEXm	Amp
2743	*Hs*CTNS(Δ2–23/7NA/Δ363–367)	pLEXm	Amp
2744	*Hs*CTNS(Δ2–23/7NA/N301A/Δ363–367)	pLEXm	Amp
2742	*Hs*CTNS(Δ2–115)	pLEXm	Amp
2745	*Hs*CTNS2	pLEXm	Amp
2279	*Hs*LAMP2A	pLEXm	Amp
2280	*Mm*LAMP2A	pLEXm	Amp

b, Constructs bearing a C-terminal Flag-tag.

Identifier	Construct	Plasmid	Selection
2746	*Hs*CTNS-Flag	pLEXm	Amp
2814	*Hs*CTNS(7NA)-Flag	pLEXm	Amp
2748	*Hs*CTNS(Δ2–23/7NA/Δ363–367)-Flag	pLEXm	Amp
2749	*Hs*CTNS(Δ2–23/7NA/N301A/Δ363–367)-Flag	pLEXm	Amp
2747	*Hs*CTNS(Δ2–115)-Flag	pLEXm	Amp
2750	*Hs*CTNS2-Flag	pLEXm	Amp

c, Constructs bearing a C-terminal StrepII-tag.

Identifier	Construct	Plasmid	Selection
2815	*Hs*CTNS-StrepII	pLEXm	Amp
2817	*Hs*CTNS(7NA)-StrepII	pLEXm	Amp
2816	*Hs*CTNS(Δ2–115)-StrepII	pLEXm	Amp
2818	*Hs*CTNS2-StrepII	pLEXm	Amp

restriction sites. Important constructs mentioned in this thesis are listed in **Table 2.3 a**, less relevant binders are shown in **Table A.4**.

To generate tag-free binders better suitable for co-crystallization, constructs were subcloned into pBXNPHM3 (Addgene #110099) using the fragment exchange (FX) method (Geertsma, 2013; Zimmermann *et al.*, 2020). Target vector pBXNPHM3 (Amp resistance) and Syb sequences contained in pSb_init (chloramphenicol (CAM) resistance) were mixed in a 1:3 (w:w) ratio (40 ng μL^{-1} total) and digested for 1 h at 50 °C using 1 U μL^{-1} BspQI (New England Biolabs). After deactivation of the restriction enzyme (20 min, 80 °C), a final concentration of 0.5 U μL^{-1} T4 DNA ligase (Thermo Scientific) and 1 mM ATP (Sigma-Aldrich) were added, and the mixture was incubated for 1 h at 37 °C, then 10 min at 65 °C. Subsequently, 5 µL of the mixture was transformed into 50 µL chemically competent *E. coli* MC1061 cells (Lucigen) using the heat-

Table 2.3 – Constructs for Bacterial Cell Culture.
All constructs contain an N-terminal pelB leader sequence. Newstead lab plasmid collection identifiers are annotated for each construct.

a, Constructs bearing a C-terminal Myc- and His-tag.

Identifier	Construct	Plasmid	Selection	Induction
1950	α-*At*CTNS Nb1-His-Myc	pADL-23c	Amp	IPTG
1953	α-*At*CTNS Nb4-His-Myc	pADL-23c	Amp	IPTG
2933	α-*At*CTNS Syb1-Myc-His aka G4b	pSb_init	CAM	L-(+)-arabinose
2934	α-*At*CTNS Syb7-Myc-His aka G1b	pSb_init	CAM	L-(+)-arabinose
2935	α-*At*CTNS Syb12-Myc-His aka G4a	pSb_init	CAM	L-(+)-arabinose
2936	α-*At*CTNS Syb39-Myc-His aka B1c	pSb_init	CAM	L-(+)-arabinose
2937	α-*At*CTNS Syb41-Myc-His aka B5b	pSb_init	CAM	L-(+)-arabinose
2983	α-*At*CTNS Syb54-Myc-His	pSb_init	CAM	L-(+)-arabinose
2985	α-*At*CTNS Syb57-Myc-His	pSb_init	CAM	L-(+)-arabinose
2938	α-*Hs*CTNS(Δ2–115) Syb20-Myc-His	pSb_init	CAM	L-(+)-arabinose
2939	α-*Hs*CTNS(Δ2–115) Syb25-Myc-His	pSb_init	CAM	L-(+)-arabinose
2940	α-*Hs*CTNS(Δ2–115) Syb30-Myc-His	pSb_init	CAM	L-(+)-arabinose
2941	α-*Hs*CTNS(Δ2–115) Syb32-Myc-His	pSb_init	CAM	L-(+)-arabinose
2942	α-*Hs*CTNS(Δ2–115) Syb41-Myc-His	pSb_init	CAM	L-(+)-arabinose

b, Constructs bearing an N-terminal, 3C-cleavable His-MBP-tag.

Identifier	Construct	Plasmid	Selection	Induction
1951	α-*At*CTNS Nb1	pBXNPHM3	Amp	L-(+)-arabinose
1954	α-*At*CTNS Nb4	pBXNPHM3	Amp	L-(+)-arabinose
2272	α-*At*CTNS Syb1 aka G4b	pBXNPHM3	Amp	L-(+)-arabinose
2273	α-*At*CTNS Syb7 aka G1b	pBXNPHM3	Amp	L-(+)-arabinose
2274	α-*At*CTNS Syb12 aka G4a	pBXNPHM3	Amp	L-(+)-arabinose
2275	α-*At*CTNS Syb39 aka B1c	pBXNPHM3	Amp	L-(+)-arabinose
2276	α-*At*CTNS Syb41 aka B5b	pBXNPHM3	Amp	L-(+)-arabinose
2277	α-*At*CTNS Syb54	pBXNPHM3	Amp	L-(+)-arabinose
2278	α-*At*CTNS Syb57	pBXNPHM3	Amp	L-(+)-arabinose

shock method. The reaction was plated on LB-agar-Amp, ensuring that only constructs fused with the target vector yield colonies (undigested pBXNPHM3 contains a ccdB suicide cassette). Nbs from pADL-23c do not contain BspQI restriction sites, and furthermore, Amp resistance is already granted. As a workaround, Nb constructs were amplified by PCR, adding BspQI restriction sites in the process (primers see **Table A.2**). DNA fragments were then purified by agarose gel extraction (GeneJET Gel Extraction Kit, Thermo Scientific). Subsequently, a modified FX cloning reaction was set up, combining Nb DNA and target vector pBXNPHM3 at a 1:1 (w:w) ratio (20 ng μL^{-1} total). The ensuing constructs are listed in **Table 2.3 b**.

2.2 Recombinant Protein Production

2.2.1 Yeast Transformation

10–15 mL YPD medium (1% (w/v) yeast extract, 2% (w/v) peptone, and 2% (w/v) D-glucose) was inoculated with *S. cerevisiae* strain BJ5460 (ATCC) and grown O/N at 30 °C and 250 RPM. Whilst in the exponential growth phase, 600 µL of culture was harvested at 16,000 × *g* for 15 s and resuspended in 100 µL 100 mM Li-acetate, 10 mM Tris-HCl, pH 7.4. A mixture of 10 µL heat-denatured (95 °C, 5 min, then ice) salmon sperm DNA at 10 mg mL^{-1}, used as a carrier (Schiestl & Gietz, 1989), and ~500 ng of plasmid containing the gene of the desired protein (**Table 2.1**) was added to the cells, followed by 400 µL 50% (w/v) polyethylene glycol (PEG) 3350, 100 mM Li-acetate, 10 mM Tris-HCl, pH 7.4, to induce chemical competence (Ito *et al.*, 1983). The solution was vortexed and rotated at room temperature (RT) for 25–30 min. After addition of 50 µL dimethyl sulfoxide (DMSO) to increase the transformation efficiency (Hill *et al.*, 1991), the cells were incubated at 42 °C for 15 min and subsequently pelleted by centrifugation at 500 × *g* for 30 s. The cell pellet was resuspended in 40 µL Milli-Q water (mQ-H_2O), plated on –Ura, and incubated at 30 °C for 2–3 days. Yeast colonies were then picked and re-streaked on –Leu, followed by another 3-day incubation.

2.2.2 Expression Trials in Yeast Cells

Transformed *S. cerevisiae* BJ5460 cells were grown O/N in 10 mL synthetic complete medium minus leucine (–Leu) supplemented with 2% (w/v) D-glucose (30 °C, 250 RPM). 5 mL of that culture was diluted into 40 mL –Leu supplemented with 2% (v/v) DL-lactate, pH 5.1, and grown to an optical density at 600 nm wavelength (OD_{600}) of 1.5–2.5. Protein expression was induced by adding 1.5% (w/v) D-galactose in –Leu, and expression was carried out O/N, usually ~16 h. Expression levels were estimated from 8 mL of culture by resuspending the cell pellet in 150 µL PBS and measuring the relative GFP fluorescence from 100 µL of that suspension using a SpectraMax M3 Microplate Reader (Molecular Devices). Cells from the remaining culture were harvested by centrifugation for 5 min at 4,000 × *g* and 4 °C and immediately proceeded to small-scale membrane fraction preparation or frozen until further use.

2.2.3 Small-Scale Membrane Fraction Preparation

Cells from 20 mL of culture were thawed, if necessary, and resuspended in 750 µL PBS. Cells were added to ~250 µL silica beads in a 1.5 mL tube and lysed by vortexing for 15 min at 4 °C. After pelleting of cell debris and unlysed cells for 5 s at 21,000 × *g*, the SN was collected. 500 µL cold PBS was added to the pellet, briefly vortexed, and after a repeated centrifugation step, the SN was added to the one collected previously. Membranes were pelleted by centrifugation for

1 h at 30,000 × *g* and 4 °C, resuspended in 200 µL PBS, and used for fluorescent size exclusion chromatography (FSEC) analysis.

2.2.4 Fluorescent Size Exclusion Chromatography

Membranes were thawed, if necessary, and 150 µL solubilization reactions were prepared in PBS supplemented with 1% (w/v) detergent (Anatrace) and 10% (v/v) glycerol. After 1 h of end-to-end rotation at 4 °C, detergent-solubilized protein was isolated from insoluble fractions through centrifugation for 1 h at 30,000 × *g* and 4 °C. GFP-fluorescence was measured from 100 µL of sample both before and after the spin to determine the efficiency of the solubilization. 100 µL of the SN was then injected onto a Superose 6 10/300 GL column (Cytiva) pre-equilibrated in Tris-buffered saline (TBS), composed of 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, and supplemented with 0.03% (w/v) *n*-dodecyl-β-D-maltopyranoside (DDM), whilst measuring GFP-fluorescence ($\lambda_{\text{Excitation}}$ = 488 nm, $\lambda_{\text{Emission}}$ = 512 nm).

2.2.5 Expression in Yeast Cells

Transformed *S. cerevisiae* BJ5460 cells were cultivated for 8–10 h in 100 mL –Leu supplemented with 2% (w/v) D-glucose (30 °C, 270 RPM). This primary culture was used to inoculate ~100 mL of secondary culture for every litre of total expression volume (600 mL for 6 L volume, 1.4 L for 15 L) and grown for 22–24 h at 30 °C and 270 RPM. The secondary culture was then diluted ninefold into –Leu containing 2% (v/v) DL-lactate, pH 5.1. Proteins used for structural studies or binder generation (*At*CTNS, *Hs*CTNS(Δ2–115) or Avi-tagged versions thereof) were expressed using a 15 L fermentation vessel (Eppendorf BioFlo 415) set at 30 °C, 600 RPM, 6 L m^{-1} airflow, and pH controlled at 5.1, whilst variants of *At*CTNS were grown in baffled flasks using 6 L volumes (30 °C, 200 RPM). Following overnight cultivation, expression was induced at an OD_{600} of ~2.5 by adding 1.5% (w/v) D-galactose, dissolved in –Leu, and cultivation continued for 20–23 h (altered settings: fermenter: 1,000 RPM and 10 L m^{-1} airflow; flasks: 250 RPM). Yeast cells were harvested by centrifugation for 12 min at 6,000 × *g* and 4 °C, resuspended in cold PBS (5 mL g^{-1} of cells), and proceeded to membrane fraction preparation.

2.2.6 Membrane Fraction Preparation

Cells were thawed, if necessary, and lysed using a pre-cooled TS Series Continuous Cell Disruptor (Constant Systems) in line with a Neslab ThermoFlex 1400 Recirculating Chiller (Thermo Scientific) set at 38 kpsi and 5 °C (two passes). Following a low-speed centrifuge spin at 30,000 × *g* and 4 °C to remove cell debris and unlysed cells, membranes were isolated through ultracentrifugation for 1.5 h at 200,000 × *g* and 4 °C. Membranes were resuspended in cold 20 mM HEPES-NaOH, pH 7.5, 1 M K-acetate using a Dounce Homogenizer (Carl Roth) whilst

splitting the total volume in about half. Ultracentrifugation was repeated, and membranes were resuspended in PBS (5 mL g^{-1}) and snap frozen using liquid nitrogen.

2.2.7 Membrane Protein Purification

Wild type (WT) *At*CTNS, variants thereof, and *At*CTNS-Avi were solubilized in DDM (Anatrace) and purified to homogeneity using standard immobilized metal affinity chromatography (IMAC) and size exclusion chromatography (SEC) steps. Solubilization was carried out for 1 h at 4 °C under gentle agitation. To do so, membranes were thawed and diluted in a final concentration of 1% (w/v) DDM, 1 × PBS, 150 mM NaCl, and 10% (v/v) glycerol. Following ultracentrifugation for 1 h at 200,000 × *g* and 4 °C, the soluble fraction was supplemented with 25 mM imidazole and bound to HisPur Ni-nitrilotriacetic acid (NTA) resin (Thermo Scientific) in batch for 2.5–3 h at 4 °C under gentle agitation. The resin was transferred into a Glass Econo-Column (Bio-Rad), washed with 25–30 column volumes (CV) of solubilization buffer supplemented with 0.1% (w/v) DDM and 35 mM imidazole, and subsequently eluted with 5 CV containing 250 mM imidazole. The GFP mass present in the elution fraction was determined by fluorescence measurement (Drew *et al.*, 2008) and 0.35–0.4 mg of TEV was added for every mg of GFP. The mixture was dialysed in TBS supplemented with 0.02% (w/v) DDM at 4 °C O/N, reducing the imidazole concentration to ~5–10 mM. Subsequently, a reverse IMAC step was performed using the same resin as before after regenerating it according to the manufacturer's instructions (10–20 CV 20 mM MES, 100 mM NaCl at pH 5.0, then mQ-H_2O) and equilibrating it in dialysis buffer. After 1–1.5 h binding at 4 °C using gentle agitation, the suspension was again transferred to a gravity flow column, and the flow-through (FT) was collected. The resin was washed with 1.5 CV of dialysis buffer and, combined with the FT, concentrated using a 50 kilodalton (kDa) molecular weight cut-off (MWCO) Amicon Ultra Centrifugal Filter Unit (Millipore). The sample was then applied to a Superdex 200 Increase 10/300 GL column (Cytiva) pre-equilibrated in MES-buffered saline (MBS), composed of 20 mM MES-NaOH, pH 6, 150 mM NaCl, and supplemented with 0.03% (w/v) DDM. For crystallization, the detergent concentration was reduced to 0.013% (w/v) DDM. For proteoliposome reconstitution, SEC was performed in MBS supplemented with 0.3% (w/v) *n*-decyl-β-D-maltopyranoside (DM). Peak fractions were pooled and concentrated, if necessary, and finally aliquoted and snap frozen. *Hs*CTNS(Δ2–115) and its Avi-tagged version were solubilized and purified in a 10:1 (w:w) mixture of lauryl maltose neopentyl glycol (LMNG) and cholesteryl hemisuccinate (CHS), instead of DDM, and HEPES-buffered saline (HBS), composed of 20 mM HEPES-NaOH, pH 7.5, 150 mM NaCl, supplemented with 0.003% (w/v) 10:1 (w:w) LMNG:CHS, was used as size exclusion buffer.

2.2.8 Expression in Bacterial Cells

Sybodies and nanobodies used in structural and functional studies (where mg amounts are needed) were expressed in bacterial cell cultures using 1–12 L volumes. α-*At*CTNS binders

(mainly used for co-crystallization) were expressed from the pBXNPHM3 plasmid yielding tag-free molecules, whilst α-*Hs*CTNS(Δ2–115) sybodies (for cryogenic electron microscopy (cryo-EM)) were expressed as C-terminally Myc-His-tagged binders from pSb_init. To do so, ~100 ng of vector containing the gene of the desired binder (**Table 2.3**) was transformed into 25 µL of MC1061 cells using the heat-shock method. Where pSb_init was transformed, 400 µL LB was added, and cells were recovered for 45 min at 37 °C and 180 RPM. Cells were then pelleted by centrifugation for 30 s at 2,000 × *g*, and most of the SN was removed. Subsequently, the cells were resuspended in the remaining volume (target ~40 µL), plated on LB-agar supplemented with 25 µg mL^{-1} CAM, and grown O/N at 37 °C. For pBXNPHM3 transformations, recovery was skipped, and cells were plated directly on LB-agar supplemented with 120 µg mL^{-1} Amp. A Terrific Broth (TB) starter culture (1 mL per litre of expression culture) supplemented with the appropriate antibiotic was inoculated and grown for 6 h at 37 °C and 220 RPM and subsequently used to inoculate the pre-warmed TB-antibiotic expression culture. Cells were grown for 2.5–3 h until an OD_{600} of 0.20–0.25 was reached, at which point the temperature setting was changed to 22 °C and agitation reduced to 180 RPM. Cultivation was continued for 1–1.5 h, and expression was induced at an OD_{600} of 0.6–0.7 by adding 0.02% (w/v) L-(+)-arabinose. After O/N cultivation, cells were harvested by centrifugation, and binders were purified as described hereafter.

2.2.9 Purification of Sybodies & Nanobodies

α-*At*CTNS sybodies and nanobodies were purified after complete cell lysis. Harvested cells were resuspended in PBS (5 mL g^{-1} of cells), supplemented with lyophilized deoxyribonuclease I (Sigma-Aldrich), and disrupted at 5 °C using a high pressure (25 kpsi, three passes). The lysate was cleared by centrifugation for 30 min at 30,000 × *g* and 4 °C followed by ultracentrifugation for 30 min at 200,000 × *g* and 4 °C. The SN was supplemented with 20 mM imidazole, incubated for 1 h in batch using HisPur Ni-NTA resin under gentle agitation, and after wash (W) steps using PBS supplemented with 20 mM (15 CV) and 50 mM (10 CV), bound protein was eluted with 300 mM imidazole (5 CV). After addition of 1 mg human rhinovirus 3C protease (3C) per 100 mg of protein, as determined using absorbance at 280 nm (A_{280}), the eluate was dialysed in TBS at 4 °C O/N. The dialysate was applied consecutively to Amylose Resin (New England Biolabs) and HisPur Ni-NTA, each time collecting the FT and a 1.5 CV W using dialysis buffer. After concentration using a 3 kDa MWCO Amicon Ultra Centrifugal Filter Unit (Millipore), SEC was performed using a Sepax SRT-C SEC-300 column (Chromex Scientific) pre-equilibrated in MBS. Peak fractions were pooled and concentrated, if necessary, and finally aliquoted and snap frozen.

α-*Hs*CTNS(Δ2–115) sybodies were purified from the periplasmic extract using a protocol adapted from the small-scale purification of 50 mL cultures described by Zimmermann *et al.* (2020). Harvested cells from 1 L of culture were resuspended in 100 mL ice-cold periplasmic

extraction buffer (100 mM Tris-HCl, pH 8, 20% (w/v) sucrose, 0.5 mM ethylenediaminetetraacetic acid (EDTA)) and incubated for 1 h at 4 °C under gentle agitation. 400 mL ice-cold periplasmic extraction supplement (25 mM Tris-HCl, pH 8, 5% (w/v) sucrose, 5 mM $MgCl_2$) was added, and incubation was extended by 20 min. After centrifugation for 20 min at 5,000 × *g* and 4 °C, the SN was supplemented with 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 20 mM imidazole, and Ni-NTA resin. IMAC was performed as described before, albeit using TBS instead of PBS. Due to high yields, only a fraction of the eluate was concentrated, injected on the SRT-C SEC-300 column pre-equilibrated in HBS, pooled, again concentrated, and snap frozen. The remaining eluate was dialysed in TBS at 4 °C O/N, concentrated, and snap frozen to be gel filtrated whenever fresh sybodies were needed.

2.3 Membrane Protein Characterization

2.3.1 Heat-Fluorescent Size Exclusion Chromatography

Heat-FSECs were carried out as stated in **Sec. 2.2.4**. An additional incubation step was executed at various temperatures (4–70 °C for 10 min) followed by centrifugation for 20 min at 90,000 × *g* and 4 °C. The SN was recovered and injected onto the column.

2.3.2 Polyacrylamide Gel Electrophoresis & Western Blotting

Standard sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) analysis was performed using gels with a 13.5 or 15% (w:v) acrylamide content (37.5:1 (w:w) acrylamide:bisacrylamide ratio), individually cast as required. Gels were typically run at 120–180 V for 45–70 min and stained using the Coomassie-based InstantBlue Protein Stain (Expedeon), unless stated otherwise. In-gel fluorescence images were taken prior to staining using an FLA-7000 Fluorescent Image Analyzer (Fujifilm) and the FAM detection channel, typically at 60% intensity. Western blotting was performed by transfer onto PVDF membrane (Thermo Scientific). Protein bands were detected after subsequent incubation with mouse α-Flag (Sigma-Aldrich) and goat α-mouse horseradish peroxidase (HRP)-conjugated (Dako, Agilent) antibodies using chemiluminescence (Amersham ECL Detection Kit, Cytiva).

Blue native (BN)-PAGE analysis was performed using precast 4–16% Bis-Tris Gels (Invitrogen) following the manufacturer's instructions (~3 µg protein was loaded per lane, and the Dark Blue Cathode Buffer was used).

2.3.3 Circular Dichroism Spectroscopy

CD spectroscopy measurements were taken using a J-815 spectropolarimeter (Jasco) in combination with a 1 × 10 mm quartz cuvette (Starna Scientific). Purified and concentrated protein

was diluted at least 100-fold into 10 mM K-phosphate, pH 7.5, 50 mM Na-sulfate, also containing 3 × critical micelle concentration (CMC) of the desired detergent. The final protein concentration used was 0.10–0.26 mg mL^{-1}. CD spectra ranging from 200–240 nm wavelength were recorded at 20–86 °C in 2 °C increments. To determine the melting temperature (T_m), ellipticity data near the minima at 208–210 and 220–222 nm were plotted against the temperature and fitted using the ''Sigmoidal, 4PL (four parameter logistic)'' equation in GraphPad Prism (Dotmatics).

2.3.4 Nano Differential Scanning Fluorimetry

NanoDSF profiles were generated using a Prometheus NT.48 (NanoTemper), operated at 10–50% intensity. Samples were prepared in standard capillaries and contained protein at 0.2–0.5 mg mL^{-1}, buffered at various pH values ranging from pH 5.5–8.0 using 20 mM MES-NaOH, HEPES-NaOH, or Na-phosphate, respectively, whilst also containing 150 mM NaCl and 3 × CMC of the respective detergent. Potential ligands were added at up to 10 mM concentration and loaded after a 5 min incubation step at RT. The fluorescence emitted at 350 and 330 nm ($\lambda_{Excitation}$ = 280 nm) was monitored, and its ratio was plotted against a temperature gradient from 20–90 °C, incremented at 1.0–1.5 °C min^{-1}. T_ms were calculated using the included analysis software (NanoTemper).

2.4 Generation of Sybodies & Nanobodies

2.4.1 Target Protein Biotinylation

Avi-tagged constructs were biotinylated using the BirA Biotin-Protein Ligase Kit (Avidity) with small adjustments made to the protocol. 400 µL biotinylation reactions were set up containing 40 µM purified *At*CTNS-Avi or *Hs*CTNS(Δ2–115)-Avi, 1 × Biomix A, 2 × Biomix B, and 12 µg, *i.e.* three times the recommended amount, GST-BirA, all diluted in mQ-H_2O (final NaCl concentration was ~50 mM). The biotin-protein ligase fused to glutathione-*S*-transferase (GST) was prepared in the lab; instructions are provided elsewhere (Fairhead & Howarth, 2015). The reaction mix was incubated on ice O/N, and subsequently, SEC was repeated in the same manner as during membrane protein purification. Peak fractions were pooled, combined, aliquoted, and snap frozen. Biotinylation efficiency was assessed using a streptavidin shift assay: A sample of biotinylated protein was incubated with 1.5 × molar excess streptavidin (Sigma-Aldrich) for 10 min on ice and subsequently analysed by SDS-PAGE, performed at 120 V and 4 °C.

2.4.2 Sybody Screening

Target-specific sybodies were generated following the steps outlined in Zimmermann *et al.* (2020) using C-terminally Avi-tagged and biotinylated *At*CTNS (in a DDM detergent micelle), as well as the dimeric and monomeric species of *Hs*CTNS(Δ2–115), solubilized in 10:1 (w:w) LMNG:CHS. The procedure is described hereafter based on a single mRNA library. However, selections were performed in parallel using all three libraries of concave, loop, and convex sybodies, colour-coded green (G), blue (B), and red (R), respectively. Considerable changes to the protocol are highlighted with a † followed by the altered detail, if applicable.

Ribosome Display, Reverse Transcription & Amplification of cDNA

10 µL *in vitro* translation reactions were assembled using the PURExpress 2.1 kit (GeneFrontier), including 0.7 µL of the mRNA sybody library (1.6×10^{12} mRNA strands), and incubated for 30 min at 37 °C to form ribosomal complexes. These were diluted in ice-cold 100 µL WTB-BSA-D (50 mM Tris-acetate, pH 7.5, 150 mM NaCl, 50 mM Mg-acetate, 0.5% (w/v) bovine serum albumin (BSA), 0.1% (w/v) detergent), also containing 5 mg mL^{-1} heparin and 40 U RNasin Ribonuclease Inhibitor (Promega), and centrifuged for 5 min at 20,000 × *g* and 4 °C. The SN was mixed with 50 nM biotinylated target protein and incubated on ice for 20 min (panning). 12 µL Dynabeads MyOne Streptavidin T1 (Invitrogen) were washed twice using 500 µL WTB-BSA and a DynaMag-2 magnetic rack (Invitrogen) to immobilize the beads. After blocking in 500 µL WTB-BSA for 20 min, the beads were washed three times in 500 µL WTB-BSA-D, subsequently combined with the panning mix, incubated for 5 min on ice to pull down the biotinylated target protein and bound ribosomal complexes, and then washed again three times using 500 µL ice cold WTB-D (3 × CMC detergent). The RNA was eluted by resuspending the beads in 100 µL 50 mM Tris-acetate, pH 7.5, 150 mM NaCl, 50 mM EDTA, 100 µg mL^{-1} yeast RNA, and purified using the RNeasy Micro Kit (Qiagen) according to the manufacturer's instructions (elution in 15 µL nuclease-free water). Reverse transcription was carried out as follows: First, a 20 µL mix containing the eluted RNA, 10 µM ''RT_Primer'', and 2 mM of each dNTP was heated to 65 °C for 5 min, then cooled on ice. Subsequently, components of the AffinityScript Reverse Transcription Kit (Agilent) were used to add 4 µL buffer, 4 µL DTT, 3.2 µL dNTP mix at 10 mM each († 100 mM), 5.8 µL nuclease-free water, 1 µL RNasin at 40 U µL^{-1}, and 2 µL reverse transcriptase. The 40 µL reaction was then incubated at 37 °C for 1 h, followed by enzyme deactivation at 95 °C for 5 min. The cDNA was purified using the NucleoSpin Gel and PCR Clean-up kit (Machery-Nagel), eluting in 30 µL mQ-H_2O, and subsequently PCR amplified using the GoTaq G2 DNA Polymerase (Promega) in the buffer it is supplied with. 100 µL reactions containing 1 × green buffer, 200 µM of each dNTP († 800 µM), 0.5 µM of each primer, the purified cDNA, and 2.5 U polymerase († 1.25 U) were prepared, and PCR was carried out as stated in the protocol. Amplified DNA was run on a 1.5% (w/v) agarose gel, extracted, and purified using the NucleoSpin Gel and PCR Clean-up kit.

Cloning of the Phagemid Libraries & Electrotransformation

PCR amplified DNA fragments and pDX_init vector (Addgene #110101) were digested separately for 1.5 h at 50 °C using the BspQI restriction enzyme, followed by a 20 min deactivation step at 80 °C. The DNA fragments were purified directly using the aforementioned kit; the linearized vector was gel extracted from a 1% (w/v) agarose gel and then purified. Ligation was carried out at a 4:1 (w:w) vector:insert ratio (1.5 µg DNA total) using 12.5 U T4 DNA ligase (Thermo Scientific) in its compatible buffer. After a 1.5 h incubation at 37 °C, followed by 10 min at 65 °C, the 50 µL ligation reaction was chilled, mixed with 350 µL ice-cold *E. coli* SS320 cells (Lucigen), and electroporated at 2,400 V, 36 µF († 25 µF), and 150 Ω († 750 Ω) using an ECM 399 (BTX, †) equipped with a 0.2 cm gap cuvette. For the sybody selection against *Hs*CTNS(Δ2–115), the ligation was performed differently: Vector:insert ratio was changed to 7:1 (w:w) using 2.35 µg DNA in total, only 7.5 U T4 ligase were added, and the reaction was incubated for 2 h at 37 °C followed by 20 min at 65 °C. Electroporated cells were recovered in 25 mL SOC medium (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 0.4% (w/v) D-glucose) for 30 min at 37 °C and 160 RPM. From this culture, the electrotransformation efficiency was determined by plating 5 µL of a 1:100 and 1:1,000 dilution on LB-agar supplemented with 120 µg mL^{-1} Amp, and counting the number of colonies after O/N incubation at 37 °C. The remainder of the culture was used to inoculate 225 mL 2YT medium (1.6% (w/v) tryptone, 1% (w/v) yeast extract, 0.5% (w/v) NaCl), supplemented with 2% (w/v) D-glucose and 200 µg mL^{-1} Amp, and grown O/N. A glycerol stock was prepared as a backup by mixing 1 mL of the culture with 0.8 mL 50% (v/v) glycerol and snap freezing.

Phage Rescue, Production & Purification

A 50 mL 2YT-glucose-Amp culture was inoculated using 1 mL of the O/N culture and grown for ~1.5–2 h at 37 °C and 160 RPM to an OD_{600} of 0.6. For phage rescue, 2.7×10^{10} plaque-forming units (PFU) M13KO7 helper phage (New England Biolabs) were placed in a 50 mL conical tube, and 10 mL of culture was added, mixed, and incubated for 30 min at 37 °C without shaking. Infected cells were harvested by centrifugation for 10 min at 4,200 × *g* and RT, resuspended in 50 mL 2YT-Amp, also supplemented with 25 µg mL^{-1} kanamycin (Kan), and grown O/N at 37 °C and 160 RPM to produce phages. The O/N culture was centrifuged for 30 min at 4,200 × *g* and 4 °C, 40 mL of the SN was recovered and mixed with 10 mL 20% (w/v) PEG 6000, 2.5 M NaCl, then incubated for 30 min on ice. Precipitated phage was pelleted by centrifugation for 30 min at 4,200 × *g* and 4 °C, resuspended in 1 mL ice-cold PBS, and centrifuged for 5 min at 20,000 × *g* and 4 °C twice to pellet residual bacteria and aggregates. The SN was recovered in a fresh 1.5 mL tube each time, and finally, the phage concentration was determined by ultraviolet (UV) absorption using two different wavelengths and the following equation:

$$\frac{\text{Phages}}{\text{mL}} = \frac{(A_{269} - A_{320}) \times 6 \times 10^{16}}{4{,}900} \tag{2.1}$$

Phage Displays & Enrichment Factor Analysis by qPCR

For the first round of phage display, half the wells of a Nunc-Immuno 96-well plate (Sigma-Aldrich) were coated O/N with 100 µL 67 nM neutravidin (Thermo Scientific) in PBS per well, subsequently washed with 250 µL TBS (containing 40 mM Tris-HCl during sybody and nanobody screening), then blocked with 250 µL TBS-BSA (0.5% (w/v)) for 30 min, and washed once with 250 µL TBS-BSA-D (0.1% (w/v) detergent). 4.9 mL ice-cold TBS-BSA-D containing 1×10^{12} phages mL^{-1} was supplemented with 50 nM biotinylated target protein, incubated for 20 min on ice (panning), and subsequently distributed into 47 of the 48 wells (100 µL per well). To the final well, an equivalent mix containing 50 nM biotinylated negative control (NC) protein was added, here peptide transporter from *Shewanella oneidensis* ($PepT_{So}$). After a 5 min incubation, all wells were washed three times with 250 µL TBS-D (3 × CMC detergent), each time incubating the solution for 2 min before removing it. Subsequently, 100 µL PD elution buffer (TBS supplemented with 0.25 mg mL^{-1} trypsin) was added to each well and incubated for 10 min at RT to cleave the phage from the bound sybodies. The PD elution was pooled and mixed with 40 µL TBS supplemented with 7.5 mg mL^{-1} 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF) hydrochloride (Sigma-Aldrich) to inhibit the trypsin (0.8 µL for the NC). 10 µL 1:10 (v:v) dilutions in mQ-H_2O were prepared for quantitative polymerase chain reaction (qPCR) analysis. 4 mL of an O/N SS320 culture in 2YT supplemented with 10 µg mL^{-1} tetracycline was used to inoculate 200 mL of 2YT and grown to an OD_{600} of 0.6. 45 mL of that culture was mixed with the 4.7 mL eluted phage and incubated for 30 min at 37 °C without shaking (the NC was not infected). Subsequently, the suspension was added to 200 mL 2YT-glucose-Amp and grown O/N at 37 °C and 160 RPM. Phage was produced as described before.

The second round of phage display was performed similarly to the first, albeit with a few modifications. Dynabeads MyOne Streptavidin C1 (Invitrogen) were used for immobilization instead of the plate: 12 µL were washed and blocked using TBS-BSA, then washed again in TBS-BSA-D as described for the ribosome display. The beads were resuspended in 100 µL panning solution containing 5×10^{12} phages mL^{-1} and 50 nM biotinylated target protein, incubated for 5 min on ice, and washed once with 500 µL TBS-BSA-D. The same was done for the NC. The washed beads of each reaction were resuspended in 100 µL competition buffer consisting of 5 µM non-biotinylated target protein in TBS-BSA-D, incubated for 3 min on ice, and subsequently washed twice with 500 µL TBS-D. The phage was eluted from the beads, and a qPCR sample was prepared as described for the NC in the first phage display. The 100 µL eluted phage was mixed with 1.4 mL of a freshly grown SS320 2YT culture for infection and then added to 10 mL 2YT-glucose-Amp for O/N cultivation.

For the sybody generation against *Hs*CTNS(Δ2–115), the second phage display was repeated once more. Hereby, two subsequent competition steps at 5 µM unbiotinylated protein were

performed instead of one. Additionally, the number of washes performed after was increased to five.

Enrichment factors were calculated by qPCR. 2 µL of the collected samples were analysed in 10 µL reactions using the SYBR Select Master Mix (Applied Biosystems) supplemented with 5% (v/v) DMSO and 0.3 µM of primers that anneal to the pDX_init vector backbone (''qPCR_PD_pDX_for'' and ''qPCR_PD_pDX_rev''). DNA was amplified on a C1000 Touch Thermal Cycler equipped with a CFX96 module (Bio-Rad) for 40 cycles (10 s at 95 °C, then 30 s at 63 °C). CFX Maestro (Bio-Rad) was used to compare the accumulated DNA fragments of the target and NC to a standard curve established using empty pDX_init vector.

Single Clone Expression, ELISA & Sequence Analysis

The O/N culture was harvested by centrifugation for 10 min at 4,200 × *g* and 4 °C, and phagemids were isolated by miniprep. An additional purification step (†) was performed by transforming 5 µL of the eluted DNA into 100 µL OmniMAX by heat shock. Cells were recovered for 40 min in 0.5 mL LB and subsequently used to inoculate 8 mL of LB-glucose-Amp. After O/N cultivation, the DNA was again purified by miniprep and subsequently used for subcloning into the pSb_init expression vector using the FX method (analogous to **Sec. 2.1.3**). Transformations were recovered and plated on multiple LB-agar-CAM plates. 96 individual colonies were picked and used to inoculate 1.2 mL TB-CAM, each using a deep-well plate (pre-culture). Cells were grown for 4–5 h at 37 °C and 300 RPM until turbid and subsequently diluted 20-fold into 96 × 1 mL TB-CAM (expression culture). Both cultures were cultivated for 2 h, then the temperature setting was changed to 22 °C, and cultivation was continued for 1.5 h. The expression culture was induced at an OD_{600} of 0.4–0.8 by adding 0.02% (w/v) L-(+)-arabinose, and cultivation was continued O/N. Cells were pelleted by centrifugation for 20 min at 4,000 × *g* and 4 °C. The pre-culture was frozen, cell pellets of the expression culture were resuspended in 100 µL well^{-1} ice-cold periplasmic extraction buffer supplemented with 0.5 µg mL^{-1} lysozyme (Sigma-Aldrich), resuspended by vortexing, and incubated for 30 min on ice. 900 µL TBS supplemented with 1 mM $MgCl_2$ was added, and centrifugation was repeated. The ensuing SN was used as periplasmic extract during ELISA.

Individual sybody binders were identified by ELISA, carried out using two Nunc-Immuno 96-well plates coated O/N with 100 µL of 5 µg mL^{-1} protein A (Sigma-Aldrich) in PBS per well. After initial washing in 250 µL well^{-1} TBS, the plates were blocked using 100 µL well^{-1} TBS-BSA for 30 min, followed by three more washes. To the blocked and washed plates, the following molecules were added successively: 1 µg mL^{-1} α-c-Myc antibody (Sigma-Aldrich), 20% (v/v) periplasmic extract (adding the same extract to consecutive columns), 50 nM biotinylated target or NC protein (alternating by column), and 0.2 µg mL^{-1} streptavidin-peroxidase (Supelco). Each time, 100 µL of the respective molecule diluted in ice-cold TBS-BSA-D was added per well and incubated for 20 min on ice. Each incubation was followed by three washes using 250 µL well^{-1}

cold TBS-D. The ELISA was finally developed by adding 100 µL 51.5 mM Na_2HPO_4, 24.3 mM citric acid, 0.006% (v/v) H_2O_2, 1 mg mL^{-1} 3,3',5,5'-tetramethylbenzidine (TMB) (Sigma-Aldrich) per well. After incubation for 15 min at RT, the absorbance at 650 nm was measured using a SpectraMax M3 Microplate Reader. The plasmids of ELISA-positive wells (criterion: A_{650} ratio of target/NC > 1.5) were recovered by miniprep using the frozen cell pellets of the pre-culture and subsequently sent for Sanger sequencing requesting primer ''PBADF'' (Source BioScience). Sybody sequences were phylogenetically analysed using ETE v3 (Huerta-Cepas *et al.*, 2016) as implemented on the GenomeNet, accessible online *via* https://www.genome.jp/tools/ete/. Groups of unique sequences were numbered from top to bottom.

Small-Scale Recombinant Production of Tagged Sybodies

Individual sybody plasmids were transformed into MC1061 cells (~50–100 ng in ~20 µL) by heat shock, recovered, and plated on LB-agar-CAM. 1 mL TB-CAM starter cultures were inoculated, grown for 2 h at 37 °C and 200 RPM, and used to inoculate 50 mL TB-CAM expression cultures. After cultivation for 2 h at 37 °C and 200 RPM, the temperature setting was changed to 22 °C, and cultivation continued. Expression was induced at an OD_{600} of ~0.6 using a final concentration of 0.02% (w/v) L-(+)-arabinose and carried out O/N. Cells were harvested by centrifugation for 20 min at 4,000 × *g* and 4 °C, resuspended in 5 mL ice-cold periplasmic extraction buffer supplemented with 0.5 µg mL^{-1} lysozyme, and incubated for 30 min on ice. 20 mL TBS supplemented with 1 mM $MgCl_2$ was added, and cells were pelleted again by centrifugation. The SN was recovered and supplemented with 150 mM NaCl, 15 mM imidazole, and 100 µL HisPur Ni-NTA beads. Incubation for 1 h under gentle agitation was followed by centrifugation for 2 min at 500 × *g* to settle the beads and subsequent transfer into a Pierce Micro-Spin Column (Thermo Scientific). Spinning at 50 × *g* for 10–20 s, the beads were washed four times using 500 µL TBS supplemented with 150 mM NaCl and 30 mM imidazole. Elution was carried out using 300 mM imidazole in the same buffer after a 5 min incubation step. The eluate was injected onto a Sepax SRT-C SEC-300 column or, in the case of sybodies generated against *Hs*CTNS(Δ2–115), a Sepax SRT-C SEC-100 column (Chromex Scientific). Peak fractions were collected and concentrated, if necessary, then snap frozen.

2.4.3 Nanobody Screening

To identify *At*CTNS-specific nanobodies, a library was raised through the immunization of a llama using reconstituted protein and screened using C-terminally Avi-tagged and biotinylated protein. Immunization, cloning, and library construction were executed by Prof. Raymond Owens' group (The Rosalind Franklin Institute), specifically Dr. Jiandong Huo, under the authority of the project license PPL 70/8108, and using a protocol modified from Pardon *et al.* (2014). The library was received in the group of Prof. Simon Newstead, and α-*At*CTNS

nanobodies were selected using a modernized version of Pardon *et al.* (2014) by implementing reagents and equipment from the sybody selection (Zimmermann *et al.*, 2020).

Llama Immunization, Repertoire Cloning & Library Construction

Nanobodies were raised in a llama following intramuscular immunization with purified protein reconstituted into liposomes (proteoliposome reconstitution is described in **Sec. 2.6.2**) and using Gerbu LQ 3000 as the adjuvant. 150 mL blood was collected, and peripheral blood mononuclear cells were prepared using Ficoll-Paque PLUS (Cytiva) according to the manufacturer's instructions. Total RNA was extracted using TRIzol, and nanobody cDNAs were generated by reverse transcription PCR using primers, as detailed previously (Parker *et al.*, 2021b). The pool of nanobody encoding sequences was amplified by two rounds of nested PCR: firstly with ''CALL_001'' and ''CALL_002'', followed by ''VHH_For'' and ''VHH_Rev_IgG2'' and ''VHH_Rev_IgG3'', and cloned into the SfiI sites of the phagemid vector pADL-23c. Electrocompetent *E. coli* TG1 cells (Lucigen) were transformed with the recombinant pADL-23c vector, resulting in a nanobody library of about 2×10^8 independent transformants. The resulting TG1 library stock was used to obtain a library of nanobody-presenting phages.

Phage Rescue, Production & Purification

A 50 mL 2YT-glucose-Amp (100 µg mL^{-1}) culture was inoculated using a 1 mL glycerol stock representing the complete diversity of the nanobody library. The culture was grown for ~1.5 h at 37 °C and 200 RPM to an OD_{600} of 0.5. For phage rescue, 4×10^{10} PFU M13KO7 helper phage was placed in a 50 mL conical tube and 20 mL of culture was added, mixed, and incubated for 30 min at 37 °C without shaking, followed by 30 min at 37 °C and 80 RPM. Infected cells were harvested by centrifugation for 10 min at 2,800 × *g* and RT, resuspended in 400 mL 2YT-Amp-Kan, and grown O/N at 25 °C and 200 RPM to produce phages. The O/N culture was centrifuged for 15 min at 4,000 × *g* and 4 °C, and the SN was recovered and mixed with 100 mL 20% (w/v) PEG 6000, 2.5 M NaCl, then incubated for 60 min on ice. Precipitated phage was pelleted by centrifugation for 10 min at 12,000 × *g* and 4 °C, resuspended in 1 mL ice-cold PBS, and centrifuged for 1 min at 20,000 × *g* and 4 °C. The SN was mixed with 250 µL of the PEG-NaCl solution to re-precipitate the phage and incubated for 10 min on ice. Phage was pelleted again (15 min, 20,000 × *g*, 4 °C), resuspended in 1 mL ice-cold PBS, again centrifuged for 1 min, and finally, the SN was recovered. The concentration of infective phage was determined by mixing 90 µL of freshly grown TG1 cells (at OD_{600} = 0.5, in LB) with 10 µL of a serial dilution of the phage. After incubation for 15 min at 37 °C without shaking, 5 µL drops were placed on LB-agar-glucose-Amp plates, air-dried, and incubated O/N at 37 °C.

Phage Displays & Enrichment Factor Analysis

12 µL Dynabeads MyOne Streptavidin C1 were washed three times using 500 µL TBS-BSA, then blocked for 2 h. Subsequently, the beads were washed five times using 500 µL TBS-D (0.1% (w/v) DDM), then resuspended in 100 µL 50 nM biotinylated *At*CTNS or $PepT_{So}$, each in TBS-BSA-D (0.05% (w:v) BSA, 0.1% (w/v) DDM), and incubated for 15 min at RT. Beads were again washed five times, resuspended in 100 µL TBS-BSA-D pre-incubated (30 min at RT) with 1×10^{11} infective phage, and incubated for 2 h at RT. Subsequently, the beads were washed 15 times using 500 µL TBS-D, resuspended in 100 µL PBS supplemented with 0.25 mg mL^{-1} trypsin, and incubated for 30 min at RT. Beads were immobilized, and the SN recovered and mixed with 3 µL PBS supplemented with 7.5 mg mL^{-1} AEBSF hydrochloride. 50 µL of the phage elution was added to 3 mL freshly grown TG1 cells (at OD_{600} = 0.5, in LB) and incubated for 30 min at 37 °C without shaking. The mixture was added to 7 mL LB, supplemented with D-glucose and Amp, and finally grown O/N at 37 °C and 170 RPM. 0.5 mL of the O/N culture was then used to inoculate the 50 mL 2YT-glucose-Amp culture, which is subsequently used for phage rescue, production, and purification as described above. A second round of phage display was performed in an identical manner to the first. After both displays, unused eluted phage was used to determine the concentration of infective phage, both for the target and NC protein, as described above, thus determining the factors of enrichment.

Single Clone Expression, ELISA & Sequence Analysis

After the second round of phage display, a fraction of the 10 mL O/N culture was diluted 1,000- and 10,000-fold and plated on LB-agar-glucose-Amp, then incubated O/N at 37 °C. 191 individual colonies were picked and expressed as described in the sybody screening (**Sec. 2.4.2**), albeit with some modifications, detailed hereafter. Position A1 of the first 96-deep-well plate was inoculated using Syb39 from the pSb_init vector and transformed in MC1061 cells (the protocol was left unchanged for this well). The medium for the pre-culture was changed to 2YT-glucose-Amp, also supplemented with 10% (v/v) glycerol. Furthermore, antibiotics were changed to Amp throughout the expression, and isopropyl β-D-1-thiogalactopyranoside (IPTG) was used for induction (100 µL at 10 mM in TB).

ELISA and sequence analysis were carried out as for the sybodies (see **Sec. 2.4.2**), with the sole difference being the primer used during sequencing: For the nanobodies, primer ''Phagemid_Seq_fw'' (**Table A.2**) was used.

Small-Scale Recombinant Production of Tagged Nanobodies

Individual nanobodies were purified after expression in chemically competent *E. coli* WK6 cells (ATCC) using the same protocol as for the sybodies, but using Amp as the antibiotic marker and IPTG (1 mM final concentration) as the induction agent.

2.4.4 Size Exclusion Chromatography Co-Elutions

1.25 nmol *At*CTNS and 1.88 nmol sybody (1:1.5 molar ratio) were incubated in 100 µL MBS supplemented with 0.03% (w/v) DDM for 1 h on ice. Subsequently, the mixture was applied to a Sepax SRT-C SEC-300 column pre-equilibrated in the same buffer. Peak fractions were collected and analysed by SDS-PAGE. Gels were stained using the ProteoSilver Plus Silver Stain Kit (Sigma-Aldrich) according to the manufacturer's instructions.

2.4.5 Pulldowns Using His-Tagged Binders

150 pmol *At*CTNS or *Hs*CTNS(Δ2–115) were incubated with 150 pmol His-tagged sybody or nanobody in 50 µL HBS at pH 7.5 or MBS at pH 6.5, each supplemented with 3 × CMC detergent, for 1 h on ice. 10 µL samples were removed to serve as loading controls. To the remaining 40 µL reactions, 12.5 µL HisPur Ni-NTA (25 µL of a 50% (v/v) slurry equilibrated in the aforementioned buffer) and 15 mM imidazole (0.975 µL of a 1 M stock at pH 8) were added, and tubes incubated for 30 min at 4 °C under continuous end-over-end rotation. The resin was settled by centrifugation for 30 s at 500 × *g*, and the SN was removed using gel-loading tips. The resin was washed three times using 500 µL of the aforementioned buffer supplemented with 15 mM imidazole and transferred into a fresh 1.5 mL tube during the last wash. The resin was finally resuspended in a mixture of 12 µL 1 M imidazole stock and 6 µL 4 × SDS Sample Buffer (Millipore), incubated for 5 min on ice, and analysed by SDS-PAGE.

2.4.6 Bio-Layer Interferometry

Specific sybody and nanobody binding was measured on an Octet RED384 (Sartorius) using Streptavidin (SA) Biosensors (ForteBio). Experiments were performed at RT in MBS supplemented with 0.03% (w/v) DDM (for *At*CTNS) or HBS supplemented with 0.003% (w/v) 10:1 (w:w) LMNG:CHS (for *Hs*CTNS(Δ2–115)). Sensors were equilibrated in buffer for at least 20 min. After a 5 min baseline step, sensors were loaded with 100 or 20 nM biotinylated protein (specified each time in the figure legend) for 5 min, followed by another baseline step. Association and dissociation steps of the desired binder were each performed for 5–10 min. Data were fitted and k_{off} values calculated using the ''Association then dissociation'' equation in GraphPad Prism. Only reasonable fits were used ($R^2 > 0.95$); reported error: standard deviation (SD). To calculate the K_D, a dilution series of the binder was used, and data were fitted in the Octet v9.0 software package, extracting the response after equilibration ($Response_{eq}$) for each concentration. These were imported into GraphPad Prism and fitted using the ''One site -- Specific binding'' equation. K_D values were calculated from a minimum of three independent experiments, reported error: SD. For epitope binning, multiple association steps were performed successively using a mixture of a new and the previous binder to minimize dissociation of the previous binder.

2.5 Structural Analysis of CTNS

2.5.1 Crystallization

Crystallization was performed using protein at ~15–43 mg mL^{-1} final concentration, as determined using absorbance at 280 nm. Protein complexes were formed by incubating *At*CTNS and tag-free binder(s) at a 1:1.2 molar ratio for 1 h on ice and subsequently applying the sample to a Sepax SRT-C SEC-300 column pre-equilibrated in MBS supplemented with 0.013% (w/v) DDM. Peak fractions were pooled and concentrated using a 50 kDa MWCO Amicon Ultra Centrifugal Filter Unit. L-Cystine was added to a final concentration of 0.5 mM, and the protein was left on ice for >2 h prior to crystallization using the lipidic cubic phase (LCP) method (Caffrey & Cherezov, 2009). L-Cystine-*S*,*S*-dioxide was added at 10 mM 30 min prior to the binders, and SEC was performed using MBS-DDM supplemented with 10 mM of the alternative ligand. Protein-laden mesophase was obtained by mixing protein with monoolein or a 10:1 (w:w) mix of monoolein and cholesterol (both Anatrace) in a 40:60 (w:w) ratio using a coupled syringe device (Art Robbins). 50 nL boli (drops) were placed on hydrophobically coated (Rain-X) glass plates (127.8 × 85.5 × 1.0 mm, Thermo Scientific) with an attached perforated double-stick spacer (112 × 77 × 0.14 mm, Saunders) in a defined 96-well format, and 800 nL precipitant solution was added. Both steps were performed using a Crystal Gryphon LCP robot (Art Robbins). Precipitant solutions originated from MemMeso (Molecular Dimensions) and a set of screens kindly made available by Prof. Nureki of the University of Tokyo (screens were assembled using the services of Molecular Dimensions). Optimization screens around promising conditions were prepared in-house. Cover plates (112 × 77 × 0.2 mm, Thermo Scientific) were affixed, and the plate sandwiches were incubated at either 20, 19, or 4 °C. Initial crystals typically appeared after 1–3 days and matured after 5–10 days. Bright-field (BF) and UV images were taken using a JANSi UVEX Imaging System (UVEXm, SWISSCI) with an exposure time of 0.5 s and level 10 edge lighting. Wells were opened using a tungsten carbide glass cutter, and the crystals were harvested using 35–100 µm MicroMounts (MiTeGen). Crystals were cryo-cooled directly in liquid nitrogen and stored in unipucks (MiTeGen).

2.5.2 Heavy Atom Soaking

Wells containing crystals to be soaked were opened as described above. 1 µL of the appropriate precipitant solution, also containing 0.5 mM $Hg(OAc)_2$, was added to the bolus/mother liquor, and the well was sealed again using a glass cover slip. After incubation for 8 h at 20 °C, the well was opened again, and the crystals were harvested as before.

2.5.3 X-Ray Diffraction Data Collection

Conventional X-ray diffraction data were collected at microfocus beamlines I24 and I04, Diamond Light Source (DLS), UK, equipped with PILATUS3 6M (Kraft *et al.*, 2009) and EIGER2 XE 16M (Casanas *et al.*, 2016) detectors (DECTRIS), respectively. Crystals were initially centred through grid scans performed at the following settings: I24: 10 × 10 µm box size, 0.05 s exposure, 15–20% transmission; I04: 11 × 5 µm, 0.005 s, 100%. At I24, typically datasets of 180° (0.5° rotation increment, 0.05 s exposure, 15–20% transmission, 7 × 7 µm beam size) were collected, whilst at I04, 360° datasets (0.1°, 0.005–0.008 s, 100%, 11 × 5 µm) were generated. In each case, a wavelength between 0.9762–0.9999 Å was used. Examples of typical diffraction images were converted to .tif format using Adxv (Arvai, 2021) and included in this thesis. For those collected using the EIGER2 XE 12M detector, 10 slabs of consecutive 0.1° rotation were summed together to increase the visibility of the reflections.

Sulfur-single-wavelength anomalous dispersion (SAD), *i.e.* S-SAD, data were collected in collaboration with Dr. Kamel El Omari and Dr. Armin Wagner at I23, DLS, UK, on a PILATUS2 12M detector (DECTRIS). 360° datasets (0.1° rotation increment, 0.1 s exposure, 100% transmission, beam size ranging from 50 × 50 to 100 × 140 µm) were collected at a temperature close to 60 K and a wavelength of 2.7552 Å.

Hg-SAD data were collected at I24 in 180° sweeps (0.25° rotation increment, 0.05 s exposure, 10% transmission, 7 × 7 µm beam size) using a wavelength of 1.008 Å.

2.5.4 X-Ray Diffraction Data Processing

The data were processed with xia2 DIALS, xia2 3dii (Winter, 2010), or autoPROC (Vonrhein *et al.*, 2011), either initiated automatically through the DLS processing pipelines, which provide processed files and statistics *via* the Information System for Protein Crystallography Beamline (ISPyB)'s online interface (Delagenière *et al.*, 2011; Fisher *et al.*, 2015), or run locally using the departmental server. These pipelines utilize DIALS (Winter *et al.*, 2018) or XDS (Kabsch, 2010) for indexing and integration. Subsequent symmetry analysis was performed in POINTLESS (Evans, 2006), scaling and merging in AIMLESS (Evans & Murshudov, 2013), followed by data reduction in CTRUNCATE (Evans & Murshudov, 2013), or using the analogous tools in DIALS (Beilsten-Edmands *et al.*, 2020; Winter *et al.*, 2022). Where XDS was used, scaling was performed in XSCALE (Kabsch, 2010), followed by POINTLESS, AIMLESS (merge only), and CTRUNCATE.

2.5.5 Crystal Structure Determination

Reflection data generated during processing were used in .mtz format (free.mtz from xia2 and truncate-unique.mtz from autoPROC) during structure solution and refinement. Anomalous

data were analysed and prepared in SHELXC (Sheldrick, 2008, 2010) using unmerged reflection files in .HKL or .sca format (from XDS/XSCALE). Phase estimates for sybodies and nanobodies in the first co-crystal structure were obtained by molecular replacement (MR) in Phaser (McCoy *et al.*, 2007); initial phases for *At*CTNS were obtained by S-SAD phasing and density modification in PHENIX AutoSol (Adams *et al.*, 2010; Terwilliger *et al.*, 2009). Model building into the electron density was performed in Coot (Emsley & Cowtan, 2004; Emsley *et al.*, 2010), and the accuracy of the model was determined using S and Hg anomalous difference maps created in ANODE (Thorn & Sheldrick, 2011) using the model and fa.hkl file generated by SHELXC. Structure refinement was mainly carried out in BUSTER (Blanc *et al.*, 2004; Bricogne *et al.*, 2017). Subsequent co-crystal structures were solved by MR using the finalized *At*CTNS structure as a search model. The 1.96 Å S-SAD data of *At*CTNS-Syb1-Nb4 was solved automatically in PHENIX AutoBuild (Terwilliger *et al.*, 2008) as implemented in the DLS downstream processing pipeline, accessible *via* ISPyB. The L-cystine-bound crystal structure (PDB: 7ZKW) exhibited pseudo-merohedral twinning in space group *C*2 with twin law $-h-2l$, $-k$, l applied during refinement using REFMAC (Murshudov *et al.*, 2011). Geometry restraints for ligands L-cystine, DDM detergent, and monoolein lipids were calculated using the Grade Web Server (Smart *et al.*, 2020) with isomeric SMILES strings as input, retrieved from the PubChem website (Kim, 2021). Model validation was carried out using MolProbity (Chen *et al.*, 2010; Williams *et al.*, 2018). Images were prepared using PyMOL (Schrödinger, LLC, 2020). Polder OMIT maps were generated in PHENIX (Liebschner *et al.*, 2017). Contour maps of .mtz format were converted using the FFT program in CCP4 to allow for display in PyMOL.

2.5.6 Cryo-EM Grids Preparation & Screening

Complexes of *Hs*CTNS(Δ2–115) and His-Myc-tagged sybodies were formed as described in **Sec. 2.5.1**, but gel filtrated in HBS supplemented with 0.003% (w:v) 10:1 (w:w) LMNG:CHS, concentrated only if necessary, and used immediately for grid preparation. In collaboration with Dr. Takafumi Kato (Newstead group), 2.5 µL samples were applied onto freshly glow-discharged (15 mA for 20 s) Quantifoil R1.2/1.3 Au 300 mesh grids and vitrified in liquid ethane using a Vitrobot Mark IV (FEI) using a blot force of −5 and blot time of 3 s (100% humidity, 4 °C). Grids were screened on a Talos Arctica 200 kV cryo-TEM (FEI) using a defocus of −3.0 µm.

2.5.7 Cryo-EM Data Collection, Processing & Model Building

Data for single particle analysis were collected and processed by Dr. Takafumi Kato. Using the counting super-resolution mode on a Titan Krios G3 300 kV cryo-TEM (FEI) with a BioQuantum imaging filter (Gatan) and K3 direct detection camera (Gatan) at ×105,000 magnification, 19,555 movies were collected for the *Hs*CTNS(Δ2–115)-Syb30-Myc-His complex (0.41 mg mL^{-1} concentration) at a dose rate of 16.8 e$^-$ Å^{-2} s^{-1} and exposure time of 2.50 s, corresponding to a total dose of 42 e$^-$ Å^{-2} over 40 fractions.

Initial micrograph processing was performed in real-time using the SIMPLE pipeline (Caesar *et al.*, 2020): SIMPLE-unblur for patched (15 × 10) motion correction, SIMPLE-CTFFIND for patched contrast transfer function (CTF) estimation, and SIMPLE-picker for particle picking and extraction. All subsequent processing was performed in either cryoSPARC (Punjani *et al.*, 2017) or RELION-3.1 (Zivanov *et al.*, 2018) using the csparc2star.py script within UCSF pyem (Asarnow *et al.*, 2019) to convert between formats. Resolution estimates were derived from gold-standard Fourier shell correlations (FSCs) using the 0.143 criterion as calculated in cryoSPARC. Local resolution estimations were also calculated in cryoSPARC.

The workflow for cryo-EM image processing of the *Hs*CTNS(Δ2–115)-Syb30-Myc-His complex is shown in **Fig. 8.17**. We first used 3,016 micrographs (dataset 1) to run two rounds of 2D classification, followed by a four-class *ab initio* reconstruction in cryoSPARC using 1,367,906 particles derived from the second round of 2D classification. Subsequent non-uniform (NU) refinement (Punjani *et al.*, 2020) yielded a first putative CTNS-Syb map at 3.4 Å. Next, all 14,795,683 particles extracted from both datasets 1 and 2 were subjected to one round of 2D classification in cryoSPARC, and 7,546,005 selected particles were used for heterogeneous refinement under two junk maps (from the first *ab initio* reconstruction) and the target map. Particles belonging to the CTNS-Syb class were subjected to one round of 2D classification in cryoSPARC, and subsequently, 2,686,157 particles were used for multi-class *ab initio* reconstruction ($K = 3$). NU refinement with $C2$ symmetry expansion (8 Å initial low-pass filter) was performed using 980,414 particles, yielding a 2.7 Å map, which was Bayesian polished (Zivanov *et al.*, 2019) in RELION. Polished particles were subjected to 2D classification in cryoSPARC and subsequent alignment-free classification with a whole protein mask in RELION. 754,866 particles belonging to CTNS-Syb were used for NU and local refinements (8 Å initial low-pass filter) with $C1$ or $C2$ symmetry, and the final maps were determined at 2.60 Å and 2.45 Å, respectively, based on the FSC = 0.143 criterion.

The outward-open *Hs*CTNS model (PDB: 8DKI), deposited by Guo *et al.* (2022), was manually fitted in the map using UCSF ChimeraX (Pettersen *et al.*, 2021). Model readjustments and ligand additions were carried out in Coot as described in **Sec. 2.5.5**, followed by real-space refinement in PHENIX (Afonine *et al.*, 2018). Figures were prepared in ChimeraX and PyMOL (Schrödinger, LLC, 2020). The model was validated using MolProbity (Chen *et al.*, 2010; Williams *et al.*, 2018).

2.6 Functional Analysis of CTNS

2.6.1 Cystine Binding Assay

DDM-purified and concentrated WT, as well as mutant forms of *At*CTNS, were diluted to 10 µM (~0.32 mg mL^{-1}) using HBS containing 0.03% (w/v) DDM and up to 3.2 mM L-cystine, preset to

pH 7.20 using a pH meter (L-cystine was added from a 250 mM stock in 1 M HCl). L-Cystine binding was quantified from the tryptophan fluorescence quench observed during nanoDSF experiments, which were executed as described in **Sec. 2.3.4** (25% intensity), but using a reduced temperature slope of 0.5 °C min^{-1}. Ten subsequent data points commencing from 25 °C were averaged and subtracted from a control run without L-cystine. For each *At*CTNS variant, the averaged difference from a triplicate (using a fresh L-cystine stock for each replicate) was then normalized by dividing it by that of the average of a WT triplicate.

2.6.2 Proteoliposome Reconstitution

Purified WT and mutant forms of *At*CTNS were reconstituted into lipid vesicles consisting of a 3:1 (w:w) ratio of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE) and 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (POPG) by rapid dilution. To do so, the proteins were first exchanged into DM by SEC using MBS supplemented with 0.3% (w/v) DM and subsequently mixed with the vesicles (extruded 11 times through a 0.4 µm polycarbonate membrane; Avanti) at a 40:1 (w:w) lipid:protein ratio. The lipid:protein mix was incubated at RT for 20 min before diluting it 27-fold into ice-cold reconstitution buffer (50 mM K-phosphate at pH 7.0). Proteoliposomes were harvested by ultracentrifugation at 200,000 × *g* and 4 °C for >2 h and dialysed O/N extensively against two batches of reconstitution buffer. Proteoliposomes were recovered and resuspended to a final concentration of 0.5 µg $µL^{-1}$ as analysed by SDS-PAGE using ImageJ, comparing all bands of reconstituted protein to a known concentration of detergent-solubilized protein. Proteoliposomes were subjected to two rounds of freeze-thaw using liquid nitrogen and stored at −80 °C. For llama immunization to raise nanobodies, a total of 600 µg *At*CTNS was reconstituted at a lipid:protein ratio of 30:1, and proteoliposomes were resuspended to a final concentration of 1 mg mL^{-1}.

2.6.3 Transport Assay Using Pyranine

Proteoliposomes were pelleted by ultracentrifugation at 108,000 × *g* and 4 °C for 25 min and subsequently resuspended in INSIDE buffer consisting of 5 mM HEPES-NaOH, 120 mM KCl, 2 mM $MgSO_4$, and 2 mM pyranine (trisodium 8-hydroxypyrene-1,3,6-trisulfonate) at pH 7.2. Proteoliposomes were subjected to eight rounds of freeze-thaw before being extruded 11 times through a 0.4 µm polycarbonate membrane. The uniquely formed vesicles were then harvested through ultracentrifugation as before, but at 18 °C, and external pyranine was removed by applying the sample to a G-25 column (Cytiva) pre-equilibrated in INSIDE buffer without pyranine. Vesicles were collected again and finally resuspended in INSIDE buffer without pyranine. Transport assays were performed at 25 °C in a cuvette equipped with a small magnetic flea using a Cary Eclipse Fluorescence Spectrophotometer (Agilent), set at dual excitation at 460 and 415 nm, with emission at 510 nm. At the start of each assay, vesicles containing 5 µg *At*CTNS were diluted into 545 µL OUTSIDE buffer (5 mM HEPES-NaOH, 120 mM NaCl, 2 mM

$MgSO_4$ at pH 7.2). After 15 s, substrates were added per 100 µL from a freshly prepared stock in OUTSIDE buffer (typically 200 µM final), and transport was initiated with a final concentration of 1 µM valinomycin (v) in OUTSIDE buffer at 30 s. The ratio of the resulting fluorescence profiles was calculated (excitation at 460 nm over 415 nm) and plotted using GraphPad Prism or Excel (Microsoft); individual traces were normalized to one (using five averaged data points at the 0.02 min mark) to allow for comparison. Bar charts were generated using the response at the 3 min mark (five averaged data points) and normalized to the average response of WT *At*CTNS at 200 µM L-cystine.

2.6.4 Radioactive Transport Assay

Proteoliposomes were pelleted and resuspended in INSIDE buffer (10 mM MES-NaOH, 120 mM K-acetate, 2 mM $MgSO_4$ at pH 6.2). Four rounds of freeze-thaw were followed by extrusion through a 0.4 µm polycarbonate membrane. To increase the proteoliposome concentration, vesicles were harvested again at 18 °C and resuspended in INSIDE buffer. Starting each assay, vesicles containing 2 µg protein were diluted into 250 µL OUTSIDE buffer (10 mM MES-NaOH, 120 mM NaCl, 2 mM $MgSO_4$ at pH 6.2) also containing 48 µM L-cystine, 2 µM radiolabelled ^{14}C-L-cystine (specific activity 0.2 Ci mmol^{-1}; Hartmann Analytic), and 10 µM valinomycin (v). Assays were performed at 20 °C. 45 µL fractions were removed at different time points, diluted into 2 mL of OUTSIDE buffer, and vesicles were immediately isolated on a 0.22 µm cellulose filter (Merck) using a vacuum manifold. Filters were washed twice with 2 mL OUTSIDE buffer and subsequently transferred into Ultima Gold (PerkinElmer) scintillation liquid. Remaining radioactivity stemming from the inside of the vesicles was measured using a Wallac scintillation counter.

CHAPTER 3

Screening & Characterization of CTNS

The yeast *Saccharomyces cerevisiae* has been described as an excellent host for overexpressing eukaryotic membrane proteins, often yielding milligram amounts of well-folded and functional material that can be used for structural and biochemical investigations (Newstead *et al.*, 2007). At the time this project was started, yeast cell culture was used routinely in the lab to recombinantly produce different eukaryotic membrane proteins, including members of the PQ-loop family, like the KDEL receptor (Bräuer *et al.*, 2019), as well as representatives of the solute carrier (SLC) superfamily (Parker *et al.*, 2019). Since CTNS belongs to both of these groups, it made sense to try and use yeast as the primary expression system for CTNS. In this chapter, I will show that, among other CTNS homologues, only the plant version from *Arabidopsis thaliana* can be expressed efficiently enough in yeast to enable crystal trials for structure determination (**Chs. 4**, **6**, and **7**), as well as liposome assays (**Ch. 9**) to address mechanistic questions. *At*CTNS can be purified to homogeneity using standard purification techniques, yielding a thermally stable eukaryotic membrane protein that features a high and pH-dependent melting point of ~60–70 °C. Ligand addition increases the melting temperature only marginally but quenches the tryptophan fluorescence before the protein starts to unfold, which we exploited to establish a ligand binding assay for *At*CTNS.

3.1 Homologue Screening

3.1.1 Construct Comparison

To identify a specific homologue of CTNS most suitable for structural analyses, we first cloned *CTNS* genes from five different organisms into pDDGFP-Leu2d, the lab's optimized vector for expressing membrane proteins as C-terminal GFP-His fusion proteins (Parker & Newstead, 2014). The constructs included human CTNS (*Homo sapiens* (*Hs*), UniProt: O60931), mouse (*Mus musculus* (*Mm*), P57757), chicken (*Gallus gallus* (*Gg*), E1BXK0), zebrafish (*Danio rerio* (*Dr*), F1QM07), as well as a plant version of CTNS (*Arabidopsis thaliana* (*At*), P57758). All genes share a high sequence identity, with human and mouse CTNS matching the most residues at any position (84% identical), followed by chicken (67%), zebrafish (60%), and plant CTNS, which lacks the long N-terminal tail but retains 45% of sequence identity in the TM region. Sequences are aligned in **Fig. 3.1** with the potential location of the seven TM helices predicted

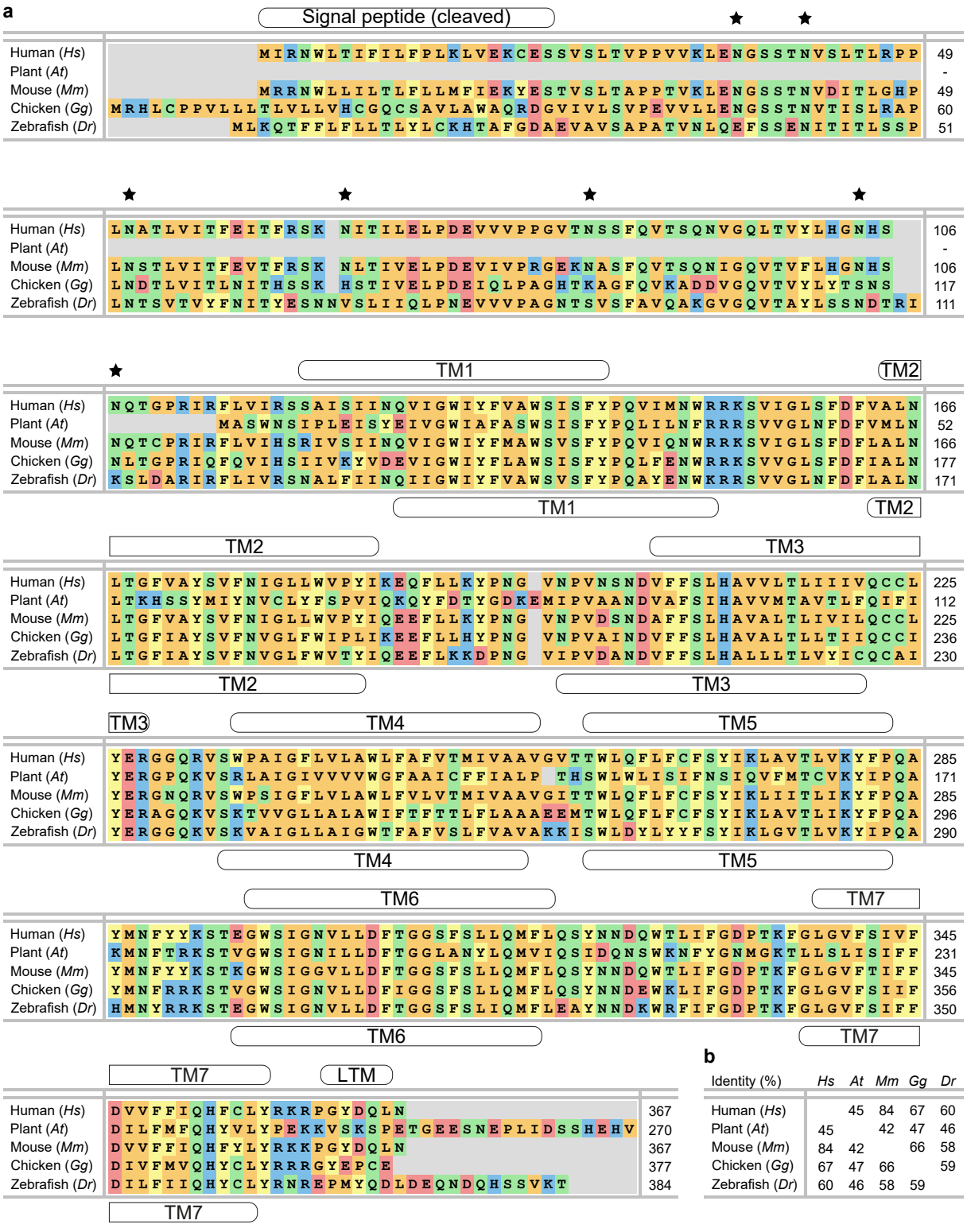


Fig. 3.1 – Sequence Alignment of CTNS Homologues.
a, Sequence alignment of CTNS homologues coloured *via* amino acid chemistry. TM helices 1–7 for human and plant CTNS, as predicted by TMHMM (Krogh *et al.*, 2001), are depicted above and below the alignment, respectively. Human CTNS includes a long N-terminal tail, which is highly glycosylated at the positions marked with a star (Nevo *et al.*, 2017). **b**, Sequence identity among the five CTNS homologues compared in **a**.

for both the human and plant homologue using the online tool TMHMM (Krogh *et al.*, 2001). The close proximity of TMs suggests a compact architecture without major domains in between individual helices. Human CTNS contains an N-terminal signal peptide (residues 1–22), which is recognized by the signal recognition particle, initiating entry into the endoplasmic reticulum (ER), where it is post-translationally modified (Keenan *et al.*, 2001). This modification primarily involves the addition of seven N-linked glycans to asparagines 36, 41, 51, 66, 84, 104, and 107 (marked with stars in **Fig. 3.1 a**). Furthermore, human CTNS bears a tyrosine-based lysosomal targeting motif (LTM) of the (G)YXXΦ type – where X is a random residue, and Φ is a bulky, hydrophobic one – at the C-terminal end. This sequence is recognized by the adaptor protein complexes and subsequently trafficked to the lysosome (Bonifacino & Traub, 2003; Andrzejewska *et al.*, 2015). Plant cells do not have functional lysosomes, but the vacuole takes its place in regulating protein degradation, one of the main functions of the lysosome (Callis, 1995). Vacuolar sorting determinants are more diverse than LTMs (Xiang *et al.*, 2013), and thus, it is more difficult to identify a specific localization signal within plant CTNS. The literature indicates, however, that CTNS homologues of organisms that lack functional lysosomes instead localize to the vacuole. Specifically, the yeast homologue of CTNS, Ers1, functions as a cystine transporter in the vacuole (Gao *et al.*, 2005), and plant CTNS from *Arabidopsis thaliana* has been found to reside in the tonoplast of the vacuole (Jaquinod *et al.*, 2007).

3.1.2 Expression Trials

The recombinant production of membrane proteins as GFP fusions, *e.g.* through the use of the pDDGFP-Leu2d vector, offers a couple of advantages that are especially useful during homologue screening: First, expression levels can be quickly monitored by measuring the GFP-fluorescence ($\lambda_{\text{Excitation}}$ = 488 nm, $\lambda_{\text{Emission}}$ = 512 nm) of a small volume of the whole cell suspension using a plate reader. Additionally, protein quality can be assessed quickly and without the need to purify the protein to homogeneity: Overexpressed membrane proteins are solubilized in detergent, membranes and other insoluble fractions are removed by centrifugation, and the sample is run on a SEC column whilst monitoring the GFP-fluorescence. These fluorescent SECs (FSECs) are ideal to test the compatibility of different detergents with the protein of interest.

CTNS homologues were expressed on a small scale of 50 mL to compare expression levels ~16 h post-induction, hereafter referred to as the time of harvest. Expression counts (F_{GFP} normalized to the OD_{600} at the time of harvest) were measured at ~2,000–4,000 relative fluorescence units (RFU) for most of the constructs except plant CTNS from *Arabidopsis thaliana* (**Fig. 3.2 a**). *At*CTNS showed a five- to 10-fold increase in GFP-fluorescence, totalling a theoretical expression of ~6 mg of GFP per litre of culture, compared to the ~0.6–1.2 mg produced using the other constructs – calculated as described by Drew *et al.* (2008).

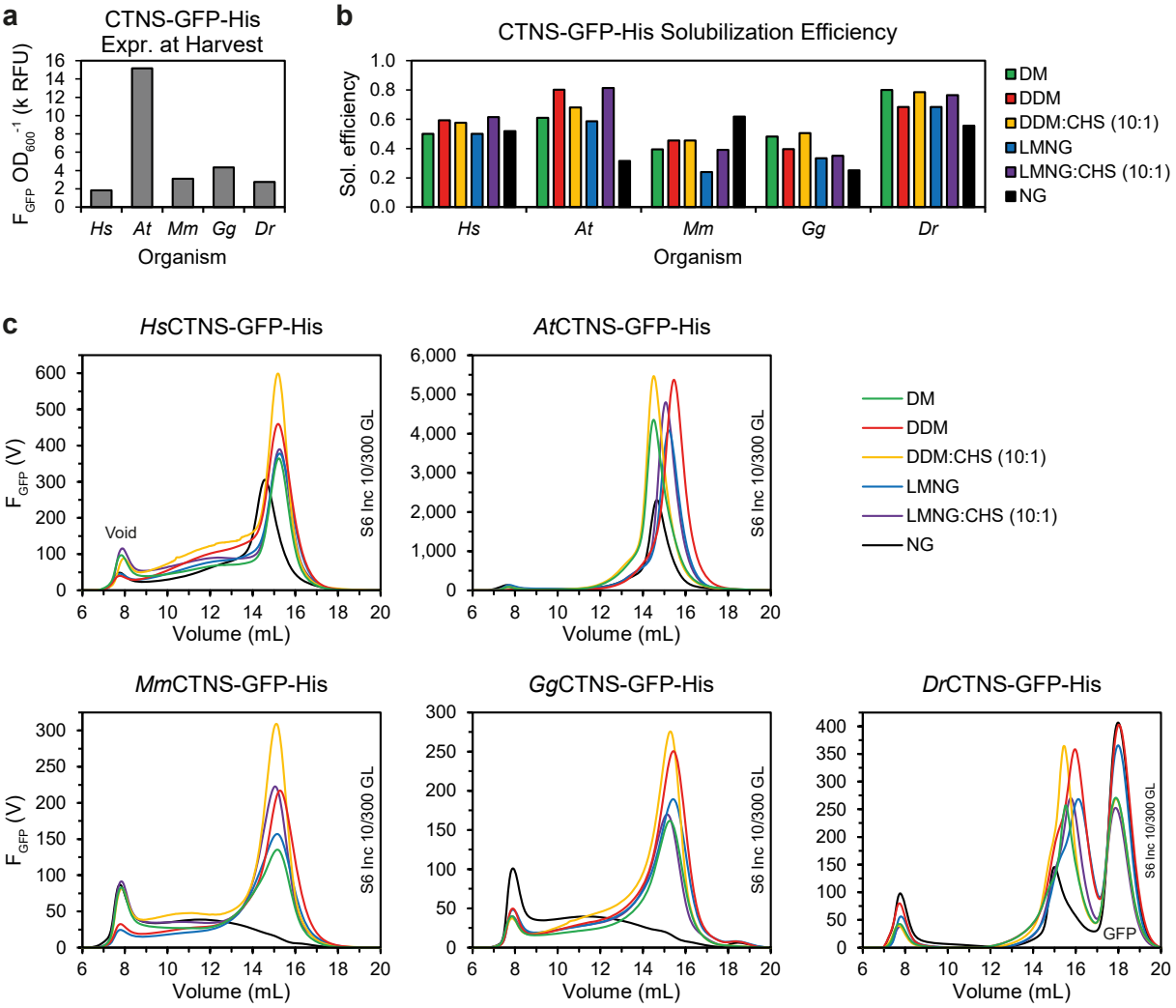
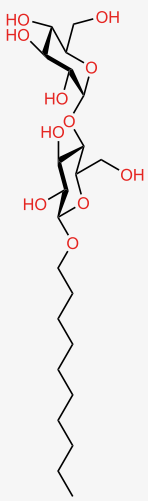
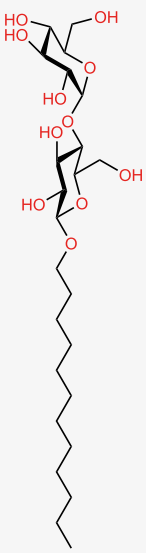
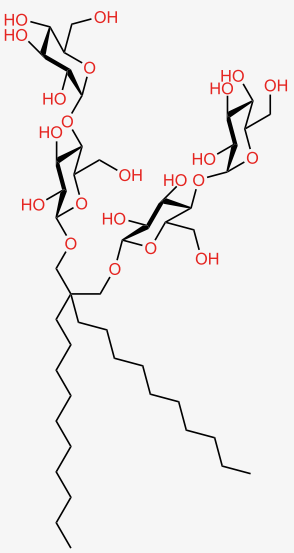
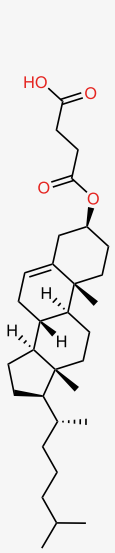


Fig. 3.2 – Expression Trials of CTNS Homologues.
a, Expression counts at the time of harvest (~16 h post-induction), **b**, solubilization efficiency, and **c**, FSEC profiles for each CTNS construct aligned in **Fig. 3.1 a** and in combination with each detergent outlined in **Table 3.1**.

As listed in **Table 3.1**, constructs were solubilized in detergents *n*-nonyl-β-D-glucopyranoside (NG) – a rather harsh nonionic detergent with a small glucose-based head group and a relatively short C9 acyl chain, chosen to test the limits to CTNS stability in solution; DM and DDM – the two most frequently used detergents in membrane protein structural biology (Stetsenko & Guskov, 2017), composed of a maltose-based head group and a C10 and C12 acyl chain, respectively; as well as 2,2-didecylpropane-1,3-bis-β-D-maltopyranoside, better known as LMNG – a much larger detergent that effectively fuses two DDM molecules. Both DDM and LMNG were also tested in combination with CHS, a soluble version of cholesterol (CHL), which is abundantly present in mammalian membranes and often found bound to membrane proteins (Fantini & Barrantes, 2013). The addition of CHS also alters the micelle properties, *e.g.* by increasing the micelle size – at least in the case of DDM (Thompson *et al.*, 2011) – thereby justifying an evaluation of its influence on solubilization efficiency and FSECs.

Table 3.1 – Main Detergents Used in This Research.
The chemical structure of the main detergents used throughout this thesis is shown. NG was mainly used as a negative control. DM was used to ease protein reconstitution into liposomes. DDM and LMNG:CHS were mainly used for purification. All detergents listed were used in the homologue screen. Chemical structures were prepared in ChemDraw 20.1 and are coloured by element. Table properties were taken from Stetsenko & Guskov (2017). Detergents are sorted by decreasing CMC values.

	NG	DM	DDM	LMNG	CHS
Chemical structure					
MW (Da)	306.4	482.6	510.6	1,069.2	486.7
CMC (%)	0.20	0.087	0.0087	0.001	N/A
Micelle size MW (kDa)	85	40	65–70	91	N/A

The human construct was solubilized at a moderate level of 50–60% in all detergents tested (**Fig. 3.2 b**), indicating that approximately half the protein was denatured or otherwise inaccessible to be engulfed into a detergent micelle, or had lost its GFP-His-tag. Solubilization efficiencies were measured as the ratio between F_{GFP} counts after centrifugation and those before. The subsequent FSECs, however, exhibited well-defined monodisperse peaks at a retention volume of 14–16 mL, albeit with slight left-hand shoulders and considerable void peaks at ~8 mL (**Fig. 3.2 c**). Generally, the more symmetrical and monodisperse the main peak of an FSEC trace, and the fewer side peaks and shoulders present, the better the quality and, thus, stability of the tested membrane protein in the detergent. Left-handed shoulders leading up to the main peak, and especially peaks in the void, typically correspond to aggregated and/or denatured protein. Compared to human CTNS, the *At*CTNS-GFP-His fusion solubilized better on average, except for NG, which was measured at only ~30% (**Fig. 3.2 b**). Additionally, all detergents produced well-shaped, symmetrical FSEC peaks without major shoulders (**Fig. 3.2 c**). Both mouse and chicken CTNS would only solubilize to a lower percentage than human or plant, and also showed worse FSEC traces with leaning main peaks and strong void signals. In each case, NG only generated a void peak and lacked a peak that would correspond to the

fusion solubilized in detergent. CTNS from *Danio rerio* solubilized best among most detergents tested, but it showed two major peaks in its FSEC profile, with the latter at ~18 mL in retention volume (**Fig. 3.2 c**), likely indicating the degradation of the fusion protein, which produced free GFP. Based on its high expression level and solubilization efficiency, as well as excellent protein quality as judged from its FSEC profile, *At*CTNS was chosen as the optimal construct for initial protein analysis. DDM was chosen as the main detergent due to its performance in these experiments, as well as its compatibility with LCP crystallization and its lower price point compared to LMNG and DM.

3.2 Preparation of *At*CTNS for Structural & Functional Studies

3.2.1 Recombinant Protein Production

The expression of initial large-scale *At*CTNS cultures was induced at an OD_{600} of ~1.5 and carried out for ~20 h. As a result, these expression cultures typically matched the GFP-fluorescence counts per OD_{600} of ~15 k RFU, which were previously obtained during small-scale expression (**Fig. 3.2 a**). However, due to the low OD_{600} at the time of harvest (~3–4), only 4–4.5 g of cells were harvested per litre of culture. To improve this in later expressions, induction was delayed to an OD_{600} of ~2.5, and expression time was increased to 22–24 h, resulting in higher yields of biomass and relative fluorescence counts, 7–8 g L^{-1} and ~18 k RFU, respectively. This optimiziation did not affect the solubilization efficiency in DDM compared to earlier large-scale expressions, retaining a percentage of 70–85%, which is an increase from the ~60% achieved on a small scale (**Fig. 3.2 b**).

*At*CTNS purification was carried out using standard IMAC steps and finally applied to a gel filtration column for SEC. Imidazole concentration during the first IMAC step was set at 25 mM (35 mM during the final wash and 250 mM for elution) and found to be sufficient to bind the *At*CTNS-GFP-His fusion with a final efficiency of 60–75% as calculated from the F_{GFP} of the IMAC FT divided by that of the solubilized membranes just before HisPur Ni-NTA resin addition. The protein mass present in the elution was measured using GFP-fluorescence (Drew *et al.*, 2008). Subsequently, TEV was added to the elution at a concentration optimized to match 35–40% (w/w) of the GFP mass, as excessive TEV was found to cause precipitation in the subsequent O/N dialysis step. The addition of TEV created two main bands on an SDS-PAGE gel, consisting of GFP-His and *At*CTNS, with the latter coincidentally running at the exact height of the protease itself (**Fig. 3.3 b**). Reverse IMAC was initially performed by passing the dialysed and filtered protein supplemented with a final concentration of 25 mM imidazole through a pre-packed HisTrap HP column (GE Healthcare). However, this method left a considerable amount of GFP-His in the FT, which would co-elute with *At*CTNS from

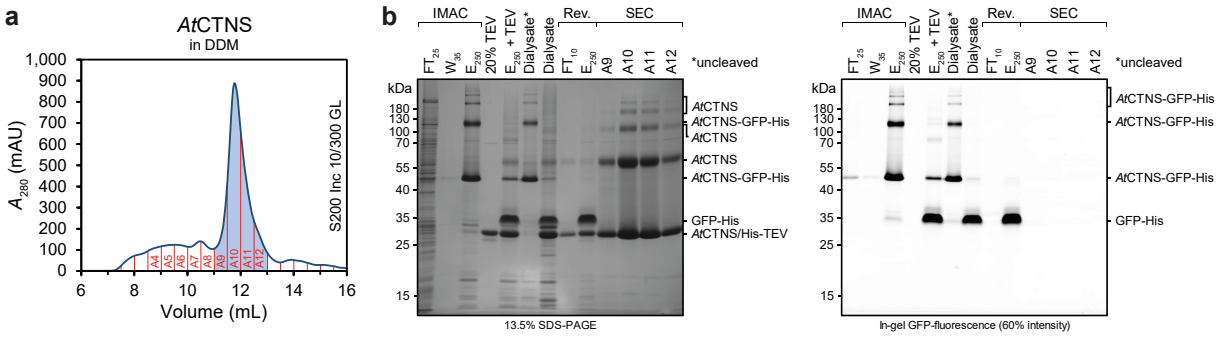


Fig. 3.3 – *At*CTNS Purification.
a, SEC profile of *At*CTNS purified in DDM. Pooled and combined fractions are shaded in blue. **b**, Relevant samples of the purification were separated on a 13.5% SDS-PAGE gel. Both in-gel fluorescence and subsequent Coomassie-stained images are shown. Label subscripts refer to imidazole concentration in mM. Dialysis* refers to a small sample that was dialysed without TEV present.

the gel filtration column, leaving a final sample that was only an estimated ~90–95% pure. To improve the final purity and, thus, the chances of successful crystallization, we optimized the purification protocol as follows. Imidazole levels during reverse IMAC were reduced to 10 mM to maximize the binding of GFP-His, His-TEV, and other minor contaminants that were present after IMAC and dialysis. Additionally, the reverse IMAC was performed in batch, reusing the HisPur resin from the IMAC after regenerating it according to the manufacturer's instructions (10–20 CV 20 mM MES, 100 mM NaCl at pH 5.0, then mQ-H_2O) and subsequently equilibrating it in dialysis buffer. As was the case during the IMAC, the GFP-fluorescence and, thus, binding efficiency could now be measured progressively, and the reverse IMAC FT was collected as soon as the SN of the binding suspension was confirmed to be GFP-free (F_{GFP} < 500 RFU), which was typically reached after 1–1.5 h. These changes resulted in an almost perfectly pure sample (**Fig. 3.3 b**) and an overall yield of 120–170 µg L^{-1} expression culture.

Interestingly, although the corresponding SEC trace only shows one well-defined and monodisperse peak at ~12 mL in retention volume (**Fig. 3.3 a**; sometimes a right-handed shoulder is visible, similar to **Fig. 3.4 a**), *At*CTNS produces multiple bands on the SDS-PAGE gel (**Fig. 3.3 b**). Visible at ~25–30, ~55–60, ~100, ~160 kDa, and even higher molecular weights (MWs), these bands potentially hint at protein oligomerization, *i.e.* monomeric, dimeric, and higher oligomeric states. Other PQ-loop proteins also show this effect, including the KDEL receptor, albeit to a lesser degree (Bräuer *et al.*, 2019), as well as the sugars will eventually be exported transporter (SWEET) from *Oryza sativa* (rice), which was cross-linked into stable dimeric and trimeric complexes (Tao *et al.*, 2015). In the case of *At*CTNS, exposure to SDS seems to cause this phenomenon, and since longer exposures, *e.g.* O/N, increase the intensity of higher MW bands and lower protein concentrations seem to lower it, this may be a gel artefact. Cross-linking *At*CTNS using different concentrations of glutaraldehyde was tried, but no effect

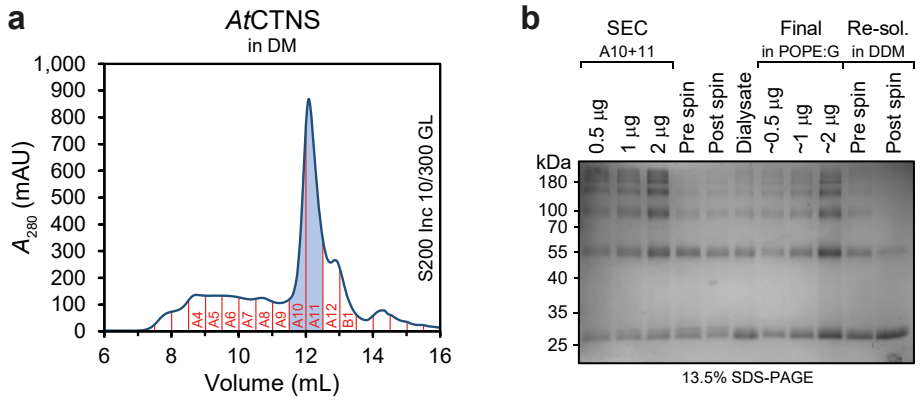


Fig. 3.4 – *At*CTNS Reconstitution into Liposomes.
a, SEC profile of *At*CTNS from the same purification as **Fig. 3.3** but detergent exchanged into DM after the reverse IMAC. Pooled and combined fractions are shaded in blue. **b**, SDS-PAGE gel of *At*CTNS from **a** reconstituted into 3:1 (w:w) POPE:POPG.

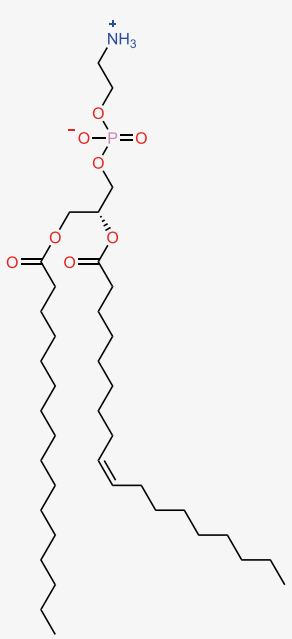
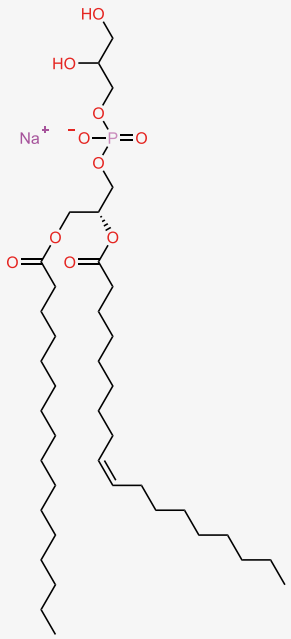
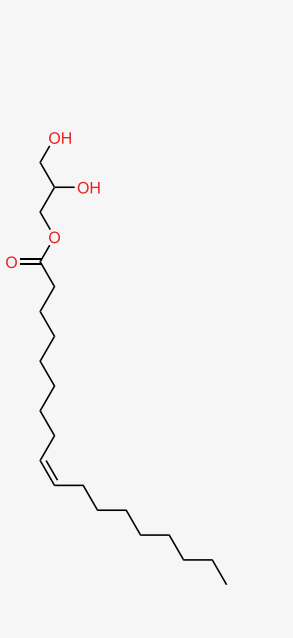
was visible on the gel (not shown). As a result, the importance of oligomerization was mostly disregarded in this chapter, however, more emphasis was put on it later (**Ch. 8**).

3.2.2 Proteoliposome Reconstitution

For reconstitution into a 3:1 (w:w) mix of phospholipids POPE and POPG (chemical structures shown in **Table 3.2**), *At*CTNS was injected onto a gel filtration column pre-equilibrated in DM instead of DDM (**Fig. 3.4 a**). DM has a 10-fold higher CMC than DDM (**Table 3.1**); thus, it dissociates much more readily, allowing for the formation of proteoliposomes through rapid dilution (Rigaud & Lévy, 2003). Here, the protein was mixed with pre-formed (see **Sec. 2.6.2**) POPE:POPG liposomes at RT and then rapidly diluted into ice-cold detergent-free buffer (50 mM K-phosphate at neutral pH), shifting the detergent concentration to below the CMC, thus allowing for spontaneous proteoliposome formation. These were then harvested by ultracentrifugation and subsequently dialysed extensively to eliminate most of the remaining detergent molecules so that proton-tight proteoliposomes could form (Parker *et al.*, 2014). After dialysis, they were harvested and resuspended in a known volume of detergent-free buffer, and the protein concentration was calculated through SDS-PAGE, comparing the band intensity to that of a known protein mass from a sample in detergent (**Fig. 3.4 b**). To further test reconstitution efficiency and rule out the possibility of protein precipitation during proteoliposome formation, *At*CTNS was also re-solubilized in DDM and analysed on gel.

As shown in **Fig. 3.4 b**, *At*CTNS was successfully incorporated into POPE:POPG liposomes. After dialysis, proteoliposomes were taken up in 90% of the volume that would have corresponded to a 100% efficient reconstitution process (see **Sec. 2.6.2**), and as analysed using the program ImageJ, the sum of the intensities of all bands was still marginally larger than that of a known amount of *At*CTNS in detergent (**Fig. 3.4 b**), meaning that the reconstitution was more

Table 3.2 – Main Lipids Used in This Research.
The chemical structure of the main lipids used throughout this thesis is shown. Phospholipids POPE and POPG (Anatrace) were used to form liposomes at a 3:1 (w:w) ratio, building the foundation for protein reconstitution. Monoolein (MO or 9.9 MAG) was used in most crystal trials. Chemical structures were prepared in ChemDraw 20.1 and are coloured by element.

	POPE	POPG	Monoolein (MO, 9.9 MAG)
Chemical structure			
MW (Da)	718.0	771.0	356.6

than 90% efficient. Re-solubilization also recovered a majority of the protein, however, the higher MW bands seem to be much fainter than before. Proteoliposomes formed in this manner were used for llama immunization to generate nanobodies specific to *At*CTNS (**Sec. 5.3**), as well as for ligand transport assays outlined in **Sec. 9.2**.

3.3 Biophysical Characterization of *At*CTNS

3.3.1 Thermal Stability Determination

The use of a properly folded and thermally stable membrane protein is essential for all subsequent experiments. Therefore, we tested the stability of *At*CTNS using various biophysical techniques. Following the homologue screening (**Sec. 3.1**), we initially analysed the thermal stability of the *At*CTNS-GFP-His fusion using heat-FSECs (Hattori *et al.*, 2012). Individual samples were heated after solubilization to either 30, 40, 50, 60, or 70 °C, subsequently ultracentrifuged, and finally applied to the SEC column. As illustrated in **Fig. 3.5 a**, monodisperse peaks were generated up to temperatures of 50–60 °C, suggesting a high melting temperature (~40–50 °C) and impressive stability of *At*CTNS in the DDM micelle.

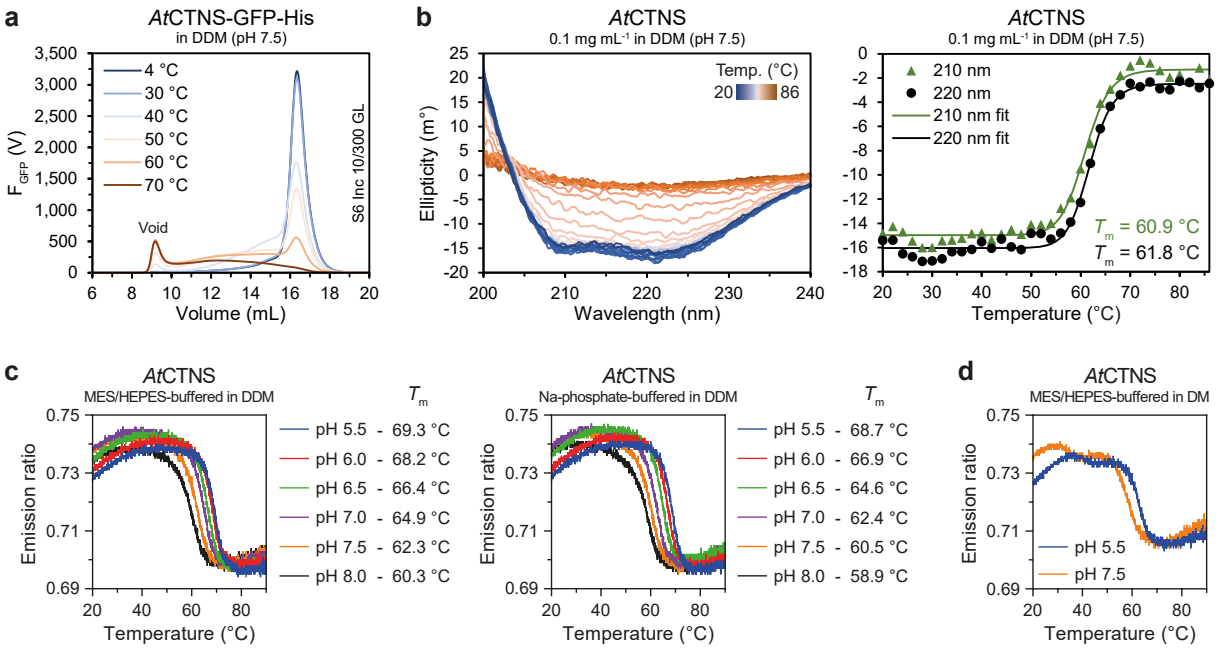


Fig. 3.5 – Thermal Stability Analyses of *At*CTNS.
a, Heat-FSEC of the *At*CTNS-GFP-His fusion in DDM at pH 7.5. **b**, CD spectrum and melt for DDM-purified *At*CTNS at pH 7.5. **c**, NanoDSF profiles of *At*CTNS in DDM at different pH, ranging from 5.5–8. **d**, NanoDSF profiles of *At*CTNS after detergent exchange from DDM to DM at pH 5.5 and 7.5.

After purifying the tag-free protein to homogeneity, we wanted to analyse this in greater detail. Thus, we determined the melting temperature of *At*CTNS using circular dichroism (CD) spectroscopy (**Fig. 3.5 b**) and nano differential scanning fluorimetry (nanoDSF) using a Prometheus NT.48 system (**Fig. 3.5 c** in DDM and **d** in DM). Both techniques offer thermal stability assessment in a label-free fashion. CD spectroscopy monitors the effect of left- and right-handed circularly polarized light of different wavelengths on chiral molecules, specifically the deviation from equal absorption, which is measured in units of ellipticity. α-Helical content is hereby indicated by absorbing left-handed circularly polarized light of wavelengths between ~205–235 nm less than right-handed circularly polarized light, creating minima near 208 and 222 nm (Woody, 1995). NanoDSF monitors the fluorescence of tryptophans in a protein, which is highly dependent on its close surroundings (Strutz, 2016). The sample is excited using light at 280 nm wavelength, and the emission is measured at both 350 and 330 nm. Thermally destabilizing the protein by heating the sample changes the protein conformation from a well-ordered to a disordered state, altering the surroundings of the tryptophan residues, resulting in a red-shift (increase in the emission ratio) or a blue-shift (decrease) melting curve.

The CD scan of 0.1 mg mL^{-1} *At*CTNS in DDM shows two clear minima at ~208–210 and ~220–222 nm, clearly indicating α-helical content (**Fig. 3.5 b**). Thermally unfolding the protein leads to a progressive loss of this conformation-specific profile. Plotting the ellipticity data collected at 210 and 220 nm against the temperature creates a well-defined transition point,

suggesting a T_m of ~61–62 °C. Interestingly, the CD melt suggests a much higher thermal stability for *At*CTNS than previously estimated by heat-FSEC.

Given the extraordinarily high melting temperature indicated by CD spectroscopy, we sought to confirm this finding using nanoDSF, also validating the Prometheus NT.48 system, as this was a new acquisition for the lab. This system promises a faster and easier readout using a smaller sample compared to CD spectroscopy whilst retaining similar accuracy. This proved to be the case, as nanoDSF revealed a very similar melting behaviour, showing a blue-shift melting curve with a T_m of ~62 °C in a HEPES-based buffer at pH 7.5 (**Fig. 3.5 c**).

Because CTNS is a proton-driven transporter (Kalatzis *et al.*, 2001) and localized to the lysosome (Cherqui *et al.*, 2001), an organelle with a very acidic environment at pH 4.5–5.5 (Han & Burgess, 2010) – although *At*CTNS is located to the tonoplast (Jaquinod *et al.*, 2007) of the vacuole which has a pH of 5.9 (Krebs *et al.*, 2010) – we wanted also to test the effect of pH on the protein. Thus, we diluted concentrated *At*CTNS in DDM into buffers at different pH, including HEPES-buffered solutions at pH 8.0–7.0, MES at pH 6.5–5.5, as well as Na-phosphate buffer covering the same pH range (**Fig. 3.5 c**). We found that protein stability increases considerably with rising proton concentrations, with T_ms increasing to ~69 °C at pH 5.5. Sulfonic acid buffers HEPES and MES additionally stabilized the protein by ~1–2 °C at most pH values tested when compared to the Na-phosphate buffered system. As a result, we chose MES at pH 6 as the buffer for most upcoming experiments, including final gel filtration during purifications as well as crystallization trials.

DM-detergent exchanged *At*CTNS was also tested using nanoDSF (**Fig. 3.5 d**), showing a slightly worse melting trace compared to DDM. A clear transition between folded and unfolded protein was still visible, though, and T_ms of ~63 °C and ~58 °C were calculated at pH 5.5 and 7.5, respectively, highlighting that the pH dependency is not detergent specific.

3.3.2 Ligand Binding Analysis

With nanoDSF established in the lab, we wanted to see if we could also use it to analyse ligand binding of *At*CTNS. Samples were prepared and briefly incubated before generating nanoDSF profiles (see **Sec. 2.3.4**). Interestingly, using the protein stability optimized conditions at pH 5.5, addition of 2 mM L-cystine only increased the T_m of *At*CTNS by 0.4 °C compared to a sample that did not contain any ligand (**Fig. 3.6 a**). Repeating the experiment in a more neutral environment doubled the ΔT_m to 0.8 °C, however, we did not consider this shift to be substantial enough to attribute it to protein stabilization caused by the ligand. Further destabilization prior to ligand addition, achieved by using systems buffered in Na-phosphate, also only helped marginally, increasing the ΔT_m from 0.4 °C at pH 6.0 to 1.4 °C at pH 8.0.

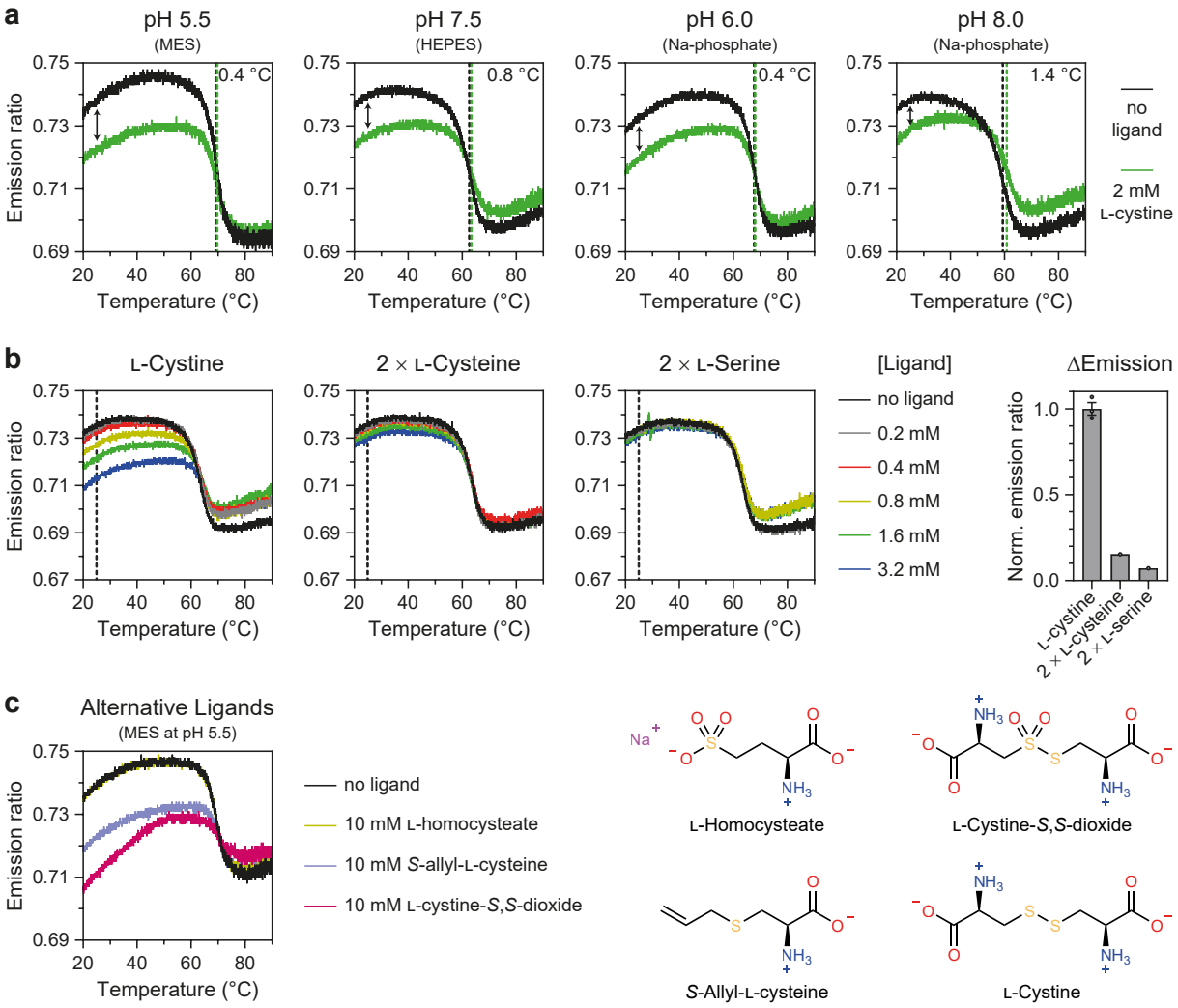


Fig. 3.6 – *At*CTNS Ligand Binding Analysis Using NanoDSF.
a, NanoDSF profile comparison of *At*CTNS in DDM without ligand and supplemented with 2 mM L-cystine, highlighting the change in melting temperature (T_ms indicated using dashed lines, ΔT_ms shown in the top right corner of each graph) as well as the F_{Trp} quench observed in the early course of the traces (indicated by arrows). The experiment was performed at different pH using MES, HEPES, or Na-phosphate as buffer. **b**, Ligand binding assay performed at pH 7.20 using *At*CTNS purified in DDM. Amino acids L-cysteine and L-serine were used as negative controls. Their concentrations were doubled in the assay to better resemble the chemical structure of L-cystine. The bar chart illustrates the normalized difference in emission ratio between no ligand and 3.2 or 6.4 mM, respectively, as measured at 25 °C (highlighted using dashed lines). **c**, NanoDSF profile of alternative ligands at 10 mM concentration and their chemical structures (prepared in ChemDraw 20.1).

Since increasing the ligand concentration beyond 2–3 mM is technically challenging due to the low solubility of L-cystine (Bannai, 1984), we initially attempted alternative techniques such as microscale thermophoresis using the uncleaved fusion protein and specifically its GFP as fluorophore for detection. Repeated scans, however, featured a fluctuating signal, which masked ligand binding, rendering this technique useless for this purpose.

Ultimately, we decided to investigate the possibility of creating a ligand binding assay using the decrease in emission ratio induced by the ligand at all conditions tested, which can be observed in the early course of the nanoDSF traces (see arrows in **Fig. 3.6 a**). Using a concentration gradient of L-cystine and amino acids L-cysteine and L-serine as negative controls, we were able to show that the F_{Trp} quench is concentration dependent and specific to L-cystine (**Fig. 3.6 b**).

*Hs*CTNS was previously shown to have narrow substrate specificity regarding transport (Ruivo *et al.*, 2012). Using our binding assay, we sought to investigate whether this holds true for ligand binding of *At*CTNS. Similar to L-selenocystine (where the sulfur atoms are replaced by selenium atoms) and L-cystathionine (one sulfur atom replaced by a methylene group), which were identified as alternative substrates by Ruivo *et al.* (2012), we found that L-cystine-*S*,*S*-dioxide (two additional carbonyl groups which bind to one of the sulfur atoms) causes a F_{Trp} quench in the same way L-cystine does (**Fig. 3.6 c**). The addition of the two carbonyl groups makes this structural analogue more soluble than L-cystine, which led us to use it in both structural and functional trials whenever higher concentrations were needed (**Secs. 7.3** and **9.3**).

L-Homocysteate and *S*-allyl-L-cysteine were also used in the binding assay. Both of these compounds have one of the two amino acid groups replaced by a sulfonate and allyl group, respectively. Interestingly, supplementing pH 5.5-buffered *At*CTNS with 10 mM *S*-allyl-L-cysteine quenches the F_{Trp} similarly to 2 mM L-cystine (compare **Fig. 3.6 a** and **c**), whilst 10 mM L-homocysteate has no effect on the fluorescence profile (**Fig. 3.6 c**).

CHAPTER 4

Crystal Trials Using *At*CTNS

The molecular structure of a protein can alleviate and advance our understanding of its function. We used X-ray crystallography (XRC) to gain structural information on the plant homologue of CTNS from *Arabidopsis thaliana*. In this chapter, I will describe our initial attempts to solve the three-dimensional structure of *At*CTNS by crystallizing the protein in the lipidic cubic phase. Crystals formed readily and quickly, however, only a single crystal yielded sufficient diffraction data to attempt structure solution by MR. Due to the lack of protein structures with high similarity in the protein data bank (PDB), we were unable to solve the phase problem at this stage and decided to generate sybodies and nanobodies (**Ch. 5**) to serve as crystallization chaperones to try and optimize the crystal quality (**Ch. 6**), which ultimately enabled us to solve the structure experimentally (**Ch. 7**). I will additionally show here that structures predicted by AlphaFold (Jumper *et al.*, 2021), a tool that uses machine learning to fold proteins, may be used successfully to estimate the missing phases of a crystallographic dataset by MR. However, AlphaFold was only launched after we had already solved the structure of *At*CTNS.

4.1 Choice of Cubic Phase Host Lipid

For crystallization in the lipidic cubic phase (LCP), purified *At*CTNS (**Sec. 3.2.1**) was mainly reconstituted into 9.9 monoacylglycerol (MAG). Alternative host lipids, such as 7.8 MAG and 8.8 MAG, were also tried but did not perform as well, frequently resulting in a cloudy lipid:protein solution after mixing, indicating suboptimal reconstitution conditions and potentially precipitated protein. MAGs are distinguished by the position of their *cis*-configured double bond: 7.8 specifies that there are seven carbons from the ester to the double bond and eight carbons in the remaining acyl chain – 9.9 MAG therefore resembles the oleoyl group of phospholipids POPE and POPG (see **Table 3.2**). Given that *At*CTNS was successfully reconstituted into POPE:POPG proteoliposomes (**Sec. 3.2.2**), it seems fitting that we identified monoolein (MO) as a suitable host lipid for LCP crystallization. The specifics regarding the formation of the cubic phase are described in **Sec. 2.5.1**.

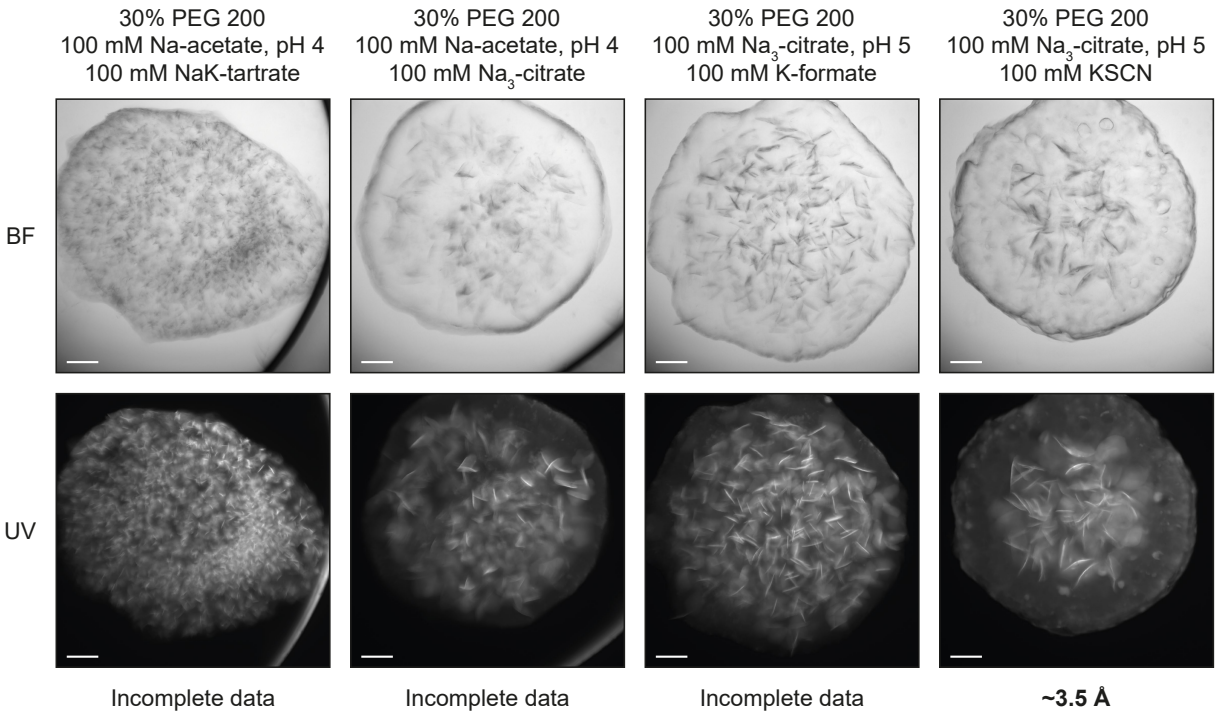


Fig. 4.1 – LCP Crystals of *At*CTNS.
BF and UV images of LCP boli containing *At*CTNS crystals are shown. *At*CTNS was crystallized in MO using protein at 43 mg mL^{-1}. Annotated at the top of each column of images is the precipitant solution. The diffraction outcome is pinned at the bottom. Scale bars 100 µm.

4.2 Crystallization

Multiple attempts were made to crystallize *At*CTNS in the LCP. Initially, purified protein was concentrated to ~15 mg mL^{-1} and crystallized (see **Sec. 2.5.1**). This resulted in showers of microcrystals that were observed in the presence of various precipitant solutions. Seeking to enhance the size of these crystals, we decided to successively increase the protein concentration from ~15 to ~22 and ~30 mg mL^{-1}, and indeed, larger crystals formed. In **Fig. 4.1**, BF and UV images of typical *At*CTNS crystals set up using protein at 43 mg mL^{-1} are shown. Nucleation events remained frequent throughout all trials, and crystals formed in the shape of fine needles, measuring up to ~80–100 µm in length but only a few µm in width and height. Among many conditions tested, these needles were often found in precipitant solutions containing PEG 200 in combination with a low-pH buffer (pH 4–5) and a monovalent alkali metal-based salt (Na^+, K^+, or both), most preferably Na_3-citrate. Crystals were harvested, and diffraction experiments were performed at the DLS synchrotron (procedure see **Sec. 2.5.3**).

4.3 Data Collection & Processing

In the diffraction experiment, protein crystals are irradiated using parallel X-rays of fixed wavelength. Upon interaction with electrons in the crystal, these waves are scattered in all

directions. Most of them cancel each other out through destructive interference. Diffracted waves that are in phase, however, experience constructive interference. Their amplitudes are combined, and the event can be captured as a diffraction spot using a photon-counting detector. According to Bragg's law (**Eq. 4.1**), individual spots arise from the sum of beams reflected from sets of parallel planes that intersect the smallest repeatable unit within the crystal: the unit cell (Bragg & Bragg, 1913; Rhodes, 2006). Using variables h, k, and l (Miller indices), these sets of planes are characterized and named based on how often they subdivide unit cell axes a, b, and c, respectively. Each spot, referred to as a reflection, results from a specific set of planes and can, therefore, be assigned a unique Miller index hkl.

$$n\,\lambda = 2\,d_{hkl}\,\sin\theta \tag{4.1}$$

Bragg's law describes the relationship of the X-ray wavelength λ to the interplanar spacing of a specific set d_{hkl} and the impact and reflection angle θ. Diffraction spots are only created when n is an integer, meaning waves are in phase and interfere constructively (Rhodes, 2006).

All spots combined create a unique diffraction pattern in reciprocal space, which can be used to infer the structure of the crystal components in real space. We need the amplitudes, frequencies, and phases of the observed reflections to calculate the electron density of the crystal components necessary for model building and structure solution. The amplitude of a reflection is proportional to the square root of its intensity and can be easily calculated directly from the diffraction images (Rhodes, 2006). Frequencies in the three directions x, y, and z can also be retrieved directly from the data, as they are proportional to Miller indices h, k, and l, respectively (Rhodes, 2006). The phases, however, cannot be measured in the experiment and must to be retrieved experimentally (see **Sec. 7.2**).

The diffraction data is processed to obtain both amplitudes and frequencies of the observed reflections. Nowadays, this step is routinely performed in an automatic fashion using pipelines xia2 (Winter, 2010) or autoPROC (Vonrhein *et al.*, 2011), which themselves utilize a sequence of programs for the calculations (see **Sec. 2.5.4**). Initially, reflections are identified, Miller indices assigned (indexing), and spot intensities measured (integration). Then, the cell symmetry is analysed, and intensities from multiple observations of reflections are put on a common scale and merged into an average intensity (scaling and merging). In the final step, intensities are converted to amplitudes and contained within structure factors F_{hkl}. A set of reflections is set aside to serve as the basis for calculating R_{free} during structure solution and refinement.

The diffraction image quality was initially inspected visually, taking into account the homogeneity and number of observable spots. For the data collected from the *At*CTNS crystals, specifically the spot count would decrease dramatically during data collection, causing the signal-to-noise ratio of the mean of all reflection intensities, the mean $I/\sigma(I)$ or $\langle I/\sigma(I)\rangle$, to approach zero after ~10–30° rotation. This was likely caused by the disintegration of the

Table 4.1 – X-Ray Data Collection & Processing Statistics of *At*CTNS.

a, Relevant sweeps of data collected at DLS (proposal number mx23459-103).

Identifier	Beamline	Wavelength (Å)	Beamsize (µm)	Transmission (%)	Exposure (s)	Oscillation (°)	Total images	Resolution max. (Å)	Detector
6589_1	DLS - I24	0.9999	7 × 7	15.62	0.0500	0.50	360	2.00	PILATUS3 6M
6589_2	DLS - I24	0.9999	7 × 7	20.37	0.0500	0.10	900	2.00	PILATUS3 6M

b, Data processing.

Pipeline	Images used	High-resolution data cut	Space group	Cell dimensions *a*, *b*, *c* (Å)	Cell angles α, β, γ (°)	PDB
ISPyB - xia2 3dii	1_35–360, 2_1–900	3.50 Å	$C222_1$	94.45, 116.30, 123.16	90.00, 90.00, 90.00	----

c, Merging statistics by resolution bin and overall. *Significant at $p = 0.01$.

Resolution (Å)	Observations	Unique reflections	Multiplicity	Completeness (%)	Mean $I/\sigma(I)$	R_{merge} (%)	R_{meas} (%)	R_{pim} (%)	$CC_{1/2}$ (%)
73.33–9.49	2,403	468	5.13	95.9	18.8	5.6	6.4	2.8	99.4*
9.49–7.54	2,526	456	5.54	99.1	14.3	8.3	9.2	3.7	99.7*
7.54–6.59	2,478	459	5.40	99.1	7.4	15.5	17.1	7.1	99.6*
6.59–5.98	2,505	442	5.67	99.3	5.3	30.2	33.2	13.5	94.9*
5.98–5.56	2,461	439	5.61	99.3	4.5	37.0	40.7	16.6	92.8*
5.56–5.23	2,572	449	5.73	99.6	5.0	34.4	37.8	15.4	92.9*
5.23–4.97	2,405	420	5.73	99.3	5.5	31.5	34.5	14.0	95.8*
4.97–4.75	2,640	448	5.89	99.6	5.5	31.9	35.0	14.1	92.8*
4.75–4.57	2,527	442	5.72	99.1	5.2	33.8	37.1	15.1	94.1*
4.57–4.41	2,366	422	5.61	98.4	6.1	28.8	31.7	13.0	93.7*
4.41–4.27	2,518	439	5.74	99.1	3.8	48.6	53.3	21.6	92.8*
4.27–4.15	2,524	433	5.83	99.5	4.0	43.2	47.4	19.1	92.2*
4.15–4.04	2,545	436	5.84	99.5	3.8	49.1	53.8	21.7	96.6*
4.04–3.94	2,475	418	5.92	99.5	2.8	67.0	73.4	29.6	85.5*
3.94–3.85	2,659	455	5.84	99.8	2.5	72.3	79.3	32.2	68.9*
3.85–3.77	2,436	416	5.86	99.8	2.0	96.9	106.4	43.1	69.8*
3.77–3.69	2,579	443	5.82	99.6	1.8	108.4	118.9	48.2	66.0*
3.69–3.63	2,540	431	5.89	100	1.8	102.2	112.0	45.2	62.3*
3.63–3.56	2,516	435	5.78	99.3	1.3	148.9	163.5	66.5	41.3*
3.56–3.50	2,440	433	5.64	99.8	1.2	151.4	166.6	68.5	44.7*
73.32–3.50	**50,115**	**8,784**	**5.7**	**99.2**	**5.2**	**30.2**	**33.2**	**13.6**	**98.9**

crystal due to radiation damage, although it could have also resulted from the crystal rotating out of the beam, given its morphology and dimensions. As a result, most datasets did not contain a complete set of reflections and failed to process using all pipelines available. However, a single crystal grown in 30% (v:v) PEG 200, 100 mM Na_3-citrate, pH 5, 100 mM KSCN (**Fig. 4.1**) withstood the high-energy radiation. Two sets of images, *i.e.* sweeps, were collected (**Table 4.1 a**), showing clear and well-defined diffraction patterns even after rotating the crystal by 90° (see **Fig. 4.2**) and more. Although only a small number of reflections were visible on each image, the data were successfully processed in orthorhombic space group $C222_1$ with unit cell angles $\alpha = \beta = \gamma = 90°$ (**Table 4.1 b**). The unit cell measured 94.45 × 116.30 × 123.16 $Å^3$, large enough to fit multiple copies of *At*CTNS in the asymmetric unit as suggested by the Matthews coefficient (V_M). The V_M for crystallized proteins of 20–60 kDa typically ranges from ~2–3.5 $Å^3\ Da^{-1}$ (Kantardjieff & Rupp, 2003). A single copy of *At*CTNS in this setup would

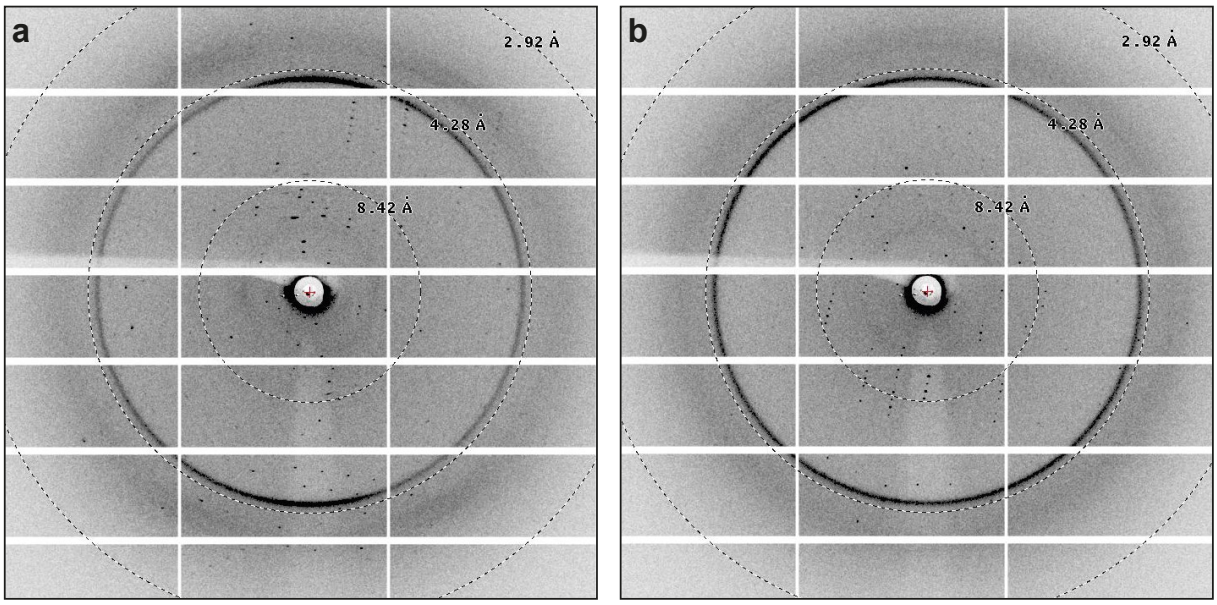


Fig. 4.2 – Best Diffraction Patterns of *At*CTNS.
Diffraction patterns obtained at $\varphi = 0°$ (**a**) and $\varphi = 90°$ (**b**), visualized using Adxv (Arvai, 2021).

result in a V_M of 5.00 Å^3 Da^{-1} and thereby account for a solvent content of 75.40%. The correct solution more likely contains two copies of *At*CTNS with a V_M of 2.50 Å^3 Da^{-1} and a solvent content of 50.81%.

4.4 Data Quality Analysis

Following processing, diffraction data quality was assessed using parameters such as the mean $I/\sigma(I)$, R_{meas}, and R_{pim}, as generated by AIMLESS during the scaling and merging steps (Evans & Murshudov, 2013). $\langle I/\sigma(I)\rangle$ reports the averaged signal-to-noise ratio of all reflection intensities compared to the background. R_{meas} and R_{pim} are optimizations of R_{merge} and measure the precision of the individual measurements of the reflections, *i.e.* how well the data agrees internally. R_{merge} (first suggested by Arndt *et al.* (1968); **Eq. 4.2**) was reported to be flawed, as it rises with increasing multiplicity, making the data appear worse statistically when, in fact, more independent measurements are available, yielding a more precise average reflection. The multiplicity-corrected version, R_{meas} (Diederichs & Karplus, 1997; **Eq. 4.3**), and the precision-indicating version, R_{pim} (Weiss, 2001; **Eq. 4.4**), were introduced to eliminate this caveat. R_{meas} reports a value independent of the multiplicity, whilst R_{pim} decreases with higher multiplicity (Evans & Murshudov, 2013). The R-factors are defined as follows:

$$R_{merge} = \frac{\sum_{hkl}\sum_{i=1}^{n} |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl}\sum_{i=1}^{n} I_i(hkl)}, \tag{4.2}$$

$$R_{\text{meas}} = \frac{\sum_{hkl} \sqrt{\frac{n}{n-1}} \sum_{i=1}^{n} |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^{n} I_i(hkl)}, \tag{4.3}$$

$$R_{\text{pim}} = \frac{\sum_{hkl} \sqrt{\frac{1}{n-1}} \sum_{i=1}^{n} |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^{n} I_i(hkl)}, \tag{4.4}$$

where $I_i(hkl)$ is the i-th measurement of the intensity of a reflection hkl, $\bar{I}(hkl)$ is their average, and n is the number of observations per reflection, *i.e.* its multiplicity.

The statistics generated during processing of the collected *At*CTNS diffraction data are listed in **Table 4.1 c**. The data appear to be almost complete at 99.2%. The mean $I/\sigma(I)$ of 5.2 suggests that the overall reflection intensities are weaker than ideal (should be close to, if not greater than 10). The R_{meas} of 33.2% and R_{pim} of 13.6% across all resolution shells are also at the edge of their optimal range, yet still in line with many published structures, including some of other PQ-loop proteins (Bräuer *et al.*, 2019; Tao *et al.*, 2015). All three metrics depend on the resolution range of the data and are negatively influenced by low-intensity, high-resolution data as present in this dataset. R_{meas} is an excellent data quality indicator for high-intensity, low-resolution, *i.e.* early resolution shells, where a value of <10% is pursued (Evans & Murshudov, 2013) – the *At*CTNS dataset achieves this with 6.4% in the first bin. Overall, the data were cut to a resolution of 3.50 Å, resulting in a mean $I/\sigma(I)$ of 1.2, R_{meas} of 166.6%, and R_{pim} of 68.5% in the highest resolution bin. Traditional crystallography guidelines suggest truncating the data earlier, *e.g.* at ~3.8–4.1 Å, to avoid the mean $I/\sigma(I)$ dropping below 1.5–2.0, and so the R_{meas} does not rise above 50–60%. However, it was shown that inclusion of the weak high-resolution data in many cases improves the ensuing electron density (Diederichs & Karplus, 2013; Wang & Wing, 2014; Karplus & Diederichs, 2015), and in any case ''seems to do no harm for the purposes of either automatic or manual model building'' (Evans & Murshudov, 2013). It was argued that because R_{meas} approaches infinity as the data become weaker, a sensible cut-off cannot be determined, and an arbitrary value should not be used to truncate the data (Karplus & Diederichs, 2012; Evans & Murshudov, 2013). Instead, ''paired-refinement'' tests should be carried out to establish whether a high-resolution shell containing weak data improves the electron density and/or statistics of the atomic model or not. A useful data quality indicator that accurately reflects whether this is the case is the correlation coefficient of randomly halved datasets, termed $CC_{1/2}$ (Karplus & Diederichs, 2012; Diederichs & Karplus, 2013). A perfect correlation is reflected by a $CC_{1/2}$ of 100%; no correlation by 0%. Whether a high-resolution shell contains valuable information is determined by the Student's t-test (Student, 1908), which is performed to analyse if the correlation is significantly different from 0 at $p = 0.01$. Significance is indicated in the $CC_{1/2}$ column of the statistics table using an asterisk (**Table 4.1 c**). $CC_{1/2}$ was used as the primary indicator to justify data truncation throughout this work. In the *At*CTNS

dataset, the highest-resolution shells retained a significant $CC_{1/2} > 40\%$ and were therefore included in the data.

4.5 Structure Solution Attempts Using Homology Models

In XRC, structure solution is based on the interpretation of the electron density map $\rho(xyz)$, achieved through model fitting. The map represents a complex three-dimensional wave function generated by Fourier synthesis using the specific characteristics of the observed reflections (Rhodes, 2006). It contains amplitudes $|F_{hkl}|$, frequencies h, k, and l, and phases α_{hkl} for each reflection hkl. The electron density can be expressed mathematically as

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} |F_{hkl}| \, \mathrm{e}^{-2\pi \mathrm{i} (hx + ky + lz - \alpha_{hkl})}, \tag{4.5}$$

where V is the volume of the unit cell and the imaginary number i is the square root of −1. Amplitudes and frequencies were recovered from the diffraction images (see **Sec. 4.3**), however, the phases cannot be measured in the experiment, yet they carry more structural information than the other two parameters (Taylor, 2003; Rhodes, 2006). This is called the phase problem.

Amplitudes and phases are only linked *via* the electron density or molecular structure – there is no direct relationship (Taylor, 2003). Therefore, if the protein's structure is unknown, the only way to solve the phase problem is through experimental calculations (see **Sec. 7.2**). If the target structure is known, or close homologues thereof, initial phase values can be estimated by MR using the known structure as a search model. Homologous structures may be used as search models, provided they share high structural similarity with the target. Generally, the search model and target need to superimpose with a root mean square deviation (RMSD) of <2 Å across all overlaid C_α atoms for MR to work (Abergel, 2013). This correlates to a sequence identity of ~30–35% between the model and target (Chothia & Lesk, 1986), not considering larger conformational changes, which may result in a higher RMSD between highly similar structures than suggested by the sequence identity.

MR is a trial-and-error technique that places a model in a plethora of different orientations and positions to find the best fit within the crystal structure, *i.e.* it aims to find the best match between the observed diffraction and the diffraction predicted for the model at any place in the unit cell (Evans & McCoy, 2008). Once found, the phases are copied, and an initial electron density map is calculated using the borrowed phases alongside the experimentally observed amplitudes. Phases are then optimized by crystallographic refinement to yield a final electron density map.

MR trials within the scope of this thesis were tackled using the program Phaser (McCoy *et al.*, 2007), operated from within the CCP4 suite (Winn *et al.*, 2011). Phaser uses the maximum likelihood method to estimate the most probable orientation and location of the model in the unit cell (McCoy, 2004). Searches are split into a rotation function and a translation function and are scored using multivariate statistics (McCoy *et al.*, 2007). The key statistics to analyse whether MR by Phaser was successful are the translation-function *Z*-score (TFZ), where *Z* represents the number of standard deviations that the log-likelihood gain (LLG) exceeds the mean, and the LLG itself (Oeffner *et al.*, 2018). Correct solutions are confidently indicated when Phaser reports a TFZ of >8 and an LLG of >64 for the placement of a first model. However, scores as low as 6 and 36, respectively, may still indicate the correct placement of a first search model.

Efforts were made to solve the structure of *At*CTNS by MR using the presented data. However, finding an appropriate search model proved difficult. The only structures of eukaryotic PQ-loop family members available in the PDB at the time were the chicken KDEL receptor, codes 6I6B, 6I6H, and 6I6J (Bräuer *et al.*, 2019), and two plant homologues of SWEETs, *At*SWEET13, code 5XPD (Han *et al.*, 2017), and rice SWEET2b from *Oryza sativa* (*Os*), codes 5CTG and 5CTH (Tao *et al.*, 2015). The structure of *At*SWEET13 was solved by MR using *Os*SWEET2b as a search model in Phaser (Han *et al.*, 2017). Their sequence identity is 38% over a 220-residue overlap, and the structures superimpose to an RMSD of 1.711 Å over 205 C_α atoms (**Fig. 4.3 a**). In contrast, local alignment with *At*CTNS only accounts for 17–20% sequence identity using any of the three proteins mentioned above, well below the ~30–35% identity minimally required for successful MR. Unsurprisingly, MR using these protein structures failed.

In an attempt to increase the MR success probability, we generated homology models for *At*CTNS based on *At*SWEET13 and *Os*SWEET2b, respectively, using the SWISS-MODEL pipeline (Waterhouse *et al.*, 2018). As a result, the models share 100% sequence identity with *At*CTNS, however, their overall three-dimensional shape remains widely unaltered and biased towards the base, as they superimpose well over large areas with the respective SWEET (**Fig. 4.3 b**). Placement of a single ensemble consisting of either model in Phaser resulted in a TFZ of 4.0 and 3.8 as well as an LLG of 30 and 32, respectively, indicating a unique solution was not found for either model. Placement of a second ensemble resulted in slightly higher TFZ scores (5.1 and 5.0) but lower LLG. When superimposing the models with the finalized *At*CTNS structure from **Ch. 7** (PDB: 7ZK1, **Fig. 7.6**), we understand why MR was unsuccessful, as the RMSDs amount to 6.623 and 6.129 Å, respectively (**Fig. 4.3 c**).

Better search models for MR would have been the structures predicted by AlphaFold (Jumper *et al.*, 2021), however, they were not available at the time, as AlphaFold was only launched about two months after we had solved the structure experimentally (**Ch. 7**). Interestingly, the AlphaFold prediction (UniProt: P57758) pictures *At*CTNS in a more occluded conformation

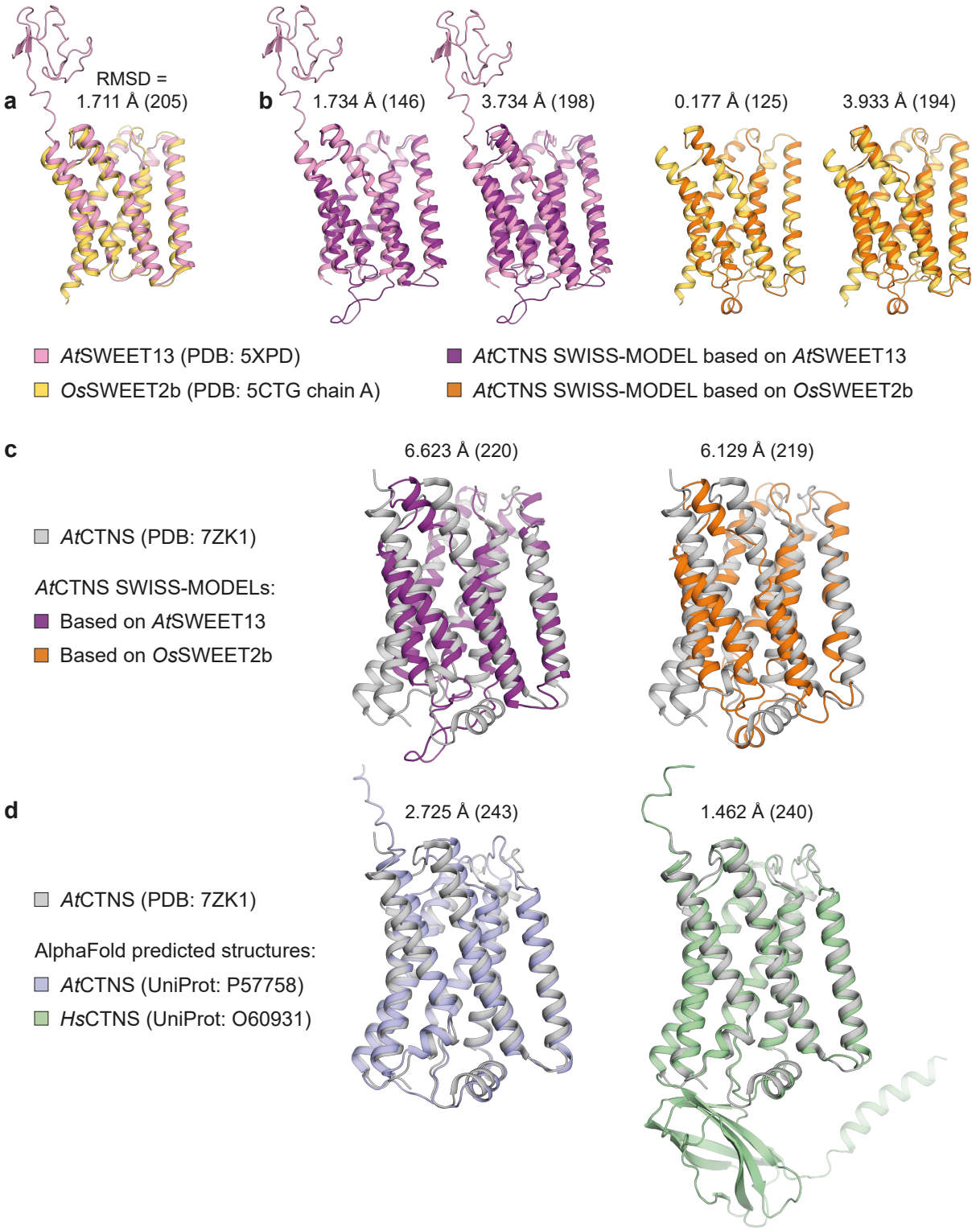


Fig. 4.3 – Homology Models for Molecular Replacement.
Cartoon representation of superimposed structures highlighting the RMSD over a span of C_α atoms, shown in parentheses. **a**, Superimposition of *At*SWEET13 and *Os*SWEET2b. **b**, Superimposition of the *At*CTNS SWISS-MODELs generated based on the SWEET structures in **a**. **c**, Superimposition of the homology models of **b** with the finalized *At*CTNS structure from **Fig. 7.6**. **d**, Superimposition of the AlphaFold predictions of *At*CTNS (UniProt: P57758) and *Hs*CTNS (O60931) with the finalized *At*CTNS structure. The AlphaFold predictions are available under the identifier AF-''UniProt code''-F1-v4 from the database accessible online at https://alphafold.ebi.ac.uk (Varadi *et al.*, 2022, 2024).

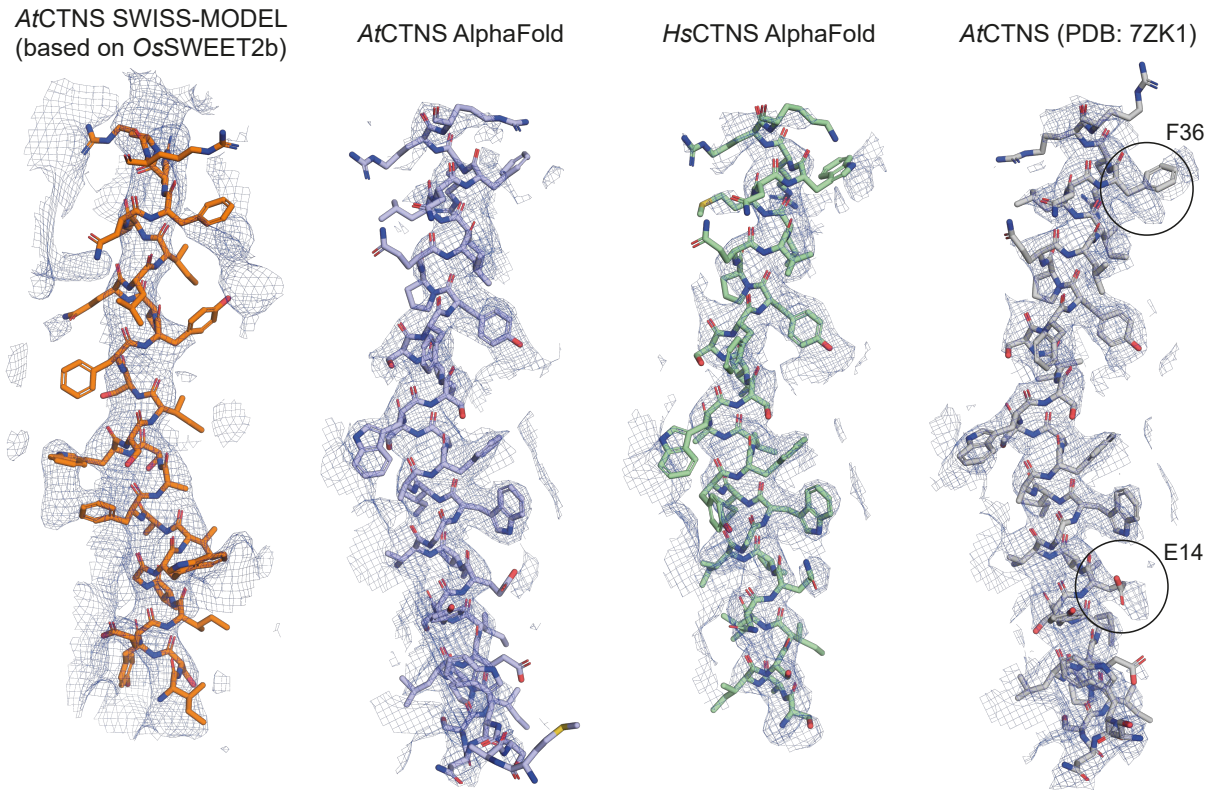


Fig. 4.4 – Exemplary Electron Density Maps After MR.
$2mF_o$–DF_c electron density maps obtained after MR in Phaser (McCoy *et al.*, 2007) with 10 subsequent rounds of refinement in REFMAC (Murshudov *et al.*, 2011). Maps are shown for TM1 of each search model used during MR, each contoured at 1 σ with a carve of 5 Å for the SWISS-MODEL and 3.5 Å for the remaining structures.

than all our crystal structures do, leading to an RMSD of 2.725 Å over all 243 C_α atoms when superimposed (**Fig. 4.3 d**). Using the entire model (residues 1–244, omitting the long disordered region after TM7) in Phaser yields TFZ scores of 5.3 and 6.3 as well as LLGs of 54 and 57 for the first and second molecule placed in the unit cell, respectively. Phaser positions the model correctly in the unit cell, however, due to the disagreement of the model and target structure over large portions, the phases are misjudged slightly, leading to a suboptimal electron density map with little to no coverage for many side chains (**Fig. 4.4**), that needs to be optimized through crystallographic refinement. The AlphaFold prediction of *Hs*CTNS (UniProt: O60931) is in an inward-open conformation (as are our crystal structures), making it a better search model for MR (**Fig. 4.3 d**). Using residues 121-352 (omitting the N-terminal domain (NTD) and C-terminal disordered region), Phaser scores TFZs of 6.8 and 12.4 and LLGs of 72 and 195, indicating that a unique solution has been found. The resulting map achieves better side chain coverage but still requires optimization, this time due to the fact that *Hs*CTNS only shares about 45% of its amino acid sequence with *At*CTNS. An exemplary map for TM1 is shown in **Fig. 4.4** and compared to the map created when using the finalized crystal structure of *At*CTNS obtained in **Ch. 7** (PDB: 7ZK1, **Fig. 7.6**) in Phaser (TFZ = 10.2 and 20.2, LLG = 203 and 539). For

the latter, good side chain coverage is achieved for most residues, and the map additionally enables the orientation optimization for those side chains that do not fit optimally, *e.g.* those of residues E14 and F36 (highlighted in **Fig. 4.4**).

CHAPTER 5

Generation of Sybodies & Nanobodies against *At*CTNS

Sybody (Syb) screening was inspired by past projects carried out in the lab, especially that of the KDEL receptor, a small membrane protein similar to CTNS, where a single Syb enabled the structure to be determined through XRC, and cell localization experiments (Bräuer *et al.*, 2019). Sybs were designed on the basis of different nanobodies (Nbs) with varying CDR3 lengths, creating the three types of concave, loop, and convex Sybs, colour-coded green (G), blue (B), and red (R), respectively (Zimmermann *et al.*, 2018; **Fig. 1.7 a**). Loop Sybs originate from a Nb that binds the β_2-adrenergic receptor (**Fig. 1.6 c**), one of many GPCRs where structural investigation was facilitated by specific Nbs (Manglik *et al.*, 2017). Due to the structural similarity between *At*CTNS and GPCRs (**Sec. 9.1.1**), and since *At*CTNS can be reconstituted into liposomes (**Fig. 3.4**), we also decided to generate Nbs against the transporter by immunizing a llama. Ultimately, we needed both the sybodies and nanobodies to solve the three-dimensional structure of *At*CTNS. This chapter details how these binders were generated and how those best suited for structural and functional studies were selected. Emphasis is placed on Sybs 1 and 39 as well as Nbs 1 and 4, which later occur in co-crystal trials and structures (**Chs. 6** and **7**). For those binders, we determined the binding characteristics with *At*CTNS in detail.

5.1 Purification & Biotinylation of *At*CTNS-Avi

*At*CTNS was first re-cloned to generate a construct that bears a C-terminal Avi-tag, adding amino acids GLNDIFEAQ**K**IEWHE, which upon purification allows for specific *in vitro* biotinylation of the lysine residue (highlighted in bold) using the biotin-protein ligase BirA (Beckett *et al.*, 1999). Purification of *At*CTNS-Avi (**Fig. 5.1 a–c**) was carried out as for *At*CTNS without the Avi-tag (**Fig. 3.3**), peak fractions were then pooled and biotinylated in an O/N reaction using a commercially available kit (BirA Biotin-Protein Ligase Kit, Avidity) along with the GST-BirA fusion protein purified in the lab (Fairhead & Howarth, 2015). Subsequently, the sample was re-applied to the gel filtration column to separate out excess reagents of the biotinylation kit, including uncoupled D-biotin and adenosine phosphates – the latter cause a major A_{280} peak at ~20–21 mL (not shown). The monodispersity of the main protein peak, as well as its retention

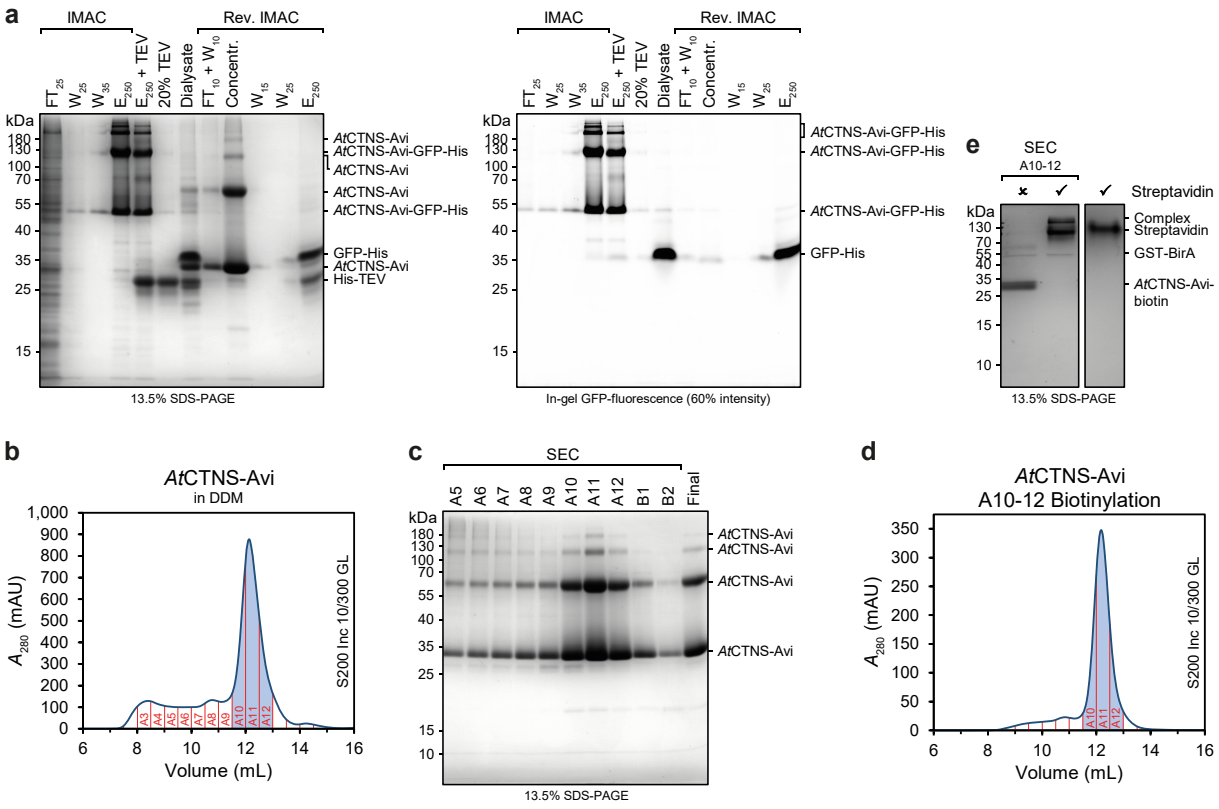


Fig. 5.1 – Purification & Biotinylation of *At*CTNS-Avi.
a, Relevant samples collected during IMAC and reverse-IMAC were separated on a 13.5% SDS-PAGE gel. Both in-gel fluorescence and Coomassie-stained images are shown. Label subscripts reflect imidazole levels in mM. The concentrated FT_{10} and W_{10} were gel filtrated. SEC profile is shown in **b**, with corresponding fractions run on gel (**c**). **d**, A portion of the pooled fractions of **b** (shaded in blue) was biotinylated and re-applied to the SEC column. **e**, Streptavidin shift assay using the pooled fractions of **d**, showing a complete transition of *At*CTNS-Avi, creating a higher MW band after streptavidin addition, indicating near-perfect biotinylation efficiency.

volume at ~12–12.5 mL, remained unchanged compared to the SEC profile of the purification (**Fig. 5.1 d** and **b**, respectively). Biotinylation efficiency was found to be nearly 100%, as tested by streptavidin shift, illustrated in **Fig. 5.1 e**. *At*CTNS-Avi-biotin was not passed through GSH-agarose after biotinylation, leaving minor GST-BirA contamination in the final sample used throughout the Syb and Nb screening, however, since only biotinylated molecules would be immobilized in the screening, this was not believed to interfere significantly.

N-terminally Avi-tagged *At*CTNS (Avi-*At*CTNS) was also cloned to enable the presentation of epitopes located closer to the C-terminus, which otherwise might be blocked by the avidin-biotin binding site. However, Avi-*At*CTNS expression was reduced to ~25% compared to *At*CTNS-Avi, which expressed similar to tag-free *At*CTNS. Given that TMHMM predicted a longer linker at the C-terminus after TM helix 7 (**Fig. 3.1**), we were hopeful that epitopes located closer to the C-terminus would still be accessible for Sybs and Nbs to interact with.

5.2 Sybodies

5.2.1 Screening

Syb generation was performed as outlined by Zimmermann *et al.* (2020). Ribosome display makes up the first of three selection phases: mRNA libraries are translated *in vitro* but stalled on the ribosome, resulting in ribosome-mRNA-Syb complexes that are presented to the immobilized target protein. After washing, the mRNA from the preserved complexes is eluted, reverse transcribed, and then ligated into a phagemid (pDX_init) and electroporated into bacterial cells (SS320) for the production of phages. These ligation and transformation steps are of major importance in the pipeline, as a low efficiency will reduce the library diversity and thus restrict the performance of the following phage displays. The protocol specifies that the number of transformants required to minimize diversity loss should be greater than 5×10^6 colony-forming units (CFU) mL^{-1} of cell culture. In a first attempt, this was only achieved for one of the three libraries. The transformation output was calculated at 2.8, 5.6, and 1×10^6 CFU mL^{-1} for G, B, and R libraries, respectively. Ligation and transformation were therefore repeated, and overall efficiency was improved, totalling 1.4, 8, and 7×10^6 CFU mL^{-1} for the same libraries. Although less colonies were found for the concave (G) library, and some diversity may have been lost, the greater number of transformants for the other two libraries led us to use the second bacterial cultures to proceed with the Syb generation (the amount of cDNA was insufficient to repeat the ligation and transformation a third time).

For the remaining selection phases, two rounds of phage display, Syb-presenting phages were purified and incubated with immobilized *At*CTNS-Avi-biotin. After washing, phages coupled to Sybs that bound the target were recovered and used to infect a bacterial culture, facilitating phage production for the subsequent display. In addition to *At*CTNS, phage displays were simultaneously performed using a NC protein, here $PepT_{So}$. During the second round of phage display, unbiotinylated *At*CTNS was used to help dissociate weaker binders. The abundance of phage eluted from both targets (its DNA) was then analysed using qPCR to assess the target-specific enrichment of each library. After the first phage display, enrichment factors were found to be 0.851, 3.62, and 1.25 for G, B, and R libraries, respectively. After the second round, they were calculated at 3.78, 139, and 0.227, indicating a highly enriched loop (B) library, a mildly enriched concave (G) library, and a convex (R) library that was not enriched compared to the NC.

Identification of individual Syb binders was carried out by ELISA, exposing each clone to the target protein as well as a NC (again, $PepT_{So}$ was used). In preparation, the phage display output was first cloned into a dedicated expression vector (pSb_init) using the highly efficient FX method (Geertsma, 2013). Subsequently, the mixed Syb constructs were transformed into chemically competent cells (MC1061) and plated. 96 individual colonies were then picked for

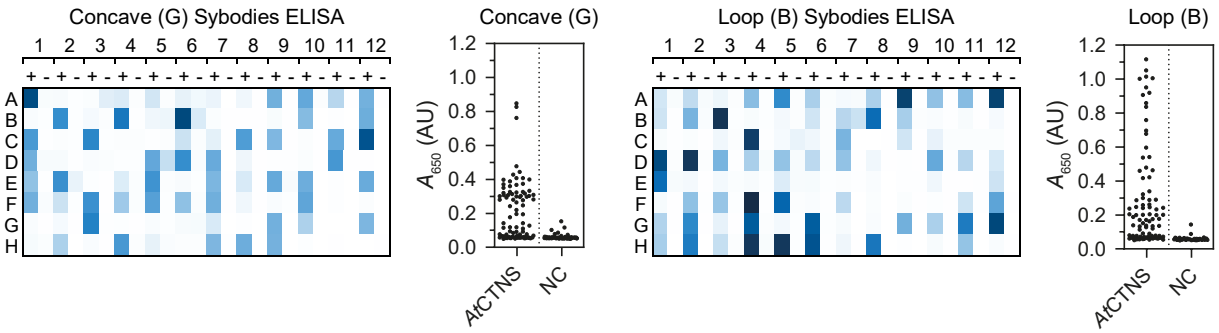


Fig. 5.2 – ELISA of *At*CTNS-Specific Sybodies.
Gradient plot illustration of ELISA scores, visualizing binding of concave (G) and loop (B) sybodies to *At*CTNS-Avi-biotin (+) and $PepT_{So}$-Avi-biotin (NC, -), respectively. Scatter plots of the same data demonstrate the abundance of different absorbance intensities.

each of the libraries, expressed, and extracted from the periplasm. Sybs expressed from pSb_init bear an N-terminal pelB leader (amino acid sequence: MKYLLPTAAAGLLLLAAQPAMA), which causes periplasmic secretion (Better *et al.*, 1988); the pelB sequence is cleaved in the process (Lei *et al.*, 1987). Additionally, Sybs contain a C-terminal Myc-His-tag, which is used for immobilization during ELISA using an α-Myc antibody. ELISA readout was facilitated through a streptavidin-HRP polymer, which binds the biotinylated target protein or NC in case they interact with the presented Syb. TMB was used as HRP substrate, which, upon oxidation, yields a blue colour with an absorption maximum of ~650 nm (Bos *et al.*, 1981).

ELISA results are illustrated in **Fig. 5.2**. Absorbance levels at 650 nm were generally low for the NC, with only a few wells exceeding background levels of ~0.05–0.07. A_{650} was increased for many wells that contained biotinylated *At*CTNS, with levels reaching as high as 0.85 and 1.12 for the concave (G) and loop (B) Sybs, respectively. None of the wells in the convex (R) library ELISA illuminated (not shown), suggesting the generation of *At*CTNS-specific Sybs of that type failed. In total, 111 wells showed an absorbance ratio (*At*CTNS/NC) of 1.5 or higher (**Table A.5**), which, according to the Syb generation protocol (Zimmermann *et al.*, 2020), may indicate a specific binder. The Sybs stemming from these wells were sequenced, phylogenetically analysed based on their full sequence, and subsequently numbered from top to bottom (for concave (G) binders, see **Fig. 5.3**, loop (B) Sybs in **Fig. 5.4**). These analyses reveal 18 unique concave and 34 loop Sybs, although multiple Sybs of either library were identical in their CDR and only differed phylogenetically due to a framework mutation that was likely introduced during one of the pipeline's reverse transcription or PCR amplification steps (see **Table A.5**). Excluding those, the Syb selection yielded 44 unique sybodies in total.

Comparison of the unique CDR sequences of concave and loop Sybs reveals a dominance of polar and aromatic residues over nonpolar aliphatic ones at almost every hypermutated position (**Fig. 5.5**). The only exception to note are some alanine residues that preferably populate the first half of the CDR1 of concave binders and the end of the CDR3 of loop Sybs. The remaining

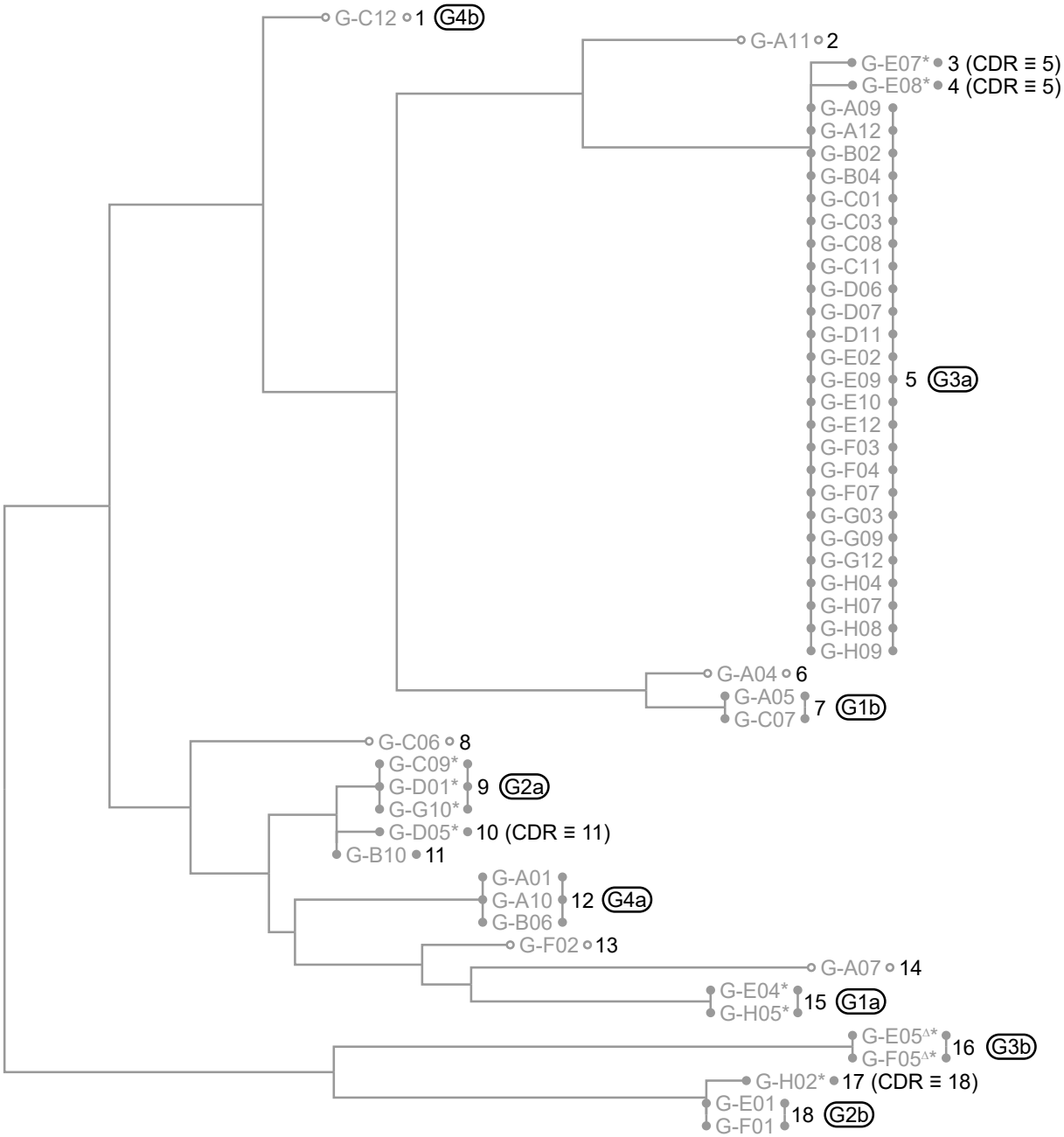


Fig. 5.3 – Phylogenetic Analysis of *At*CTNS-Specific Concave (G) Sybodies.
Groups are formed using the full-length Syb sequences. Unique sequences are numbered from top to bottom. Branches of Sybs with a reoccurring CDR end in a solid bullet point. For binding analyses, Sybs are arranged in classes (encircled) based on their ELISA score. The original ELISA location is given in grey font (compare with **Fig. 5.2**, left side). *Sequence contains a backbone mutation. $^{\Delta}$Sequence contains a deletion in one of the CDRs. For more information, see **Table A.5**.

nonpolar residues, such as valine, leucine, isoleucine, and methionine, are underrepresented throughout all CDR residues. The dominance of aromatic residues, mainly tryptophans and tyrosines, is focused on the centre of the CDR3 of both Syb types, however, they are also present in either CDR2, albeit to a lesser degree. Concave Sybs are most conserved at positions 33 (glycine) and 35 (glutamate) in the CDR1, as well as positions 97 (glutamate), and 99 and 103

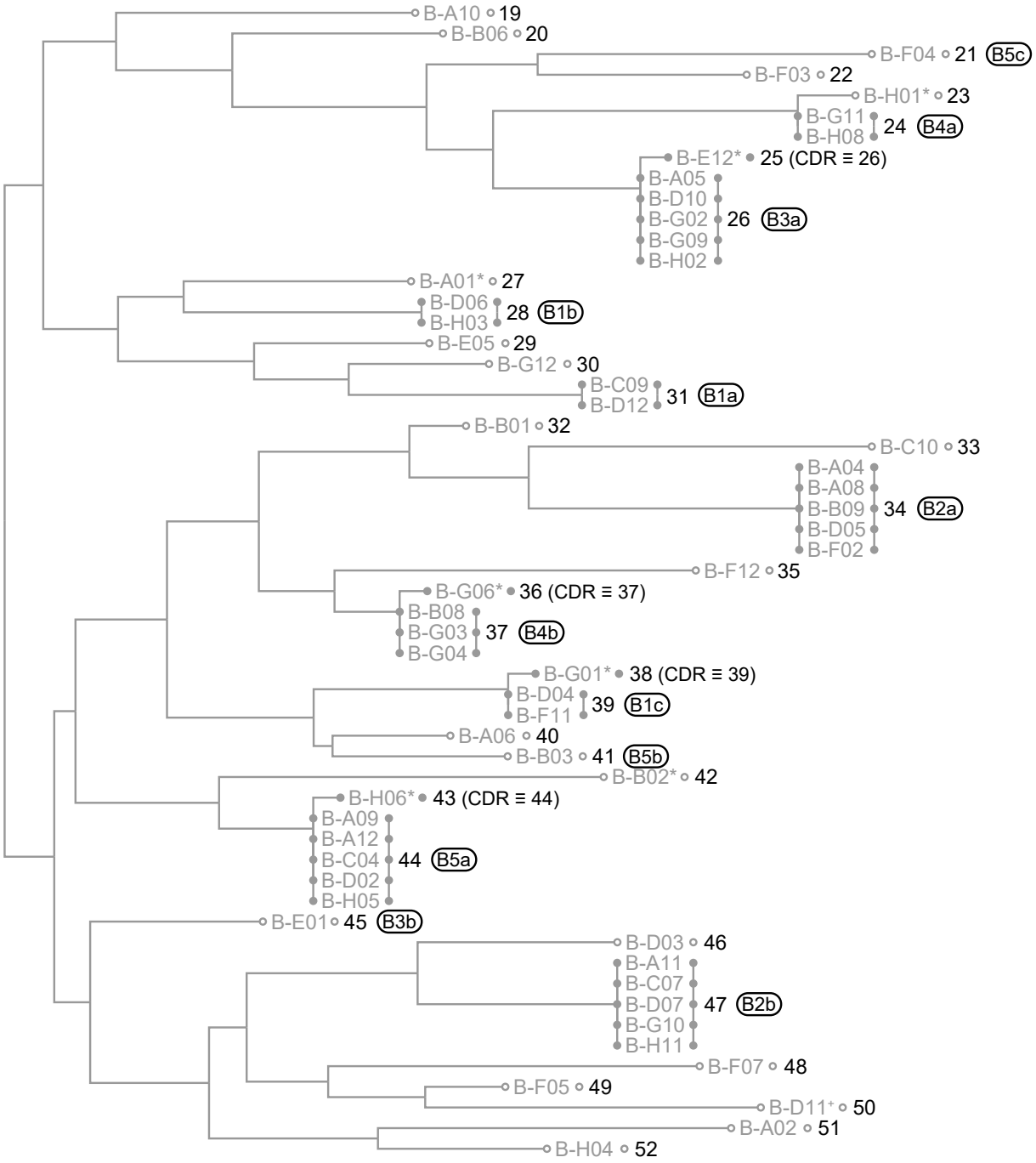


Fig. 5.4 – Phylogenetic Analysis of *At*CTNS-Specific Loop (B) Sybodies.
Groups are formed using the full-length Syb sequences. Unique sequences are numbered from top to bottom. Branches of Sybs with a reoccurring CDR end in a solid bullet point. For binding analyses, Sybs are arranged in classes (encircled) based on their ELISA score. The original ELISA location is given in grey font (compare with **Fig. 5.2**, right side). *Sequence contains a backbone mutation. +Sequence contains an insertion in one of the CDRs. For more information, see **Table A.5**.

(each a tryptophan) of the CDR3. The presence of multiple conserved charged residues within the CDRs may indicate that most Sybs interact with the target protein in a similar way. Multiple conserved glutamates could indicate that most concave Sybs bind to an epitope comprised of multiple residues with positive charge, such as the transition of TM1–2 and TM5–6, each

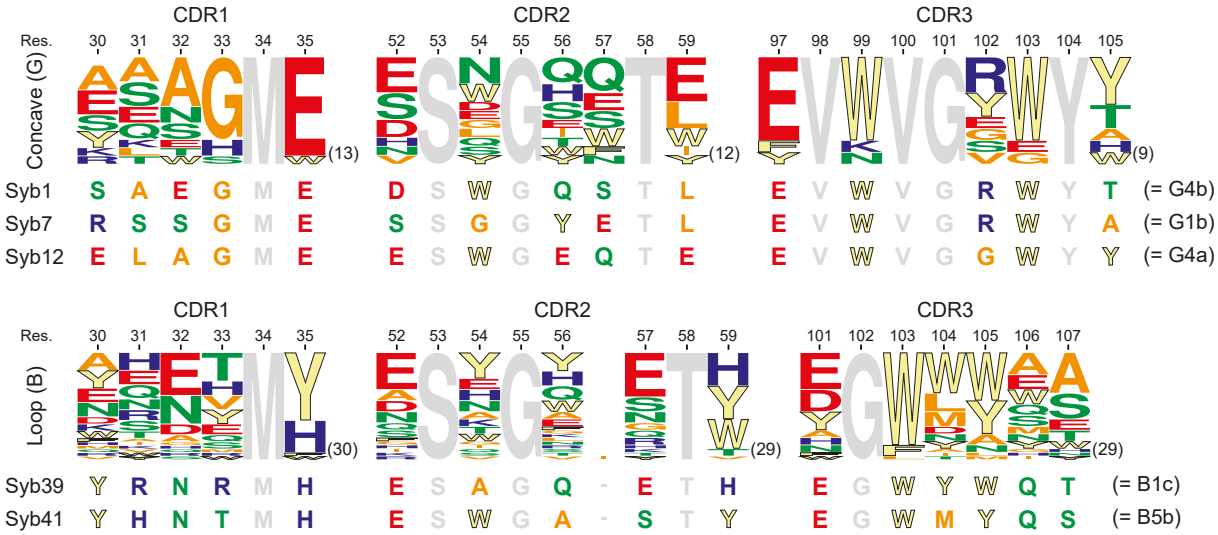


Fig. 5.5 – α-*At*CTNS Sybody CDR Sequence Comparison.
Frequency plot logos using unique CDRs only (number of sequences shown in parentheses), generated using WebLogo (Crooks *et al.*, 2004). The CDR sequences of Sybs 1, 7, 12, 39, and 41 are highlighted, and individual residues are coloured *via* amino acid chemistry. Greyed letters indicate CDR residues that are constant by design. One of the 29 unique loop Syb CDR2 sequences contains an additional isoleucine inserted in between residues 56 and 57.

of which is rich in arginines and lysines (**Fig. 3.1 a**). Loop Sybs are overall more diverse in the CDR1 and 2 as compared to the concave Sybs. The CDR3, however, is highly enriched in bulky hydrophobic residues, mostly tryptophans, some tyrosines, leucines, methionines, and phenylalanines, which may indicate interactions with more hydrophobic areas of *At*CTNS, potentially including its inner core.

5.2.2 Biochemical & Biophysical Analyses of *At*CTNS-Sybody Interactions

With no major correlation expected between ELISA signal ratios and binding efficiencies and kinetics of individual unique Sybs (according to Zimmermann *et al.* (2020), signals as low as 1.5-fold over background can indicate high-affinity binders), we wanted to evaluate the binding capabilities in detail using a set of biochemical and biophysical experiments. Therefore, 20 of the 44 unique Sybs were expressed on a small scale (50 mL cultures) and purified using periplasmic extraction followed by standard IMAC and SEC (traces shown in **Fig. A.1**, retention volume at peak maximum annotated in **Fig. 5.6**). The number of Sybs was limited due to experimental constraints, Sybs containing a backbone mutation were avoided where possible. Sybs with multiple representatives in the phylogenetic analysis (**Figs. 5.3** and **5.4**) were favoured, and the remaining Sybs were chosen at random but in a way that all ranges of ELISA signal ratios were covered. For easier comparison, Sybs were classified based on their ELISA score: Class 1 is made up of Sybs with ELISA signal ratios ranging from 1.5 to ~3, class 2 from ~3 to ~5,

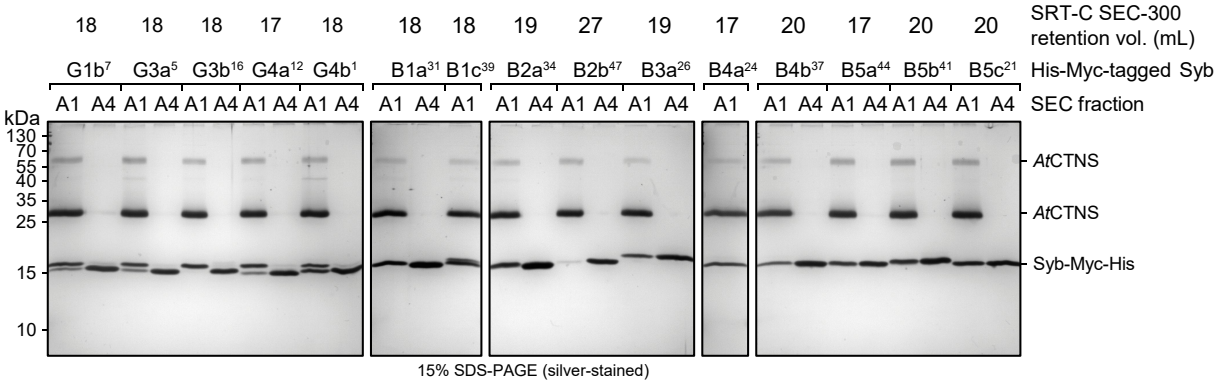


Fig. 5.6 – SEC Co-Elutions of *At*CTNS & Sybodies.
SEC fractions A1 (first fraction of *At*CTNS peak) and A4 (first fraction of Syb excess) for each *At*CTNS-Syb complex were analysed by SDS-PAGE. Corresponding SEC traces are shown in **Fig. A.2**. Superscripts denote sybody identification number, as based on the phylogenetic analysis. For each Syb, the retention volume of the elution peak (as observed during the purification) is annotated.

class 3 from ~5 to ~8, class 4 from ~8 to ~15, and class 5 higher than 15 (see **Figs. 5.3** and **5.4** and **Table A.5**). The expression yield of 25% of Sybs was not sufficient to include them in the following experiments. Those either did not elute from the SEC column or were not injected in the first place.

To test specific sybody binding, we first performed SEC co-elutions using the optimized conditions for *At*CTNS at pH 6. *At*CTNS and individual Sybs were incubated and run on the SEC column. Major peak shifts were not observed (**Fig. A.2**), therefore, peak fraction contents were separated by SDS-PAGE and analysed for the presence of Syb in the *At*CTNS peak (specifically in the A1 fraction), indicating stable complex formation. As shown in **Fig. 5.6**, many sybodies co-eluted with *At*CTNS, most prominent bands are visible for concave sybody G4b (Syb1) as well as loop sybodies B1a (31), B1c (39), B2a (34), B5a (44), B5b (41), and B5c (21). Concave (G) Sybs show an additional band at the height of loop (B) sybodies, which persisted even after the samples were boiled and re-run on gel (not shown). These may stem from a loop Syb that exhibited strong interactions with the matrix of the SRT-C SEC-300 column during purification, causing it to elute late (like SybB2b (47), which eluted at ~27 mL) or not at all. Repeated application of *At*CTNS likely led to the capture of some of this material from the column, resulting in a ‘‘contaminated’’ A1 fraction. Assuming the concave sybodies bind a similar epitope on *At*CTNS as this loop sybody does, this would make sense when comparing the band intensities of the higher MW Syb band with the lower one. SybG3b (16) did not co-elute with *At*CTNS at all, yet shows the strongest higher MW Syb band (compare excess sybody in A4 with A1).

Taking advantage of the sybodies’ C-terminal His-tag, we next performed pulldowns using Ni-NTA resin. Buffered at pH 7.5 to ensure proper binding of the His-tag to the divalent nickel ion immobilized on the resin, *At*CTNS-Syb complexes were captured and analysed by SDS-

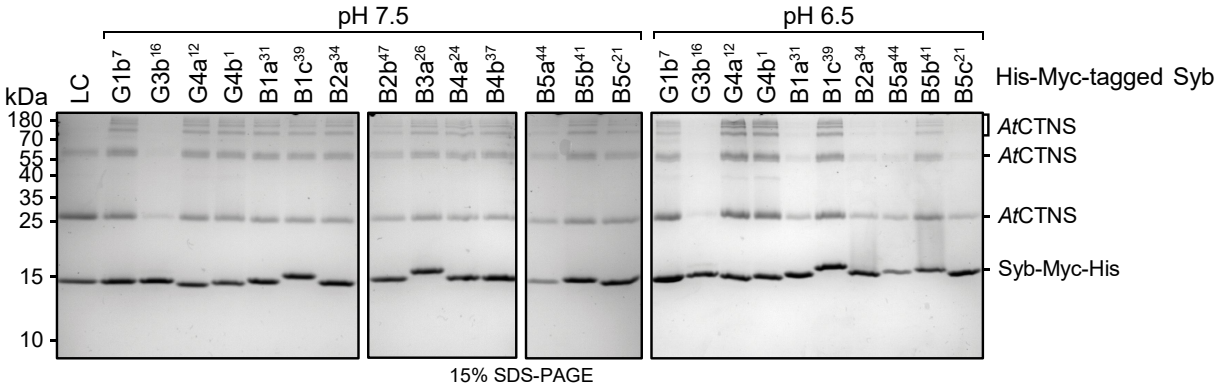


Fig. 5.7 – Pulldowns of *At*CTNS Using His-Tagged Sybodies.
Pulldowns performed at pH 7.5 and 6.5 were analysed by SDS-PAGE. Superscripts denote sybody identification number, as based on the phylogenetic analysis. LC = loading control.

PAGE (**Fig. 5.7**). Interestingly, most sybodies retained a similar amount of *At*CTNS in these experiments – only the pulldown using concave SybG3b (Syb16) resulted in much fainter bands for *At*CTNS, consistent with the SEC co-elutions. Ranking the sybodies based on their binding capabilities proved difficult under these conditions. Therefore, we repeated the pulldowns in a more acidic buffer at pH 6.5, keeping the pH higher than the pK_a of histidine, which is ~6.0, to ensure that most remain deprotonated to bind the resin. The goal was to improve the stability of *At*CTNS and potentially shift its confirmation to enhance the interaction with some sybodies and loosen it with others, ultimately identifying few sybodies more qualified for co-crystallization than others, considering that most *At*CTNS crystals grew at acidic pH (see **Fig. 4.1**). Indeed, only five sybodies, including concave sybodies G1b (7), G4a (12), and G4b (1), as well as loop sybodies B1c (39) and B5b (41), produced strong *At*CTNS bands in these experiments.

The prioritized Sybs were FX cloned (Geertsma, 2013) into the pBXNPHM3 vector and re-purified as tag-free sybodies on a large scale to be used for co-crystallization (**Ch. 6**). Using this plasmid, Sybs were initially expressed as maltose-binding protein (MBP) fusions but cleaved during purification to yield sybodies that omit any tags, which may negatively influence crystal contact formation. Concave Sybs G1b (Syb7) and G4b (1), as well as loop Sybs B1c (39) and B5b (41), were successfully purified, however, G4a (12) was found to be unstable at higher molarities, as it precipitated during the concentration process – only a small amount was recovered. SEC traces and SDS-PAGE analysis of Syb1, 39, and 41 are shown later in **Fig. 6.1 a** and **c**, respectively.

To further analyse which sybodies are best suited for crystal trials, we compared their association and dissociation rate constants (k_{on} and k_{off}, respectively) using bio-layer interferometry (BLI). BLI uses white light that is reflected off the tip of a fiber-optic biosensor, which is coated with biomolecules, here streptavidin, to facilitate the immobilization of biotinylated proteins.

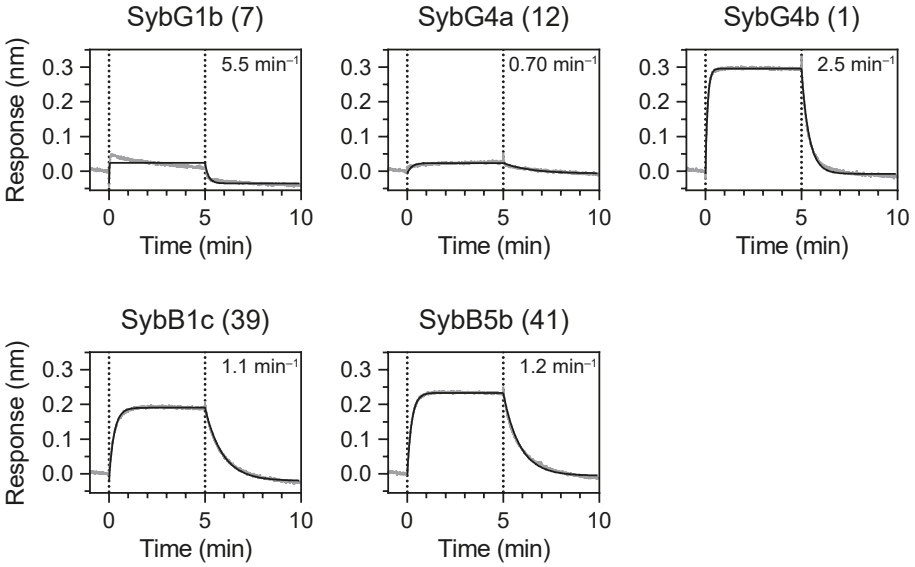


Fig. 5.8 – *At*CTNS-Sybody Binding Analysis Using BLI.
BLI experiments performed at pH 6.0, showing association and dissociation steps. Traces were fitted using GraphPad Prism. The resulting k_{off} of each binder is printed at the top of the dissociation step. *At*CTNS-Avi-biotin was immobilized at 100 nM, and sybodies were bound at 200 nM. Numbers in parentheses identify the sybodies, as based on the phylogenetic analysis.

Specific binding (to the streptavidin coat or the immobilized protein) thickens the biomolecular layer and, therefore, the path length of light reflected from this layer, causing a change in the interference pattern compared to an internal reference. The overall wavelength shift (in nm) is monitored as a function of time (Sultana & Lee, 2015). Three of the five Sybs tested demonstrated adequate binding as indicated by their pronounced association signals (**Fig. 5.8**). The response of the other two, SybG1b (Syb7) and SybG4a (Syb12), was found too low at 200 nM concentration to be considered high-affinity binders. The rapid on-rates observed for Sybs G4b (Syb1), B1c (39), and B5b (41) are linked to fast off-rates of 2.5, 1.1, and 1.2 min^{-1}, respectively. Nevertheless, these three sybodies were used in subsequent co-crystal trials (**Ch. 6**). Interestingly, they originated from both higher and lower ELISA scores, emphasizing the importance of testing all Sybs identified by ELISA.

Two of the sybodies proved vitally important in the pursuit of determining the structure of *At*CTNS: Concave Syb1 (SybG4b) and loop Syb39 (SybB1c) – see **Chs. 6** and **7**. To analyse their biomolecular interactions with the transporter in more detail, we determined the K_Ds for both binders using BLI (**Fig. 5.9**). The K_Ds were found to lie in the mid-to-high nanomolar range, averaging ~560 and ~760 nM across three and four replicates for Syb1 and Syb39, respectively. Among the Syb39 repetitions, large fluctuations were observed despite similar executions, resulting in a K_D that ranged from ~200–1,200 nM, causing a substantial standard deviation of 502 nM (see **Fig. 5.9 b**). These experiments also enabled us to determine the dissociation rate constant (k_{off}) of each sybody more precisely. Using all available data, the Syb1 dissociation rate was calculated at 5.8 ± 1.6 min^{-1}; for Syb39, a value of 2.0 ± 0.6 min^{-1} was determined, each

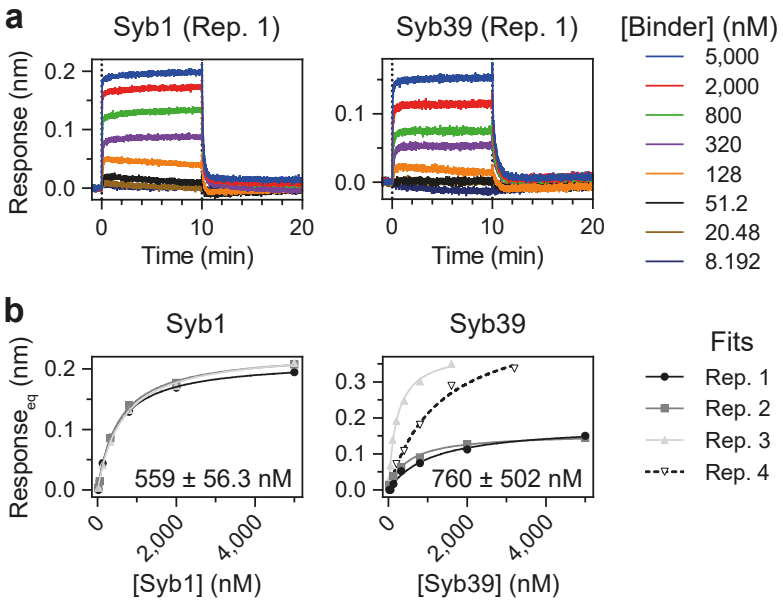


Fig. 5.9 – K_D Analysis of Important Sybodies Using BLI.
a, BLI traces of the first replicates of Syb1 (SybG4b) and Syb39 (SybB1c) are shown, highlighting association and dissociation steps using various binder concentrations. *At*CTNS-Avi-biotin was immobilized at 20 nM (100 nM for Syb39 replicates 3 and 4). **b**, K_D analysis of multiple replicates using the response at equilibrium (from **a**), as calculated using the integrated analysis tool of the Octet RED384 system. Data were fitted using GraphPad Prism. K_D values are shown in the bottom right corner of each plot.

showing an approximately twofold increase compared to the prior definitions based on only one dissociation trace. Sybodies obtained in other projects (Zimmermann *et al.*, 2020) often featured a K_D in the low-nanomolar range as well as slower off-rates – the affinities presented in this work are relatively poor in comparison. Despite this, I will demonstrate the usefulness of our binders for structural and functional studies in the following chapters.

5.2.3 Repeated Screening at pH 5.5

Given the increased thermal stability of *At*CTNS in more acidic environments (see **Fig. 3.5**), we also repeated the sybody screening for the concave (G) and loop (B) libraries at pH 5.5, aiming to identify novel sybodies with improved binding characteristics. Preserved glycerol stocks from after the ribosome display of **Sec. 5.2.1** were revived, and phage displays were carried out buffered at pH 5.5 instead of pH 7.5. Enrichment factors compared to PepT$_{So}$ used as NC were found to be significantly lower than in the conventional screening. Factors of 1.41 and 3.16 were calculated for the concave (G) and loop (B) libraries, respectively. The following ELISA (binding and washing steps also performed at pH 5.5) only yielded 24 hits in total. Sequencing revealed 11 novel unique binders, which were numbered Sybs 53–63. One of these binders is from the concave (G) library, and the remaining 10 are from the loop (B) library. Sybs 27, 37 (SybB4b), 40, 41 (B5b), and 45 (B3b) were also re-identified. Upon analysis of the *At*CTNS-Syb interactions, only Sybs 54 and 57 were deemed promising *At*CTNS binders, however, they were

not utilized in any structural or functional studies in this work, as we prioritized the superior sybodies identified in the conventional screening (data not shown).

5.3 Llama Derived Nanobodies

5.3.1 Screening

The procedure for generating α-*At*CTNS Nbs was adapted from Pardon *et al.* (2014) and Zimmermann *et al.* (2020). In collaboration with Prof. Raymond Owens (The Rosalind Franklin Institute), we immunized a llama with *At*CTNS reconstituted into POPE:POPG liposomes (see **Fig. 3.4** and **Sec. 2.6.2**). Following several rounds of immunization, Prof. Owens' group generated a library containing the Nb genes amplified from the peripheral blood mononuclear cells, which were isolated from the llama blood. The library was cloned into phagemid pADL-23c (Pardon *et al.*, 2014) and subsequently used to generate phages, which were used to screen for binders to *At*CTNS. Consecutive phage displays were carried out similarly to the second round performed in the Syb selection (Zimmermann *et al.*, 2020; **Sec. 5.2.1**), although without unbiotinylated competition protein and using 15 washing steps instead of three. Rather than using qPCR analysis, enrichments were calculated by infecting bacterial cells (TG1) using the eluted phage from the display. Comparison of the number of colonies yielded from phage eluted from the *At*CTNS display to that of the NC yielded enrichment factors of ~5 and ~60 after a first and second round, respectively. ELISA was performed using 191 individual clones to identify potential Nb binders (Syb39 was inoculated in 1-A01 as a positive control).

ELISA results are illustrated in **Fig. 5.10 a** (raw data in **Table A.6**). Overall, only three wells were visibly illuminated, showing A_{650} values of ~0.25–0.4. Excluding the positive control, five more wells elicited absorbances significantly higher than background levels at >0.10. In total, nine clones with signal-over-background ratios higher than 1.5 were identified. Those, along with an additional three clones with ratios as low as 1.35, were sequenced and phylogenetically analysed (**Fig. 5.10 b**). Unique groups were numbered from top to bottom. Only Nb2 was represented multiple times in the ELISA. Nbs 3 and 4, however, share identical CDRs and only differ in two residues positioned towards the N- and C-terminus, respectively. Full-length Nb sequences are listed in **Table A.3**.

Individual residues of the three CDRs are compared in **Fig. 5.11**. Only few trends are observed, which may be a result of the fact that only a small number of unique Nbs were generated; six unique CDR1 and 2, and seven unique CDR3 sequences were identified. Further enrichment of the *At*CTNS-specific Nb pool through a third round of phage display could have helped to increase the number of positive clones identified by ELISA. However, it is possible that this would have only increased the number of representatives for each sequence already identified rather than yielding more and novel sequences.

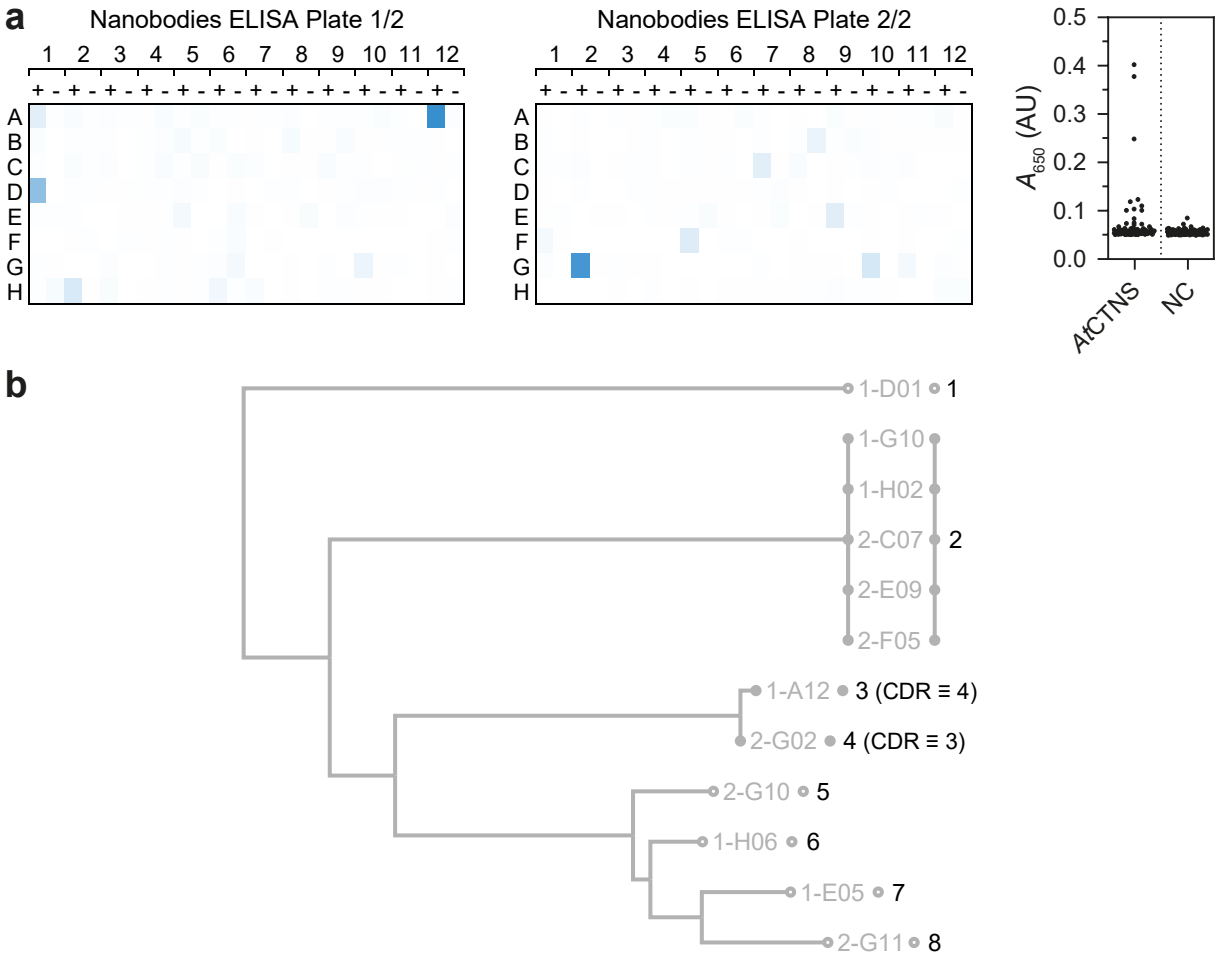


Fig. 5.10 – ELISA & Phylogenetic Analysis of *At*CTNS-Specific Nanobodies.
a, Gradient plot illustration of ELISA scores from **Table A.6**, visualizing nanobody binding to *At*CTNS-Avi-biotin (+) and PepT_{So}-Avi-biotin (NC, -), respectively. Scatter plot of the same data demonstrates the abundance of different absorbance intensities. **b**, Groups are formed using the full-length Nb sequences. Unique sequences are numbered from top to bottom. Branches of Nbs with a reoccurring CDR end in a solid bullet point. The original ELISA location is given in grey font (compare with **a**).

Among the few trends present in the Nb CDRs, many result from residues that are not expected to be target specific, further highlighting the diversity within the Nb sequences. Those include residues that are also present in the Syb CDRs, where they were left constant by design (Zimmermann *et al.*, 2018). Note that we defined the Nb CDRs in **Fig. 5.11** to stretch from one constant to the next constant residue (greyed), while the Syb CDRs in **Fig. 5.5** range from the first to the last randomized residue (constant residues in between still greyed), making the Nb CDRs seem longer than those of the Sybs (full sequences in **Table A.3**). The CDR1s of concave and loop Sybs end in MXW (where X is a randomized residue), those of convex Sybs in XYLGW, explaining the prevalence of MGW residues at the end of the Nb CDR1s. Syb CDR2s contain SXGXXT and TXXGXT, respectively, matching the widespread serine, glycine, and threonine residues in the Nb CDR2s. The CDR3s also contain similar residues towards their start and end. Nb1's CDR3 begins with CNV, identical to loop Sybs, while Nb4 and most other

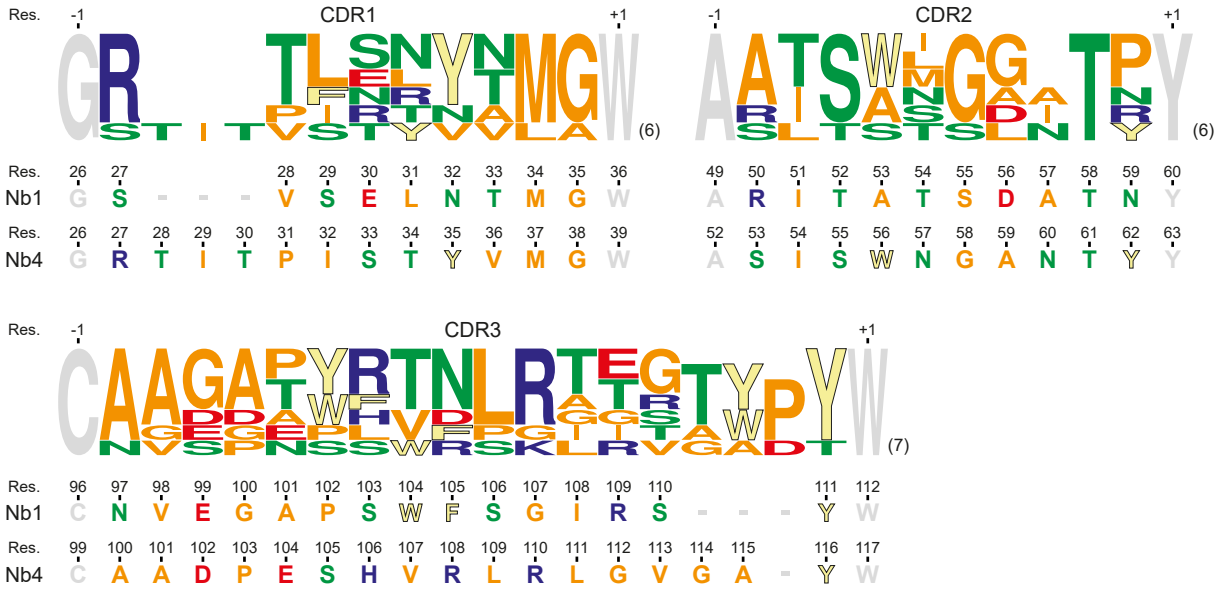


Fig. 5.11 – α-*At*CTNS Nanobody CDR Sequence Comparison.
Frequency plot logos using unique CDRs only (number of sequences shown in parentheses), generated using WebLogo (Crooks *et al.*, 2004). The CDR sequences of Nb1 and 4 are highlighted, and individual residues are numbered and coloured *via* amino acid chemistry. Greyed letters indicate constant backbone residues.

Nbs use CAA, constant residues of convex Sybs. Consistent with the varying CDR3 lengths, this suggests that Nb1 may be more structurally related to loop Sybs, whilst the remaining Nbs, including Nb4, may be more similar to convex Sybs. Either Syb type contains a tyrosine followed by the tryptophan, which is a constant at the end of the CDR3 of loop and convex Sybs, as well as Nbs. The level of CDR diversity observed between Nb1 and Nb4 may indicate that they interact with different *At*CTNS epitopes.

Overall, the Nb CDRs contain a higher percentage of polar neutral (coloured green) and, more strikingly, nonpolar aliphatic (orange) residues as compared to the Syb sequences (compare **Figs. 5.5** and **5.11**). Especially leucine, isoleucine, valine, and glycine are used much more frequently. They replace aromatic (yellow) and charged amino acids (blue and red), which are featured less often compared to the Syb CDRs, most noticeably tryptophans, tyrosines, and glutamates.

5.3.2 Biochemical & Biophysical Analyses of *At*CTNS-Nanobody Interactions

To analyse the specific binding towards *At*CTNS, His-Myc-tagged versions of the Nbs were expressed on a small scale (50 mL cultures). The pADL-23c vector enables the production of such constructs without the need for prior subcloning (Pardon *et al.*, 2014). Here, the sequences of the tagged Nbs are connected to the pIII gene, an integral component in the assembly of

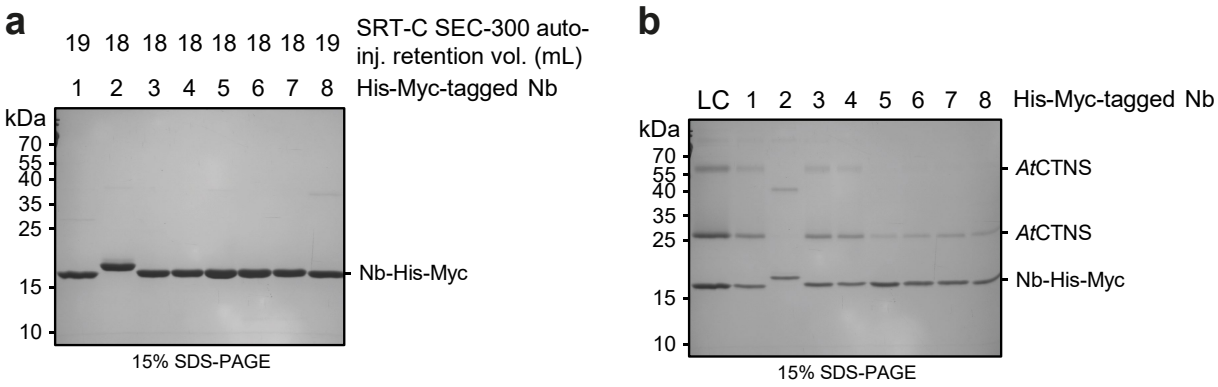


Fig. 5.12 – Purification of His-Tagged Nanobodies & *At*CTNS Pulldowns.
a, Final SDS-PAGE analysis of tagged Nbs, purified from 50 mL bacterial cultures. Corresponding SEC traces are shown in **Fig. A.3**. **b**, Pulldowns, performed at pH 7.5, were analysed by SDS-PAGE. LC = loading control.

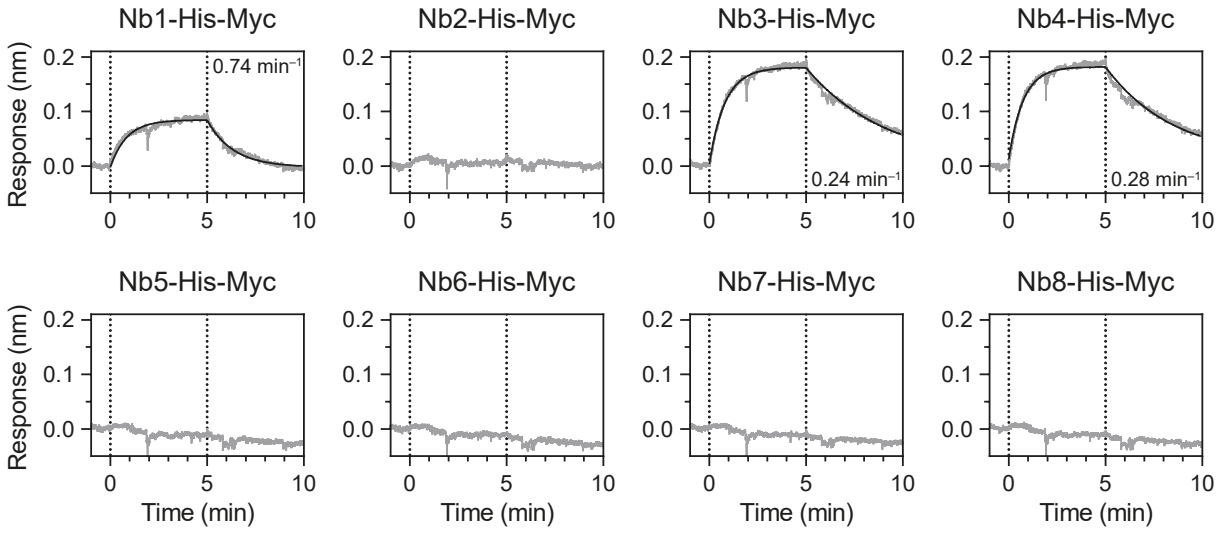


Fig. 5.13 – *At*CTNS-Nanobody Binding Analysis Using BLI.
BLI experiments performed at pH 6, showing association and dissociation steps. Traces were fitted using GraphPad Prism. Where appropriate, resulting k_{off}s are printed at the top or bottom of the dissociation step. *At*CTNS-Avi-biotin was immobilized at 100 nM, and nanobodies were bound at 100 nM.

M13 bacteriophages, *via* the rare amber stop codon using bases TAG. Expression in amber-suppressant bacterial cells, such as TG1, suppresses the amber stop codon and yields the fusion protein of tagged Nb and pIII, which is used during the phage displays. Non-suppressant bacterial strains, such as WK6, on the other hand, terminate translation at the amber stop codon.

The purified Nbs (**Fig. 5.12 a**, with SEC peak volumes annotated; traces shown in **Fig. A.3**) were used to capture *At*CTNS in a pulldown experiment (**Fig. 5.12 b**), highlighting Nbs 1, 3, and 4 as more potent binders compared to the other Nbs tested. These results were confirmed through BLI (**Fig. 5.13**), with the tagged Nbs 3 and 4 showing the slowest k_{off} rates of any identified

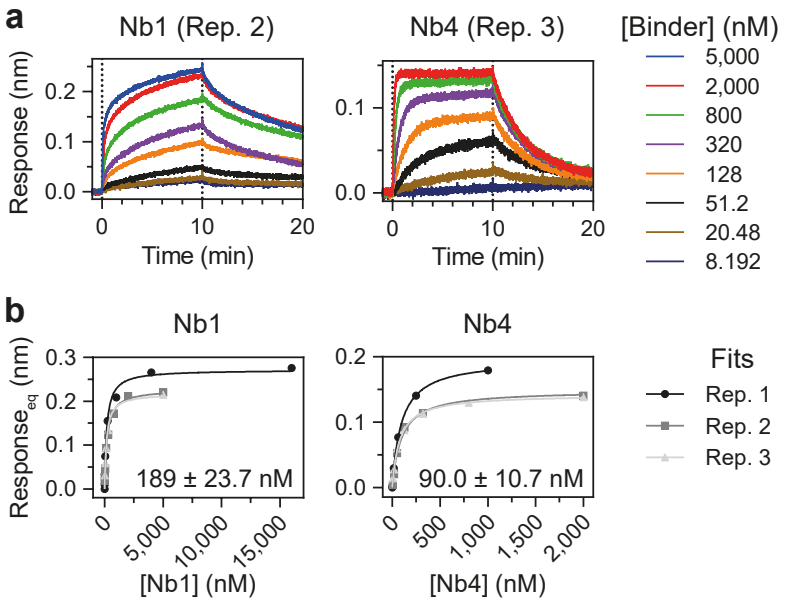


Fig. 5.14 – K_D Analysis of Important Nanobodies Using BLI.
a, BLI traces of Nb1 replicate 2 and Nb4 replicate 3 are shown, highlighting association and dissociation steps using various binder concentrations. *At*CTNS-Avi-biotin was immobilized at 20 nM. **b**, K_D analysis of multiple replicates using the response at equilibrium (from **a**), as calculated using the integrated analysis tool of the Octet RED384 system. Data were fitted using GraphPad Prism. K_D values are shown in the bottom right corner of each plot.

binder at 0.24 and 0.28 min^{-1}, respectively. Nb1-His-Myc also featured a defined association peak, coupled with a k_{off} of 0.74 min^{-1}, however, all other Nbs did not show association peaks at 100 nM Nb concentration using BLI and were therefore not considered for structural and functional studies.

In addition to Sybs 1 and 39, Nbs 1 and 4 also emerged as integral puzzle pieces to help build well-diffracting crystals and solve the structure of *At*CTNS – see **Chs. 6** and **7**. I note that Nb4 was preferred over Nb3 for no particular reason, as they share identical CDRs. To better understand the binding characteristics of Nbs 1 and 4, we performed K_D analyses for each in triplicate after re-purifying tag-free versions of the binders. We found that the *At*CTNS-Nb interactions are defined by lower K_D values compared to the *At*CTNS-Syb interactions described earlier, ~190 and 90 nM for Nbs 1 and 4, respectively, indicating that our nanobodies are higher-affinity binders (**Fig. 5.14**). The superior binding behaviour compared to the Sybs is also reflected in the slowed dissociation trends apparent in these experiments. The k_{off} for Nb4 across all replicates was calculated at 0.19 ± 0.034 min^{-1}, in line with the prior estimation. In the case of Nb1, the dissociation behaviour has changed dramatically after removal of the His-Myc-tag, decreasing the k_{off} almost 10-fold to 0.081 ± 0.012 min^{-1}.

CHAPTER 6

Crystal Optimization Using Sybody- & Nanobody-Bound *At*CTNS Complexes

In an attempt to optimize the native *At*CTNS crystals from **Ch. 4**, complexes were formed using the sybodies and nanobodies with the slowest k_{off} rates (see **Ch. 5**) and subsequently co-crystallized in the LCP. In this chapter, I will describe the recombinant production of tag-free versions of these binders, the formation of complexes with *At*CTNS, and the individual LCP crystals that were formed using these complexes. I will demonstrate that the use of crystallization chaperones enabled us to grow more voluminous crystals. Furthermore, I will analyse the X-ray diffraction data obtained from relevant crystals and compare the data quality. In summary, the three-dimensionality of the crystals was improved successively: Co-crystallization with Syb39 yielded thin and still almost one-dimensional rods; using Nb4 resulted in fine but two-dimensional plates, as well as small bipyramidal crystals; and finally, the combination of both Syb39 and Nb4 co-crystallized with *At*CTNS as a triple complex gave large, cuboidal crystals. Data quality improved with crystal volume, however, none of the datasets enabled the structure solution for *At*CTNS using a sybody/nanobody frame as the search model for MR. The structure was ultimately solved experimentally using the *At*CTNS-Syb39-Nb4 crystals owing to their size and reproducibility. Solution and resulting structures are the topic of **Ch. 7**.

6.1 Co-Crystal Trials Using Sybodies

6.1.1 Purification of Tag-Free Sybodies

Tag-free versions of the most promising sybodies, Sybs 1, 39, and 41 (**Sec. 5.2.2**), were expressed using multiple litres of bacterial cell culture (see **Sec. 2.2.8**). Expression from the pBXNPHM3 vector initially generates sybodies with an N-terminal His-MBP tag. The addition of the tag is supposed to improve the solubility of the fusion compared to Syb alone (Kapust & Waugh, 1999) and simplify purification by enabling MBP-specific affinity chromatography (AC) steps using amylose resin (di Guan *et al.*, 1988). Initially, both IMAC and MBP-AC steps were performed consecutively to capture the fusion protein and clear it from other substances present in the cell lysis. O/N dialysis was then performed to remove the D-(+)-maltose used to elute the MBP

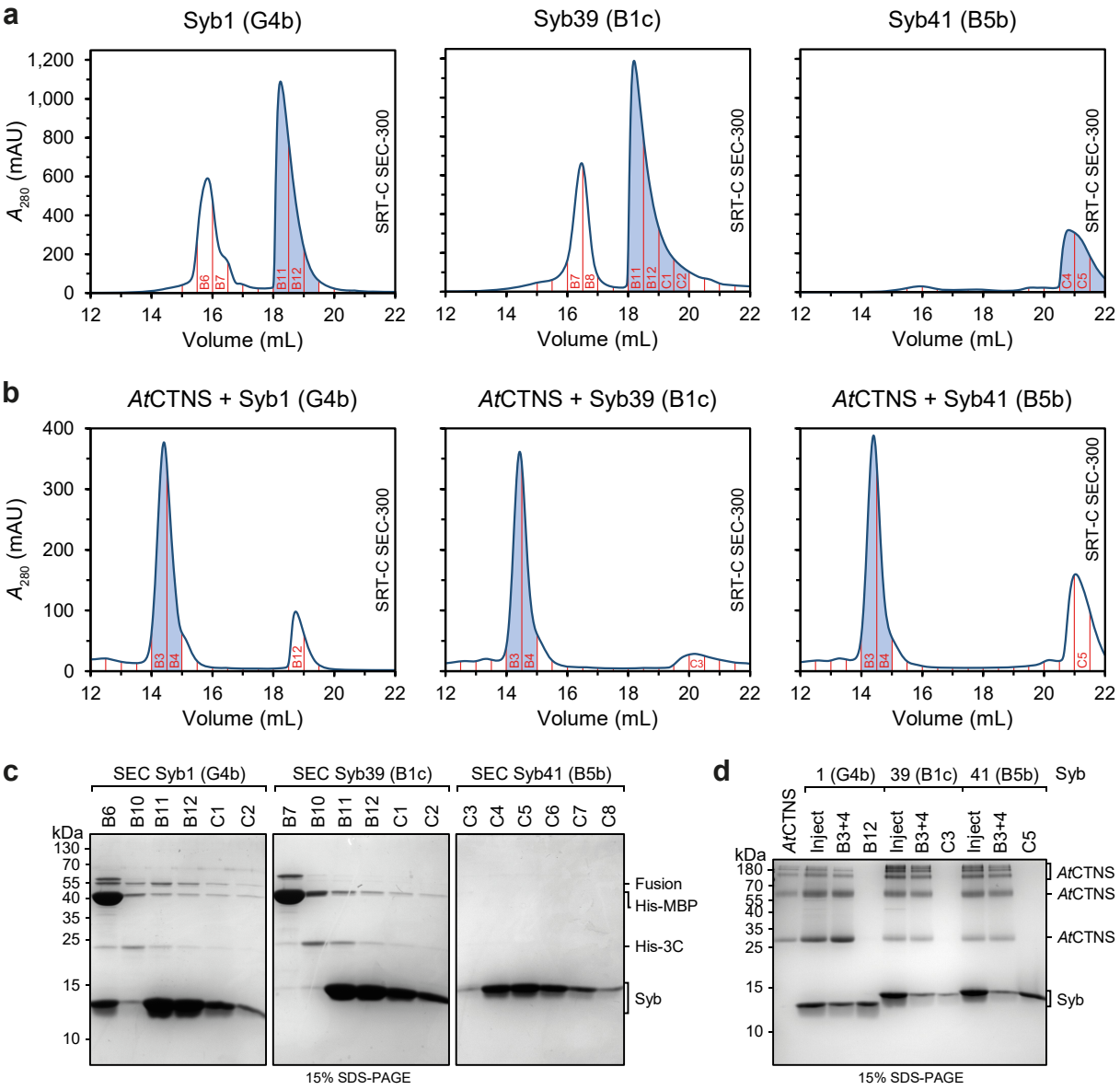


Fig. 6.1 – *At*CTNS-Sybody Complex Formation for Co-Crystallization.
a, Purification SEC traces of sybodies used for co-crystallization. Pooled fractions (shaded in blue) were combined and used to form complexes with purified *At*CTNS. **b**, SEC co-elution traces of complexes. Pooled fractions were used for LCP co-crystallization. **c**, SEC fractions of **a** analysed by SDS-PAGE. **d**, SEC fractions of **b** analysed by SDS-PAGE.

fusion from the resin, cleaving the tags with 3C along the way. Surprisingly, complete tag removal from the protein solution proved difficult in the next step, despite using both reverse MBP-AC and IMAC steps, resulting in minor His-MBP and His-3C contamination in the final sample eluted from the SEC column (see Sybs 1 and 39 in **Fig. 6.1 a** and **c**). This was likely caused by the incomplete dissociation of D-(+)-maltose from MBP during dialysis, causing the reverse MBP-AC to fail. To improve the purity in future experiments, the fusion protein was captured solely using IMAC. After dialysis and cleavage, the protein solution was then passed through the amylose resin to significantly reduce the amount of His-MBP present in

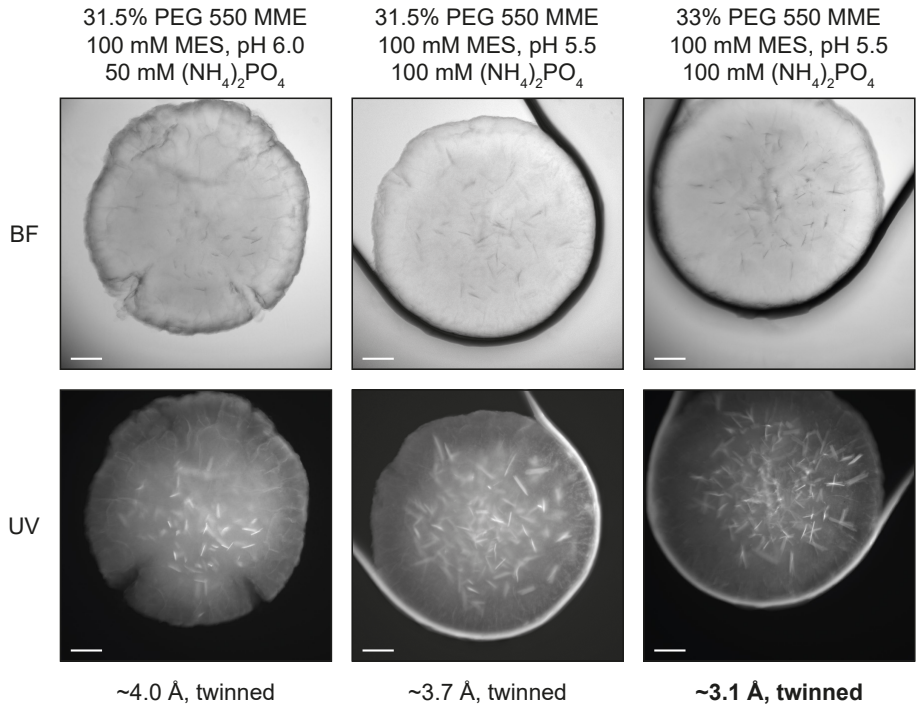


Fig. 6.2 – LCP Crystals of *At*CTNS in Complex with Syb39.
BF and UV images of LCP boli containing *At*CTNS-Syb39 crystals are shown. Complexes were co-crystallized in 10:1 (w:w) MO:CHL using protein at 26.4 mg mL^{-1}. Annotated at the top of each column of images is the precipitant solution. The diffraction outcome is pinned at the bottom. Scale bars 100 µm.

the FT. Subsequent reverse IMAC was performed to remove His-3C as well as any remaining His-MBP. The optimized procedure, resulting in an improved purity, as shown later for Nbs 1 and 4 (**Fig. 6.6 a** and **b**), is described in detail in **Sec. 2.2.9**. The overall yields for Syb1, 39, and 41 were calculated at ~1.5, ~1.5, and ~0.3 mg L^{-1} bacterial cell culture, respectively. Note that the reduced yield of Syb41 was the likely reason it was purified to a higher purity in **Fig. 6.1 a** and **c**.

6.1.2 Complex Formation & Co-Crystallization

Individual complexes were formed using purified *At*CTNS and a 20% molar excess of each sybody and isolated from unbound sybody by SEC (**Fig. 6.1 b**). The addition of a Syb did not alter the shape of the *At*CTNS peak observed during SEC using the Sepax SRT-C SEC-300 column, consistent with the SEC co-elutions that were performed to analyse *At*CTNS-Syb interactions for the tagged binders (**Fig. A.2**). Furthermore, the peak was not shifted in either direction by a significant margin and remained at 14.5 ± 0.2 mL elution volume. The presence of Syb in the *At*CTNS peak, *i.e.* the successful formation of a complex, was therefore proven by SDS-PAGE (**Fig. 6.1 d**). The indicated fractions were pooled, combined, and concentrated to ~20–30 mg mL^{-1}. LCP co-crystallization was performed using both MO and MO:CHL (10:1). Complexes using Syb1 and 41 did not crystallize in this format. *At*CTNS-Syb39 nucleated

Table 6.1 – X-Ray Data Collection & Processing Statistics of *At*CTNS-Syb39.

a, Relevant sweep of data collected at DLS (proposal number mx23459-103).

Identifier	Beamline	Wavelength (Å)	Beamsize (μm)	Transmission (%)	Exposure (s)	Oscillation (°)	Total images	Resolution max. (Å)	Detector
6553_2	DLS - I24	0.9999	7 × 7	15.62	0.0500	0.50	360	2.00	PILATUS3 6M

b, Data processing.

Pipeline	Images used	High-resolution data cut	Space group	Cell dimensions *a*, *b*, *c* (Å)	Cell angles α, β, γ (°)	PDB
xia2 DIALS	2_1–360	Automatic	*C*2	286.83, 64.13, 56.09	90.00, 102.30, 90.00	----

c, Merging statistics by resolution bin and overall. *Significant at $p = 0.01$.

Resolution (Å)	Observations	Unique reflections	Multiplicity	Completeness (%)	Mean $I/\sigma(I)$	R_{merge} (%)	R_{meas} (%)	R_{pim} (%)	$CC_{1/2}$ (%)
62.52–8.44	2,968	967	3.07	99.8	9.1	14.4	18.2	10.8	92.8*
8.44–6.70	2,849	920	3.10	99.2	4.1	16.4	19.9	11.0	95.7*
6.70–5.85	2,894	915	3.16	100	2.3	24.6	29.8	16.5	93.9*
5.85–5.32	2,959	912	3.24	100	2.1	26.3	31.5	17.2	97.1*
5.32–4.94	2,962	916	3.23	100	2.7	22.5	27.1	14.8	95.5*
4.94–4.65	2,981	909	3.28	100	2.8	22.7	27.2	14.8	94.3*
4.65–4.41	3,000	918	3.27	99.8	2.7	24.3	29.1	15.8	94.2*
4.41–4.22	2,968	897	3.31	100	2.2	29.0	34.6	18.6	86.0*
4.22–4.06	2,934	889	3.30	100	1.8	35.6	42.5	22.9	84.8*
4.06–3.92	3,093	933	3.32	100	1.6	41.2	49.2	26.6	78.0*
3.92–3.80	3,033	901	3.37	100	1.3	48.9	58.0	31.1	72.0*
3.80–3.69	2,921	890	3.28	100	1.0	57.7	68.8	37.2	63.3*
3.69–3.59	2,958	893	3.31	100	0.9	65.8	78.4	42.2	61.7*
3.59–3.50	3,001	906	3.31	100	0.9	75.6	90.3	49.0	65.0*
3.50–3.42	3,079	913	3.37	99.9	0.8	83.9	99.9	53.8	52.5*
3.42–3.35	3,021	903	3.35	100	0.7	96.6	115.1	62.1	48.1*
3.35–3.28	2,729	888	3.07	100	0.5	107.2	129.3	71.5	31.4*
3.28–3.22	2,884	913	3.16	99.4	0.5	142.1	171.6	95.1	12.4*
3.22–3.16	2,789	885	3.15	99.9	0.4	132.0	159.0	87.6	17.4*
3.16–3.11	2,896	901	3.21	100	0.4	190.8	229.2	125.7	10.9*
62.51–3.11	**58,919**	**18,169**	**3.2**	**100**	**2.0**	**34.9**	**42.0**	**23.1**	**84.3**

frequently and formed thin rod-shaped crystals appearing in PEG 550 monomethyl ether (MME) at slightly acidic conditions and in the presence of ammonium phosphate. Precipitant solutions were optimized around this condition, which led to the formation of crystals up to ~60–80 μm in length (**Fig. 6.2**).

6.1.3 Data Collection, Processing & Analysis

Many of the *At*CTNS-Syb39 crystals (**Fig. 6.2**) diffracted to a resolution of ~3–4 Å, similar to the range achieved with previous *At*CTNS crystals (**Ch. 4**), however, demonstrating a level of consistency and reproducibility that was not achieved before. As a result, numerous sweeps of diffraction images were collected and successfully processed, primarily using xia2 (Winter, 2010) within the DIALS (Winter *et al.*, 2018) pipeline. Data collection and processing statistics for the best dataset, reaching a maximum resolution of 3.11 Å, are listed in **Table 6.1**, with typical

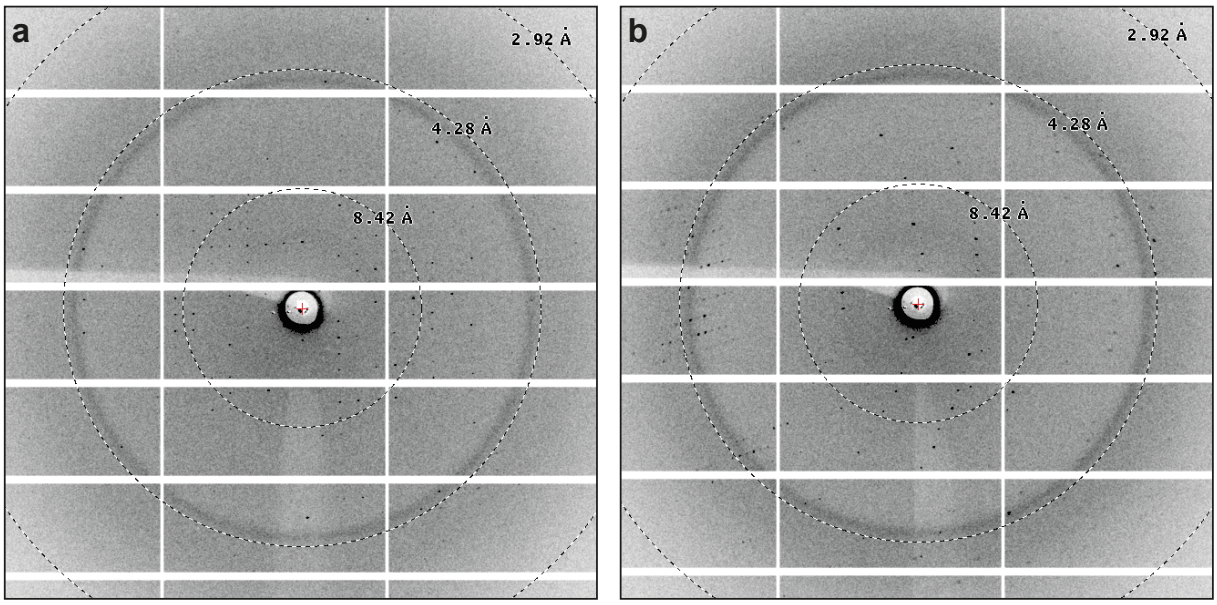


Fig. 6.3 – Typical Diffraction Patterns of *At*CTNS-Syb39.
Diffraction patterns obtained at $\varphi = 0°$ (**a**) and $\varphi = 90°$ (**b**), visualized using Adxv (Arvai, 2021).

diffraction images of the corresponding sweep shown in **Fig. 6.3**. The data were processed in monoclinic space group $C2$ with a unit cell resembling a parallelepiped with one long and two short axes ($a = 286.83$ Å, $b = 64.13$ Å, and $c = 56.09$ Å), and angles $\alpha = \gamma = 90°$ and $\beta = 102.30°$ (**Table 6.1 b**). Based on a V_M of 2.71 Å^3 Da^{-1}, the asymmetric unit most likely contained two heterodimers of *At*CTNS in complex with Syb39.

Despite the increased maximum resolution for the best dataset, the overall data quality decreased slightly compared to the best data collected from single-component crystals (compare **Tables 4.1** and **6.1**). The main statistic affected was the comparatively low signal-to-noise ratio in every resolution shell, which led to an overall mean $I/\sigma(I)$ of 2.0 across all bins (**Table 6.1 c**). We tried to optimize the reflection intensities by increasing the radiation dose, however, the crystals were found to be damaged easily when using a stronger beam (transmission >25% at I24), manifested by dropping $\langle I/\sigma(I)\rangle$ values and rising R-factors over the course of the data collection, and overall leading to incomplete datasets, especially in higher-resolution shells. The dataset presented in **Table 6.1** was collected using a minimal radiation dose, achieved by setting the transmission to 15.62% and limiting the data to only 180°, with a high rotation increment of 0.5° and an exposure time of 0.05 s per image. R_{meas} and R_{pim} values were found to exceed the optimal range in the lowest-resolution shell at 18.2% and 10.8%, respectively, but bear a resemblance to the previous dataset in mid-range shells up to ~4.5 Å. In higher-resolution bins up to ~3.5 Å, they improved in comparison. Additional shells, ultimately reaching 3.11 Å, were included based on the $CC_{1/2}$, despite high R-factors and a low $\langle I/\sigma(I)\rangle$ for the reasons mentioned earlier (**Sec. 4.4**).

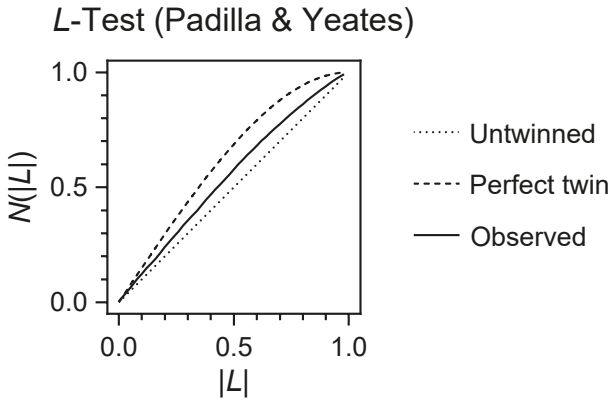


Fig. 6.4 – Twinning Detection Using the *L*-Test.
The *L*-test is demonstrated for the *At*CTNS-Syb39 data. *L* is defined as the local differences between two pairs of reflections, *h*1 and *h*2, which are close in reciprocal space but not related by a twin law. It is defined as $|L| = |(I_{h1} - I_{h2}) / (I_{h1} + I_{h2})|$ and plotted against its cumulative probability distribution $N(|L|)$. Untwinned crystals with twin fraction $\alpha = 0.0$ follow the dotted line, perfect twins with $\alpha = 0.5$ approach the dashed line, and partial twins lie in between (Padilla & Yeates, 2003).

Data processing also revealed that the *At*CTNS-Syb39 crystals were often twinned, meaning they comprised multiple building blocks of the same species aligned unequally (Parsons, 2003). Diffraction images show an overlay of different orientations of the unique diffraction pattern, correlated through an additional symmetry operator referred to as the twin law. Twinned crystals, therefore, have higher symmetry than their building blocks, explaining why the data processing initially suggested space group *I*222 instead of *C*2. Here, the monoclinic crystal lattice resembled an orthorhombic system, thus, we were dealing with pseudo-merohedral twinning. Several tests are automatically performed during data processing to identify possible twins (Evans, 2011), including the *L*-test (Padilla & Yeates, 2003), which is displayed in **Fig. 6.4**. For the presented *At*CTNS-Syb39 dataset, the twin fraction α was estimated at 0.236–0.277 and twin law $-h-2l, -k, l$ was assigned.

6.1.4 Structure Solution Attempt by Molecular Replacement

The crystal structure of the KDEL receptor (KDELR) in complex with a sybody was solved by MR using a sybody frame (CDRs removed) as the search model (Bräuer *et al.*, 2019). Additional electron density was observed, revealing α-helical content. This allowed for the placement of a poly-alanine model, and upon refinement of the electron density maps, additional density was observed, which led to the addition of side chains, ultimately solving the structure. Attempting to repeat this for *At*CTNS, we performed MR in Phaser (McCoy *et al.*, 2007) using the frame of the α-*Mm*KDELR1 sybody (Bräuer *et al.*, 2019), a loop sybody similar to Syb39, as the search model. Two sybody molecules were placed in the asymmetric unit, indicated by a TFZ of 6.1 and 15.4, as well as an LLG of 114 and 317 for the first and second placement, respectively. The ensuing electron density map covered the sybodies well (**Fig. 6.5**), however, clear density

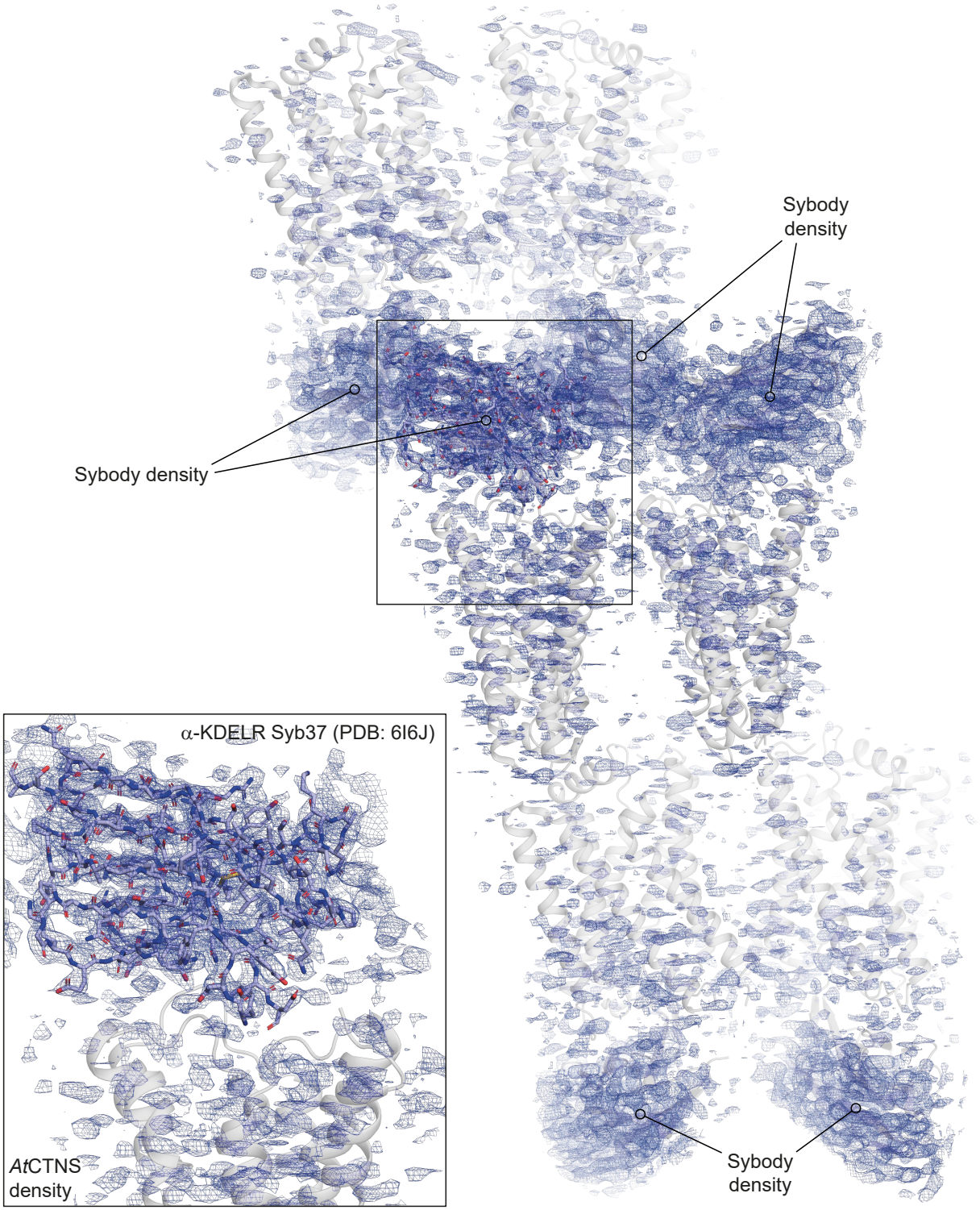


Fig. 6.5 – Electron Density After MR Using a Sybody Frame.
$2mF_o$–DF_c electron density map obtained after MR in Phaser with 10 subsequent rounds of refinement in REFMAC. The map is contoured at 1.2 σ with a carve of 5 Å, applied to six sybodies (*i.e.* the equivalent of three asymmetric units) and the position of six molecules of *At*CTNS, obtained by superimposing the sybodies with the finalized L-cystine-bound structure (PDB: 7ZKW, shown in **Fig. 7.16**), which crystallized in the same space group with identical unit cell parameters.

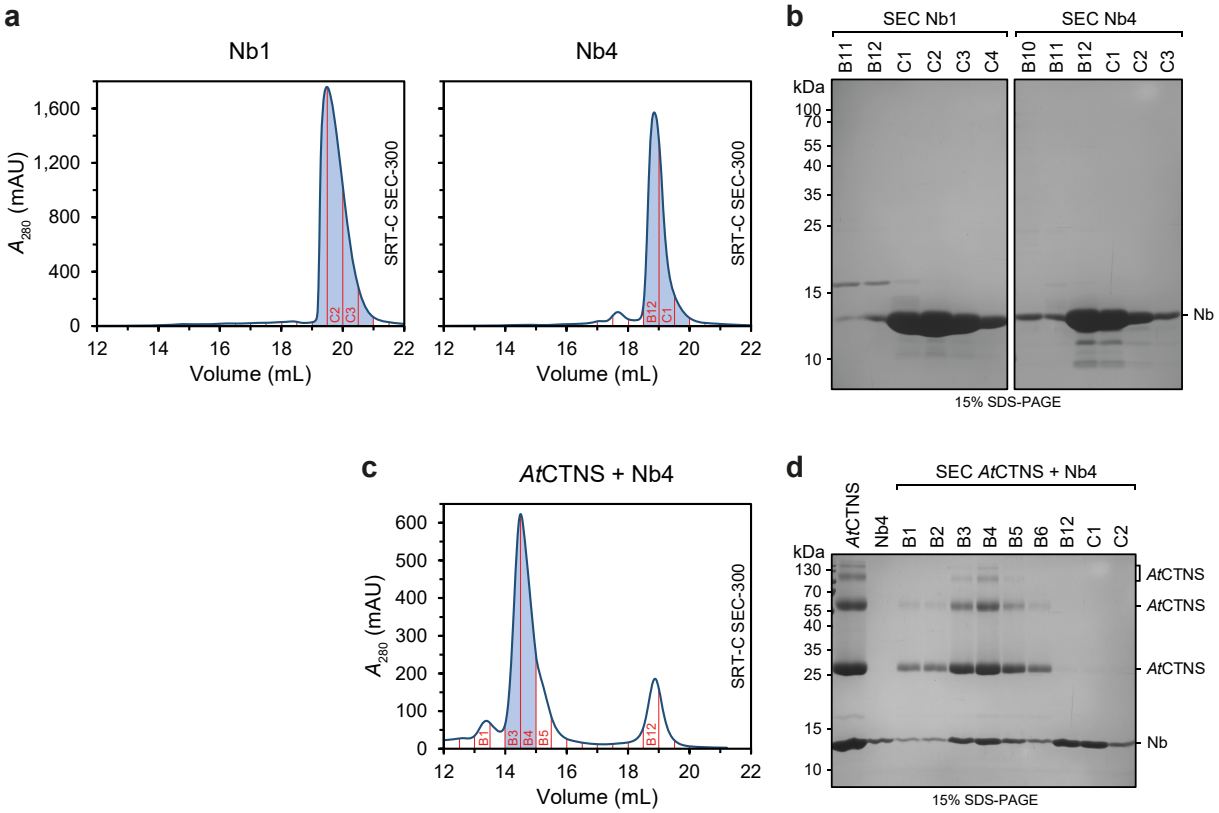


Fig. 6.6 – *At*CTNS-Nanobody Complex Formation for Co-Crystallization.
a, Purification SEC traces of nanobodies used for co-crystallization. Pooled fractions (shaded in blue) were combined and used to form complexes with purified *At*CTNS. **b**, SEC fractions of **a** analysed by SDS-PAGE. **c**, SEC co-elution traces of an *At*CTNS-Nb4 complex. Pooled fractions were used for LCP co-crystallization. **d**, SEC fractions of **c** analysed by SDS-PAGE.

outside of the placed models, which may suggest α-helical content, was not observed. With a molecular weight of ~13.6 kDa (~11.0 kDa without CDR residues), Syb39 only accounts for ~30% (~24%) of the overall weight of the complex (MW_{AtCTNS} = 31.9 kDa). For a complete structure solution by MR, *i.e.* the assignment of good approximations for phases of all reflections, the search model should cover at least 50% of the crystal contents (Abergel, 2013). As this was not achieved, and since the homology models of *At*CTNS did not yield a correct solution by MR, we were unable to solve the structure at this stage and continued to screen for more suitable crystals using our nanobodies.

6.2 Co-Crystal Trials Using Nanobodies

6.2.1 Nanobody Production & Complex Formation

Similarly to the sybodies, we also considered the most promising nanobodies, Nbs 1 and 4 (see **Sec. 5.3.2**; note that Nb3 is identical to Nb4 in its CDR, as shown in **Table A.6**), to serve as crystallization chaperones for *At*CTNS. The nanobody genes were amplified by PCR,

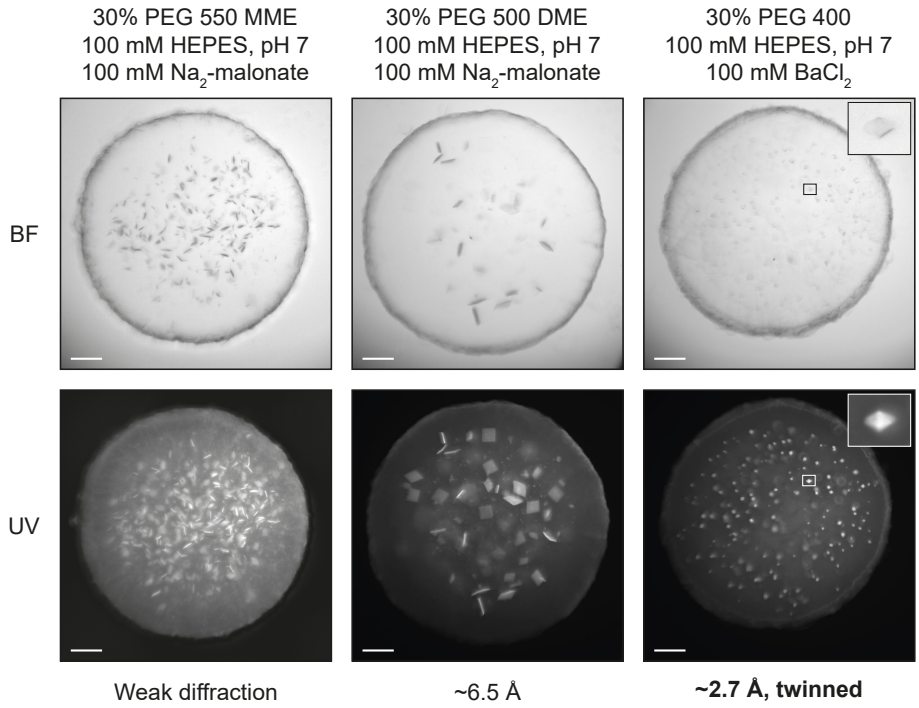


Fig. 6.7 – LCP Crystals of *At*CTNS in Complex with Nb4.
BF and UV images of LCP boli containing *At*CTNS-Nb4 crystals are shown. Complexes were co-crystallized in MO using protein at 21.5 mg mL^{-1}. Annotated at the top of each column of images is the precipitant solution. The diffraction outcome is pinned at the bottom. Scale bars 100 µm.

adding BspQI restriction sites during the process to enable FX cloning into pBXNPHM3 for the production of tag-free binders. Using the optimized protocol from the Syb large-scale purification (**Sec. 6.1.1**), both Nbs were purified to homogeneity, with only minor degradation impurities observed during SDS-PAGE analysis (**Fig. 6.6 a** and **b**). The overall yield, however, remained surprisingly low and was typically found to be within 1.5–2 mg L^{-1} bacterial cell culture. Both Nbs were then used to assemble individual complexes with *At*CTNS, co-eluting from the Sepax SRT-C SEC-300 column, as shown for Nb4 in **Fig. 6.6 c** and **d**, and used for LCP co-crystallization.

6.2.2 Co-Crystallization & Data Analysis

*At*CTNS did not co-crystallize after complexation with Nb1 using the precipitant solutions available to us (**Sec. 2.5.1**). Using Nb4, however, led to the formation of many crystals of different morphologies (**Fig. 6.7**). Unlike before, these would form predominantly at neutral pH, but still in the presence of higher MW PEGs. *At*CTNS-Nb4 crystals were generally found to be more three-dimensional than the needles and rods obtained before (**Figs. 4.1** and **6.2**, respectively). Despite this, most crystals did not generate visible reflections beyond ~6.5 Å on the detector, except for a few smaller crystals of bipyramidal shape, highlighted in the last column of **Fig. 6.7**, which at best diffracted to ~2.7 Å (**Table 6.2** and **Fig. 6.8**).

Table 6.2 – X-Ray Data Collection & Processing Statistics of *At*CTNS-Nb4.

a, Relevant sweep of data collected at DLS (proposal number mx23459-111).

Identifier	Beamline	Wavelength (Å)	Beamsize (μm)	Transmission (%)	Exposure (s)	Oscillation (°)	Total images	Resolution max. (Å)	Detector
6711_2	DLS - I24	0.9797	7 × 7	18.64	0.0500	0.50	360	2.50	PILATUS3 6M

b, Data processing.

Pipeline	Images used	High-resolution data cut	Space group	Cell dimensions *a*, *b*, *c* (Å)	Cell angles α, β, γ (°)	PDB
autoPROC	2_1–360	Based on $CC_{1/2}$	$P6_3$	80.01, 80.01, 150.58	90.00, 90.00, 120.00	----

c, Merging statistics by resolution bin and overall. *Significant at $p = 0.01$.

Resolution (Å)	Observations	Unique reflections	Multiplicity	Completeness (%)	Mean $I/\sigma(I)$	R_{merge} (%)	R_{meas} (%)	R_{pim} (%)	$CC_{1/2}$ (%)
19.83–7.27	5,721	745	7.68	100	13.8	10.5	11.3	4.1	99.6*
7.27–5.82	6,164	748	8.24	100	9.4	21.2	22.6	7.9	98.2*
5.82–5.10	6,071	731	8.31	100	8.9	22.4	23.9	8.3	97.6*
5.10–4.64	6,048	729	8.30	100	9.9	19.5	20.8	7.3	98.3*
4.64–4.31	6,111	731	8.36	100	9.5	20.2	21.6	7.5	98.6*
4.31–4.06	5,901	731	8.07	100	8.7	23.0	24.6	8.6	97.7*
4.06–3.86	6,131	730	8.40	100	7.0	29.8	31.7	10.9	97.3*
3.86–3.69	6,309	742	8.50	100	6.1	35.8	38.1	13.0	94.8*
3.69–3.55	6,005	731	8.22	100	5.0	42.8	45.6	15.8	93.2*
3.55–3.43	6,057	722	8.39	100	4.4	52.2	55.7	19.2	90.0*
3.43–3.32	6,105	726	8.41	100	3.4	67.6	72.1	24.7	85.3*
3.32–3.23	6,133	725	8.46	100	3.1	73.3	78.0	26.7	84.8*
3.23–3.14	6,268	741	8.46	99.9	2.4	96.5	102.8	35.2	76.2*
3.14–3.07	6,094	721	8.45	100	2.3	101.8	108.4	37.1	73.1*
3.07–3.00	6,222	730	8.52	100	1.8	132.7	141.2	48.1	67.9*
3.00–2.93	6,218	730	8.52	100	1.6	151.9	161.7	55.2	60.4*
2.93–2.87	6,305	743	8.49	100	1.3	179.6	191.2	65.3	48.7*
2.87–2.82	6,097	720	8.47	100	1.2	200.4	213.4	73.0	52.7*
2.82–2.77	6,111	715	8.55	100	1.1	214.2	227.9	77.5	43.1*
2.77–2.72	6,394	748	8.55	100	0.9	276.4	294.1	100.1	40.1*
19.83–2.72	**122,465**	**14,639**	**8.4**	**100**	**5.1**	**35.9**	**38.3**	**13.2**	**98.6**

Processing of the data collected from the bipyramidal crystals indicated both hexagonal space groups $P6_3$ and $P6_322$ as possible crystal systems, with cell dimensions $a = b = 80.01$ Å and $c = 150.58$ Å, and angles $\alpha = \beta = 90°$ and $\gamma = 120°$ (**Table 6.2 b**). Based on unusual intensity statistics in both space groups, $P6_3$ was likely the correct space group, and the additional twofold symmetries were a result of merohedral twinning. Specifically, the *L*-test suggested a twin fraction of 0.15 and twin law $h, -h-k, -l$ for the present data processed in $P6_3$.

The merging statistics obtained for the best *At*CTNS-Nb4 dataset (**Table 6.2 c**) indicated that the data quality improved compared to the *At*CTNS-Syb39 data (**Table 6.1 c**) and was similar to the native *At*CTNS data (**Table 4.1**). R_{merge} and R_{meas} were close to 10% in the lowest-resolution bin; R_{pim} was considerably lower (4.1%) due to the high multiplicity resulting from collecting 180° of reflection data arranged in a hexagonal lattice. The mean $I/\sigma(I)$ over all resolution

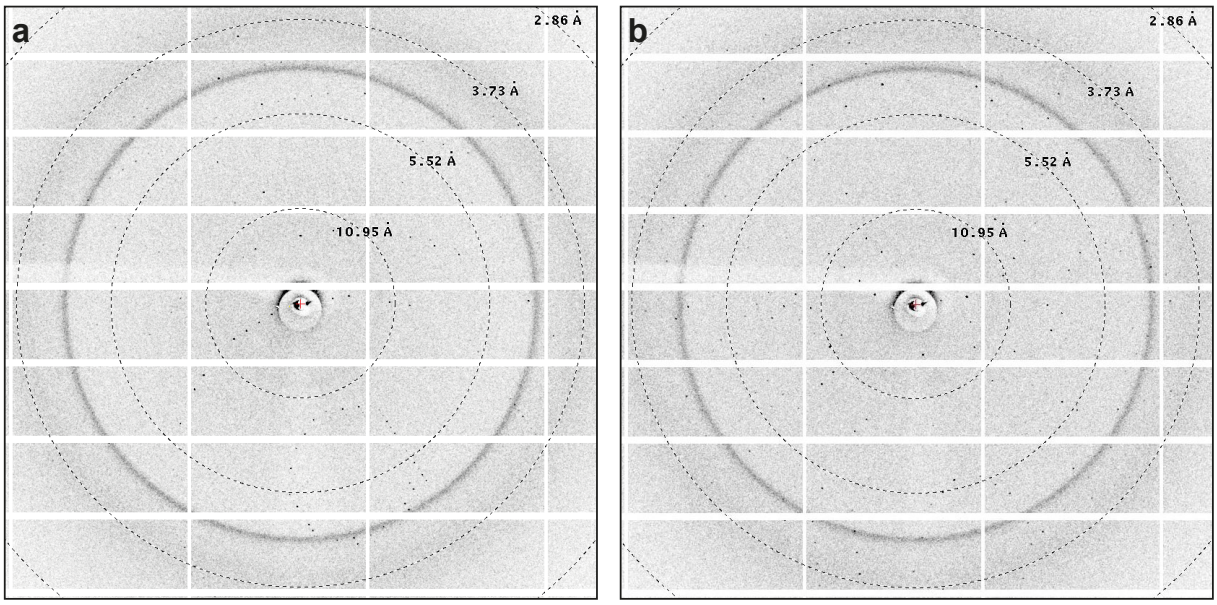


Fig. 6.8 – Best Diffraction Patterns of *At*CTNS-Nb4.
Diffraction patterns obtained at $\varphi = 0°$ (**a**) and $\varphi = 90°$ (**b**), visualized using Adxv (Arvai, 2021).

shells was 5.1, and the $CC_{1/2}$ ranged from 99.6–40.1% from low-resolution to high-resolution shells with an overall value of 98.6%.

6.2.3 Structure Solution Attempt by Molecular Replacement

Similar to the MR trials for the *At*CTNS-Syb39 data, we tried to use the frame of a homologous nanobody from the PDB to retrieve initial estimates for the lost phases. Unlike sybodies, nanobody frames are not identical by design. Here, we used α-GPCR Nb80 (PDB: 3P0G, **Fig. 1.6 c**), which, after removal of its CDRs, shares ~90% of its sequence with α-*At*CTNS Nb4 over 92 residues. With a V_M of 3.03 $Å^3\ Da^{-1}$, only a single copy of both Nb4 and the transporter were expected in the asymmetric unit. Placement of the truncated α-GPCR Nb80 using Phaser, however, only generated a TFZ of 4.5 and an LLG of 25, indicating the correct position for the nanobody in the unit cell could not be found. Yet again, we were unable to solve the crystal structure using MR. Furthermore, we were unable to re-grow these well-diffracting bipyramidal crystals, hindering us from attempting experimental phasing techniques to solve the structure.

6.3 Co-Crystal Trials Using Multiple Binders

6.3.1 Complex Formation & Co-Crystallization

Having so far been unable to solve the three-dimensional structure of *At*CTNS using individual sybodies and nanobodies, we sought to explore the feasibility of forming an *At*CTNS complex

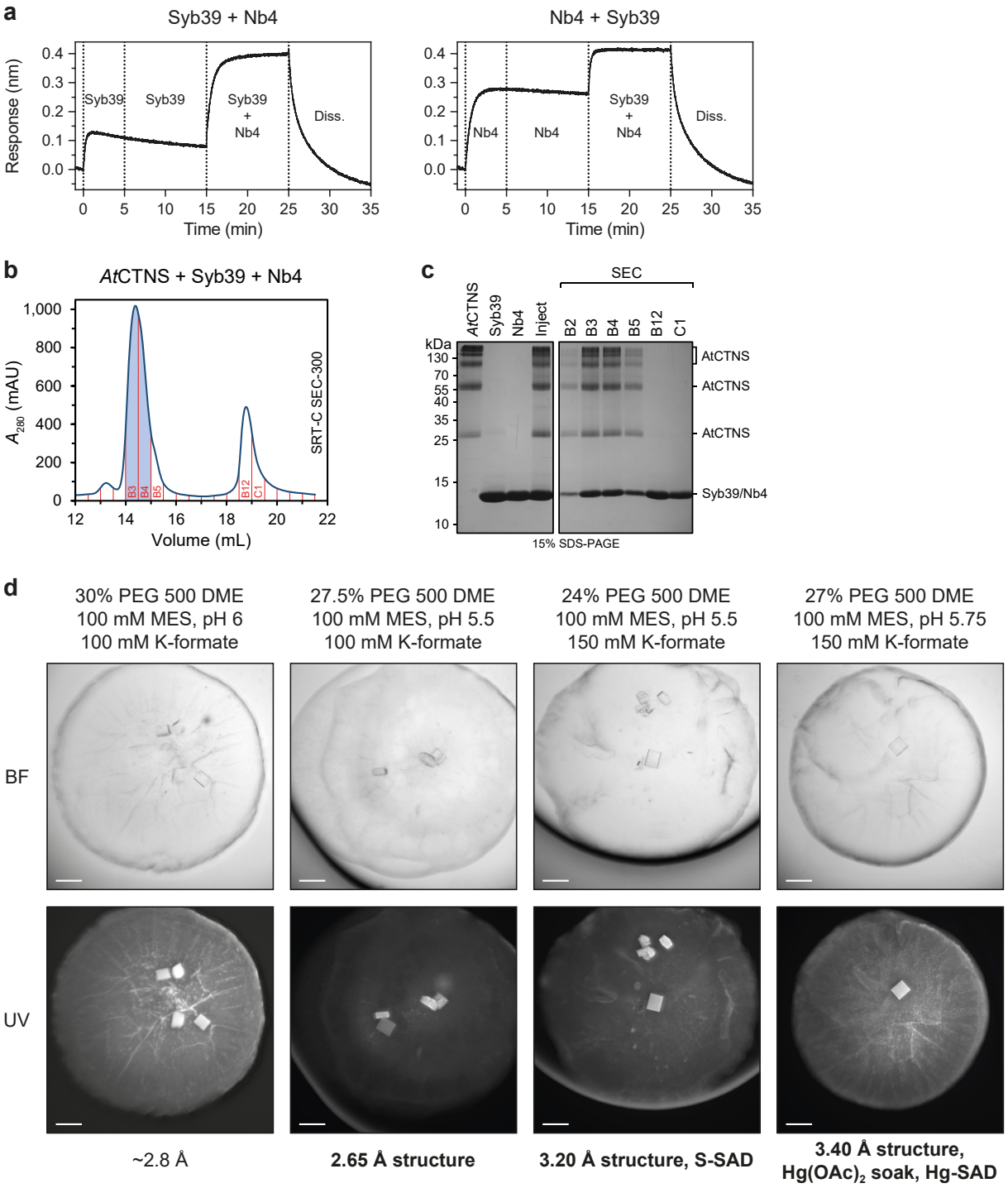


Fig. 6.9 – Complex Formation & Co-Crystallization of *At*CTNS-Syb39-Nb4.
a, BLI epitope binning of *At*CTNS using Syb39 and Nb4. Multiple association steps, using the binders individually or in combination, are shown. Binders are dissociated in the final step. *At*CTNS-Avi-biotin was immobilized at 100 nM, and Syb39 and Nb4 were bound at 400 nM. **b**, SEC co-elution trace of the *At*CTNS-Syb39-Nb4 complex. Pooled fractions were used for LCP co-crystallization. **c**, SEC fractions of **b** analysed by SDS-PAGE. **d**, BF and UV images of LCP boli containing *At*CTNS-Syb39-Nb4 crystals (all isomorphous) are shown. Complexes were co-crystallized in MO using protein at 30.5 mg mL^{-1}. Annotated at the top of each column of images is the precipitant solution. The diffraction outcome is pinned at the bottom. Scale bars 100 µm.

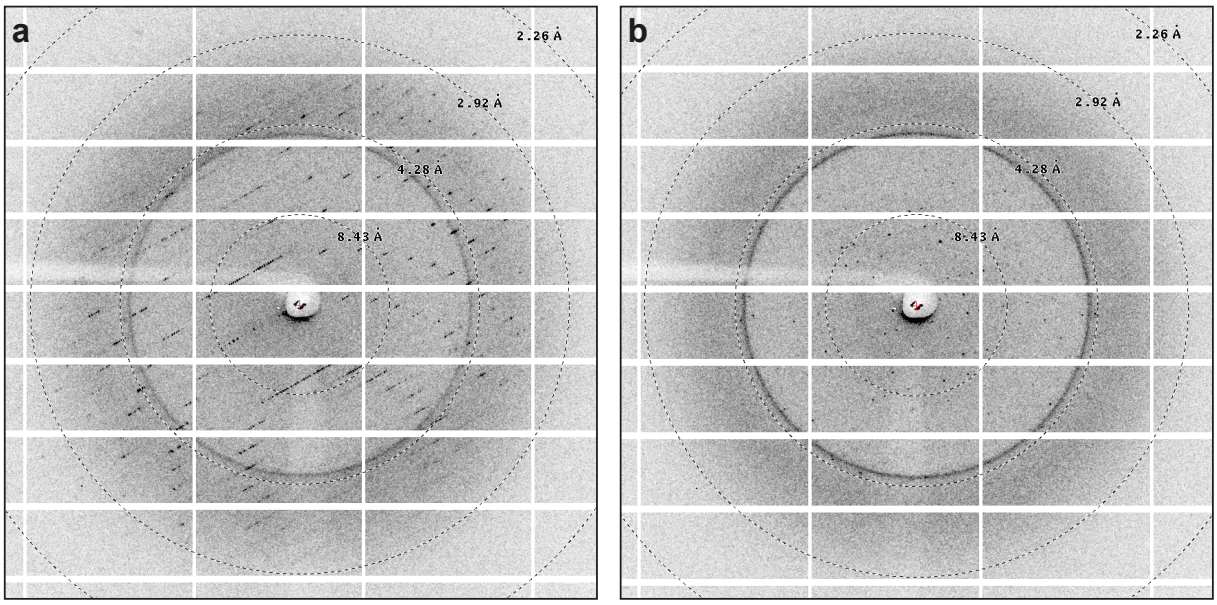


Fig. 6.10 – Typical Diffraction Patterns of *At*CTNS-Syb39-Nb4.
Diffraction patterns obtained at $\varphi = 0°$ (**a**) and $\varphi = 90°$ (**b**), visualized using Adxv (Arvai, 2021).

using multiple binders simultaneously. In doing so, additional *At*CTNS epitopes would be covered, thereby minimizing potentially weak intermembrane protein crystal contacts in the LCP, which may lead to larger and better diffracting crystals. We first tested what binders could interact with *At*CTNS simultaneously using BLI epitope binning experiments. This proved to be true for Syb39 and Nb4 (**Fig. 6.9 a**), both of which formed crystals when in complex with *At*CTNS, as shown before. Subsequently, a triple complex using those binders was formed and run on the SRT-C SEC-300 column to isolate it from excess binders. Because both binders eluted from the column at a similar retention volume and ran at an identical MW on the gel, the successful formation of a triple complex could not be shown through these measures (**Fig. 6.9 b** and **c**). However, after LCP setup, crystals of novel morphology, mostly cuboidal in shape, formed (**Fig. 6.9 d**). Appearing again at slightly acidic conditions (pH 5.5–6.5) and in the presence of ~25–30% (v:v) PEG dimethyl ether (DME) 500 and ~50–150 mM K-formate, these crystals emerged quickly and were easily regenerated. They typically matured after 5–7 days and then featured a base area of ~60–80 × 40–60 µm^2.

6.3.2 Data Collection, Processing & Analysis

The *At*CTNS-Syb39-Nb4 crystals diffracted consistently to ~2.5–3.5 Å at the synchrotron. They were also found to be less prone to radiation damage compared to the crystals used before, allowing us to collect multiple sweeps of data from the same crystal without noticeable loss in diffraction quality. The main drawback associated with these crystals was that the collected data was moderately anisotropic. This was already observed when inspecting the unique diffraction patterns in **Fig. 6.10** with the naked eye: Reflections along one axis of the reciprocal

Table 6.3 – X-Ray Data Collection & Processing Statistics of *At*CTNS-Syb39-Nb4.
This table continues on the next page.

a, Relevant sweeps of data collected at DLS (proposal number mx23459-130).

Identifier	Beamline	Wavelength (Å)	Beamsize (μm)	Transmission (%)	Exposure (s)	Oscillation (°)	Total images	Resolution max. (Å)	Detector
6774_1	DLS - I24	0.9999	7 × 7	15.62	0.0200	0.10	2,000	2.00	PILATUS3 6M
6774_2	DLS - I24	0.9999	7 × 7	20.37	0.0200	0.10	2,000	2.50	PILATUS3 6M
6774_3	DLS - I24	0.9999	7 × 7	24.30	0.0200	0.10	3,600	2.40	PILATUS3 6M

b, Data processing.

Pipeline	Images used	High-resolution data cut	Space group	Cell dimensions *a*, *b*, *c* (Å)	Cell angles α, β, γ (°)	PDB
autoPROC	All of sweeps 1–3	Based on $CC_{1/2}$	$P2_12_12$	62.79, 319.96, 45.65	90.00, 90.00, 90.00	7ZK1

c, Merging statistics by resolution bin and overall. *Significant at $p = 0.01$.

Resolution (Å)	Observations	Unique reflections	Multiplicity	Completeness (%)	Mean $I/\sigma(I)$	R_{merge} (%)	R_{meas} (%)	R_{pim} (%)	$CC_{1/2}$ (%)
79.99–7.18	34,855	1,574	22.14	100	18.0	14.1	14.5	3.4	99.8*
7.18–5.70	36,821	1,449	25.41	100	14.3	25.3	25.8	5.1	99.7*
5.70–4.98	37,285	1,454	25.64	100	13.3	30.0	30.6	6.1	99.4*
4.98–4.52	35,594	1,395	25.52	100	12.4	36.0	36.7	7.4	99.5*
4.52–4.20	35,985	1,409	25.54	100	11.3	43.6	44.5	8.8	99.4*
4.20–3.95	36,360	1,416	25.68	99.9	9.2	49.6	50.6	10.1	99.4*
3.95–3.75	36,884	1,404	26.27	100	7.3	67.3	68.6	13.4	98.9*
3.75–3.59	36,820	1,393	26.43	100	5.9	79.7	81.3	15.8	98.4*
3.59–3.45	36,680	1,395	26.29	99.9	4.8	100.0	102.0	19.9	98.1*
3.45–3.33	36,277	1,374	26.40	100	3.9	126.2	128.7	25.1	97.1*
3.33–3.23	36,101	1,375	26.26	100	3.1	164.8	168.1	32.7	95.1*
3.23–3.14	36,965	1,384	26.71	100	2.5	221.6	225.9	43.7	90.6*
3.14–3.05	36,731	1,385	26.52	100	2.0	281.7	287.3	55.7	85.8*
3.05–2.98	36,717	1,371	26.78	99.9	1.7	330.2	336.7	65.1	87.8*
2.98–2.91	36,844	1,381	26.68	99.9	1.4	483.7	493.2	95.7	78.5*
2.91–2.85	35,054	1,335	26.26	100	1.3	488.8	498.5	97.4	71.2*
2.85–2.79	30,442	1,411	21.57	99.9	0.8	654.2	669.7	140.7	54.5*
2.79–2.74	18,684	1,356	13.78	99.9	0.5	689.9	717.0	193.7	34.0*
2.74–2.69	18,240	1,335	13.66	99.9	0.5	636.8	662.3	180.2	39.5*
2.69–2.65	18,755	1,413	13.27	99.9	0.4	796.4	829.0	227.6	35.7*
27.99–2.65	**668,094**	**28,009**	**23.9**	**100**	**5.9**	**54.2**	**55.4**	**11.4**	**99.8**

lattice reach a maximum resolution of ~3 Å. Those are narrowly spaced, which indicates that the corresponding unit cell axis in real space is long. Reflections along the perpendicular axis are spaced more widely but only range to ~4 Å. After rotating the crystal by 90°, fewer reflections are observed altogether, and only few are observed past the 4.28 Å mark, as shown in **Fig. 6.10 b**.

The improved resistance to radiation damage enabled us to collect multiple sweeps of data whilst also reducing the rotation increment from 0.5 to 0.1° per image. Standard practice at most synchrotrons, fine φ-slicing implies that each reflection is measured multiple times, which improves the estimation of its amplitude and minimizes the risk of overlaps of multiple spots, especially when reflections are close together (Mueller *et al.*, 2012). Multiple datasets were

Table 6.3 (Continued)

d, Merging statistics by resolution bin and overall as determined by STARANISO. *Significant at $p = 0.01$.

Resolution (Å)	Observations	Unique reflections	Multiplicity	Compl. (%) spher. | ellip.	Mean $I/\sigma(I)$	R_{merge} (%)	R_{meas} (%)	R_{pim} (%)	$CC_{1/2}$ (%)
79.99–8.32	22,295	1,046	21.32	100 | 100	18.2	13.4	13.9	3.4	99.8*
8.32–6.48	25,371	1,046	24.26	100 | 100	16.5	18.5	19.0	3.9	99.7*
6.48–5.63	26,913	1,046	25.73	100 | 100	13.8	27.9	28.5	5.6	99.5*
5.63–5.09	26,902	1,045	25.74	100 | 100	13.4	29.7	30.3	6.0	99.7*
5.09–4.72	26,716	1,047	25.52	100 | 100	12.9	33.4	34.1	6.8	99.2*
4.72–4.43	26,277	1,044	25.17	99.9 | 99.9	12.2	39.8	40.7	8.1	99.4*
4.43–4.19	27,127	1,047	25.91	100 | 100	10.8	44.6	45.5	9.0	99.5*
4.19–4.01	26,638	1,045	25.49	99.8 | 99.8	9.2	48.9	49.9	10.0	99.4*
4.01–3.85	27,360	1,047	26.13	100 | 100	8.1	59.8	61.0	12.0	99.2*
3.85–3.71	27,629	1,046	26.41	100 | 100	6.8	70.0	71.4	13.9	98.5*
3.71–3.59	27,640	1,045	26.45	100 | 100	5.7	85.6	87.3	17.0	98.3*
3.59–3.48	27,727	1,046	26.51	99.6 | 99.6	4.9	97.3	99.2	19.2	98.1*
3.48–3.38	27,269	1,046	26.07	94.6 | 94.6	4.4	109.7	111.9	21.9	97.9*
3.38–3.28	27,744	1,047	26.50	79.1 | 82.6	4.2	116.0	118.3	23.0	97.2*
3.28–3.17	27,993	1,046	26.76	70.2 | 81.7	3.5	143.8	146.6	28.2	95.2*
3.17–3.06	27,937	1,044	26.76	59.3 | 78.7	3.4	148.0	150.9	29.0	93.7*
3.06–2.95	28,435	1,047	27.16	50.2 | 75.7	2.9	172.0	175.3	33.5	94.4*
2.95–2.81	28,481	1,046	27.23	32.9 | 70.7	2.9	184.5	188.0	35.9	91.1*
2.81–2.64	29,112	1,046	27.83	22.9 | 77.7	2.3	249.4	254.0	47.9	83.0*
2.64–2.31	27,908	1,046	26.68	7.54 | 75.1	1.7	325.2	331.5	63.5	71.1*
79.99–2.31	**543,474**	**20,918**	**26.0**	**49.9 | 90.4**	**7.9**	**44.3**	**45.3**	**9.0**	**99.8**

processed and merged using the autoPROC pipeline (Vonrhein *et al.*, 2011; **Table 6.3**). The orthorhombic space group $P2_12_12$ was assigned with cell dimensions $a = 62.79$ Å, $b = 319.96$ Å, and $c = 45.65$ Å, and the data were cut automatically based on the $CC_{1/2}$ to 2.65 Å. The underlying anisotropy was addressed using STARANISO (Tickle *et al.*, 2016), yielding the following estimations for the maximum resolution along each reciprocal lattice axis: $a^* = 3.39$ Å, $b^* = 2.30$ Å, and $c^* = 2.91$ Å. STARANISO truncates the normally spherical reciprocal space along the border of an ellipsoid to remove areas without reflections and provides an additional .mtz file containing the updated structure factors.

Processing statistics for the isotropic data and those generated after applying anisotropic corrections are listed in **Table 6.3 c** and **d**, respectively. Supported by the extraordinarily high multiplicity of >20, resulting from merging multiple fine-sliced datasets, R_{pim}s smaller than 10% were achieved at all resolution bins under 4 Å. The agreement of the data across all three datasets was indicated by an overall $CC_{1/2}$ of 99.8%. The overall mean $I/\sigma(I)$ was increased compared to previous datasets and further elevated after anisotropy treatment (from 5.9 to 7.9). The overall completeness was 100%, however, STARANISO decreased the spherical and elliptical completeness considerably for high-resolution shells starting from 3.5 Å (see **Table 6.3 d**).

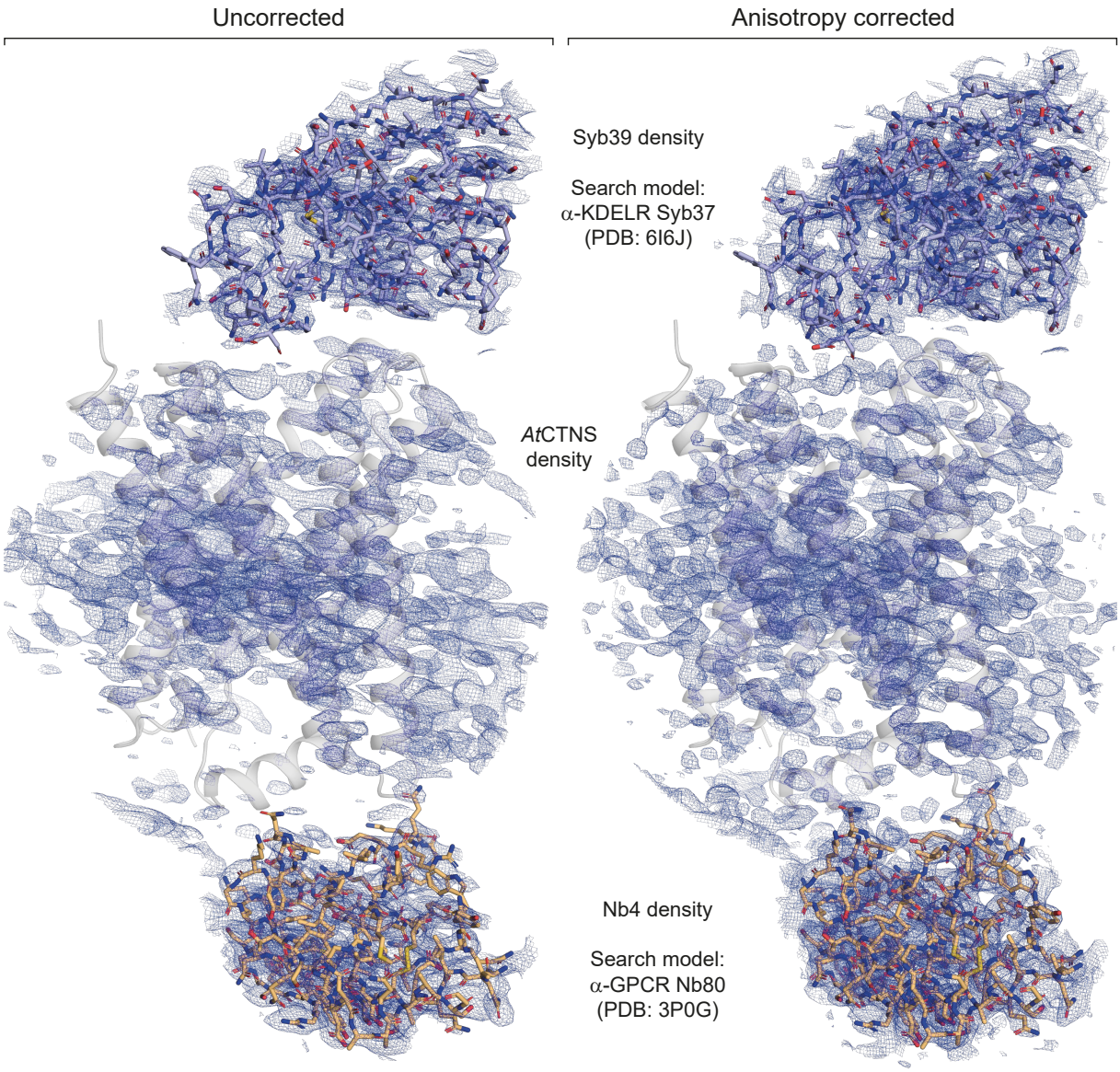


Fig. 6.11 – Electron Density After MR Using Sybody & Nanobody Frames.
$2mF_o$–DF_c electron density map obtained after MR in Phaser with 10 subsequent rounds of refinement in REFMAC. The maps are contoured at 1.2 σ with a carve of 3 Å around the binders and a 60 Å diameter sphere to cover the position of *At*CTNS, obtained by superimposing the binders with the finalized *At*CTNS-Syb39-Nb4 structure (PDB: 7ZK1, shown in **Fig. 7.6**).

6.3.3 Structure Solution Attempt by Molecular Replacement

Again, efforts were made to solve the structure of *At*CTNS by MR using the binder frames as the search models. This time, given that multiple binders could be placed, a higher percentage of the overall unit cell contents was covered. However, the recommended minimum of ~50% for successful MR (Abergel, 2013) was still not reached as the CDR-less Nb4 only adds ~10.2 kDa, thus covering ~36% of the total weight of the complex. Placement of a single copy of α-KDELR Syb37 (extracted from PDB: 6I6J) and a single copy of α-GPCR Nb80 (3P0G) in Phaser was again straightforward for both the untreated as well as the anisotropy corrected reflection intensi-

ties (using autoPROC generated files truncate-unique.mtz and staraniso_alldata-unique.mtz, respectively). In each case, a TFZ ≥ 9.0 and an LLG ≥ 99 were scored for the first placement (Syb39 homologue), the second placement being higher. The electron density fitted the search models well whilst highlighting differences between α-*At*CTNS binders and the homologous search models, especially regarding side chains and their orientation (**Fig. 6.11**). However, additional density, revealing α-helical content that would enable us to start building a model for *At*CTNS, was not found. As a result, MR-based structure solution attempts were halted, and we shifted our focus towards solving the structure experimentally using the voluminous *At*CTNS-Syb39-Nb4 crystals that could be easily re-grown.

CHAPTER 7

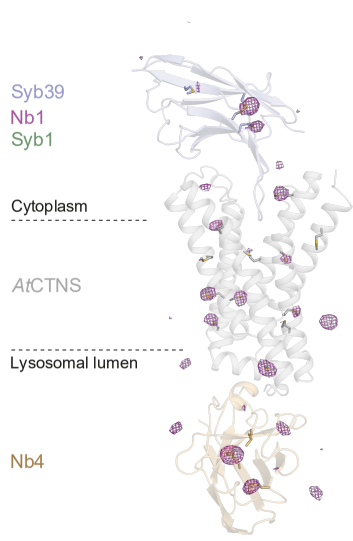
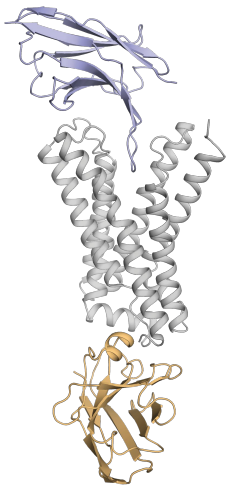
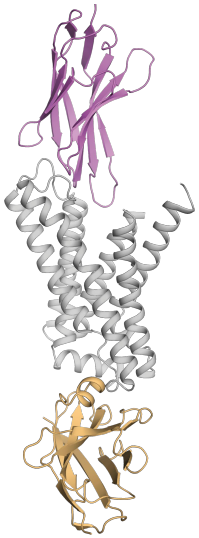
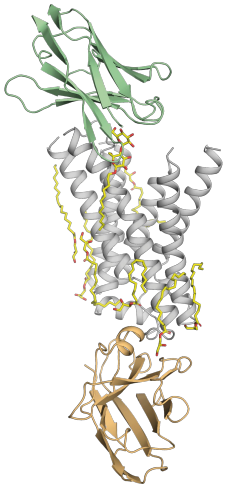
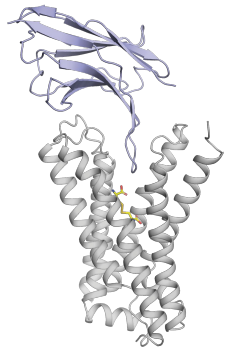
*At*CTNS Structure Solution & Co-Crystal Structures

Having so far been unable to solve the three-dimensional structure of *At*CTNS by MR (**Chs. 4** and **6**), it was the realization that we could bind multiple binders to *At*CTNS simultaneously and the discovery that such a triple complex would yield voluminous and easily reproducible crystals (**Fig. 6.9**) that allowed us to pursue the route of experimental phasing. In this chapter, I will discuss how we solved the first co-crystal structure of *At*CTNS in complex with Syb39 and Nb4 using SAD and then present several additional co-crystal structures that were solved by MR using the model of the first structure. Particular focus is placed on the different datasets obtained, the analysis of the data, and the resulting models in regard to transporter-binder interactions, packing, and chemical quality. The structural and functional analysis of CTNS follows in **Ch. 9**, after addressing the human transporter in **Ch. 8**.

7.1 Overview

Table 7.1 lists the different *At*CTNS co-crystal structures in the order they appear in this chapter. The first structure of a CTNS homologue, obtained from an *At*CTNS-Syb39-Nb4 crystal, was solved by SAD using the increased sulfur anomalous signal at longer wavelengths and the anomalous signal obtained from crystals soaked in mercury acetate (**Sec. 7.2**). Structure refinement was performed using the conventional, higher-resolution data (2.65 Å; **Sec. 6.3**), yielding a complete model for *At*CTNS in the inward-open (binding site facing the cytoplasm) conformation, described in **Sec. 7.2.5**. In the attempt to solve the ligand-bound state of *At*CTNS, we solved the structure of the transporter again using different binders. Both the *At*CTNS-Nb1-Nb4 structure (2.33 Å; **Sec. 7.3.1**) and the *At*CTNS-Syb1-Nb4 structure (1.61 Å; **Sec. 7.3.2**) validate the previous structure, showing the transporter in the same conformation as before. However, appropriate electron density for the ligand was not found in these structures, and anomalous S-SAD data collected for the *At*CTNS-Syb1-Nb4 composition also failed to locate the disulfide-containing ligand (**Sec. 7.3.3**). Instead, the structure with Syb1 revealed *At*CTNS as a crystallographic dimer, sparking another interesting research topic on CTNS transporters (**Ch. 8**). The ligand-bound state was ultimately solved from a lower-resolution dataset that only contained one binder, Syb39 (**Sec. 7.4**). Statistics for all models are summarized in **Table 7.7**.

Table 7.1 – Overview of *At*CTNS Co-Crystal Structures.
Cartoon representation of all *At*CTNS structures and heavy atom substructures discussed in this chapter. Ligands and relevant side chains are shown as sticks. S-SAD (purple mesh) and Hg-SAD (orange mesh) maps are contoured at 5 σ (initial S-SAD of *At*CTNS-Syb39-Nb4 at 3.5 σ).

	***At*CTNS-Syb39-Nb4**			***At*CTNS-Nb1-Nb4**	***At*CTNS-Syb1-Nb4**		***At*CTNS-Syb39**
							
	S-SAD	**Hg-SAD**	**Apo**	**Apo**	**Apo**	**S-SAD**	**L-Cystine Bound**
Description	Initial S-SAD data used for structure determination and validation.	Hg-SAD data used for structure determination and validation.	First structure of *At*CTNS reveals 7TM architecture.	*At*CTNS structure of higher resolution confirms previous model.	High-resolution *At*CTNS structure reveals a crystallographic dimer and potential ligand binding sites.	S-SAD data used for structure validation and ligand localization trials (unsuccessful).	Ligand bound structure of *At*CTNS highlights transporter-ligand interactions.
Max. resolution	3.20 Å	3.40 Å	2.65 Å	2.33 Å	1.61 Å	1.96 Å	3.47 Å
Solved by	----	----	MR-SAD	MR	MR	SAD	MR
PDB code	----	----	7ZK1	7ZKZ	----	----	7ZKW
Identifier	6870 + 6871	6883	6774	7589	7561	7658	6396

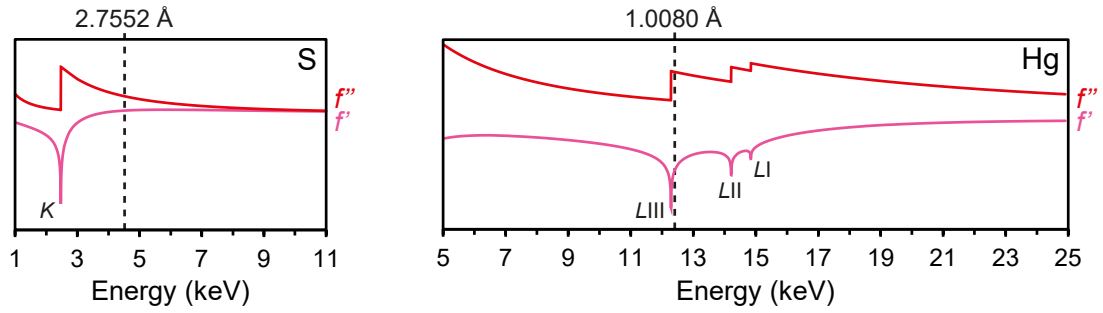


Fig. 7.1 – Energy Scans of Anomalous Scatterers S & Hg.
Sulfur- and mercury-specific X-ray absorption edges are shown. Energy proportional wavelengths used during SAD experiments are indicated. Data was extracted from the Biomolecular Structure Center at the University of Washington, available online *via* www.bmsc.washington.edu.

7.2 Structural Solution of *At*CTNS

Initial phase values for *At*CTNS were determined by SAD, which makes use of the anomalous scattering effect that causes the absorption and immediate re-emission of X-ray photons without change to their energy (Rhodes, 2006). The contribution to anomalous scattering of a specific element is wavelength specific and most observable near the absorption edge of the element (see **Fig. 7.1**). In contrast to normal diffraction, the phase of the anomalously scattered X-ray is shifted by a specific amount, expressed in a real (f') and an imaginary component (f''). Viewed as a vector, f'' always faces the same direction and therefore breaks the centrosymmetry of Friedel pair reflections hkl and $-h-k-l$, where amplitudes $|F_{hkl}|$ and $|F_{-h-k-l}|$ are identical, and their phases ϕ_{hkl} and ϕ_{-h-k-l} are exactly opposite (Friedel's law). This information can be used to locate the marker atoms that make up the crystal contents' substructure and, furthermore, to calculate the phases of all remaining reflections, thereby solving the phase problem.

Selenium is one of the most prevalent anomalous scatterers used in SAD and similar techniques (Walden, 2010). Typically, it is incorporated into the protein by replacing methionine residues with selenomethionine. A routine procedure using bacterial expression systems, this is thought to be difficult using yeast culture and was therefore avoided. Instead, we located protein-intrinsic sulfur atoms using their increased anomalous signal at longer wavelengths. Additionally, protein crystals were soaked in mercury acetate, and SAD experiments were performed to locate the Hg atoms that reacted with thiol groups of *At*CTNS.

7.2.1 Anomalous Data Collection, Processing & Analysis

S-SAD experiments were executed at DLS beamline I23 and in collaboration with Dr. Kamel El Omari and Dr. Armin Wagner. I23 is optimized to maximize the relatively weak S-SAD signal by reducing background scattering and absorption by air particles through operation in a vacuum (Wagner *et al.*, 2016; El Omari *et al.*, 2023). Sample cooling is achieved conductively

Table 7.2 – Anomalous Data Statistics of *At*CTNS-Syb39-Nb4.

a, S-SAD data statistics generated by SHELXC.

Resolution (Å)	Unique reflections	Mean $I/\sigma(I)$	Anomalous compl. (%)	Anomalous multiplicity	R_{pim} (%)	R_{anom} (%)	Mean $d''/\sigma(d'')$	Anom. $CC_{1/2}$ (%)
40.00–17.14	107	58.6	70.4	37.4	1.17	4.96	2.92	88.9
17.14–10.55	393	59.9	98.2	42.7	1.04	4.57	2.38	85.7
10.55–7.95	633	53.2	99.1	45.5	1.22	4.26	1.82	71.3
7.95–6.50	888	37.5	99.3	48.1	1.84	5.70	1.48	50.5
6.50–5.56	1,171	24.7	99.6	48.8	3.26	7.33	1.08	29.9
5.56–4.89	1,442	21.7	99.7	48.9	3.88	6.74	0.87	7.5
4.89–4.39	1,714	19.9	99.8	48.3	5.32	9.12	0.79	2.2
4.39–4.00	1,967	13.0	99.6	47.1	9.41	15.60	0.73	-0.5
4.00–3.68	2,246	6.4	99.8	44.9	20.35	24.63	0.68	-9.1
3.68–3.42	2,551	3.2	99.8	41.7	45.87	55.56	0.68	-15.1
3.42–3.20	2,780	1.5	99.0	38.7	107.20	123.20	0.69	-19.5

b, Hg-SAD data statistics generated by SHELXC.

Resolution (Å)	Unique reflections	Mean $I/\sigma(I)$	Anomalous compl. (%)	Anomalous multiplicity	R_{pim} (%)	R_{anom} (%)	Mean $d''/\sigma(d'')$	Anom. $CC_{1/2}$ (%)
58.87–18.22	125	23.1	91.9	3.8	2.40	5.76	0.99	49.9
18.22–11.21	339	23.7	99.4	4.9	2.59	8.14	1.27	51.1
11.21–8.44	575	21.1	99.8	5.4	2.79	6.88	1.05	32.7
8.44–6.90	792	14.6	99.9	5.9	4.16	10.19	0.99	16.1
6.90–5.90	1,007	7.9	100	6.1	8.70	19.40	0.93	13.8
5.90–5.20	1,237	5.8	99.9	6.2	11.75	25.98	0.87	5.5
5.20–4.67	1,470	5.0	100	6.3	13.17	30.12	0.88	5.5
4.67–4.25	1,735	3.3	99.7	6.3	21.22	48.49	0.92	14.9
4.25–3.91	1,970	2.0	99.7	6.2	37.96	86.93	0.86	12.5
3.91–3.63	2,249	0.9	99.8	6.3	88.41	189.50	0.78	6.2
3.63–3.40	2,409	0.5	99.8	6.3	151.70	361.10	0.54	-6.8

using special sample holders made from copper. The beamline is furthermore equipped with a curved detector to compensate for the increased diffraction angles that result from irradiation using X-rays of longer wavelengths (following Bragg's law, **Eq. 4.1**), and thus enables the capture of diffraction spots that would otherwise miss the detector.

Using the optimized *At*CTNS-Syb39-Nb4 crystals shown in **Fig. 6.9**, we collected multiple datasets at I23 using a wavelength close to the absorption edge of sulfur (**Fig. 7.1 a**). Processing in XDS and merging in XSCALE (Kabsch, 2010) revealed that the data were of lower maximum resolution but slightly better quality when compared to the conventional data collected at I24: 3.20 Å *versus* 2.65 Å; overall mean $I/\sigma(I)$ of 10.5 *versus* 5.9; R_{meas} < 10.4% *versus* 14.5% in low-resolution shells to ~7.2 Å; completeness and $CC_{1/2}$ at or near 100% for each (compare **Tables A.7** and **6.3**).

The quality of the anomalous data was assessed in SHELXC (Sheldrick, 2008, 2010) using the unmerged reflection file generated by XSCALE and searching for 19 sulfur atoms, including

two supersulfurs (the disulfide in each binder), in the unit cell. **Table 7.2** lists the various parameters output by SHELXC, most importantly the mean $d''/\sigma(d'')$ and anomalous $CC_{1/2}$, which indicate the strength of the anomalous signal per resolution shell and its correlation within randomly halved datasets, respectively. In low-resolution bins to ~8 Å, the presented S-SAD data featured a strong anomalous signal (mean $d''/\sigma(d'') > 1.8$), likely amplified by the high multiplicity resulting from merging multiple sweeps of data, and a high correlation of >70% (**Table 7.2 a**). However, the anomalous data quality decreased significantly approaching higher-resolution shells, and thus, the data were truncated to a maximum resolution of 5 Å (criteria: mean $d''/\sigma(d'') < 0.8–1.0$ or $CC_{1/2} < 20–30\%$) and used for the determination of the initial phases and structure validation, respectively.

In addition to S-SAD, we also carried out more traditional SAD experiments through heavy atom soaking using stronger anomalous scatterers. For this, crystals from **Fig. 6.9** were soaked in a solution containing the precipitant supplemented with mercury acetate ($Hg(OAc)_2$) or potassium hexachloroplatinate (K_2PtCl_6). The heavy atoms of Hg or Pt hereby diffuse into the lipophilic and hydrophilic environments of the LCP crystal and react with the thiol groups of the proteins' cysteine residues (Parker & Newstead, 2013). A wavelength close to the *L*III absorption edge of these elements was used during data collection, as was highlighted for Hg in **Fig. 7.1 b**.

The soaked crystals generally diffracted to a lower maximum resolution compared to the untreated ones. Nevertheless, 3–4 Å datasets were collected and analysed regarding their anomalous signal. Crystals soaked in K_2PtCl_6 did not generate a strong enough anomalous signal to help solve the phase problem, however, those incubated in $Hg(OAc)_2$ enabled us to locate the three cysteine residues of *At*CTNS and thus provided helpful information to refine and validate the structure. Data collection and processing statistics, as well as diffraction images of the relevant dataset, are shown in **Table A.8** and **Fig. A.4**, respectively. Its anomalous signal, as analysed using SHELXC, is reflected in **Table 7.2 b**: Overall, the statistics indicated a weak anomalous signal manifested in a low mean $d''/\sigma(d'')$, listed at <1.3 for all resolution shells, and a relatively low anomalous $CC_{1/2}$.

7.2.2 Initial Phase Determination & Model Building

Similar to the conventional data collected for *At*CTNS-Syb39-Nb4 (**Sec. 6.3.3**), sybody and nanobody frames were used as search models in Phaser to position the binders in the unit cell, this time using the software package PHENIX (Adams *et al.*, 2010). The models were again placed with confidence (TFZ = 8.9 and LLG = 87 for the initial placement of the sybody frame; TFZ = 16.4 and LLG = 218 for the following nanobody frame). Subsequently, AutoSol (Terwilliger *et al.*, 2009) was run to identify the heavy atom sites, calculate the phases, and carry out density modification (Adams *et al.*, 2011), ultimately generating an experimental $2mF_o–DF_c$

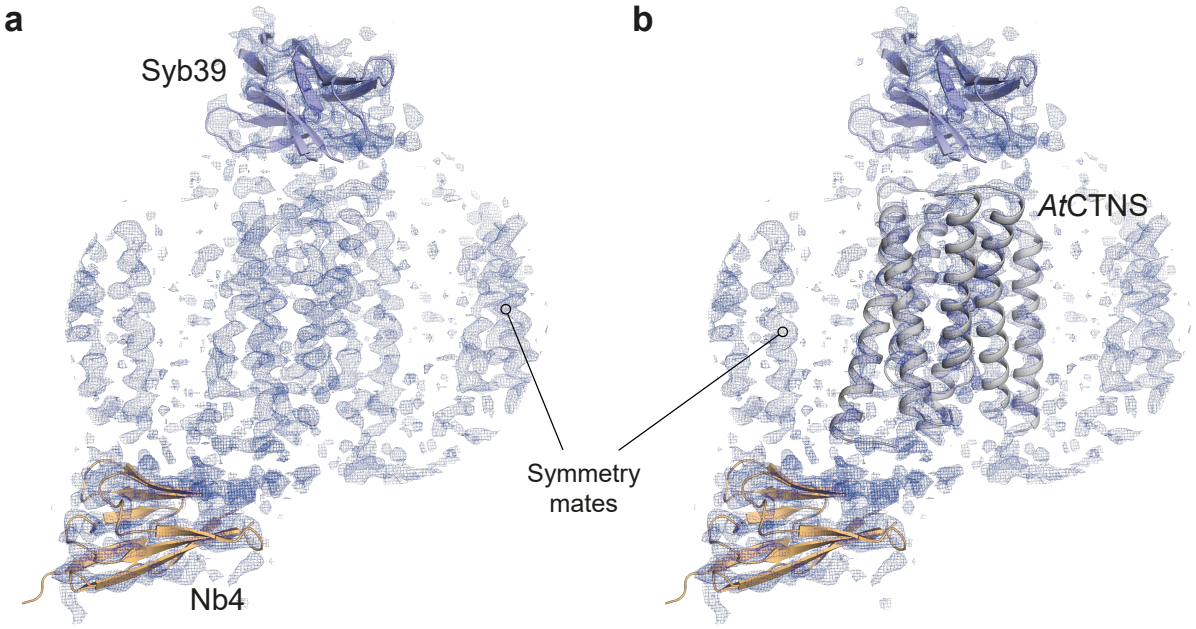


Fig. 7.2 – Placement of the First *At*CTNS Model.
a, Initial $2mF_o–DF_c$ map generated by MR-SAD using PHENIX AutoSol, contoured at 1.8 σ and carved around the binders at 3 Å and within a 100 by 60 Å diameter ellipsoid to cover the transporter. **b**, Initial poly-alanine model positioned in the α-helical density.

map through MR-SAD that showed α-helical content for the first time (**Fig. 7.2 a**). AutoSol automatically placed individual poly-alanine models for all seven TM helices in the density (**Fig. 7.2 b**). Superimposition with *Os*SWEET2b (PDB: 5CTG), using the inversion helix (TM4, see **Fig. 9.3 b**) as the anchor point, furthermore allowed us to establish the connectivity of the helices, leading to a complete poly-alanine model of *At*CTNS. Appropriate side chain densities were not observed at this stage, and only about half of the substructure's heavy atoms were located by AutoSol, hampering efforts to identify individual residues of the transporter.

7.2.3 Structure Refinement

To determine the individual residues of the poly-alanine model obtained using MR-SAD, we reverted to the higher-resolution and isomorphous *At*CTNS-Syb39-Nb4 dataset collected at DLS microfocus beamline I24 (**Fig. 6.10** and **Table 6.3**). MR in Phaser was trivial using the models for *At*CTNS, Syb39, and Nb4 as a combined search ensemble (TFZ = 11.9, LLG = 183), verifying their arrangement. The electron density, generated using the 2.65 Å data, highlighted additional density that enabled us to extend the CDRs of the binders (**Fig. 7.3 a**). Regarding the orientation of the transporter's side chains, only limited information was revealed initially, however, this was improved substantially through crystallographic refinement.

Structure models and maps are optimized through an iterative process of model fitting and refinement that aims to minimize the differences between the original intensities observed in the

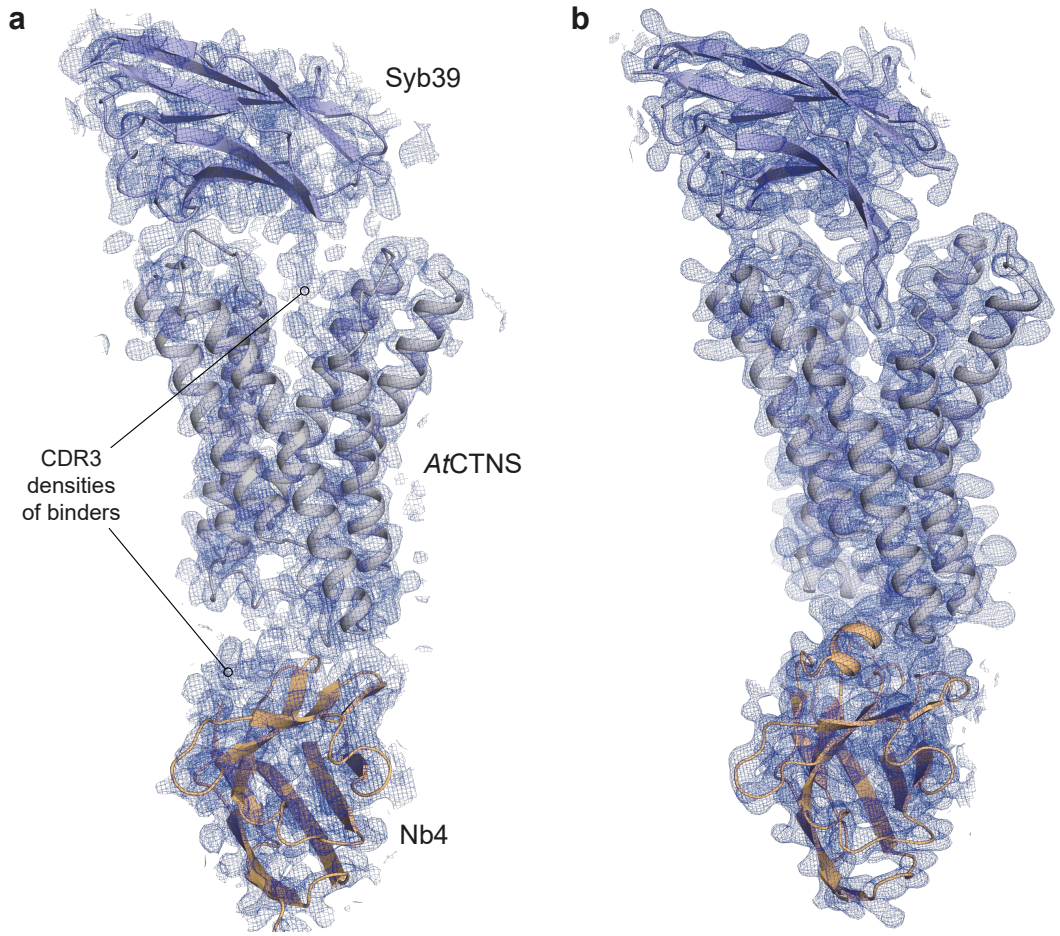


Fig. 7.3 – *At*CTNS Model & Map Refinements.
a, $2mF_o$–DF_c map obtained after MR using the conventional 2.65 Å data (**Table 6.3**) and the models from **Fig. 7.2 b** as a combined search ensemble. **b**, Refined $2mF_o$–DF_c map, showing clear side chain densities. All maps are contoured at 1 σ and carved around the models at 3 Å.

diffraction experiment and those that can be calculated directly from the model (Rhodes, 2006). Their agreement is mainly evaluated by the refinement *R*-factor, which is determined from the amplitudes of all structure factors (R_{work}) or, to reduce model bias, only using reflections that were omitted in the modelling and refinement processes (R_{free}; Brünger, 1992). Defined as

$$R = \frac{\sum \big| |F_o| - |F_c| \big|}{\sum |F_o|}, \tag{7.1}$$

where F_o and F_c are the observed and calculated structure factors, respectively, *R*-values of $R_{work} \leq 20–25\%$ and $R_{free} \leq 25–30\%$ are desired upon finalization of the refinement process.

After MR, the R_{work} and R_{free} were reported at 45 and 47%, respectively. The refinement process was then started by (re)building the model in the electron density. This was achieved manually in Coot (Emsley & Cowtan, 2004; Emsley *et al.*, 2010) and initially focused on the completion and positional optimization of the individual main chains. The resulting model was then used in combination with the original reflection data to perform automatic refinement in

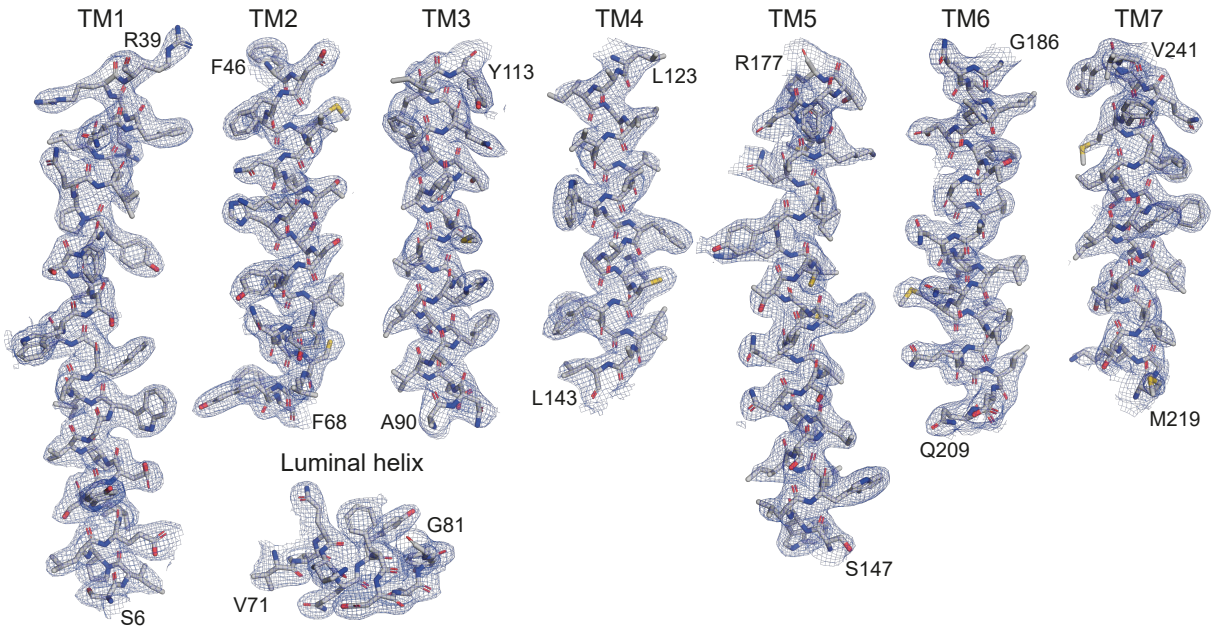


Fig. 7.4 – Essential Electron Density Maps of *At*CTNS-Syb39-Nb4 at 2.65 Å.
$2mF_o–DF_c$ maps around all *At*CTNS helices are shown, contoured at 1 σ with a carve of 2 Å.

BUSTER (Bricogne *et al.*, 2017) using rigid-body constraints. BUSTER is a maximum likelihood refinement program that scales the observed and calculated structure factor amplitudes and computes the most likely position of the model whilst satisfying set restraints/constraints (Blanc *et al.*, 2004). As a result, the $2mF_o–DF_c$ electron density and $mF_o–DF_c$ difference maps are updated using the phase angles calculated from the improved atomic model along with the observed and calculated structure factor amplitudes (*via* Fourier synthesis, see **Eq. 4.5**), producing a more detailed map in case the new model reflects the correct structure of the crystal contents more accurately (Rhodes, 2006). The maps are hereby weighted using a figure of merit (m), applied to each phase, and an estimate of error (D) for the atomic coordinates of the model.

Upon rigid-body refinement, the maps became more detailed, allowing us to model the side chains of *At*CTNS. For orientation, anomalous difference maps were calculated in ANODE (Thorn & Sheldrick, 2011) using the S-SAD and Hg-SAD data in combination with the current model, respectively, to create maps that indicate the location of the heavy atoms (see **Fig. 7.5** and **Sec. 7.2.4**). TM2, which was predicted to comprise the most sulfur-containing residues of any helix (**Fig. 3.1 a**), was chosen as the starting point for side chain fitting. The anomalous signal, overlaid with the $2mF_o–DF_c$ and $mF_o–DF_c$ maps, enabled confident placement of all three sulfur-containing residues on TM2 (M50, M60, and C65). Further side chains within TM2 and beyond were placed accordingly, and the model's placement in the density was evaluated and adjusted where necessary. Once the observed density was accounted for or no more side chains could be fitted sensibly, automated refinement was performed in BUSTER, and the process was repeated. Refinement was generally performed with translation, libration, and

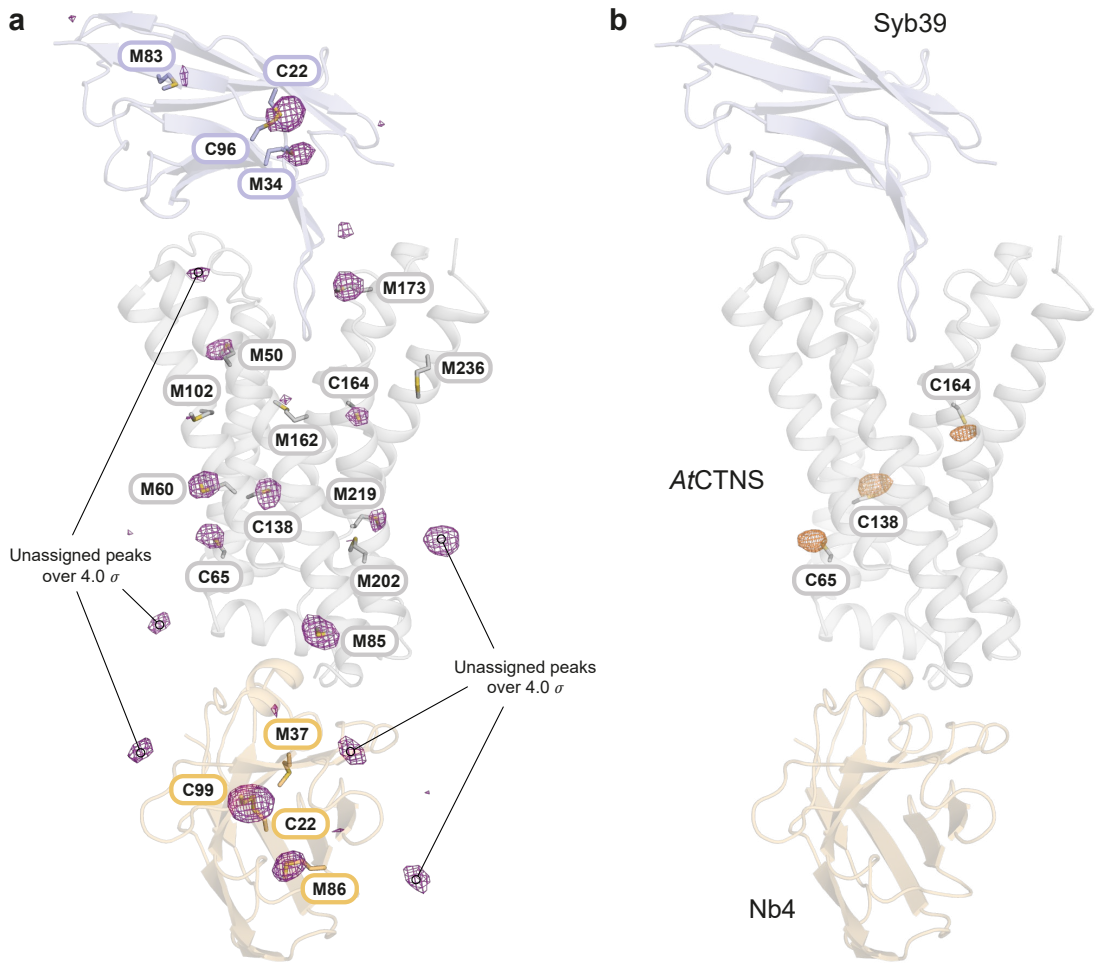


Fig. 7.5 – Anomalous Difference Maps of *At*CTNS-Syb39-Nb4.
Anomalous difference maps calculated using the finalized *At*CTNS-Syb39-Nb4 model (PDB: 7ZK1, **Fig. 7.6**) with the S-SAD data (in **a**) and the Hg-SAD data (in **b**) in ANODE, indicating the location of S (purple mesh, contoured at 3.5 σ) and cysteine-bound Hg atoms (orange mesh, 5 σ), respectively. Anomalous peaks and their closest atom in the model are listed in **Table 7.3**.

screw (TLS) parameterization, which retains a realistic data-to-parameter ratio by forming TLS groups, each treated as a rigid body with 20 parameters (instead of six per atom in individual anisotropic refinement), enabling a form of constrained anisotropic refinement for lower-resolution data (Winn *et al.*, 2001). Additionally, BUSTER was used to automatically add water molecules into the density. The iterative process was stopped when no additions or alterations could be justified for the model and when the *R*-values converged to a minimum. The final map for *At*CTNS-Syb39-Nb4 is shown in **Fig. 7.3 b**, and the fit around all helical side chains is further emphasized in **Fig. 7.4**. The final R_{work} was reported at 23.4%, R_{free} at 27.5% (see **Table 7.7** at the end of this chapter).

Table 7.3 – Anomalous Peaks of *At*CTNS-Syb39-Nb4 (S-SAD & Hg-SAD).

a, S-SAD peaks generated by ANODE using the finalized *At*CTNS-Syb39-Nb4 model (PDB: 7ZK1, **Fig. 7.6**) and data truncated to 5 Å. Peaks with a minimum height of 3.5 σ are listed.

Atom peak	Peak height (σ)	Distance to nearest atom (Å)	Atom	Residue	Chain	Molecule
S1	8.40	0.990	SG	C22	C	Syb39
S2	8.36	0.983	SG	C99	B	Nb4
S3	6.42	0.611	SD	M85	A	*At*CTNS
S4	6.22	0.726	SD	M86	B	Nb4
S5	6.19	1.328	SD	M60	A	*At*CTNS
S6	5.98	4.996	OH	Y199	A	*At*CTNS
S7	5.93	0.746	CG	M173	A	*At*CTNS
S8	5.90	1.142	SG	C138	A	*At*CTNS
S9	5.89	0.958	SD	M34	C	Syb39
S10	5.17	1.325	SD	M50	A	*At*CTNS
S11	5.09	0.911	SG	C65	A	*At*CTNS
S12	4.98	0.887	SG	C164	A	*At*CTNS
S13	4.70	1.183	SD	M219	A	*At*CTNS
S14	4.52	2.516	CD	E47	B	Nb4
S15	4.50	1.468	N	Y63	B	Nb4
S16	4.40	3.632	CG	R37	A	*At*CTNS
S17	4.30	3.584	NE	R177	A	*At*CTNS
S18	4.18	1.031	SD	M83	C	Syb39
S19	4.04	1.834	O	D102	B	Nb4
S20	3.97	0.946	CG	M162	A	*At*CTNS
S21	3.97	1.446	SD	M162	A	*At*CTNS
S22	3.96	0.937	SD	M202	A	*At*CTNS
S23	3.86	8.335	CB	P44	B	Nb4
S24	3.85	6.395	CZ	F139	A	*At*CTNS
S25	3.81	1.383	O	HOH310	D	Water
S26	3.81	0.516	CE	M102	A	*At*CTNS
S27	3.81	6.525	O	L11	C	Syb39
S28	3.79	2.166	SD	M83	C	Syb39
S29	3.77	2.701	CB	S120	C	Syb39

b, Hg-SAD peaks generated by ANODE using the finalized *At*CTNS-Syb39-Nb4 model (PDB: 7ZK1, **Fig. 7.6**) and data truncated to 3.63 Å. Peaks with a minimum height of 4 σ are listed.

Atom peak	Peak height (σ)	Distance to nearest atom (Å)	Atom	Residue	Chain	Molecule
Hg1	8.56	1.425	SG	C138	A	*At*CTNS
Hg2	8.11	2.020	SG	C65	A	*At*CTNS
Hg3	6.56	1.307	SG	C164	A	*At*CTNS
Hg4	4.24	0.762	O	N218	A	*At*CTNS
Hg5	4.23	5.485	O	G45	B	Nb4

7.2.4 Structure Validation

To validate the structure, we again used ANODE to compare the locations of the heavy atoms predicted from our SAD data (**Table 7.2**) with those of the final model. ANODE generates

a contour map (**Fig. 7.5**) as well as a table (**Table 7.3**), highlighting the highest anomalous peaks and their respective positions. The S-SAD data (**Fig. 7.5 a** and **Table 7.3 a**) agreed well with the model, albeit with low signal intensity for many of the heavy atom sites. Only the supersulfurs, stemming from the disulfides of the binders, were recognized at a peak height of >8 σ. Most other sulfurs were depicted with anomalous intensities between 6.4–4.7 σ, and five methionines of *At*CTNS only produced signals between 4.2–3.8 σ. There were also six additional peaks at >4 σ, which could not be assigned, as they were found 1.5–5 Å distant to the closest atom of the model, none of which were sulfurs. The only sulfurs that were not found close to an anomalous peak are residues M1 and M236 of *At*CTNS, located at the N- and C-terminal ends, respectively, and M37 of Nb4.

The three cysteine residues of *At*CTNS were located unambiguously using the SAD data obtained from the $Hg(OAc)_2$-soaked crystals. As shown in **Fig. 7.5 b** and **Table 7.3 b**, only three distinct peaks arose from the Hg-SAD data at a contour level >5 σ, each located within 1.3–2.0 Å from the sulfur atom of one of the free cysteine residues in the model. The Hg-S bond length has been reported at ~2.3–2.4 Å (Manceau & Nagy, 2008), suggesting the cysteines were likely positioned too close to the anomalous peaks rather than too far away.

Further structure validation was performed using MolProbity (Chen *et al.*, 2010; Williams *et al.*, 2018), which evaluates the quality of the model based on its chemical properties (see **Table 7.7**). The score for the *At*CTNS-Syb39-Nb4 model was calculated at 2.97, ranking it in the 52nd percentile for structures with a resolution of 2.65 ± 0.25 Å (n = 6,493). This was justified by 69 poor rotamers (16.5%) and 12 Ramachandran outliers (2.47%). Only a few atom clashes were found, resulting in a ''clashscore'' of 10.69, which ranked in the 97th percentile for the same resolution range as before (n = 223).

7.2.5 Structure of *At*CTNS in the Apo & Inward-Open States

The finalized model contains one copy of *At*CTNS in the asymmetric unit sandwiched between the two binders of Syb39 and Nb4 (**Fig. 7.6 a**). As expected, these binders play a crucial role in constituting crystal contacts. Symmetry mates of the triple complex are almost entirely linked through intermolecular nanobody interactions *via* their backbones. As shown in **Fig. 7.6 b**, pairs of Nb4 are formed, connecting two sets of triple complexes in the top-to-bottom plane whilst also interacting with the neighbouring *At*CTNS molecule to build the crystal side-to-side (front-to-back after 90° rotation). Arrays of Syb39 shape the crystal front-to-back (side-to-side after rotation), alternating the orientation of triple complexes to extend the crystal lattice in the top-to-bottom plane. Overall, a crystal with big solvent channels is generated, which is manifested in an extraordinarily large V_M of 3.90 $Å^3\ Da^{-1}$ with a solvent content of ~68%.

*At*CTNS adopts a compact V-shaped structure and consists of seven TMs, as predicted earlier; its structure is discussed in detail in **Sec. 9.1**. Syb39 interacts with *At*CTNS through the

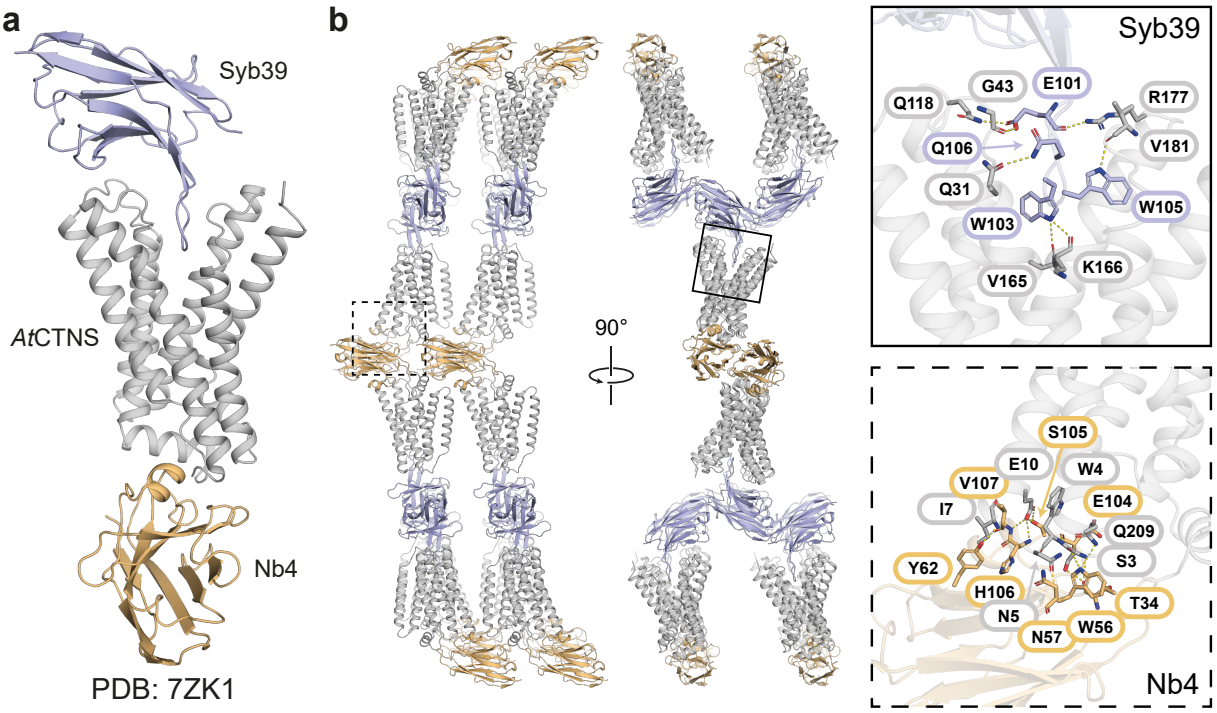


Fig. 7.6 – 2.65 Å Co-Crystal Structure of *At*CTNS-Syb39-Nb4.
a, Cartoon representation of the structure of *At*CTNS in complex with Syb39 and Nb4 (PDB: 7ZK1). **b**, Crystal packing of **a**, showing the contribution of the binders towards crystal contact formation. Specific *At*CTNS-binder interactions are shown, and intermolecular hydrogen bonds are highlighted using yellow dashed lines.

extension of its CDR3 loop into the cytoplasmic entrance of the binding site, similarly to the KDEL receptor project (compare **Figs. 1.7 b** and **7.6 a**). Several hydrogen bonds are formed, stabilizing *At*CTNS in an inward-open conformation. As shown in **Fig. 7.6 b**, two tryptophan residues, highly conserved among the Sybs generated against *At*CTNS (**Fig. 5.5**), interact with backbone carbonyl groups, docking the sybody in place. E101, also conserved, forms hydrogen bonds with residues located on all three cytoplasmic loops of *At*CTNS, possibly restricting their flexibility. Q106, which is not very well conserved, forms an additional hydrogen bond with Q31 on TM1 of the transporter, potentially boosting the binding efficiency of Syb39 compared to other sybodies.

Nb4 binding is primarily focused on the N-terminus of *At*CTNS but also involves Q209 located on the third luminal loop connecting TMs 6–7 (**Fig. 7.6 b**). In contrast to Syb39, residues of all three CDRs are involved in facilitating the binding through several hydrogen bonds, likely explaining the increased binding affinity of Nb4 compared to Syb39. The most notable interaction partners include T34 positioned on the first CDR, W56 on the second CDR, as well as *At*CTNS's E10, which binds the third CDR of Nb4. All of those residues form multiple hydrogen bonds.

7.3 *At*CTNS Structures from Unsuccessful Co-Crystal Trials Using Ligand

With the apo state of the inward-open conformation solved, we sought to gain additional structural information by crystallizing *At*CTNS in an alternative state, such as ligand-bound and/or occluded, or outward-open, to be able to characterize its transport mechanism in more detail. So far, the only crystallization setups that yielded diffraction data to sub-3 Å resolution included Nb4 (**Tables 6.2** and **6.3**). Thus, we decided to continue to use the binder in upcoming co-crystal trials. However, due to the fact that Nb4 likely stabilizes the transporter in its outward-occluded conformation, by binding to an epitope that is comprised of N-terminal residues of *At*CTNS, as well as those positioned in the first and third luminal loops (**Fig. 7.6 b**; hydrogen bonds formed between Nb4 and the first luminal loop were discovered later in **Fig. 7.13 c**), it seemed unlikely that we would be able to solve the structure in an outward-open state using this nanobody. Therefore, we shifted our primary focus to capturing the transporter in a ligand-bound state.

We first crystallized *At*CTNS after complexation with L-cystine and Nb4, keeping the transporter unbound on the cytosolic side to allow it to shift into an occluded state and close around the ligand (we also used *At*CTNS and L-cystine without any binders to try this). However, similar to previous efforts without ligand, only few crystals formed, and only a small number of datasets were collected. These were solved by MR using the individual models of the *At*CTNS-Syb39-Nb4 structure (**Fig. 7.6 a**), however, *At*CTNS was found in the same conformation as before, and no additional electron density was observed inside the transporter that would fit the ligand (data not shown).

We then reverted to the double binder composition, aiming to stabilize the ligand-bound state for structure determination by trapping the ligand inside the transporter, as was achieved earlier for another member of the SLC superfamily (Parker *et al.*, 2021a). In addition to Syb39 and Nb4, we also used Nb1 and Syb1, not knowing what *At*CTNS epitope they bound. For these co-crystallization trials, *At*CTNS was first incubated with L-cystine to promote ligand binding, as demonstrated earlier using nanoDSF experiments (**Fig. 3.6**). Subsequently, a complex was formed using Nb4 and an additional binder (Syb39, Nb1, or Syb1) and isolated from unbound material by SEC in the presence of additional ligand. Due to the low solubility of L-cystine in aqueous solutions near neutral pH (Bannai, 1984), we mainly used L-cystine-*S*,*S*-dioxide (**Fig. 7.7 a**), a more soluble analogue of L-cystine that retains a similar binding profile in the nanoDSF experiments (**Fig. 3.6 c**) and ~80% of the L-cystine transport levels, as tested using our reconstituted system (see **Fig. 9.13 h**). The ensuing LCP co-crystallization yielded large crystals of cuboidal (using Syb39), irregular planar (Nb1 and Syb1), or parallelepipedal shape (Syb1), growing mostly at mildly acidic conditions as before, except for the crystals

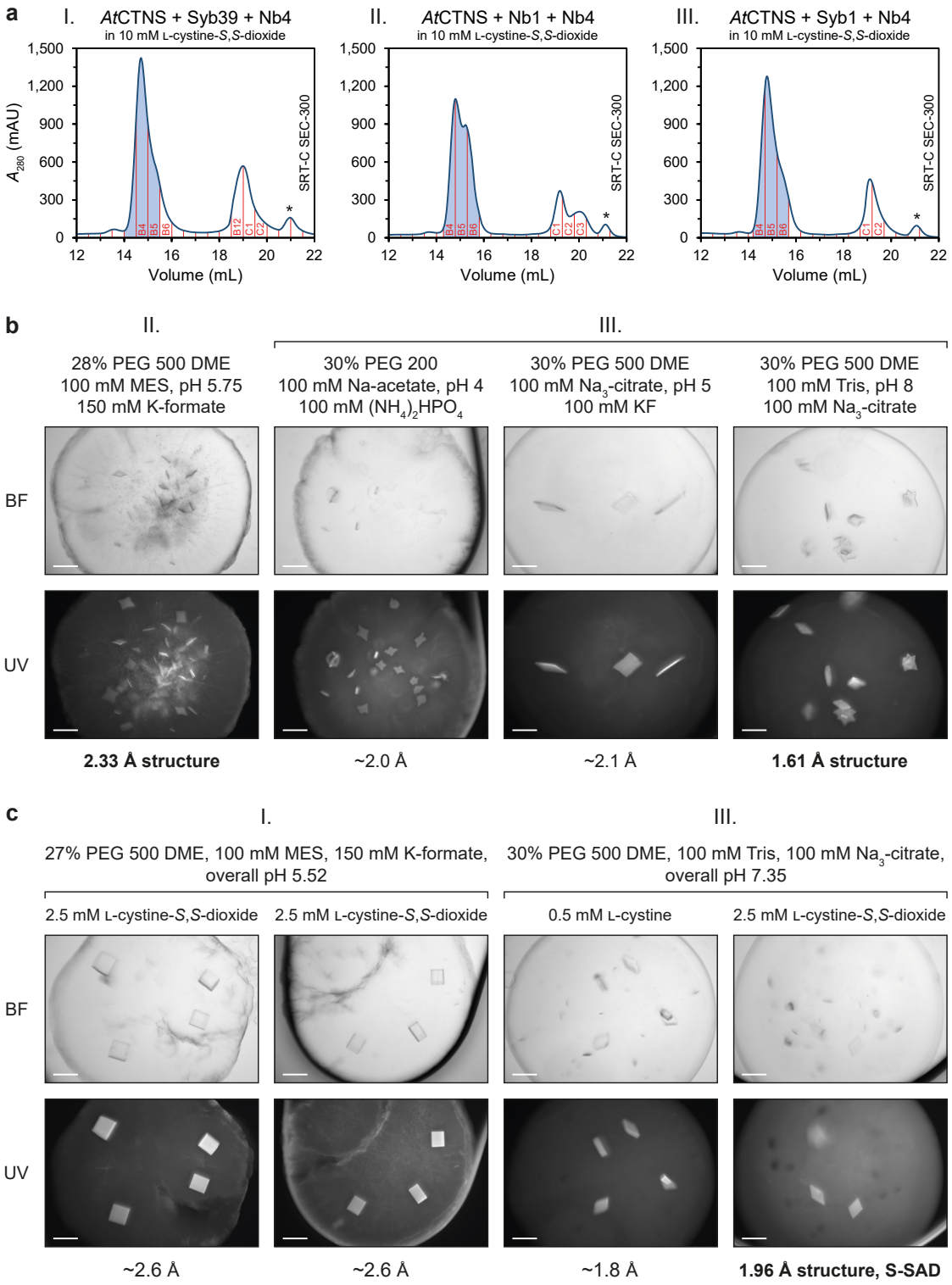


Fig. 7.7 – Complex Formation & Co-Crystallization Using Ligand-Supplemented Complexes. **a**, SEC co-elution traces of three different triple complexes, each supplemented with L-cystine-*S*,*S*-dioxide, a more soluble ligand for *At*CTNS compared to L-cystine (*highlights excess ligand eluting from the SEC column). Pooled fractions were used for LCP co-crystallization. **b** and **c**, BF and UV images of LCP boli containing crystals of *At*CTNS in complex with different binders – the exact composition is annotated *via* the SEC traces in **a**; co-crystallization trials in **c** contain additional ligand in the precipitant solution. Complexes were co-crystallized in MO using protein at 36.5–39 mg mL^{-1}. Annotated at the top of each column of images is the precipitant solution. The diffraction outcome is pinned at the bottom. Scale bars 100 µm.

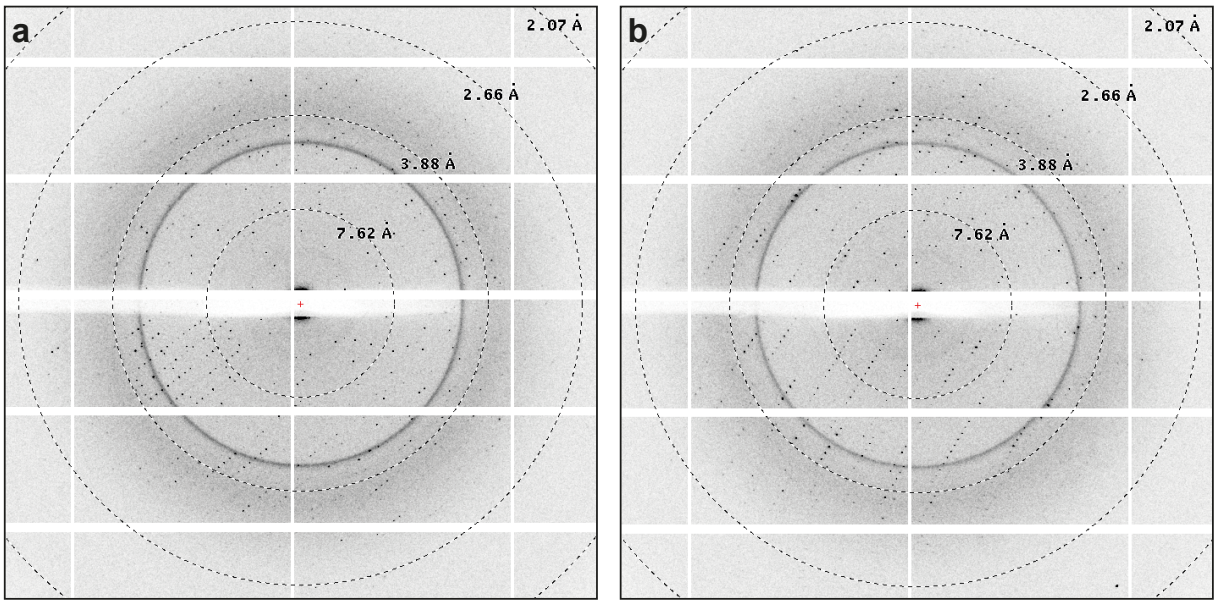


Fig. 7.8 – Typical Diffraction Patterns of *At*CTNS-Nb1-Nb4.
Diffraction patterns obtained at $\varphi = 0°$ (**a**) and $\varphi = 90°$ (**b**), visualized using Adxv (Arvai, 2021).

containing Syb1, which also grew at pH 7–8 (**Fig. 7.7 b** and **c**). Although thinner, crystals containing Nb1 or Syb1 diffracted to a higher resolution compared to all crystals obtained before. Their structures were again solved by MR using the models of the *At*CTNS-Syb39-Nb4 structure (**Fig. 7.6 a**) as individual search ensembles and subsequently refined as described before. Hereafter, I will present the datasets obtained, as well as the models built for each structure, and furthermore describe the insights we gained from them. These include the validation of the previous structure, interactions with novel α-*At*CTNS binders, potential lipid binding sites, as well as the unveiling of a crystallographic *At*CTNS dimer. However, they, too, do not include an alternative conformation of *At*CTNS or transporter-ligand interactions, as the data again did not generate additional electron density or anomalous S-SAD signal that would fit L-cystine or L-cystine-*S*,*S*-dioxide, even after supplementing the precipitant solutions with additional ligand (**Fig. 7.7 c**) to counteract potential unbinding during crystallization.

7.3.1 Higher-Resolution Structure of *At*CTNS Confirms Previous Model

The complex of *At*CTNS, Nb1, and Nb4, formed in the presence of 10 mM L-cystine-*S*,*S*-dioxide, crystallized in one of the conditions optimized for the previous binder combination (**Fig. 7.7 a** and **b**). The diffraction experiment was performed at DLS beamline I04 (**Table 7.4 a**) using an EIGER2 detector (Casanas *et al.*, 2016), allowing for rapid data collection at 0.005 s per image (see **Sec. 2.5.3**). The resulting diffraction patterns showed clear reflections to ~3 Å resolution and no signs of strong anisotropy or radiation damage (**Fig. 7.8**). The data were automatically processed to a maximum resolution of 2.33 Å using the xia2 DIALS pipeline (see **Sec. 2.5.4**).

Table 7.4 – X-Ray Data Collection & Processing Statistics of *At*CTNS-Nb1-Nb4.

a, Relevant sweep of data collected at DLS (proposal number mx23459-203).

Identifier	Beamline	Wavelength (Å)	Beamsize (μm)	Transmission (%)	Exposure (s)	Oscillation (°)	Total images	Resolution max. (Å)	Detector
7589	DLS - I04	0.9795	11 × 5	100.00	0.0050	0.10	3,600	1.80	EIGER2 XE 16M

b, Data processing.

Pipeline	Images used	High-resolution data cut	Space group	Cell dimensions *a*, *b*, *c* (Å)	Cell angles α, β, γ (°)	PDB
ISPyB - xia2 DIALS	All but 1,101–1,399	Automatic	*C*2	210.28, 77.46, 46.30	90.00, 93.25, 90.00	7ZKZ

c, Merging statistics by resolution bin and overall. *Significant at $p = 0.01$.

Resolution (Å)	Observations	Unique reflections	Multiplicity	Completeness (%)	Mean $I/\sigma(I)$	R_{merge} (%)	R_{meas} (%)	R_{pim} (%)	$CC_{1/2}$ (%)
104.87–6.32	10,566	1,659	6.37	99.9	33.7	5.8	6.4	2.6	99.7*
6.32–5.02	9,540	1,597	5.97	100	21.2	9.4	10.3	4.2	99.4*
5.02–4.39	10,498	1,614	6.50	100	24.0	9.3	10.1	3.9	99.3*
4.39–3.98	10,551	1,593	6.62	100	19.4	11.6	12.6	4.9	99.1*
3.98–3.70	10,712	1,590	6.74	100	14.0	14.6	15.8	6.1	98.8*
3.70–3.48	10,674	1,579	6.76	100	10.3	18.6	20.2	7.7	98.1*
3.48–3.31	10,368	1,570	6.60	100	7.9	22.3	24.2	9.3	97.7*
3.31–3.16	9,824	1,604	6.12	100	5.2	29.8	32.5	13.0	96.6*
3.16–3.04	9,622	1,593	6.04	100	4.2	34.9	38.3	15.4	95.1*
3.04–2.94	9,326	1,560	5.98	100	3.3	41.5	45.5	18.4	93.7*
2.94–2.84	10,015	1,602	6.25	100	2.5	54.6	59.5	23.4	90.3*
2.84–2.76	10,022	1,575	6.36	100	2.3	59.5	64.8	25.5	89.4*
2.76–2.69	10,218	1,580	6.47	100	1.8	78.1	84.9	33.0	84.3*
2.69–2.62	10,252	1,578	6.50	100	1.5	90.6	98.5	38.3	80.7*
2.62–2.56	9,994	1,533	6.52	100	1.2	105.7	114.8	44.5	76.7*
2.56–2.51	10,688	1,629	6.56	100	1.0	123.6	134.1	51.7	69.9*
2.51–2.46	10,447	1,579	6.62	100	0.8	146.4	158.8	61.0	59.2*
2.46–2.41	10,169	1,545	6.58	100	0.7	166.8	181.0	69.8	56.5*
2.41–2.37	10,583	1,593	6.64	100	0.7	183.2	198.6	76.1	52.2*
2.37–2.33	10,356	1,554	6.66	100	0.6	220.1	238.7	91.7	42.7*
104.76–2.33	**204,425**	**31,727**	**6.4**	**100**	**7.9**	**17.0**	**18.5**	**7.2**	**99.5**

Monoclinic space group *C*2 was assigned with cell dimensions $a = 210.28$ Å, $b = 77.46$ Å, and c = 46.30 Å, and angles $\alpha = \gamma = 90°$ and $\beta = 93.25°$ (**Table 7.4 b**). The merging statistics indicated a complete dataset with improved statistics compared to all data presented previously despite a comparatively low multiplicity of 6.4 (**Table 7.4 c**). R_{merge} and R_{meas} were at ~6% in the lowest-resolution shell, R_{pim} was <10% to a resolution ranging to ~3.3 Å, and the overall $CC_{1/2}$ was 99.5%, all indicative of high-quality and well-behaved data. Only the mean $I/\sigma(I)$ was relatively low, especially in the higher-resolution bins, however, the data still correlated well among random halves ($CC_{1/2}$ never <42.7%).

The structure was solved by MR in Phaser using the previous models of *At*CTNS and Nb4 as consecutive search ensembles. Additionally, a modified version of Nb4 with its CDR loops truncated was used to locate Nb1. Based on a V_M of 3.20 $Å^3\,Da^{-1}$ with a solvent content of ~62%, a single copy of the triple complex was expected in the asymmetric unit. The initial placement

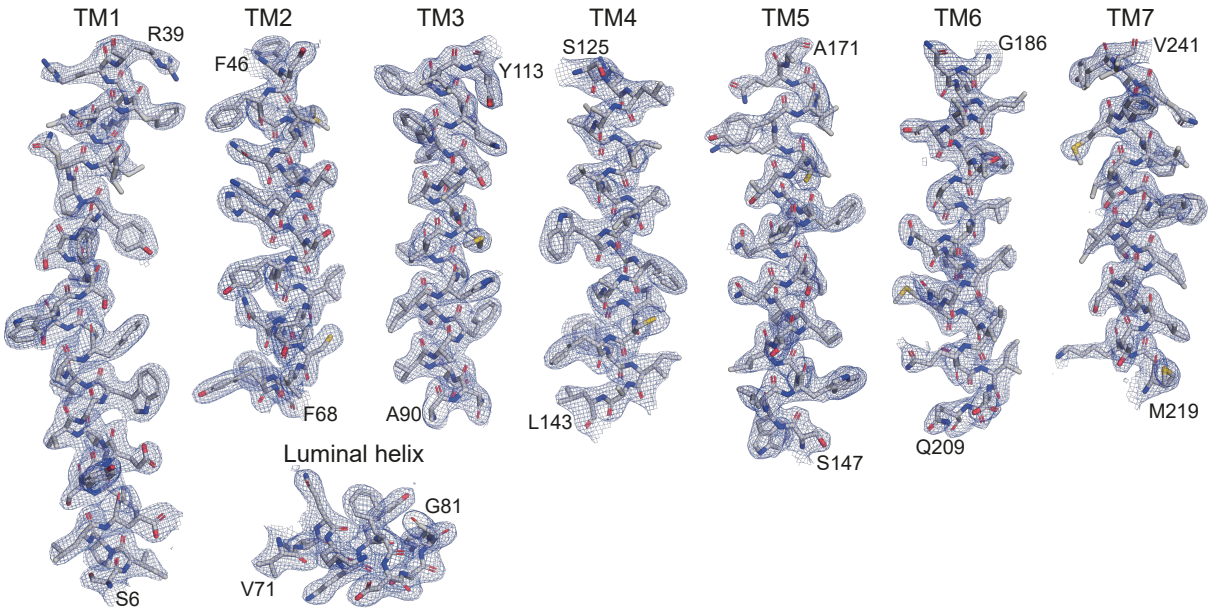


Fig. 7.9 – Essential Electron Density Maps of *At*CTNS-Nb1-Nb4 at 2.33 Å.
$2mF_o$–DF_c maps around all *At*CTNS helices are shown, contoured at 1 σ with a carve of 2 Å.

of *At*CTNS only scored a TFZ of 5.6, usually insufficient for a correct solution, however, the TFZ for a correctly placed model might be lower in the monoclinic space group (Oeffner *et al.*, 2018). The LLG for *At*CTNS of 143, as well as the subsequent placements of the binders, which yielded TFZs of 26.6 and 27.3, and LLGs of 1,103 and 2,080 for Nb4 and Nb1, respectively, were all good signs that the prior and overall placements were correct. MR was then followed up with 10 rounds of automatic refinement in REFMAC, generating initial *R*-factors of R_{work} = 30% and R_{free} = 33%. The models were adjusted manually based on the observed electron and difference densities and further refined in BUSTER using TLS parameterization, as described before. The missing CDR loops of Nb1 were built into the maps, and differing side chains from the Nb4-based search model were mutated and fitted. Parts of the *At*CTNS model (the end of TM5 and the third cytoplasmic loop) needed to be removed, as indicated by negative difference density, and re-building was not possible due to the lack of adequate electron and positive densities. The resulting $2mF_o$–DF_c map, highlighting the fit for all helical *At*CTNS residues, is illustrated in **Fig. 7.9**. Additional density that would fit the supplemented ligand was not observed.

The finalized structure featured *R*-factors of 23.5 and 26.5% and an improved F_o,F_c correlation coefficient of 94.1% compared to the previous structure (83.1%). The model was also found to be more chemically sound as analysed using MolProbity: A higher percentage of residues were found with energetically favoured peptide angles (Ramachandran *et al.*, 1963) as well as side chain rotamers (Dunbrack Jr & Karplus, 1994), leading to an improved MolProbity score of 1.80 (see **Table 7.7** for a comparison of all structures).

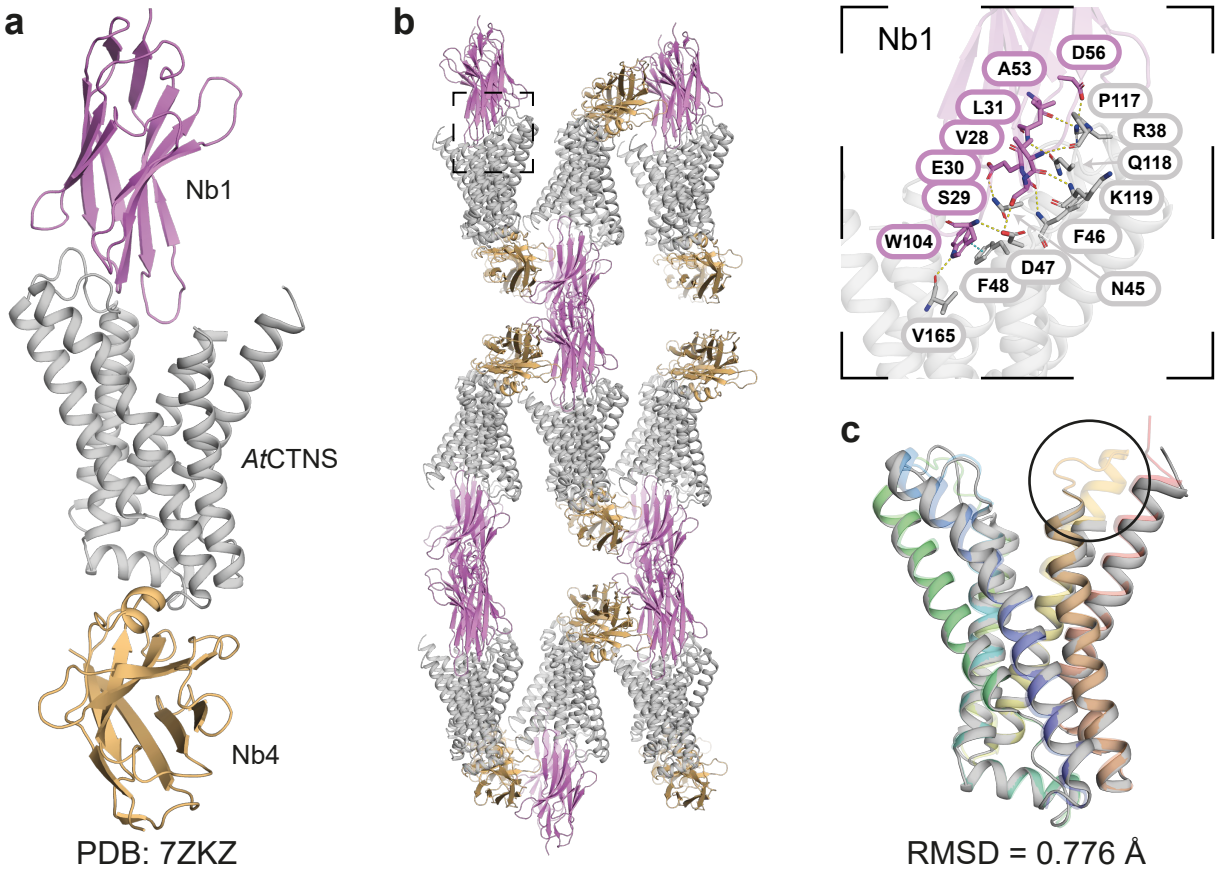


Fig. 7.10 – 2.33 Å Co-Crystal Structure of *At*CTNS-Nb1-Nb4.
a, Cartoon representation of the structure of *At*CTNS in complex with Nb1 and Nb4 (PDB: 7ZKZ). **b**, Crystal packing of **a**, showing the contribution of the binders towards crystal contact formation. Specific *At*CTNS-binder interactions are shown, with intermolecular hydrogen bonds highlighted using yellow dashed lines and π-π stacking using cyan dashed lines. **c**, Superimposition with the *At*CTNS model from the *At*CTNS-Syb39-Nb4 structure (**Fig. 7.6**; in rainbow colours), highlighting the region of disorder in the present structure.

The finalized model of *At*CTNS bound to Nbs 1 and 4 resembles the layout of the first co-crystal structure, with Nb1 taking the place of Syb39, interacting with *At*CTNS on its cytosol-facing side (**Fig. 7.10 a**). Binding is mainly mediated through CDRs 1 and 2 of Nb1, forming several hydrogen bonds with residues located at the cytoplasmic ends of TMs 1 and 2, as well as on the first and second cytoplasmic loops of *At*CTNS, connecting TMs 1–2 and 3–4 (**Fig. 7.10 b**). The third CDR loop only contributes with a single residue, W104, which interacts with F48 on TM2 through π-π stacking and a hydrogen bond with the backbone carbonyl oxygen of V165 on TM5. Nb1 does not interact with the third cytoplasmic loop of *At*CTNS, which may be the reason why no electron density was found to fit the cytoplasmic end of TM5 as well as the connection to TM6, highlighting the fact that this region may be very flexible.

The llama nanobodies are again responsible for the formation of important crystal contacts (**Fig. 7.10 b**): As before, Nb4 bridges neighbouring *At*CTNS molecules by specifically binding to one and hydrophobically interacting with the other *via* its backbone, defining the front-to-back

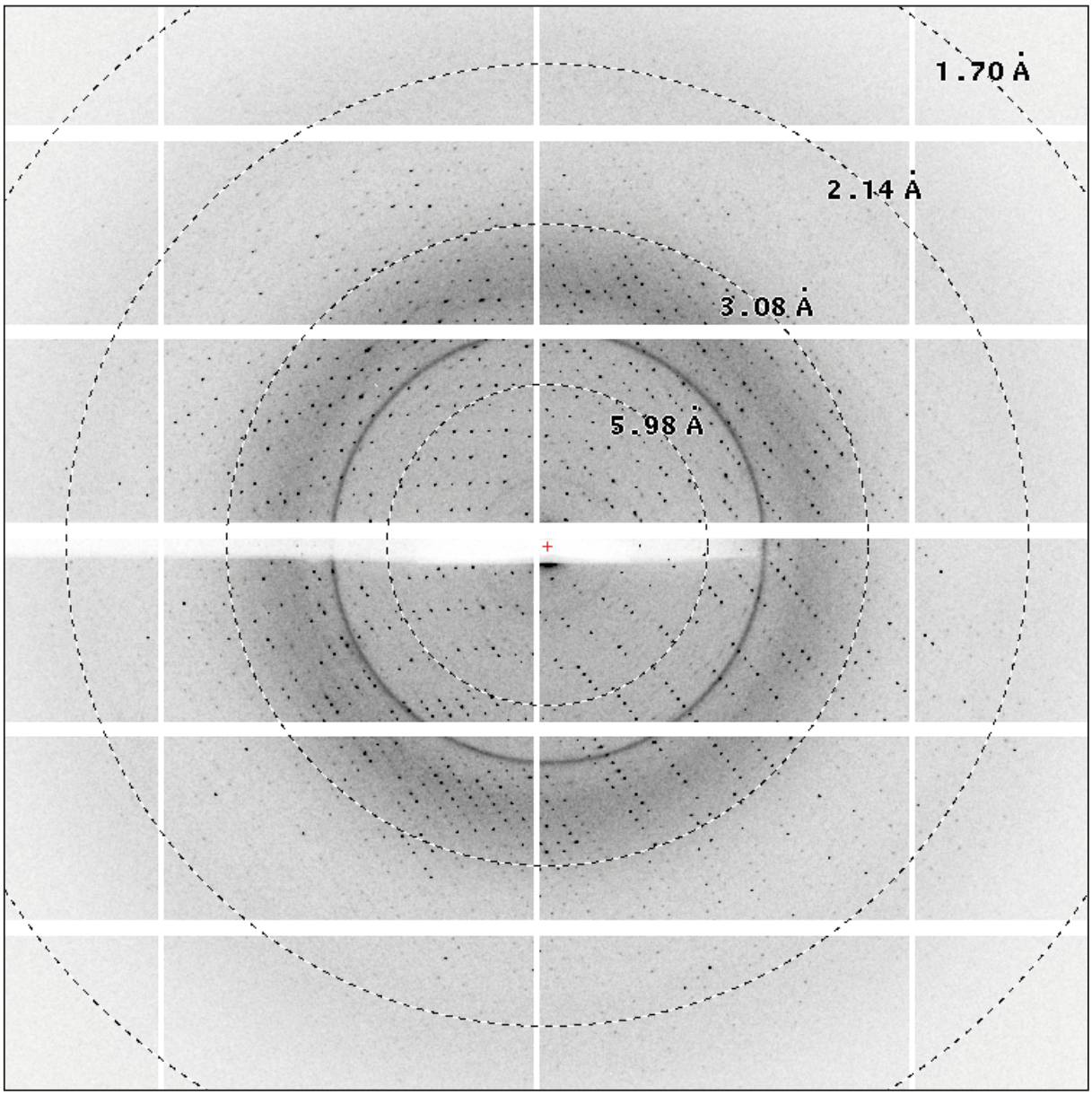


Fig. 7.11 – Typical Diffraction Pattern of *At*CTNS-Syb1-Nb4.

plane (compare to **Fig. 7.6 b**). The three-dimensional crystal structure is completed through Nb1 contacts that are formed with additional Nb1 as well as Nb4 molecules, generating a crystal composition that is denser than the one before, which is also reflected by the V_M of 3.20 Å^3 Da^{-1} and solvent content of ~62%. Overall, the structure of the transporter is in agreement with the one obtained before, superimposing with an RMSD of 0.776 Å over all 224 C_α atoms, with the only difference being the disordered region of the cytoplasmic end of TM5 and the connecting loop to TM6 (**Fig. 7.10 c**).

Table 7.5 – X-Ray Data Collection & Processing Statistics of *At*CTNS-Syb1-Nb4.

a, Relevant sweep of data collected at DLS (proposal number mx23459-203).

Identifier	Beamline	Wavelength (Å)	Beamsize (μm)	Transmission (%)	Exposure (s)	Oscillation (°)	Total images	Resolution max. (Å)	Detector
7561	DLS - I04	0.9795	11 × 5	100.00	0.0080	0.10	3,600	1.50	EIGER2 XE 16M

b, Data processing.

Pipeline	Images used	High-resolution data cut	Space group	Cell dimensions *a*, *b*, *c* (Å)	Cell angles α, β, γ (°)	PDB
ISPyB - xia2 DIALS	1–3,350	Automatic	$P2_12_12$	66.17, 92.14, 102.21	90.00, 90.00, 90.00	----

c, Merging statistics by resolution bin and overall. *Significant at $p = 0.01$.

Resolution (Å)	Observations	Unique reflections	Multiplicity	Completeness (%)	Mean $I/\sigma(I)$	R_{merge} (%)	R_{meas} (%)	R_{pim} (%)	$CC_{1/2}$ (%)
47.61–4.37	52,366	4,381	11.95	100	33.7	6.7	7.0	2.0	99.0*
4.37–3.47	52,040	4,185	12.43	100	28.2	8.1	8.5	2.4	99.8*
3.47–3.03	54,609	4,152	13.15	100	20.5	11.5	11.9	3.3	99.6*
3.03–2.75	54,916	4,112	13.36	100	14.2	15.6	16.2	4.4	99.4*
2.75–2.56	51,120	4,097	12.48	100	10.0	20.4	21.2	6.0	98.9*
2.56–2.40	49,136	4,120	11.93	100	7.1	26.6	27.8	8.0	98.3*
2.40–2.28	49,541	4,051	12.23	100	5.5	33.3	34.8	9.9	97.6*
2.28–2.19	51,565	4,075	12.65	100	4.5	39.3	40.9	11.4	96.8*
2.19–2.10	52,072	4,060	12.83	100	3.5	48.5	50.5	14.0	96.2*
2.10–2.03	52,872	4,073	12.98	100	2.7	51.2	63.7	17.6	94.4*
2.03–1.97	52,830	4,048	13.05	100	2.2	75.0	78.0	21.4	93.2*
1.97–1.91	53,476	4,068	13.15	100	1.6	100.9	104.9	28.8	89.3*
1.91–1.86	52,706	4,032	13.07	100	1.2	129.7	134.9	37.1	83.5*
1.86–1.81	48,915	4,062	12.04	100	1.0	154.2	161.0	46.0	80.8*
1.81–1.77	50,096	4,015	12.48	100	0.8	198.8	207.2	58.2	71.5*
1.77–1.73	46,989	4,037	11.64	100	0.6	233.8	244.6	71.2	58.3*
1.73–1.70	49,372	4,041	12.22	100	0.5	295.9	308.9	87.7	52.7*
1.70–1.67	50,661	4,051	12.51	100	0.5	337.4	351.8	98.9	48.9*
1.67–1.64	50,821	4,029	12.61	100	0.4	434.0	452.3	126.5	37.7*
1.64–1.61	51,341	4,029	12.74	100	0.3	516.6	538.2	149.9	31.8*
47.59–1.61	**1,027,444**	**81,718**	**12.6**	**100**	**7.1**	**16.9**	**17.6**	**4.9**	**99.5**

7.3.2 High-Resolution Structure of *At*CTNS Reveals Dimeric Arrangement

The complex of *At*CTNS, Syb1, and Nb4, formed in the presence of 10 mM L-cystine-*S*,*S*-dioxide, yielded large, three-dimensional crystals in different conditions (**Fig. 7.7 a** and **b**). Using DLS beamline I04 and crystals grown at slightly basic pH, high-resolution diffraction patterns were obtained, showing spots clearly observable to ~2 Å (**Fig. 7.11**). The data were processed to a maximum resolution of 1.61 Å in orthorhombic space group $P2_12_12$ with unit cell dimensions $a = 66.17$ Å, $b = 92.14$ Å and $c = 102.21$ Å (**Table 7.5**). The data quality was found to be at a similar high level compared to the *At*CTNS-Nb1-Nb4 dataset, with the main differences being the increased number of observed and unique reflections, as well as the increased multiplicity, resulting from the higher-symmetry space group (**Table 7.5 c**).

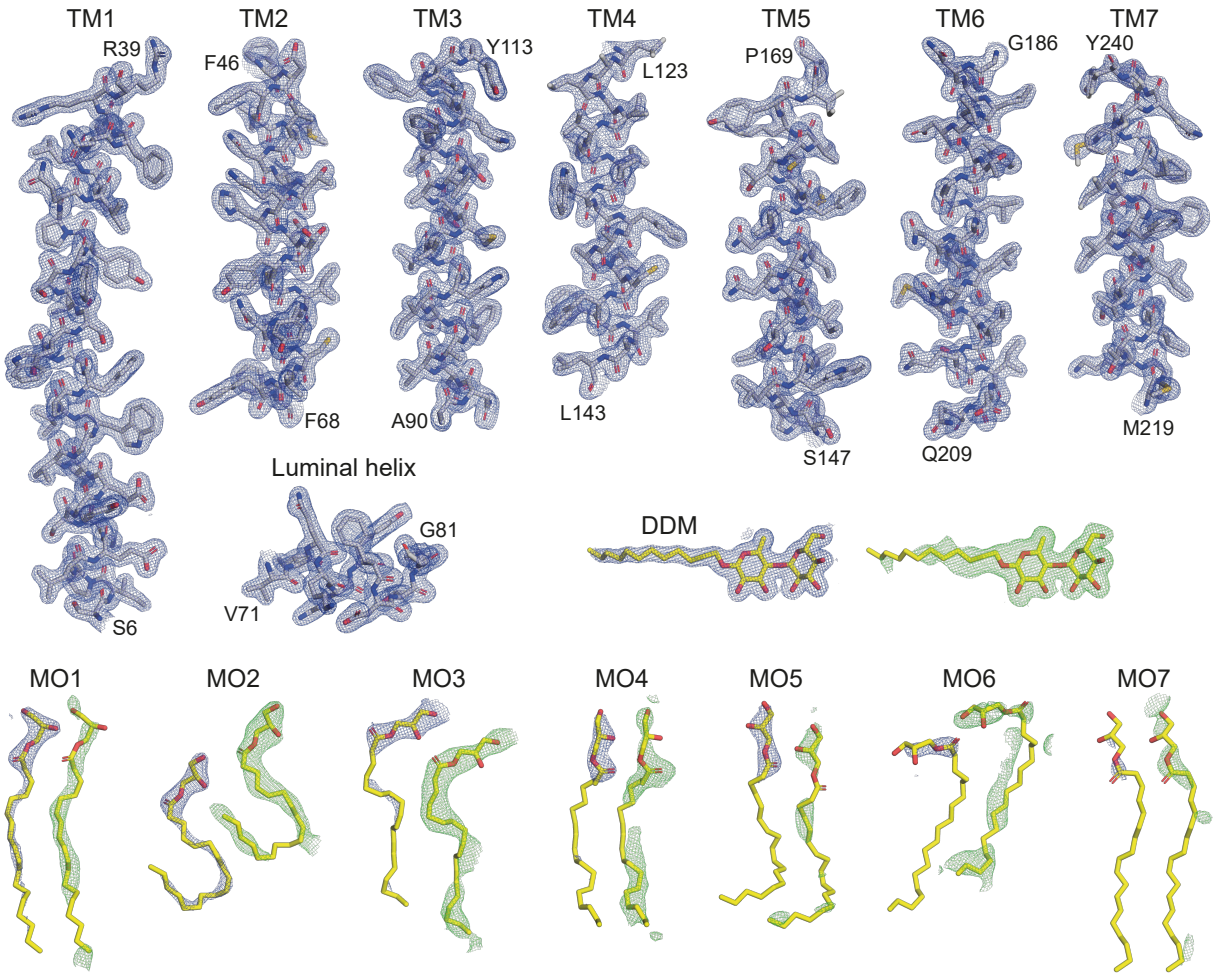


Fig. 7.12 – Essential Electron Density Maps of *At*CTNS-Syb1-Nb4 at 1.61 Å.
$2mF_o$–DF_c maps around all *At*CTNS helices as well as bound DDM detergent and MO lipid molecules are shown as blue mesh, contoured at 1 σ with a carve of 1.5 Å. Additionally, mF_o–DF_c polder OMIT maps are shown for DDM and MO molecules, illustrated as a green mesh, contoured at 3 σ with a carve of 2 Å.

Structure solution by MR, performed as described previously, turned out to be trivial as Phaser generated TFZs of 9.9, 29.6, and 43.4, as well as LLGs of 162, 1,017, and 2,625 for the consecutive placement of *At*CTNS, Nb4, and Syb1 (search models from the *At*CTNS-Syb39-Nb4 structure; CDRs of Syb39 removed; one copy of each expected in the asymmetric unit based on a V_M of 2.68 Å^3 Da^{-1} and a solvent content of ~54%). Following initial refinement in REFMAC (R_{work} = 30% and R_{free} = 34%), the structure was manually optimized in Coot and further refined in BUSTER. Positive difference density in the shape of an alkyl maltoside was observed between the surfaces of *At*CTNS and Syb1. Considering the detergent in which the transporter was solubilized, this was modelled as a DDM molecule. Additional densities around the hydrophobic surface of *At*CTNS were built as MO lipids stemming from the LCP, although they were often too large for the densities. This may be explained by a low occupancy of the MO molecule at set position, thus, they only mark potential lipid binding sites. Again,

extra density for the supplemented ligand L-cystine-*S*,*S*-dioxide was not observed. Maps are illustrated in **Fig. 7.12**.

The *R*-factors reported after completion of the refinement were R_{work} = 22.5% and R_{free} = 25.4%, the lowest for any of the co-crystal structures presented in this thesis. The *At*CTNS-Syb1-Nb4 structure additionally featured the lowest anisotropic *B*-factor for all averaged protein atoms at 36.3 Å^2, indicating the lowest uncertainty for the placement of these atoms. DDM detergent and water molecules were associated with *B*-factors of 40.8 and 47.5 Å^2, respectively, and MO lipids were linked to a *B*-factor of 84.3 Å^2, again emphasizing the reduced confidence for the atomic position of the placed lipid molecules. The *At*CTNS-Syb1-Nb4 model also scored the best MolProbity score with a value of 0.87. This resulted from only nine poor rotamers and no Ramachandran outliers, as detected by MolProbity. Refinement and validation statistics for all structures are compared in **Table 7.7**.

The finalized model once again features *At*CTNS sandwiched between two binders, this time Syb1 and Nb4 (**Fig. 7.13 a**), whereby Syb1 takes the place of Syb39 and Nb1 from the previous models and interacts with the transporter on its cytosolic side. The seven MO lipids bind the surface of *At*CTNS, and the DDM detergent molecule is embedded in the *At*CTNS-Syb1 interface. Display of symmetry-related molecules reveals a crystallographic *At*CTNS dimer, joint *via* three TM helices and a MO lipid from each symmetry mate (**Fig. 7.13 b**). The dimer interface is discussed in more detail later (**Sec. 8.1**). In the context of the presented structure, it represents a major crystal contact that is not mediated by any of the binders. As a result, the components of the *At*CTNS-Syb1-Nb4 crystal are packed much denser compared to those used previously, which is manifested in the comparatively low V_M of 2.68 Å^3 Da^{-1}. With only small gaps in between individual transporter units, the solvent content of ~54% arises from large channels, which run in between the binders that connect individual triple complex dimers (**Fig. 7.13 c**). Each binder hereby interacts with one molecule of the same type and three of the other, *i.e.* every Syb1 interacts with another Syb1 and three Nb4 molecules, and *vice versa*, thus building the crystal in all three dimensions.

The presented high-resolution structure confirms the *At*CTNS-Nb4 binding interface observed previously but also reveals additional interactions between the molecules due to a more precise placement of relevant side chains, as well as the addition of water molecules within the interface. As a result, it becomes apparent that Nb4 not only binds the N-terminus and third luminal loop of the transporter but also interacts with the first luminal loop, including the luminal helix located between TMs 2–3. CDR1 residues I29 and S33, as well as CDR3 residue E104, are involved in these interactions, forming hydrogen bonds with K74, K83, and I86 of *At*CTNS either directly or *via* water molecules (**Fig. 7.13 c**).

Syb1 binding is mediated through several interactions, some of which are formed *via* the DDM detergent molecule bound within the *At*CTNS-Syb1 interface (**Fig. 7.13 c**). Hydrogen bonds

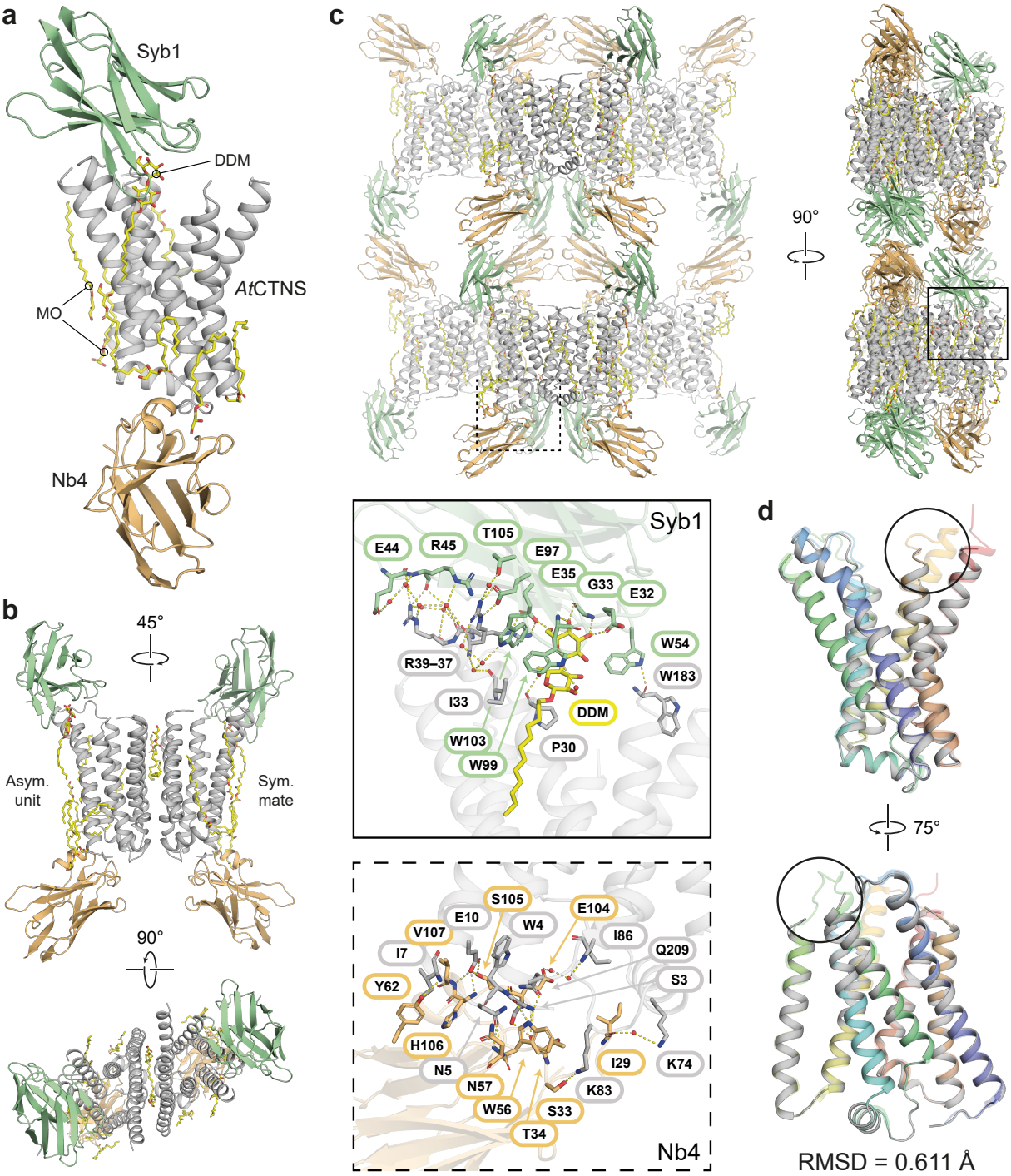


Fig. 7.13 – 1.61 Å Co-Crystal Structure of *At*CTNS-Syb1-Nb4.
a, Cartoon representation of the structure of *At*CTNS in complex with Syb1 and Nb4. Bound lipids and a DDM detergent molecule are shown as sticks. **b**, Front and top view of the dimer arrangement between *At*CTNS symmetry mates. **c**, Crystal packing of **a** and **b**, showing the contribution of the binders towards crystal contact formation. Specific *At*CTNS-binder interactions are shown, with intermolecular hydrogen bonds and salt bridge interactions highlighted using yellow dashed lines, and water molecules are represented as spheres. **d**, Superimposition with the *At*CTNS model from the *At*CTNS-Syb39-Nb4 structure (**Fig. 7.6**; in rainbow colours), highlighting the regions of disorder in the present structure.

with multiple Syb1 residues are formed, including G33 and E35 located on CDR1 as well as W99 on CDR3, all of which are highly conserved among the concave binders identified in the sybody screening (**Fig. 5.5**). With *At*CTNS, the DDM molecule forms only a single hydrogen bond (P30 located on TM1), however, additional hydrophobic contacts are made. Another important epitope for Syb1 binding is the end of TM1 of *At*CTNS, which includes a basic patch consisting of three consecutive arginines (R37–39). There are two salt bridges formed between R38 and the aforementioned E35, as well as R37, which interacts with E97, another highly conserved residue on the Syb1 CDR3. Additionally, a number of hydrogen bonds are formed both directly and indirectly *via* a defined water network. These include residues E44 and R45 from the sybody frame, as well as W103 and T105, positioned on the CDR3 loop. No interactions are formed between Syb1 and the second or third cytoplasmic loop of the transporter, causing the electron density to fade in these areas despite the high overall resolution, similar to what was observed for the triple complex structure using Nb1. The electron density in the TM5–6 loop only reappears from W183 onwards, which is positioned close to the start of TM6 and stabilized by a hydrogen bond formed with W54, located on the Syb1 CDR2. The missing cytosolic loops are highlighted in **Fig. 7.13 d**, showing the superimposition of the current model with the complete model of the *At*CTNS-Syb39-Nb4 structure. It additionally demonstrates the agreement between the models as they align with an RMSD of 0.611 Å across the remaining 217 C_α atoms.

7.3.3 High-Resolution S-SAD Data Validate *At*CTNS Model

Even though past triple complexes were formed in the presence of ligand, none of the co-crystal structures featured adequate electron density that would fit the substrate, indicating it was not bound. Because L-cystine and L-cystine-*S*,*S*-dioxide contain disulfides, we again used S-SAD to try and locate possibly bound ligands from the anomalous difference maps. Crystals were re-grown in optimized precipitant solutions that were additionally supplemented with 0.5 mM L-cystine or 2.5 mM L-cystine-*S*,*S*-dioxide to minimize potential ligand dissociation during the crystallization process. Crystals of the *At*CTNS-Syb39-Nb4 and *At*CTNS-Syb1-Nb4 compositions were successfully regenerated (**Fig. 7.7 c**). Subsequent S-SAD data collection at DLS beamline I23, again performed in collaboration with Dr. Kamel El Omari and Dr. Armin Wagner, resulted in several high-resolution datasets, one of which is described hereafter.

The *At*CTNS-Syb1-Nb4 composition (1.96 Å dataset; processing statistics are presented in **Table A.9**) solved automatically using PHENIX AutoBuild (Terwilliger *et al.*, 2008) as implemented in the DLS automatic downstream processing pipelines. The data showed a strong anomalous signal to ~3 Å resolution as indicated by the mean $d''/\sigma(d'')$ and $CC_{1/2}$ columns generated by SHELXC (**Table 7.6 a**). Using ANODE (Thorn & Sheldrick, 2011), we mapped the anomalous signal onto the previous *At*CTNS-Syb1-Nb4 structure (**Fig. 7.13**), which was solved from an isomorphous crystal. Strong anomalous peaks of up to ~19 σ were generated, and

Table 7.6 – Anomalous Data of *At*CTNS-Syb1-Nb4 (S-SAD).

a, S-SAD data statistics generated by SHELXC.

Resolution (Å)	Unique reflections	Mean $I/\sigma(I)$	Anomalous compl. (%)	Anomalous multiplicity	R_{pim} (%)	R_{anom} (%)	Mean $d''/\sigma(d'')$	Anom. $CC_{1/2}$ (%)
68.22–10.50	353	74.6	97.5	10.0	1.91	3.83	2.01	94.1
10.50–6.46	1,046	65.3	99.5	11.4	1.94	3.87	1.83	83.2
6.46–4.87	1,741	54.7	99.8	11.6	2.01	4.02	1.54	65.4
4.87–3.98	2,516	50.9	99.8	11.2	1.72	3.45	1.17	40.8
3.98–3.40	3,277	33.1	99.7	10.6	2.21	4.42	0.96	22.9
3.40–3.00	3,867	19.0	97.7	10.1	3.65	7.30	0.82	12.8
3.00–2.69	4,514	10.5	93.3	9.6	6.68	13.36	0.79	5.2
2.69–2.45	5,038	5.8	89.9	9.4	12.41	24.83	0.71	3.2
2.45–2.26	5,452	3.1	87.2	9.3	25.68	51.36	0.72	-2.5
2.26–2.10	5,999	1.8	83.6	8.6	44.35	88.69	0.66	-9.6
2.10–1.96	5,850	1.0	69.7	8.5	67.24	134.50	0.64	-19.1

b, S-SAD peaks generated by ANODE using the finalized *At*CTNS-Syb1-Nb4 model (**Fig. 7.13**). Peaks with a minimum height of 4 σ are listed.

Atom peak	Peak height (σ)	Distance to nearest atom (Å)	Atom	Residue	Chain	Molecule
S1	19.15	0.944	SG	C22	C	Syb1
S2	15.75	0.129	SD	M202	A	*At*CTNS
S3	14.61	0.443	SG	C65	A	*At*CTNS
S4	13.45	0.171	SD	M83	C	Syb1
S5	13.09	0.374	SG	C138	A	*At*CTNS
S6	13.04	0.121	SG	C99	B	Nb4
S7	12.96	0.203	SD	M85	A	*At*CTNS
S8	12.61	0.312	SG	C22	B	Nb4
S9	11.98	0.119	SD	M37	B	Nb4
S10	11.53	0.172	SG	C164	A	*At*CTNS
S11	11.47	0.341	SD	M86	B	Nb4
S12	10.85	0.764	SD	M102	A	*At*CTNS
S13	9.92	0.204	SD	M162	A	*At*CTNS
S14	9.66	0.564	SD	M219	A	*At*CTNS
S15	8.75	0.543	SD	M60	A	*At*CTNS
S16	8.30	0.984	CE	M34	C	Syb1
S17	7.21	0.460	SD	M50	A	*At*CTNS
S18	4.66	2.017	CE	M236	A	*At*CTNS
S19	4.44	1.154	O	F111	A	*At*CTNS
S20	4.41	0.744	OG	S225	A	*At*CTNS
S21	4.20	4.364	O	HOH102	W	H_2O
S22	4.09	1.407	H17	OLC1	G	MO
S23	4.06	3.745	O	HOH120	W	H_2O
S24	4.00	9.620	H18B	OLC1	E	MO

their locations were compared to the sulfur positions in the model (**Table 7.6 b** and **Fig. 7.14**). Overall, 17 peaks with a signal of >7.2 σ were located. With an average distance of ~0.4 Å, these were found to coincide well with the sulfurs in the model. M236, positioned at the C-terminal end of *At*CTNS, only generated a weak signal of 4.66 σ and was found ~2 Å away from the

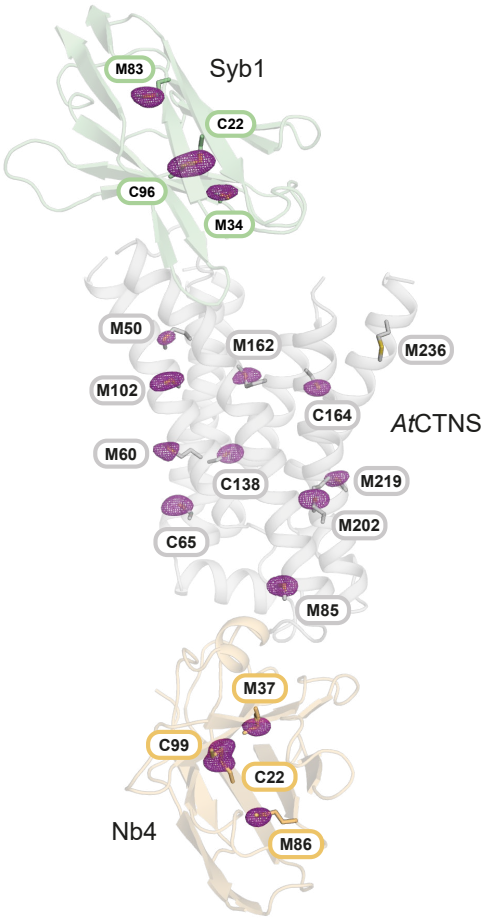


Fig. 7.14 – Anomalous Difference Map of *At*CTNS-Syb1-Nb4.
Anomalous difference maps calculated using the finalized *At*CTNS-Syb1-Nb4 model (**Fig. 7.13**) with the S-SAD data (**Table 7.6 a**) in ANODE, indicating the location of S atoms *via* a purple mesh, contoured at 5 σ. Anomalous peaks and their closest atom in the model are listed in **Table 7.6 b**.

modelled position, highlighting the probable flexibility of the C-terminal tail. The transporter's N-terminal M1 as well as M173, which is located within the third cytosolic loop, did not generate anomalous signals, consistent with the finalized *At*CTNS-Syb1-Nb4 model where they were also absent (**Fig. 7.13**). In summary, the S-SAD data provided additional validation for the model's accuracy. However, once again, no additional signal was observed that would hint at the presence of a ligand, rendering the crystal trials using ligand-supplemented complexes, at least with the presented triple complexes and the described workflow, unsuccessful for determining a ligand-bound state of *At*CTNS.

7.4 Lower-Resolution Ligand-Bound Structure of *At*CTNS

Extensive co-crystallization trials using purified *At*CTNS first incubated with ligands L-cystine or L-cystine-*S*,*S*-dioxide and subsequently bound to two binders that interact with both the cytosolic and lysosomal luminal sides of the transporter, failed to provide a crystal structure in the

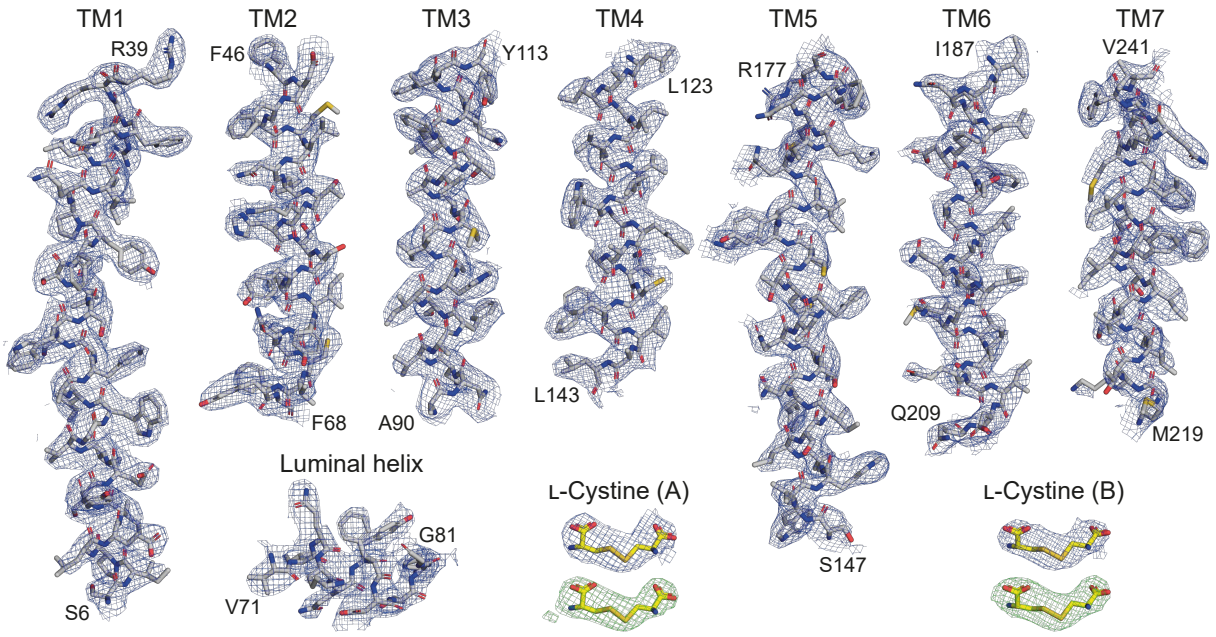


Fig. 7.15 – Essential Electron Density Maps of *At*CTNS-Syb39-Cystine at 3.47 Å.
$2mF_o–DF_c$ maps around all *At*CTNS helices of chain B as well as bound L-cystine molecules are shown as blue mesh, contoured at 1 σ with a carve of 2 Å. Additionally, $mF_o–DF_c$ polder OMIT maps, generated for the substrate bound to each transporter (chains A and B), are shown as green mesh, contoured at 5 σ with a carve of 2 Å.

ligand-bound state (**Sec. 7.3**). This was despite our attempts to minimize the dissociation of the substrate by adding it to both the SEC buffer and the precipitant solution (**Fig. 7.7**), essentially soaking the crystals from the time they nucleated until they were harvested. Ultimately, we obtained a ligand-bound structure from data collected much earlier using a different workflow. Before the nanobody screening was completed, we attempted to optimize the needle-shaped *At*CTNS-Syb39 crystals from **Fig. 6.2** by adding 0.5 mM L-cystine to the isolated complex. This mix was incubated for at least 2 h on ice and subsequently crystallized using the LCP method. The morphology of the resulting crystals and their diffraction behaviour remained unchanged, though, and thus the collected data were unprofitable in the pursuit of solving the first crystal structure of *At*CTNS. When solving these structures by MR using the models obtained in this chapter, we discovered that they indeed contained positive $mF_o–DF_c$ density that fitted the ligand well, thereby yielding the first ligand-bound structure of the transporter.

The lattice system and data quality of the *At*CTNS-Syb39-cystine crystals remained essentially unchanged from that of their ligand-free counterpart described in **Sec. 6.1.3** (compare **Tables A.10** and **6.1**): Monoclinic space group $C2$ was assigned with $\beta \approx 100°$ and one cell dimension around five times longer than each of the other two. The mean $I/\sigma(I)$ was listed at 2.0, the R_{pim} of the lowest-resolution shell was slightly above 10%, and the overall $CC_{1/2}$ was ~80%. The crystal again exhibited pseudo-merohedral twinning with twin law $-h–2l$, $-k$, l and a twin fraction of 0.156–0.311, as indicated during automatic processing. Nevertheless, the data were

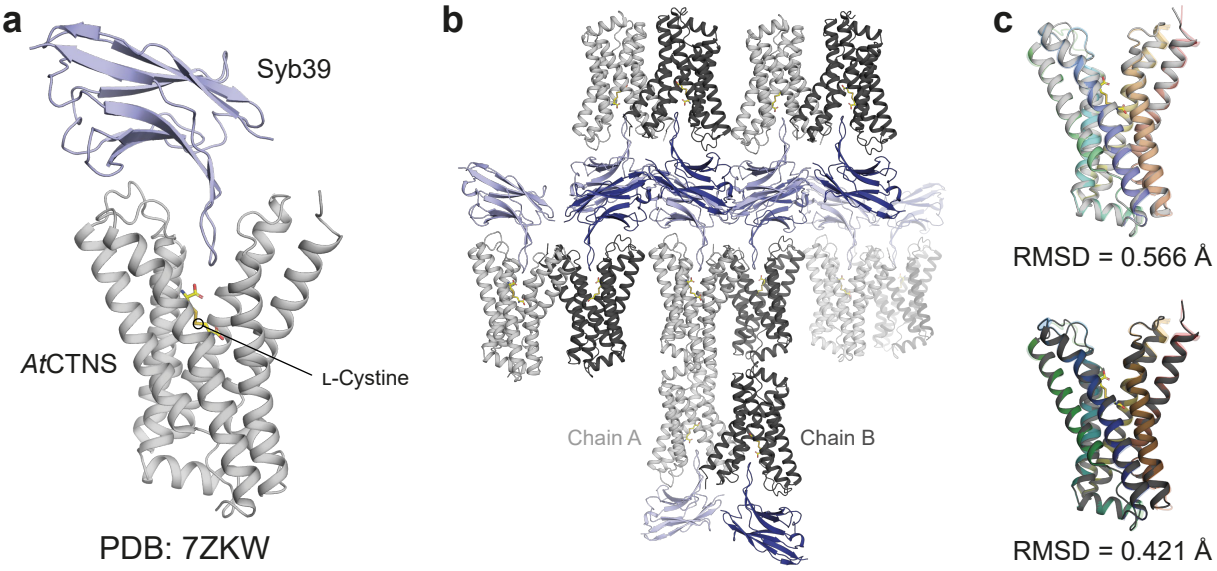


Fig. 7.16 – 3.47 Å Co-Crystal Structure of Ligand-Bound *At*CTNS in Complex with Syb39. **a**, Cartoon representation of one of the two heterodimers of *At*CTNS-Syb39 present in the asymmetric unit. The bound substrate is shown as sticks. **b**, Crystal packing of **a**, showing the contribution of the binder towards crystal contact formation. **c**, Superimposition with the *At*CTNS model from the *At*CTNS-Syb39-Nb4 structure (**Fig. 7.6**; in rainbow colours).

complete and produced a detailed electron density map considering the resolution of 3.47 Å (**Fig. 7.15**).

The structure was solved in Phaser by placing two copies of both *At*CTNS and Syb39 in the asymmetric unit, as was indicated by the calculated V_M of 2.73 Å^3 Da^{-1} with a solvent content of ~55%. As before, the search models, taken from the first co-crystal structure (PDB: 7ZK1, **Fig. 7.6**), were placed confidently, scoring consecutive TFZs of 9.1 and 18.2 (both *At*CTNS), 21.3 and 21.4 (both Syb39), and LLGs of 132 and 405, as well as 824 and 1,632, respectively. MR was followed by 10 rounds of automatic refinement in REFMAC, yielding initial *R*-factors of R_{work} = 28% and R_{free} = 31%. The model was then optimized manually in Coot, ligand L-cystine was built into the observed electron density, and the structure was refined in REFMAC using twin refinement and TLS parameterization (Murshudov *et al.*, 2011). The final twin fraction estimated by REFMAC was 0.204; final *R*-factors were reported at R_{work} = 23.5% and R_{free} = 27.7%. Essential electron density maps, highlighting the fit for all *At*CTNS helices and the ligand, are illustrated in **Fig. 7.15**. This includes mF_o–DF_c polder OMIT maps for the ligand, generated in PHENIX, whereby bulk solvent is prevented from entering the omitted region during refinement, potentially improving the map for weak ligand densities (Liebschner *et al.*, 2017).

The finalized model of *At*CTNS in complex with Syb39 and L-cystine is illustrated in **Fig. 7.16 a**, highlighting the position of the ligand within the core of the transporter. Interactions with side chain residues of *At*CTNS are analysed, discussed, and correlated with functional data on the transporter later, and led to the formulation of a mechanism for specific L-cystine recognition

Table 7.7 – Crystal Structure Refinement & Validation Statistics.
Comparison of the refinement and validation statistics generated for the four major structures described in this thesis. Ramachandran plots, generated by MolProbity (Lovell *et al.*, 2003), are illustrated in **Figs. A.5** to **A.8**.

	***At*CTNS-Syb39-Nb4 (Apo)**	***At*CTNS-Nb1-Nb4 (Apo)**	***At*CTNS-Syb1-Nb4 (Apo)**	***At*CTNS-Syb39 (L-cystine bound)**
PDB code	7ZK1	7ZKZ	----	7ZKW
Identifier	6774	7589	7561	6396
Refinement program	BUSTER	BUSTER	BUSTER	REFMAC
Resolution range (Å)	79.99–2.65	20.79–2.33	23.08–1.61	69.78–3.47
Completion for range (%)	93.4	99.3	98.4	92.8
Number of reflections	26,170	31,641	80,377	12,426
R_{work}/R_{free} (%)	23.4/27.5	23.5/26.5	22.2/25.4	23.5/27.7
Free *R*-value test set count/size (%)	1,331/5.09	1,566/4.95	4,062/5.05	631/4.80
Correlation coefficient F_o,F_c (%)	83.1	94.1	95.0	83.3
Number of atoms:				
Protein	3,925	3,784	3,666	5,898
Ligand	----	----	210	28
Water	32	108	318	----
Average *B*-factor (Å^2):				
Protein	85.0	67.0	36.3	72.4
L-Cystine	----	----	----	64.6
DDM detergent	----	----	40.8	----
Monoolein lipids	----	----	84.3	----
Water	64.4	61.9	47.5	----
RMSDs:				
Bond lengths (Å)	0.009	0.008	0.008	0.014
Bond angles (°)	1.05	0.97	0.87	1.55
MolProbity score	2.97	1.80	0.87	2.92
Clashscore, all atoms	10.69	3.33	0.26	18.2
MolProbity counts/frequencies (%):				
Poor rotamers	69/16.5	29/7.11	9/2.30	40/6.33
Favoured rotamers	298/71.3	349/85.5	371/94.6	513/81.2
Ramachandran outliers	12/2.47	3/0.64	0/0	18/2.50
Ramachandran favoured	444/91.5	462/97.9	444/98.5	646/89.7
Bad bonds	3/0.07	1/0.03	1/0.03	13/0.21
Bad angles	17/0.31	7/0.13	3/0.06	91/1.10

and transport *via* CTNS (**Secs. 9.3.1** and **9.3.3**). Interestingly, despite being ligand-bound, *At*CTNS still adopts an inward-open conformation, unchanged from the conformation of the previous structures in an apo state, and superimposes with the first co-crystal structure (PDB: 7ZK1, **Fig. 7.6**) to an RMSD of 0.566 and 0.421 Å across all 243 C_α atoms when using the first and second copy of *At*CTNS present in the asymmetric unit (chains A and B), respectively (**Fig. 7.16 c**).

The quality of the model was analysed as before and compared to the previous structures in **Table 7.7**. Despite the lower resolution, the *At*CTNS-Syb39-cystine model attained a decent MolProbity score of 2.92, similar to the first co-crystal structure, ranking it in the 91st percentile for structures of similar resolution (±0.25 Å; $n = 646$). The average anisotropic *B*-factor for all protein atoms at 72.4 Å^2 even indicated a smaller uncertainty for their placement compared to the first structure. Importantly, the average anisotropic *B*-factor for all ligand atoms was smaller

(64.6 $Å^2$) and within a similar range compared to neighbouring side chain residues of *At*CTNS, thus expressing confidence for the presence of the ligand in the co-crystal structure.

CHAPTER 8

Structural Studies on Human CTNS

Numerous PQ-loop proteins have been shown to oligomerize both *in vivo* and *in vitro*. Two members of plant SWEETs (*At*SWEET1 and 13) reportedly facilitate sugar transport as functional dimers (Xuan *et al.*, 2013; Han *et al.*, 2017). Structure solution of *Os*SWEET2b yielded a crystallographic trimer that is stable in solution (Tao *et al.*, 2015). The human KDELR was shown to form different oligomers depending on its conformational state and lipid composition (Wu *et al.*, 2022), likely regulating complex formation with other proteins (Jia *et al.*, 2024). The observation that *At*CTNS forms a crystallographic dimer (**Fig. 7.13 b**), therefore, led us to investigate whether the human transporter also oligomerizes in solution. In this chapter, I will show that both full-length *Hs*CTNS and the N-terminal truncation *Hs*CTNS(Δ2–115) can be purified from yeast cell culture, yielding thermally stable products. However, due to the low expression of the full-length protein, oligomerization and structural analyses were carried out with the truncated version. I will illustrate that *Hs*CTNS(Δ2–115) contains both a monomeric and dimeric fraction in solution and present the three-dimensional structure of the dimer, which we obtained from cryo-EM trials using the protein in complex with a sybody to enable proper particle alignment. The cryo-EM structure depicts *Hs*CTNS(Δ2–115) in the outward-open (binding site facing the lysosomal luminal) conformation, allowing for comparison with the crystal structures of **Ch. 7**, which captured *At*CTNS in the inward-open conformation (see **Ch. 9**). Interestingly, the arrangement of the *Hs*CTNS(Δ2–115) dimer differs fundamentally from that of the *At*CTNS crystal dimer. Whether full-length *Hs*CTNS also forms oligomers and the relevance to its molecular function could not be answered within the scope of this thesis and will need further investigation.

8.1 *At*CTNS Dimer as a Potential Template for *Hs*CTNS Dimerization

The *At*CTNS crystal dimer exhibits a symmetrical arrangement with a crystallographic twofold rotational axis dissecting its interface (**Fig. 8.1 a**). The contact area of the dimer comprises TMs 4, 5, and 7 and stretches over ~3,350 Å^2. The interface consists of 30 residues positioned up to 4 Å away from their counterpart, as well as two MO lipids (MO4 in **Fig. 7.12**). The dimer is stabilized through hydrophobic interactions and hydrogen bonds (**Fig. 8.1 b**). Hydrophobic

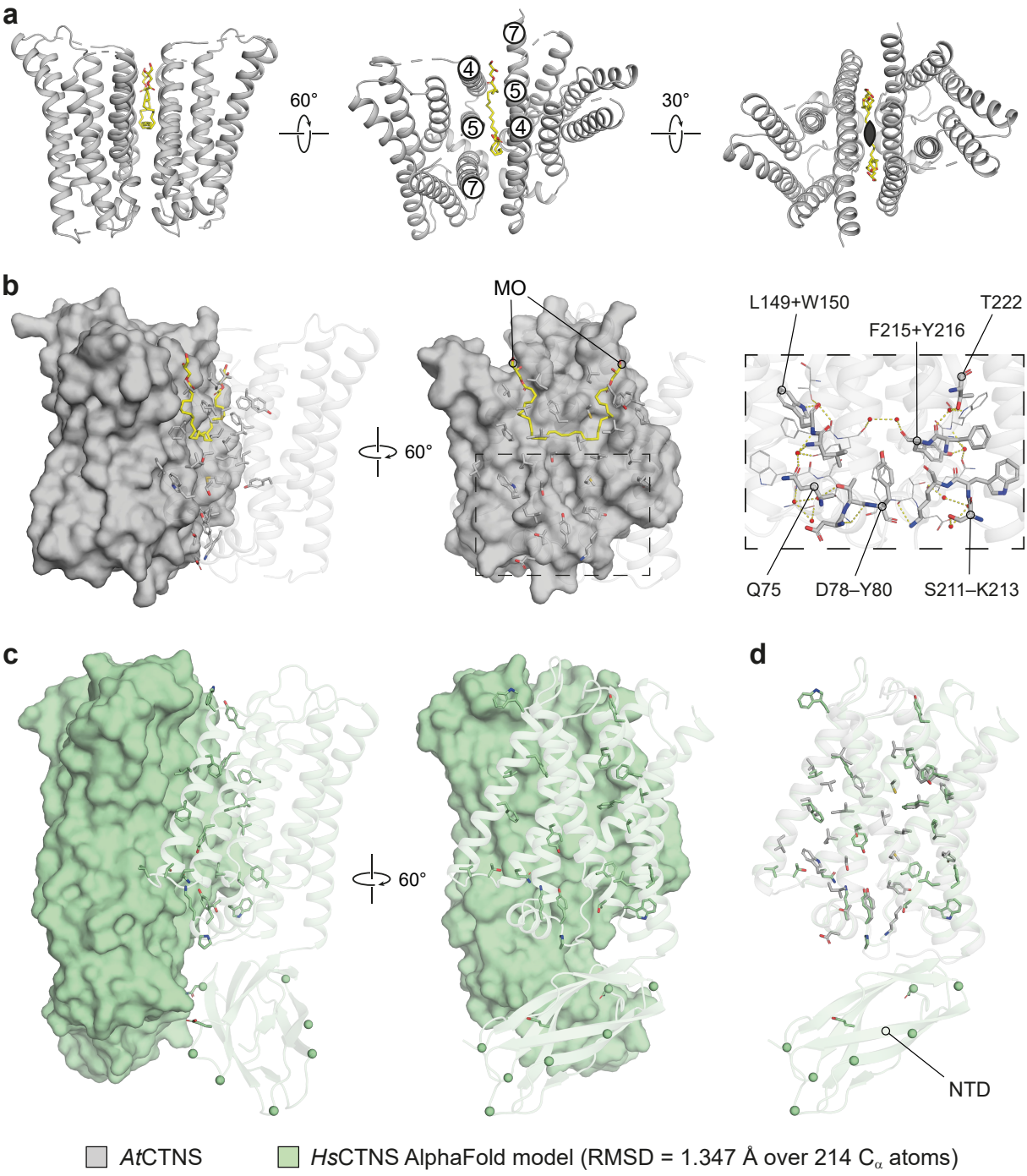


Fig. 8.1 – *At*CTNS Crystal Dimer Interface & Comparison with *Hs*CTNS.
a, Cartoon representation of the *At*CTNS crystal dimer from **Fig. 7.13**. **b**, Surface and cartoon representations of *At*CTNS highlighting side chain residues within a 4 Å distance of the dimer interface, illustrated as sticks. Hydrogen bond (yellow dashes) forming residues are illustrated as sticks and lines for the first and second monomer, respectively. **c**, The *Hs*CTNS AlphaFold model (UniProt: O60931) was used for superimposition with the *At*CTNS crystal dimer to recreate the same interface. Again, residues within 4 Å of the dimer interface are shown as sticks, and N-linked glycosylation sites are highlighted as single spheres. **d**, Comparison of **b** and **c**.

contacts are mainly formed within the top half of the transporter (cytosolic side), which is also where the MO lipids are bound. The other half of the dimer interface features several hydrogen

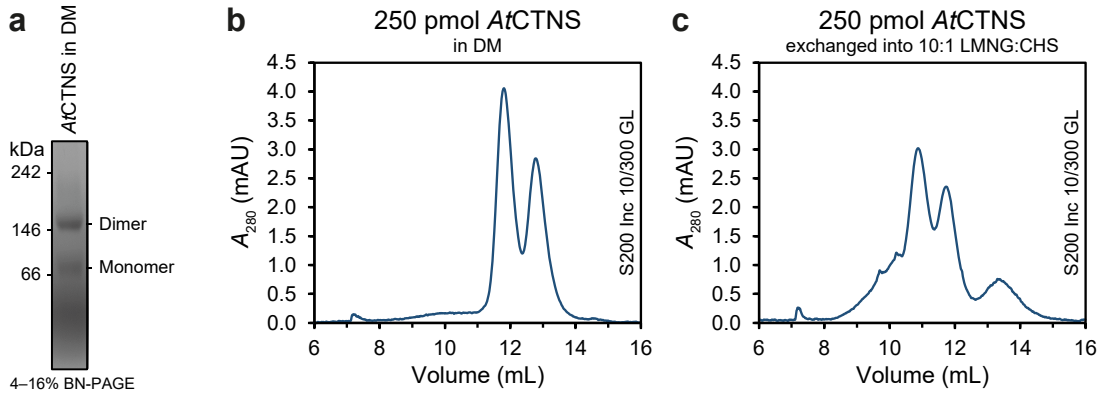


Fig. 8.2 – Oligomeric State of *At*CTNS in Solution.
a, BN-PAGE analysis of *At*CTNS purified in DM. **b**, Analytical SEC of the same sample used in **a**. **c**, Analytical SEC after a 1 h incubation in the presence of 0.1% (w:v) 10:1 (w:w) LMNG:CHS.

bonds, most of which are formed *via* bound water molecules. When superimposing the dimer with the human AlphaFold prediction (RMSD = 1.347 Å over all 214 C_α atoms), the formation of similar interactions remains feasible (**Fig. 8.1 c**). Although the formation of hydrogen bonds is difficult to predict due to the absence of solvent atoms, 24 predominantly hydrophobic residues were found within the same distance from the interface as before, showing potential van der Waals interactions. Strikingly, no side chain clashes are observed despite the presence of the NTD, which is absent in *At*CTNS (**Fig. 8.1 d**). However, N66, one of seven glycosylation sites (Nevo *et al.*, 2017), is positioned close to the dimer interface, leaving little room for a complex N-linked glycan. Overall, we deemed a *Hs*CTNS dimer based on the template of the *At*CTNS crystal dimer possible.

To learn more about the oligomeric state of *At*CTNS in solution, we performed native PAGE and analytical SEC experiments (**Fig. 8.2**). We found that two separate species co-existed in solution, irrespective of the chosen detergent, most likely representing a monomeric and dimeric form of *At*CTNS. At this point, we learned that the fractions pooled in previous purifications (**Figs. 3.3**, **3.4**, and **5.1**) mainly contained dimeric protein, and the occasional right-handed shoulder contained monomeric protein that was inseparable from the dimer when loading mg amounts of protein on the SEC column. This was also confirmed by SEC with multi-angle light scattering (not shown).

8.2 Screening & Characterization of *Hs*CTNS Constructs

8.2.1 Expression Trials

Due to the low expression of the full-length *Hs*CTNS construct (isoform 1) in yeast cells (**Fig. 3.2**), we designed a number of different constructs for testing using the experimental platform described in **Ch. 3** to identify the most suitable construct for structural studies. The

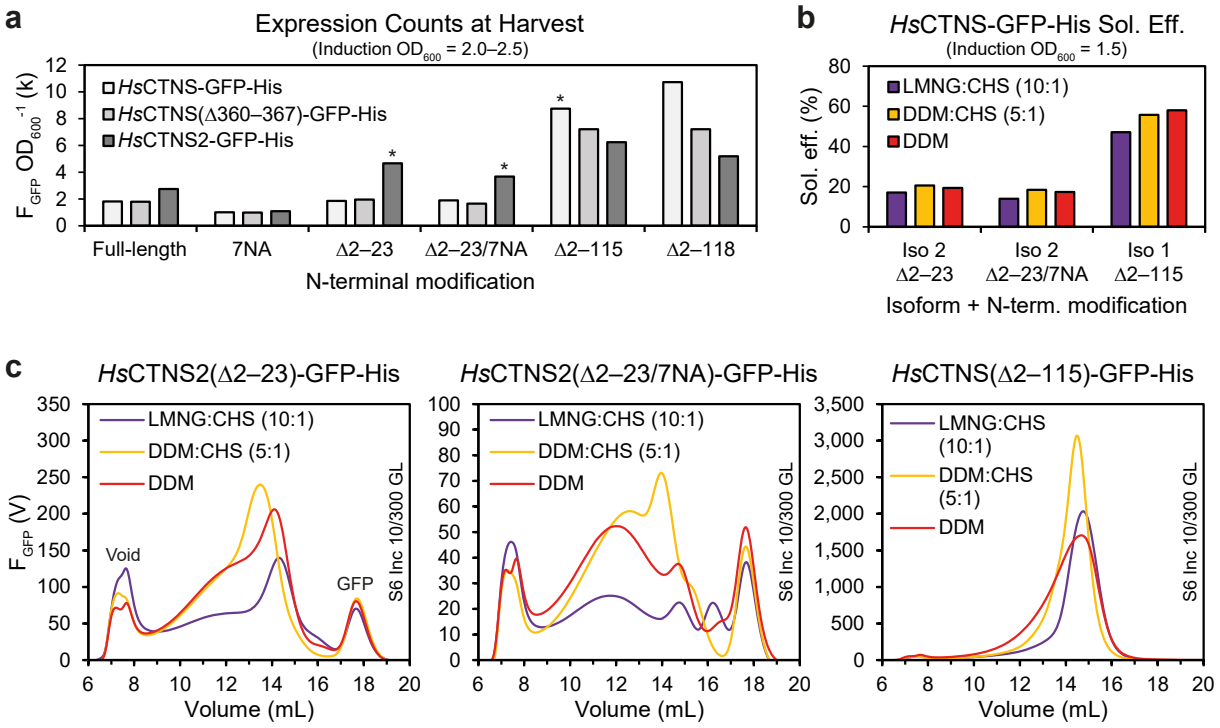


Fig. 8.3 – Expression Trials of Human CTNS Constructs in Yeast Cells.
a, Expression counts at the time of harvest (~16 h post-induction) for 18 *Hs*CTNS constructs with varying N- and C-terminal modifications. **b**, Solubilization efficiencies, and **c**, FSEC profiles for three selected constructs of **a** (*) and in combination with detergents outlined in **Table 3.1**.

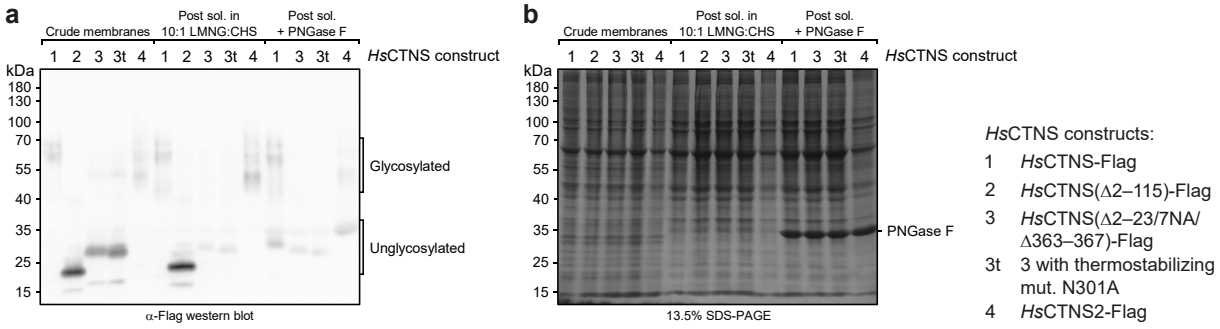


Fig. 8.4 – Expression Trials of Human CTNS Constructs in Mammalian Cells.
Gel-based analysis of five different *Hs*CTNS constructs: α-Flag western blot in **a** and subsequent Coomassie stain in **b**. Thermostabilizing mutation as reported by Guo *et al.* (2022).

combination of six N-terminal variations (full-length; mutations N36A, N41A, N51A, N66A, N84A, N104A, and N107A (7NA) to eliminate all seven N-linked glycosylation sites; Δ2–23 to remove the N-terminal signal peptide; both combined; and two constructs lacking the NTD: Δ2–115, retaining a short N-terminal tail as observed in *At*CTNS, and Δ2–118) with three C-terminal modifications (full-length isoforms 1 and 2 as well as isoform 1 without the LTM: Δ360–367) yielded 18 constructs for a systematic screening (**Fig. 8.3**). Constructs containing the NTD expressed considerably better when combined with the C-terminus of the second isoform, however, solubilization efficiencies were generally low (<20%), and FSEC traces did not show

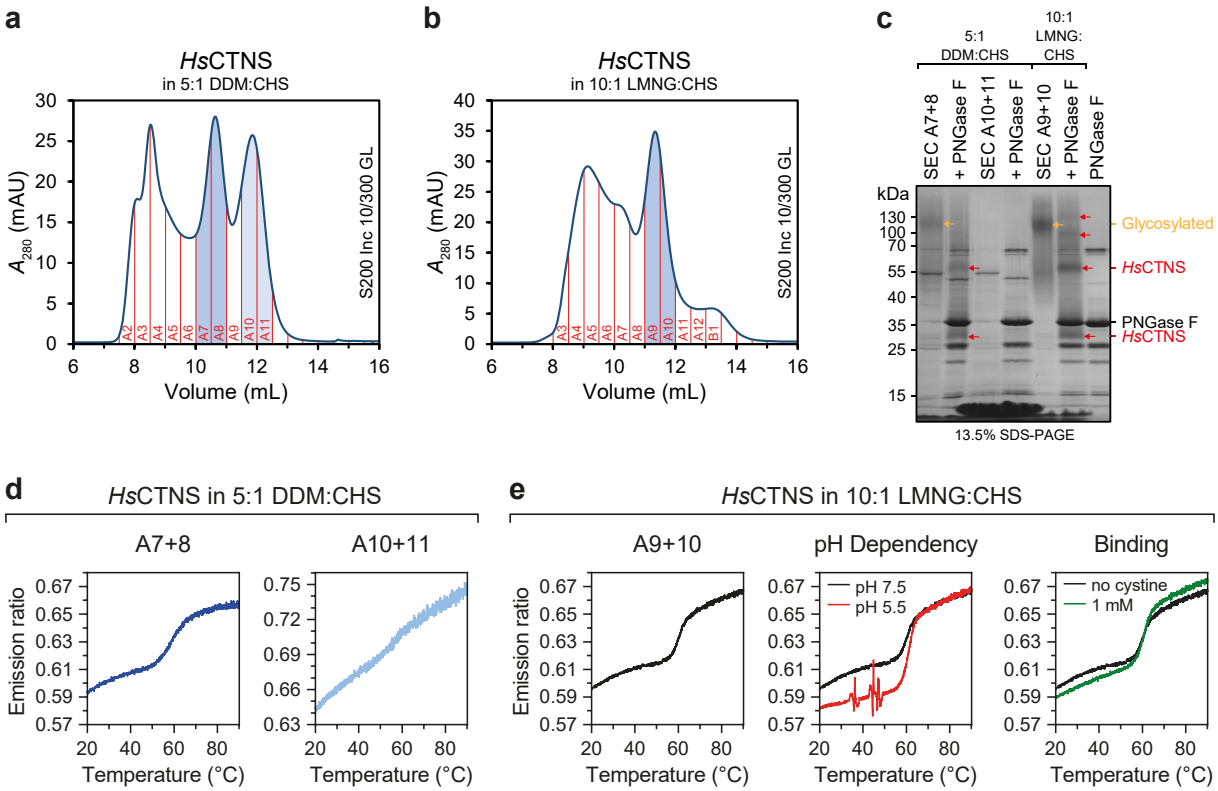


Fig. 8.5 – Purification & Thermal Stability Analysis of *Hs*CTNS Expressed in Yeast Cells. **a** and **b**, SEC profiles of *Hs*CTNS purified in different detergents. Pooled and combined fractions are shaded in blue. **c**, SDS-PAGE analysis of the final purification samples, both untreated and after a 1 h incubation in the presence of PNGase F. **d**, NanoDSF profiles of the pooled samples from **a**. **e**, NanoDSF profiles of the pooled sample from **b**, also testing its pH dependency and the influence of a supplemented ligand as described previously in **Figs. 3.5** and **3.6**, respectively.

monodisperse peaks. Constructs without the NTD expressed significantly better, approaching the levels observed for *At*CTNS, and exhibited an improved solubilization behaviour (~50–60%). FSEC analysis provided strong and nearly monodisperse signals, improving with a rising CHS concentration. We also attempted expression in mammalian HEK293F cells and found that, again, the construct without the NTD exhibited an improved expression and solubilization behaviour compared to those retaining it (**Fig. 8.4**).

8.2.2 Purification Trials & Biophysical Characterization

In an attempt to retain the NTD, initial purification trials were focused on the full-length human transporter (isoform 1) due to its superior FSEC traces compared to the alternative NTD-containing constructs (compare **Figs. 3.2** and **8.3 c**). However, we quickly noticed that the low expression would hamper downstream applications as purifications in DDM:CHS and LMNG:CHS detergents only yielded <10 µg L^{-1} yeast cell culture per SEC peak (**Fig. 8.5 a** and **b**, respectively). Interestingly, the purification products exhibited glycosylated protein, which, upon treatment with peptide:N-glycosidase F (PNGase F), generated an SDS-PAGE profile

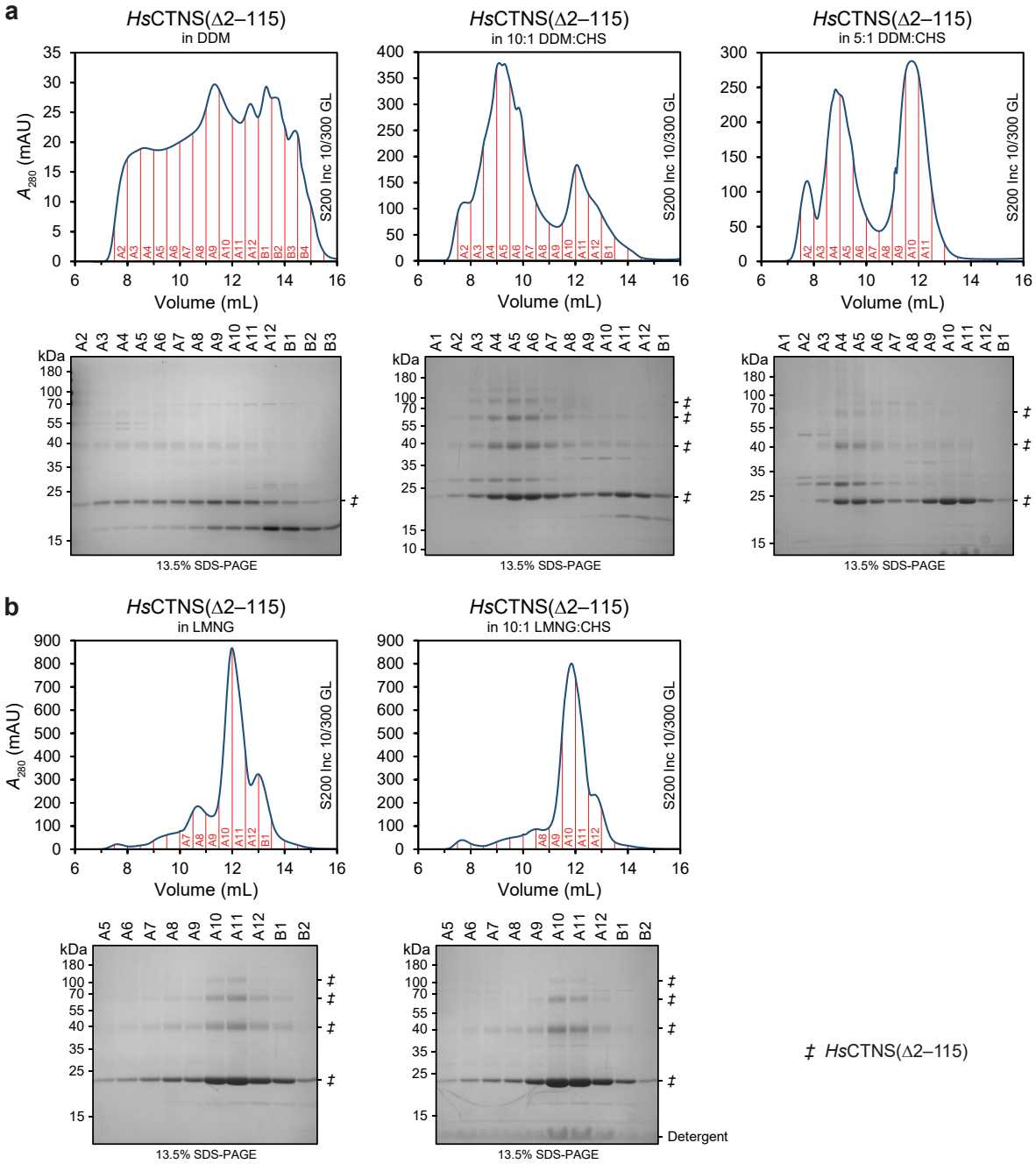


Fig. 8.6 – *Hs*CTNS(Δ2–115) Purification Detergent Screening.
SEC profiles and SDS-PAGE analysis of *Hs*CTNS(Δ2–115) purified in DDM (in **a**) and LMNG (in **b**), both with and without CHS.

similar to that of *At*CTNS (**Fig. 3.3 b**), showing a series of bands that may represent different oligomeric states (most notable for *Hs*CTNS in 10:1 LMNG:CHS in **Fig. 8.5 c**). NanoDSF experiments indicated a T_m of ~60 °C for most species, showing a red-shift melting curve (**Fig. 8.5 d** and **e**), contrary to *At*CTNS (**Fig. 3.5 c**). Further differences were observed regarding the pH dependency, where *Hs*CTNS was not stabilized significantly upon acidification of the environment (**Fig. 8.5 e**). However, L-cystine binding was again indicated through a reduction in tryptophan fluorescence before the melting point was reached, similar to *At*CTNS (**Fig. 3.6**).

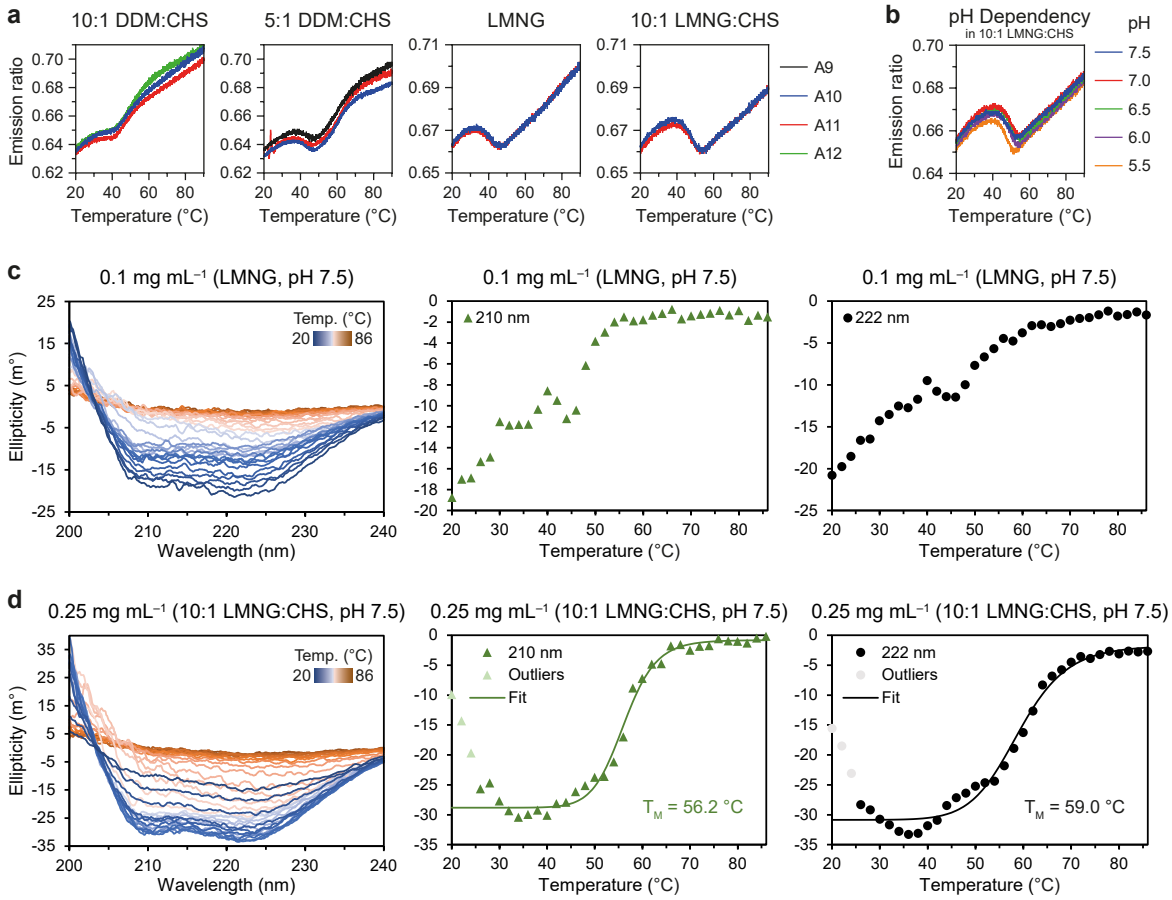


Fig. 8.7 – Thermal Stability Analyses of *Hs*CTNS(Δ2–115).
a, NanoDSF profiles using the SEC fractions from **Fig. 8.6**, pH dependency shown in **b**. In **c** and **d**, CD spectra and melts for LMNG- and 10:1 LMNG:CHS-purified *Hs*CTNS(Δ2–115) are shown, respectively.

In light of the insufficient yield for full-length *Hs*CTNS, we then shifted our focus towards purifying and characterizing the Δ2–115 truncation. Purification trials were carried out in both DDM and LMNG with various CHS concentrations, demonstrating superiority regarding yield and purity for LMNG- and 10:1 (w:w) LMNG:CHS- over DDM- and DDM:CHS-purified *Hs*CTNS(Δ2–115) (**Fig. 8.6**). Again, several bands were detected upon SDS-PAGE analysis, hinting at protein oligomerization. Biophysical analysis of these samples using nanoDSF revealed peculiar melting traces that stand out through an unusual upward trend (**Fig. 8.7 a**). Clear transitions were still observed, though, with the 10:1 LMNG:CHS condition showing the most pronounced changes in slope regarding the emission ratio. In contrast to *At*CTNS, acidification did not enhance the thermal stability of the truncated human transporter (**Fig. 8.7 b**). We performed CD spectroscopy experiments to validate the structural integrity of the purified protein (**Fig. 8.7 c** and **d**): α-Helical content was observed for both LMNG- and 10:1 LMNG:CHS-solubilized *Hs*CTNS(Δ2–115), however, the former unfolded quickly and over a broad spectrum. 10:1 LMNG:CHS instead featured a clear and rapid transition from the folded to an unfolded state (after removing initial outliers), generating a T_m of ~56–59 °C. As a result, this detergent combination was chosen for future experiments.

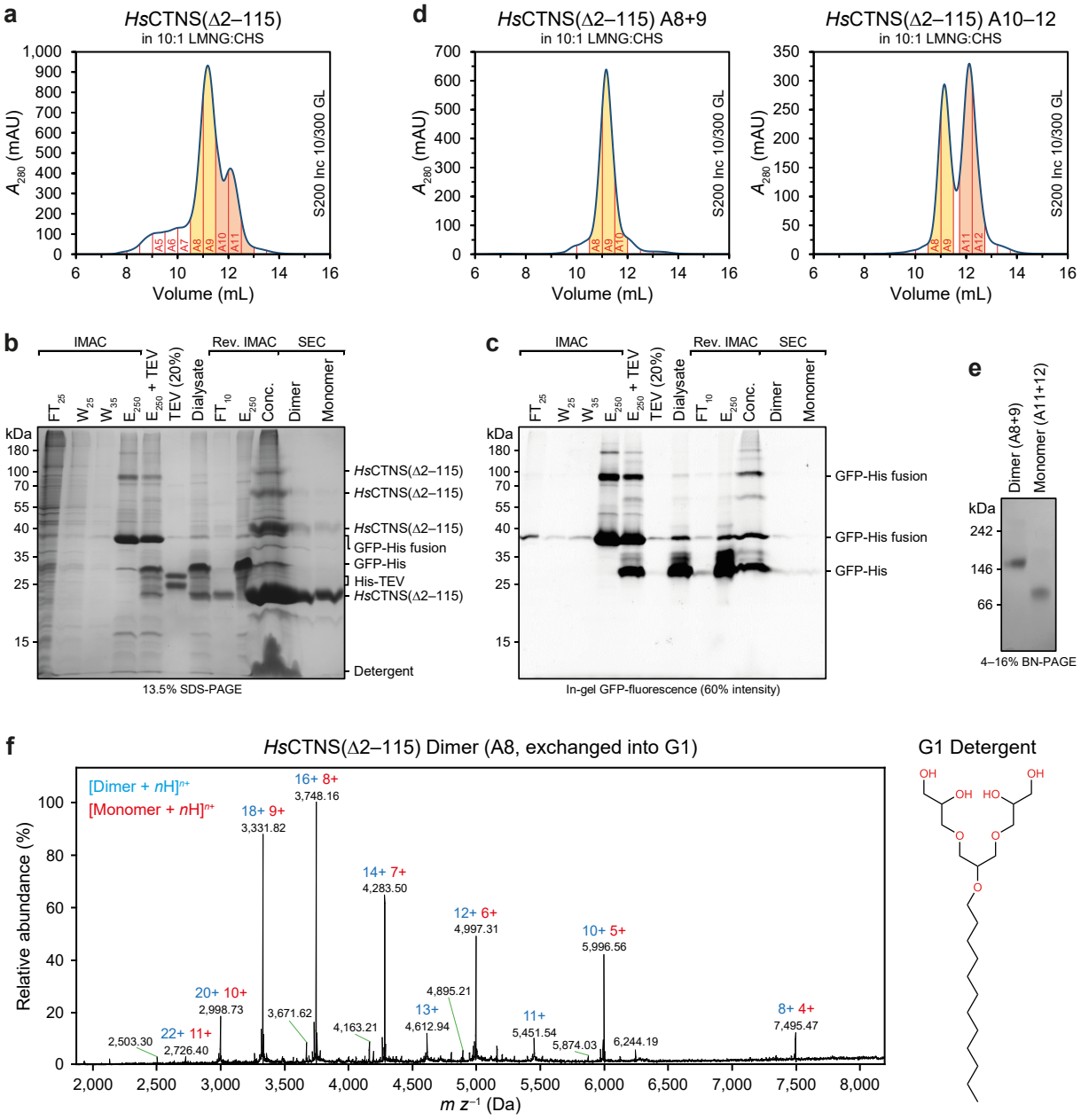


Fig. 8.8 – *Hs*CTNS(Δ2–115) Split Species Purification & Oligomeric State Analyses.
a–**d**, Initial SEC profile, SDS-PAGE analysis, and re-application of the first and second peak fractions onto the column. **e**, BN-PAGE analysis of the first and second SEC peaks of **d**. **f**, Native mass spectrometry using the dimeric species of **a** (fraction A8) after detergent exchange.

In preparation for the generation of specific sybodies against *Hs*CTNS(Δ2–115), we aimed to determine whether we could separate the protein into its individual oligomeric states. By re-applying a predominantly dimeric fraction and a mixed species onto the SEC column (a newer column compared to **Fig. 8.6**), respectively, the peaks began to split, and we were able to complete the task (**Fig. 8.8 a–d**). Spontaneous transitioning between the states was not observed. The final yields of ~200–250 µg of dimeric *Hs*CTNS(Δ2–115) per litre of yeast cell culture and ~60–100 µg L^{-1} for the monomeric species were even found to be considerably

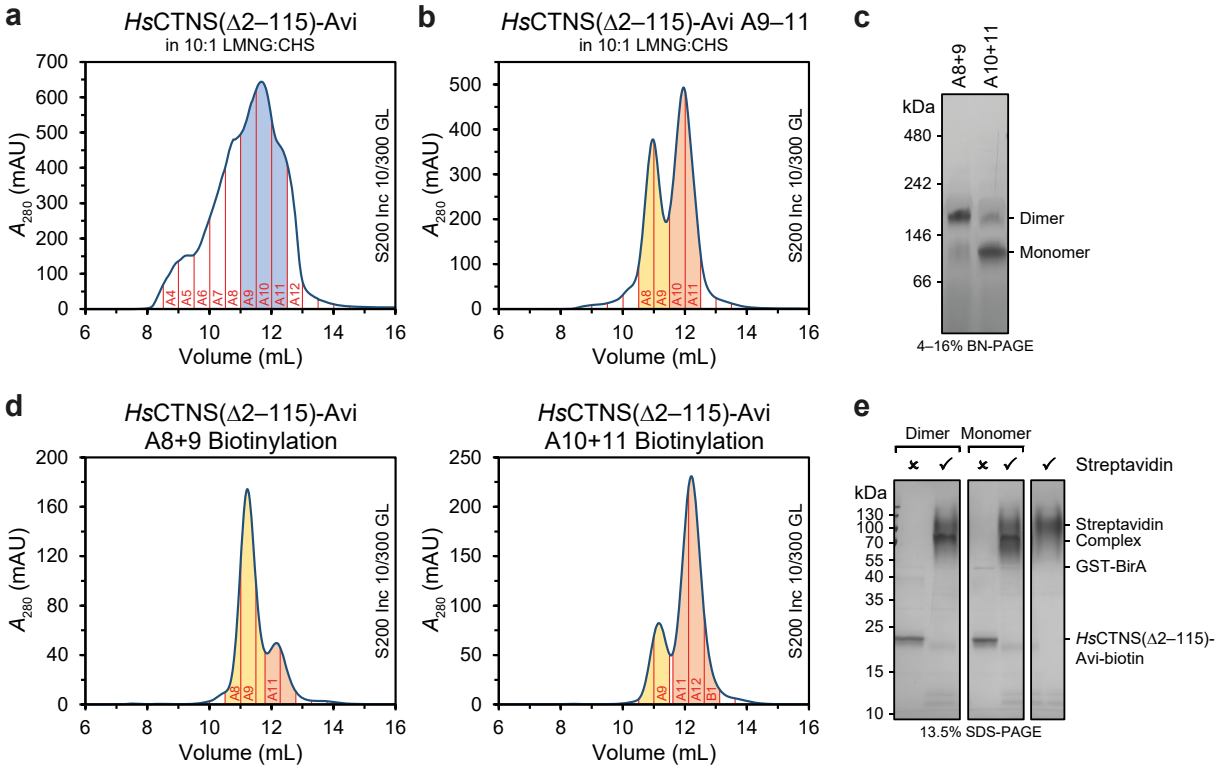


Fig. 8.9 – Purification & Biotinylation of *Hs*CTNS(Δ2–115)-Avi.
a, Initial SEC profile, and **b**, subsequent profile obtained after re-applying the pooled fractions (shaded in blue) to the SEC column. **c**, BN-PAGE analysis of the first and second SEC peaks of **b**. **d**, The pooled fractions of **b** were individually biotinylated and re-applied to the SEC column. Matching fractions were pooled and combined. **e**, Streptavidin shift assay using samples from **d**.

higher than that of *At*CTNS (120–170 µg L^{-1} for the combined species). The oligomeric states were confirmed through BN-PAGE analysis (**Fig. 8.8 e**), and in the case of the dimer, additional native mass spectrometry (nMS) was carried out in collaboration with Dr. Abraham Oluwole (Prof. Dame Carol Robinson's group, University of Oxford). Due to the incompatibility of the LMNG detergent with nMS (spectrum not shown), the sample was exchanged into G1 detergent (Urner *et al.*, 2020). Subsequent analysis revealed peaks that may originate from either dimeric or monomeric protein but also contained designated dimer signals (**Fig. 8.8 f**). Additional peaks that could not be assigned may indicate the presence of binding partners (*e.g.* lipids), impurities in the sample, or partial protein degradation.

8.3 Generation of Sybodies against *Hs*CTNS(Δ2–115)

8.3.1 Purification & Biotinylation of *Hs*CTNS(Δ2–115)-Avi

For the generation of sybodies, we expressed and purified a C-terminally Avi-tagged version of *Hs*CTNS(Δ2–115). As before, the protein was separated into monomeric and dimeric species, and individually biotinylated as described for *At*CTNS (**Fig. 8.9**; see **Secs. 2.4.1** and **5.1**).

Table 8.1 – Phage Display Enrichment Factors for *Hs*CTNS(Δ2–115) Sybody Screening.

a, Using monomeric protein.

	Concave (G)	Loop (B)	Convex (R)
PD1	7.0	0.9	2.3
PD2	2.9	2.3	0.2
PD3	74.4	19.6	3.7

b, Using dimeric protein.

	Concave (G)	Loop (B)	Convex (R)
PD1	0.6	2.1	1.1
PD2	8.7	9.1	0.7
PD3	152.7	194.3	12.3

8.3.2 Screening

Sybody screening was performed separately against dimeric and monomeric *Hs*CTNS(Δ2–115) (each at 50 nM), exposing each to the three libraries of potential binders. The selection procedure was adapted from the *At*CTNS screening (**Sec. 5.2.1**) regarding two critical steps, as outlined in **Sec. 2.4.2**. First, the diversity from the ribosome display output was maximized by using a more equimolar vector:insert ratio (1:1.4 instead of 1:2.5) during ligation, as well as a lower ligase concentration, coupled with an elongated reaction and deactivation duration, which was shown to increase the electrotransformation efficiency. This resulted in a higher number of transformants at $1.2–1.8 \times 10^7$ CFU mL^{-1} for all six reactions compared to the previously targeted 5×10^6. Additionally, triggered by the low enrichment factors after the second phage display (PD), we decided to carry out a third round by repeating the procedure of the second PD but applying two consecutive rounds of competition buffer instead of one and washing five times instead of twice after that. This resulted in a substantial increase in enrichment factors (**Table 8.1**), prompting us to perform a single-clone analysis with all six conditions.

Individual binders were identified by ELISA as described before (**Sec. 5.2.1**). Here, all binders were tested simultaneously against the monomeric and dimeric protein, generating initial data on the binding preferences for each potential Syb (**Fig. 8.10**, **Tables A.11** and **A.12**). 94 ELISA clones were identified based on an absorbance ratio (*Hs*CTNS(Δ2–115)/NC) of 1.5 and higher for the monomeric species and 296 for the dimeric species. 88 were linked to ratios >1.5 for both species and 302 for one or the other. 56 exhibited high absorbance (A_{650} > 0.1) for the NC, 21 of which retained absorbance ratios >1.5 for either *Hs*CTNS(Δ2–115) species. Those were excluded due to potential unspecific binding, yielding a total of 281 hits from the screenings. Interestingly, only seven of the 281 exhibited a higher A_{650} for the monomer than the dimer (**Table A.11**).

96 of the 281 hits were sequenced and phylogenetically analysed using a DNA string comprised of the three Syb CDR regions (**Fig. 8.11**). 52 unique sequences were identified: 17 from the concave (G), 28 from the loop (B), and seven from the convex (R) library, respectively (two of the sequences identified from the convex library turned out to belong to the other two libraries). The composition of the unique CDR sequences is analysed in **Fig. 8.12**. Overall, a similar trend

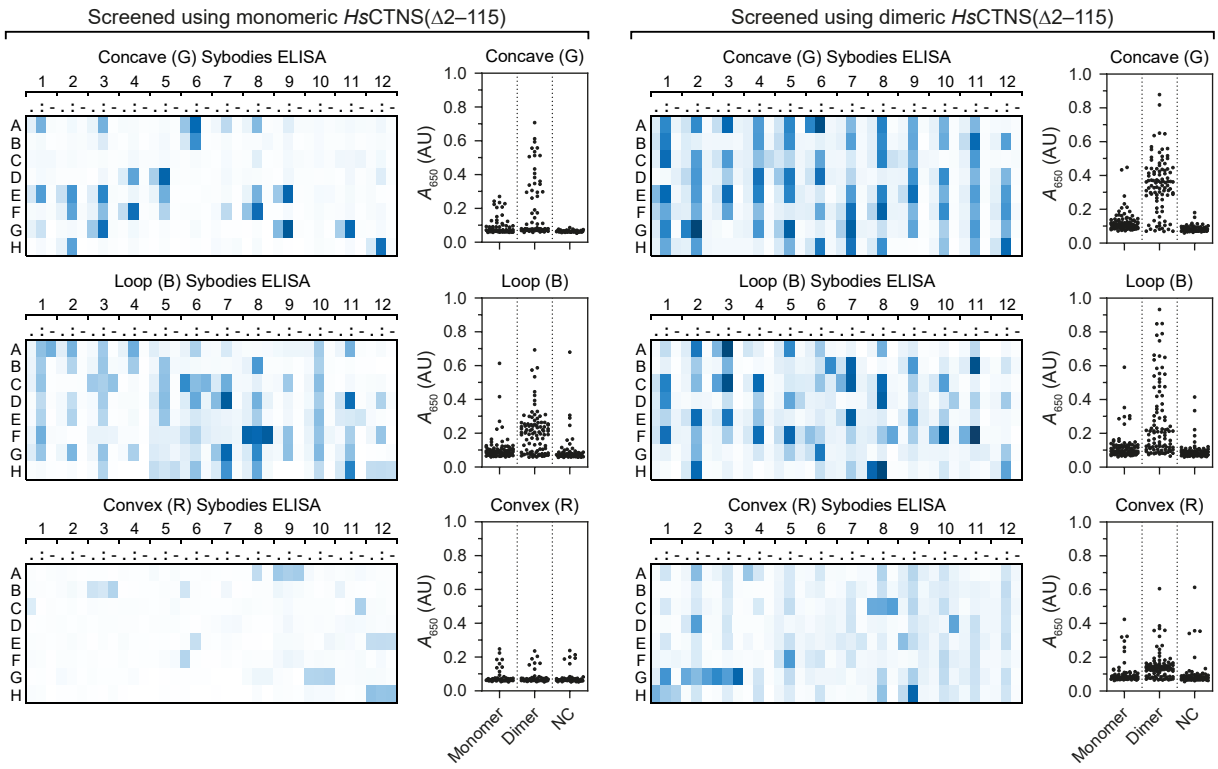


Fig. 8.10 – ELISA of *Hs*CTNS(Δ2–115)-Specific Sybodies.
Gradient plot illustration of ELISA scores from **Tables A.11** and **A.12**, visualizing binding of concave (G), loop (B), and convex (R) sybodies to the monomeric species of *Hs*CTNS(Δ2–115)-Avi-biotin (.), its dimeric species (:), and PepT_{So}-Avi-biotin (NC, -), respectively. Scatter plots of the same data demonstrate the abundance of different absorbance intensities.

to the α-*At*CTNS Syb CDRs (**Fig. 5.5**) is observed, showing a prevalence of polar and aromatic residues over nonpolar aliphatic ones. Few positions are dominated by a specific residue: Aromatic residues define the end of concave Syb CDR1s and the start of loop Syb CDR3s. The CDR2 of concave Sybs ends with glutamate, and begins with a short hydroxylic residue for the loop Sybs. Most other positions contain a more diverse set of residues. Interestingly, a few CDR2 and 3 sequences recur in both the dimer and monomer screenings, one of them even across the concave and loop Syb type (see stars in **Fig. 8.12**). Overall, the unique sequences from the dimer and monomer screenings show many similarities, suggesting that the bulk of the binders may not interact with *Hs*CTNS(Δ2–115) in an oligomer-specific way.

8.3.3 Biochemical & Biophysical Analyses of *Hs*CTNS(Δ2–115)-Sybody Interactions

36 of the 52 unique CDR Sybs were expressed on a small scale (50 mL cultures), purified, and tested for their ability to bind both the dimeric and monomeric species of *Hs*CTNS(Δ2–115). For 25% of Sybs, a sufficient yield was not achieved, and thus, they could not be included in this analysis. SEC traces of the remaining 27 are shown in **Fig. A.9**, with SDS-PAGE analysis

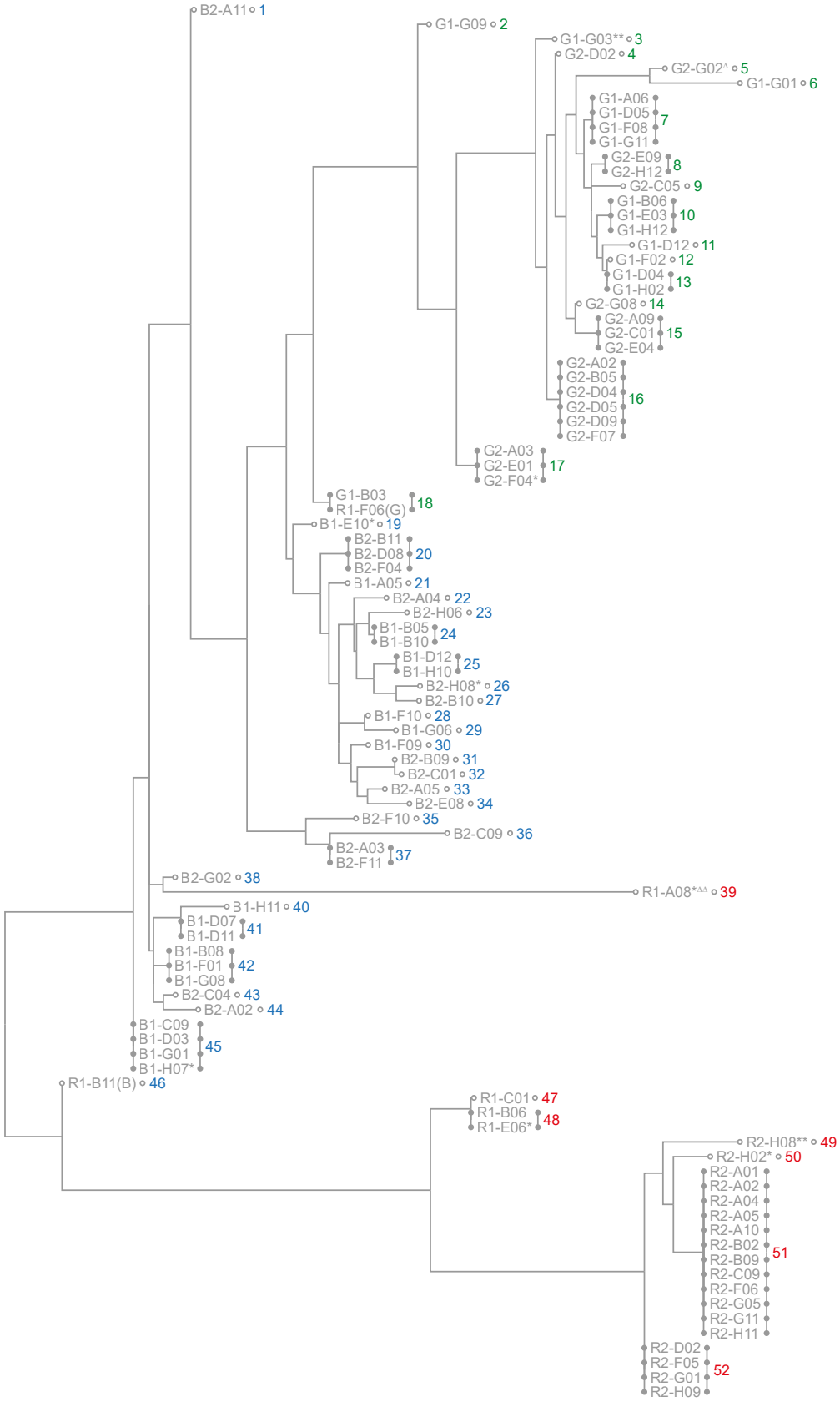


Fig. 8.11 – Phylogenetic Analysis of *Hs*CTNS(Δ2–115)-Specific Sybodies.
Groups are formed using the combined CDR1–3 sequences of 96 ELISA-positive clones. Unique sequences are numbered from top to bottom and colour-coded according to the Syb type. The original ELISA location is given in grey font (compare with **Fig. 8.10**). *Sequence contains a backbone mutation. ΔSequence contains a deletion in one of the CDRs. For more information, see **Table A.11**.

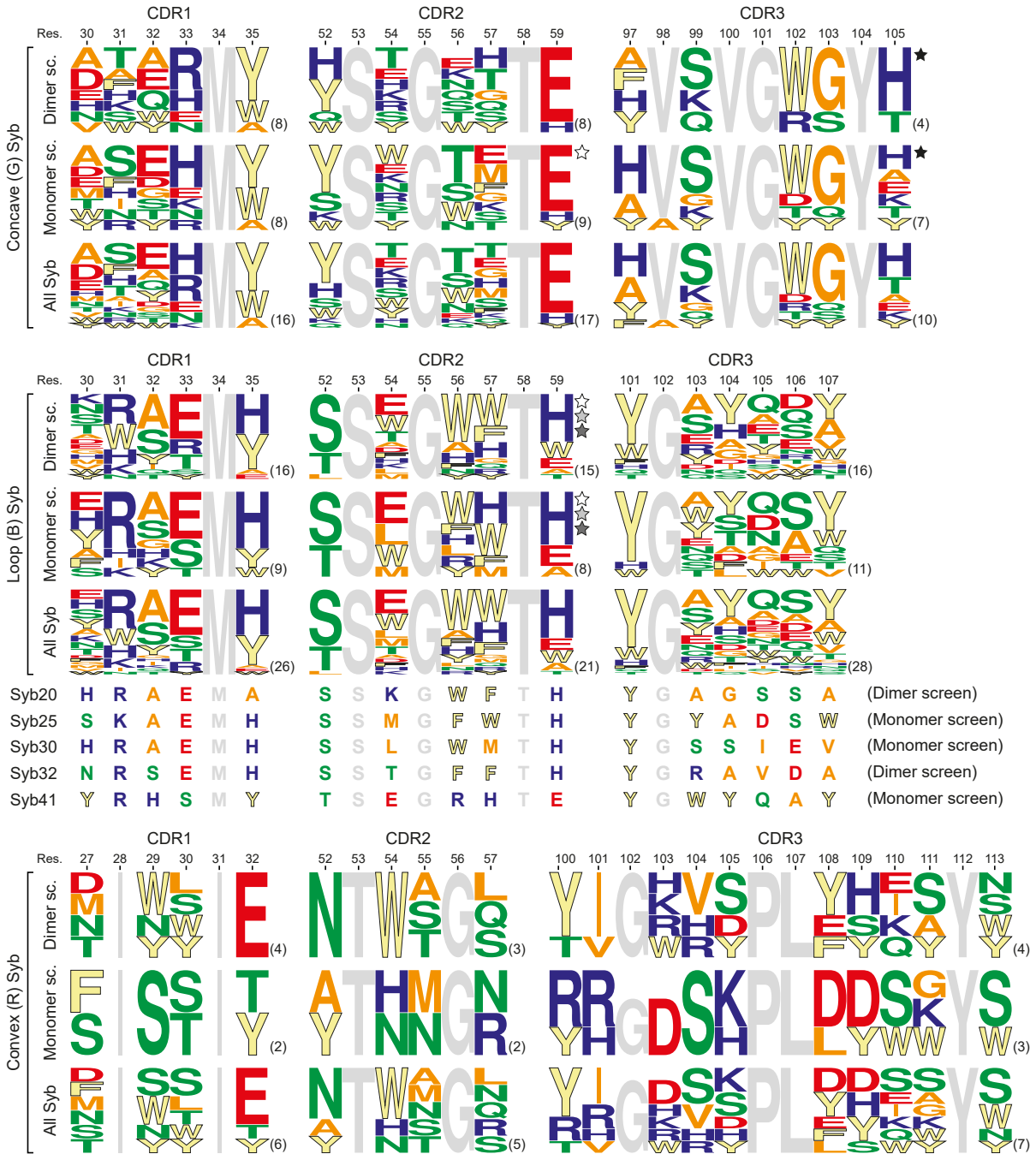


Fig. 8.12 – α-*Hs*CTNS(Δ2–115) Sybody CDR Sequence Comparison.
Frequency plot logos using unique CDRs only (number of sequences shown in parentheses), generated using WebLogo (Crooks *et al.*, 2004). The CDR sequences of Sybs 20, 25, 30, 32, and 41 are highlighted, and individual residues are coloured *via* amino acid chemistry. Greyed letters indicate CDR residues that are constant by design. Matching CDR sequences across individual screens are indicated using stars.

of the pooled and combined fractions, annotated with the SEC retention volumes, depicted in **Fig. 8.13 a**. For the purpose of facilitating structural studies on the human transporter, we were more interested in identifying Sybs that interact with the *Hs*CTNS(Δ2–115) dimer, thus, we prioritized pulldown experiments using that species first (**Fig. 8.13 b**). The most potent binders

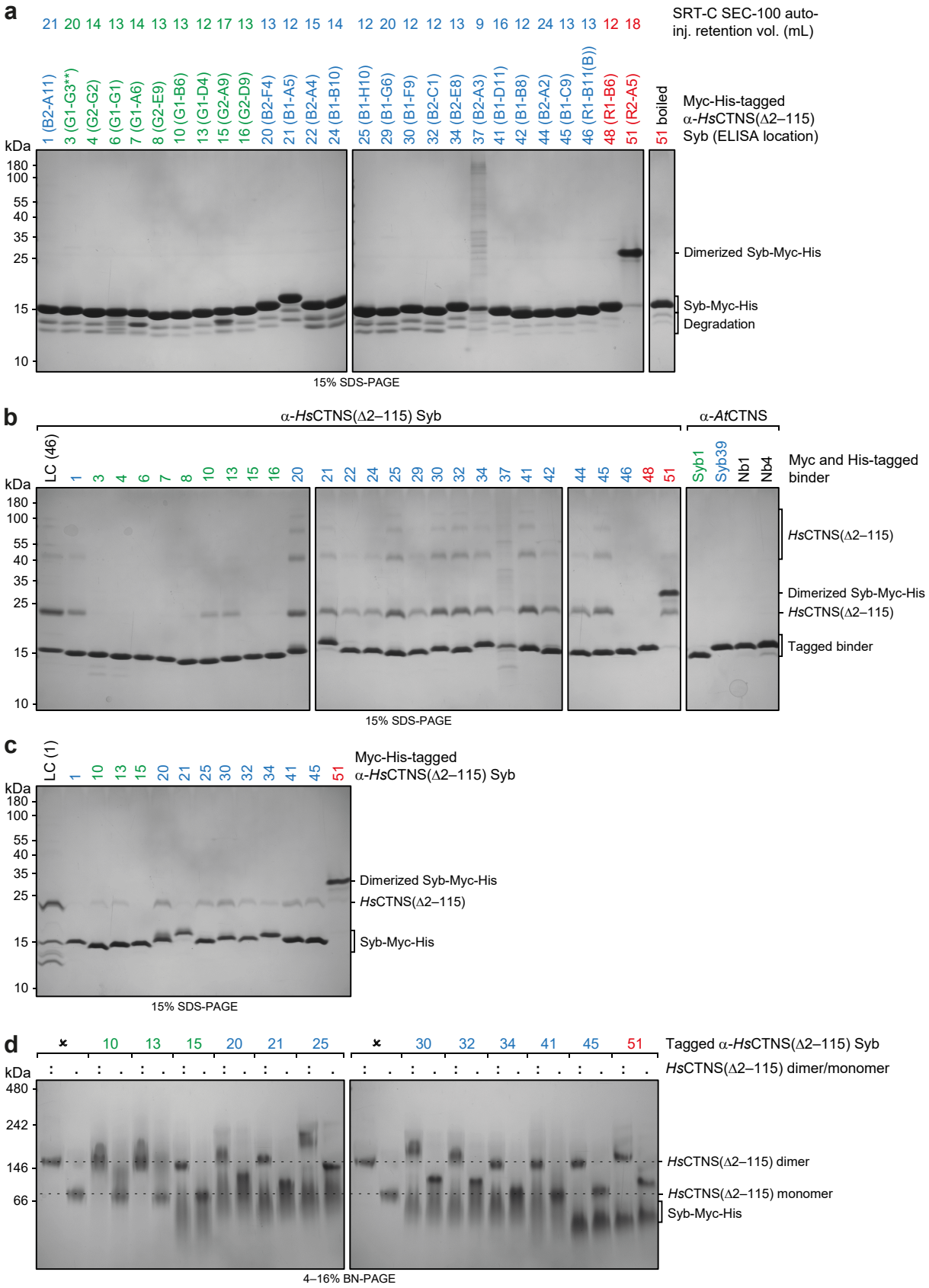


Fig. 8.13 – Purification & Biochemical Analyses of Tagged α-*Hs*CTNS(Δ2–115) Sybodies. **a**, Final SDS-PAGE analysis of tagged Sybs purified from 50 mL bacterial cultures. Corresponding SEC traces are shown in **Fig. A.9**. **b** and **c**, Pulldowns using the dimeric and monomeric species of *Hs*CTNS(Δ2–115), respectively, performed at pH 7.5 and analysed by SDS-PAGE. **d**, BN-PAGE analysis of *Hs*CTNS(Δ2–115)-Sybody complexes. *Sequence contains a backbone mutation.

from each Syb library were then also tested against the monomeric fraction (**Fig. 8.13 c**). In these experiments, the loop (B) Sybs were found to be associated with more intense bands for the transporter, compared to the concave (G) and convex (R) Sybs, indicating stronger interactions were formed. All binders tested for both oligomeric states favoured the dimeric form, consistent with the trend already observed during ELISA (**Tables A.11** and **A.12**). We also tested the α-*At*CTNS binders, identified in **Ch. 5** and used as crystallization chaperones in **Chs. 6** and **7**, for cross-reactivity with *Hs*CTNS(Δ2–115), which, unsurprisingly, was not observed (**Fig. 8.13 b**).

To further assess the suitability of the binders for structural studies using cryo-EM, where the addition of a soluble domain to a small membrane protein that is otherwise fully engulfed in a detergent micelle is critically important for particle alignment, we performed BN-PAGE shift analysis. Membrane protein-sybody complexes were formed and natively separated on gel based on their size. We hypothesized that the benefits for cryo-EM would increase proportionally with the size difference observed between the complex and the membrane protein alone. Interestingly, not all Sybs previously identified to bind *Hs*CTNS(Δ2–115) also increased the size of the complex in this experiment (**Fig. 8.13 d**). BN-PAGE shift analysis suggested that loop (B) Sybs 20, 25, 30, and 32, as well as convex (R) Syb51, were the most useful binders for structural studies using cryo-EM.

Finally, the binding was characterized biophysically using BLI. Based on two different binder concentrations, we estimated both a k_{off} and an apparent K_D for the most promising sybodies (**Fig. 8.14 a** and **b**). Association using the concave (G) and convex (R) Sybs generated only low responses, and the data could not be interpreted (fitted). Analysis of loop (B) Sybs 20, 25, 30, 32, and 41 yielded k_{off}s in the range of ~0.2–0.6 min^{-1} and apparent K_Ds as low as ~80 nM (for Syb30-Myc-His), ranking the affinities of our α-*Hs*CTNS(Δ2–115) sybodies more within the vicinity of the α-*At*CTNS nanobodies (**Figs. 5.13** and **5.14**) rather than their kinetics-inferior sybody equivalents (**Figs. 5.8** and **5.9**). Upon epitope binning experiments (**Fig. 8.14 c**), suggesting that all loop (B) Sybs tested by BLI interact with the same region of the *Hs*CTNS(Δ2–115) dimer, we chose Sybs 20, 30, and 32 for subsequent structural trials based on the combined analysis (overview in **Table A.13**).

8.4 Structural Trials Using Cryo-EM

The majority of crystal structures obtained for the plant CTNS transporter presented *At*CTNS as a monomer (**Ch. 7**), despite the presence of both a monomeric and dimeric species in the fractions pooled for LCP crystallization (**Fig. 8.2** and *e.g.* **Fig. 7.7 a**). To analyse the oligomeric state of the human CTNS transporter as it exists in solution (*i.e.* not in a crystal), we therefore decided to use single-particle cryo-EM instead of XRC.

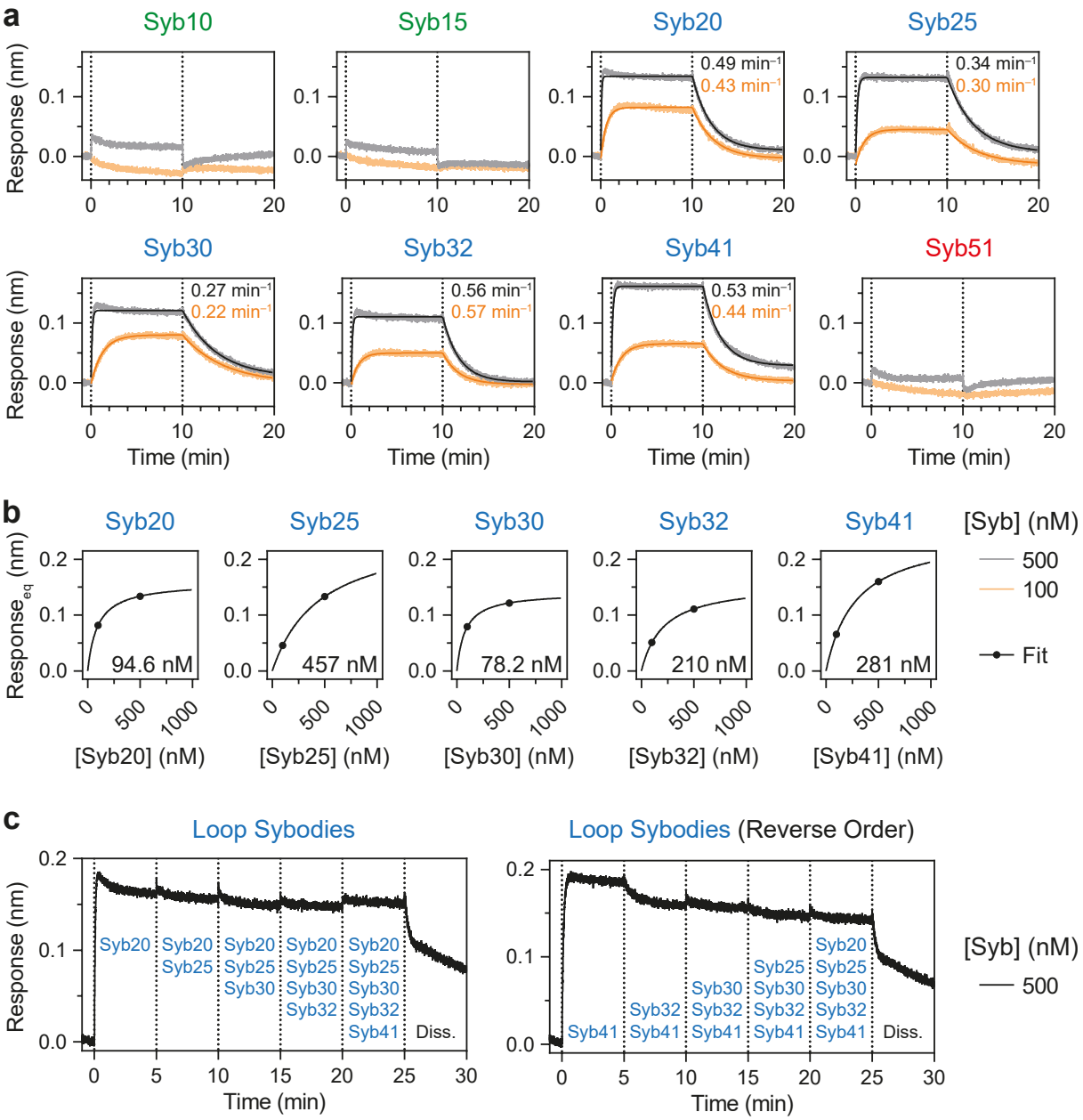


Fig. 8.14 – *Hs*CTNS(Δ2–115) Dimer-Tagged Sybody Binding Analysis Using BLI.
a, Syb association and dissociation steps are shown. The resulting k_{off} of each binder is printed at the top of the dissociation step. *Hs*CTNS(Δ2–115)-Avi-biotin was immobilized at 20 nM. **b**, Apparent K_Ds calculated using the data of **a**. **c**, Epitope binning using the most promising α-*Hs*CTNS(Δ2–115) sybodies.

8.4.1 Sample Preparation

For cryo-EM structural trials, the C-terminal Myc-His-tag was not removed from the binders. Large-scale sybody purification (typically from 500 mL to 1 L) was performed using periplasmic extraction (see **Sec. 2.2.9**), which yielded 20–35 mg of pure binder (**Fig. 8.15 a** and **e**) per litre of bacterial cell culture. *Hs*CTNS(Δ2–115)-sybody complexes were formed as described for *At*CTNS (see **Sec. 2.5.6**), avoiding unnecessary freeze-thaw cycles to ensure the highest possible

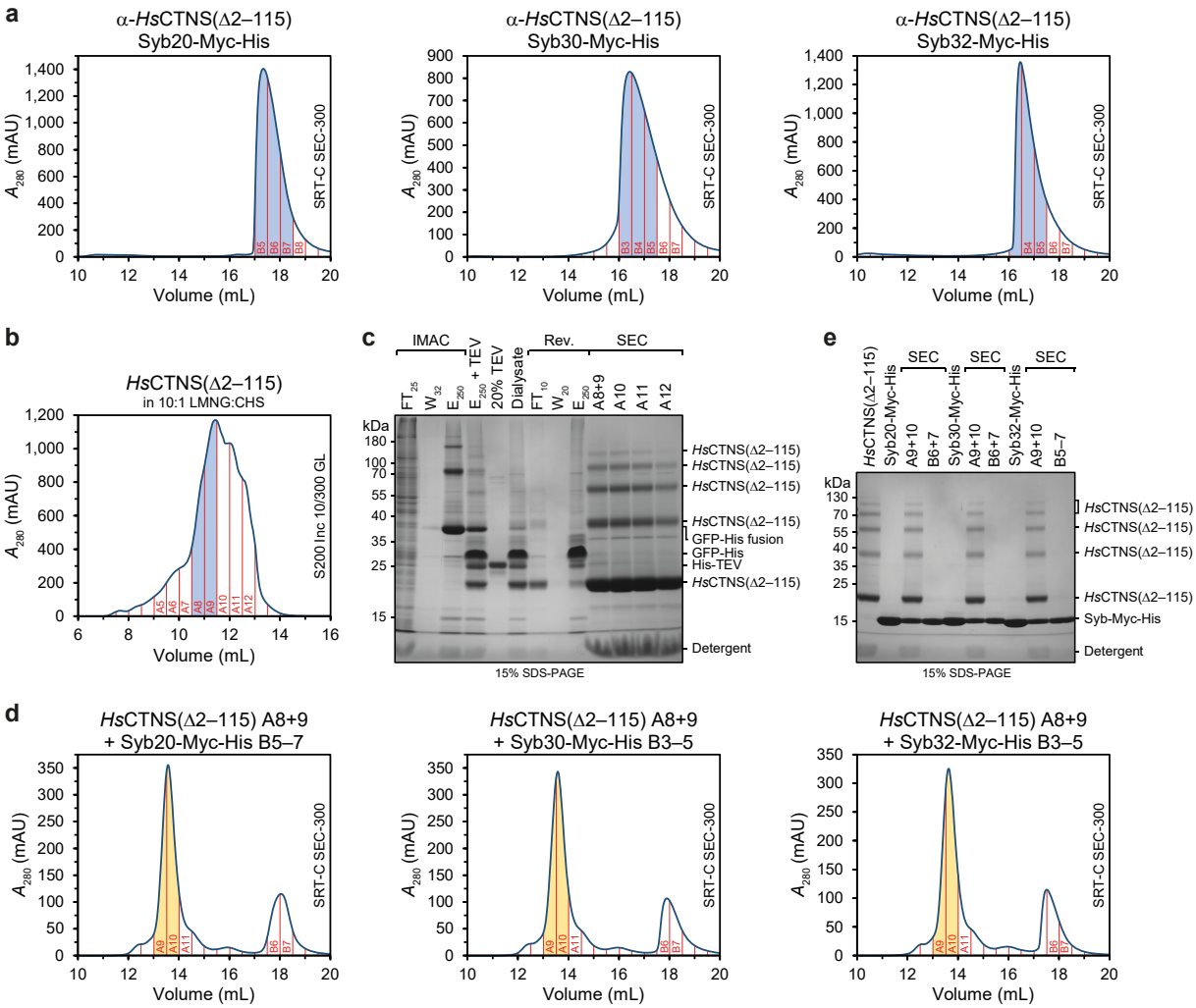


Fig. 8.15 – *Hs*CTNS(Δ2–115)-Sybody-Myc-His Complex Formation for Cryo-EM Trials.
a, Purification SEC traces of Myc-His-tagged sybodies. **b**, SEC profile, and **c**, SDS-PAGE analysis of relevant samples obtained during purification. **d** and **e**, SEC co-elution traces of the formed complexes and SDS-PAGE analysis.

protein quality. Thus, the dimeric fraction of freshly purified *Hs*CTNS(Δ2–115) (**Fig. 8.15 b** and **c**) was immediately incubated with the sybody, gel filtrated to remove unbound material (**Fig. 8.15 d** and **e**), and proceeded to cryo-EM grid preparation on the same day.

8.4.2 Grid Preparation, Data Collection & Processing

Sample vitrification, grid screening, data collection, and processing were all performed by Dr. Takafumi Kato (Newstead group), as described in **Secs. 2.5.6** and **2.5.7**, respectively. Initially, three different complexes were analysed regarding the particle distribution observed at different concentration points (**Fig. 8.16**). At >1.5 mg mL^{-1}, particles were packed very densely, exacerbating automated particle packing. At ~0.4 mg mL^{-1} (no spin concentration required), squares within very thin ice regions were empty, but those within slightly thicker ice showed better particle distribution compared to the higher concentration points. We found that

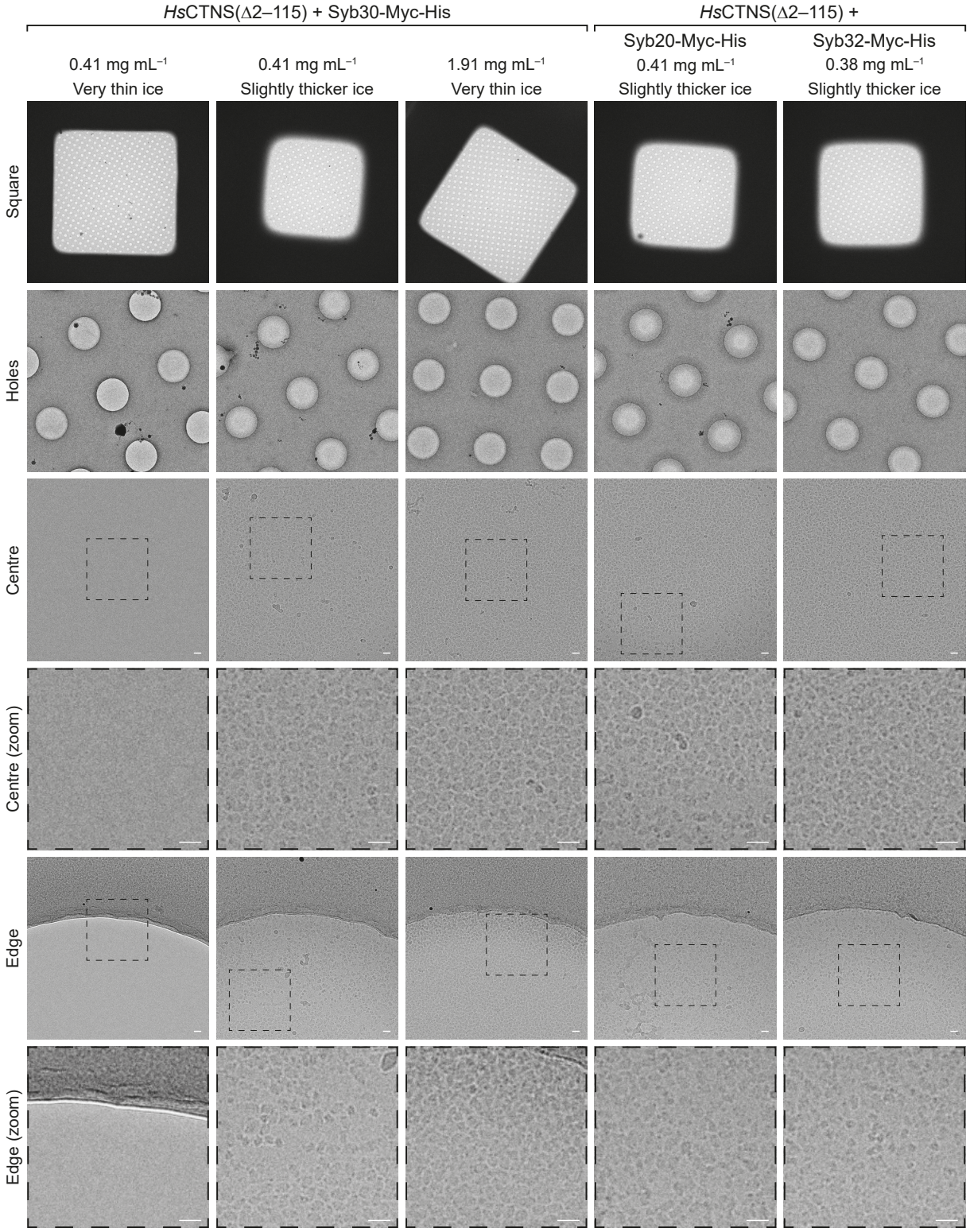


Fig. 8.16 – Cryo-EM Grid Screening.
Typical grid squares and foil holes obtained from different complexes are shown. Micrographs are depicted to highlight the particle distribution in centre and edge regions. Scale bars 200 Å.

particles of the *Hs*CTNS(Δ2–115)-Syb30-Myc-His complex were spread further apart from one another than those using Sybs 20 and 32 (**Fig. 8.16**), thus, we chose this composition for data

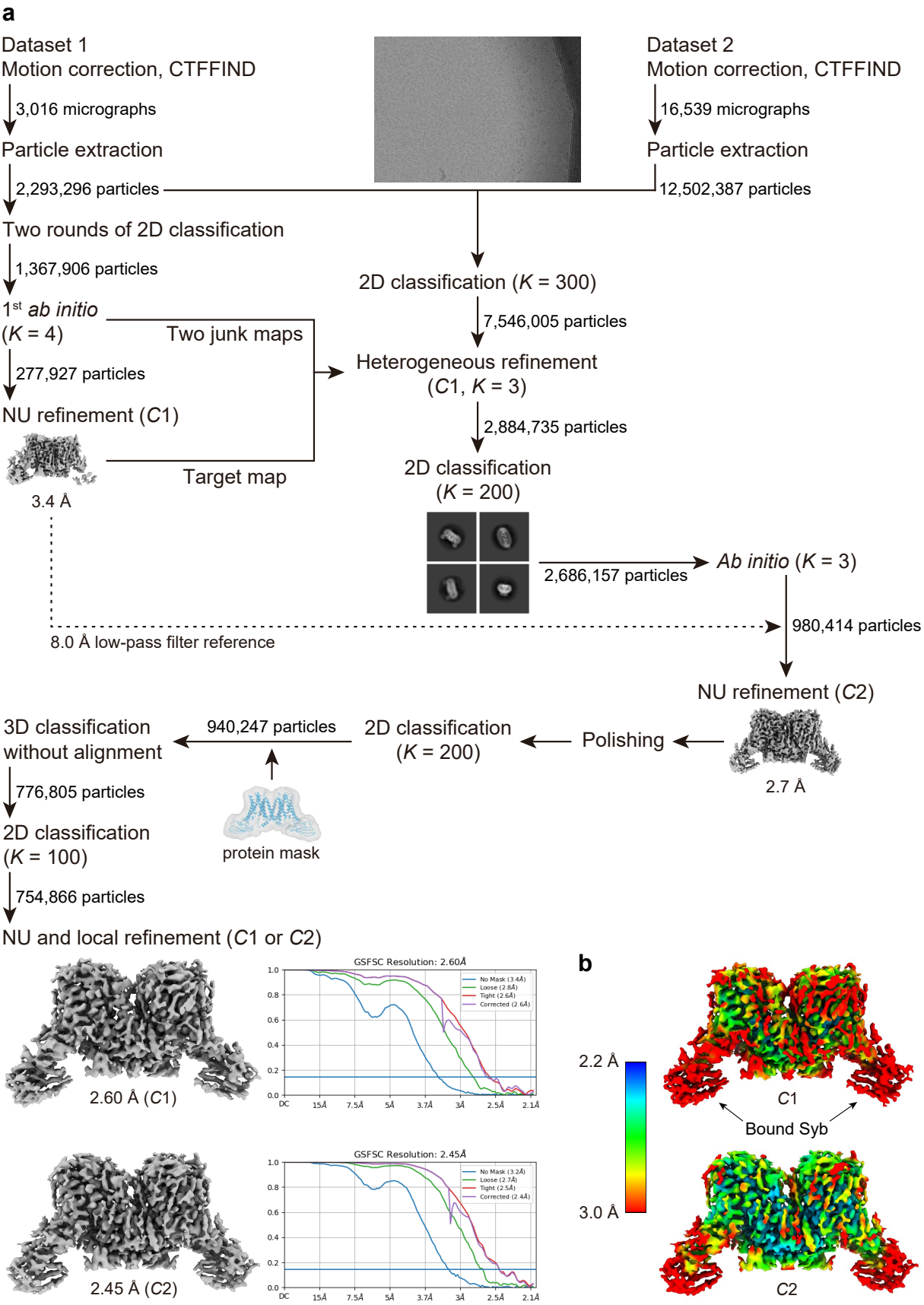


Fig. 8.17 – Cryo-EM Data Processing Workflow for *Hs*CTNS(Δ2–115)-Syb30-Myc-His. **a**, Processing workflow as performed by Dr. Takafumi Kato (Newstead group). **b**, Local resolution of the structure obtained.

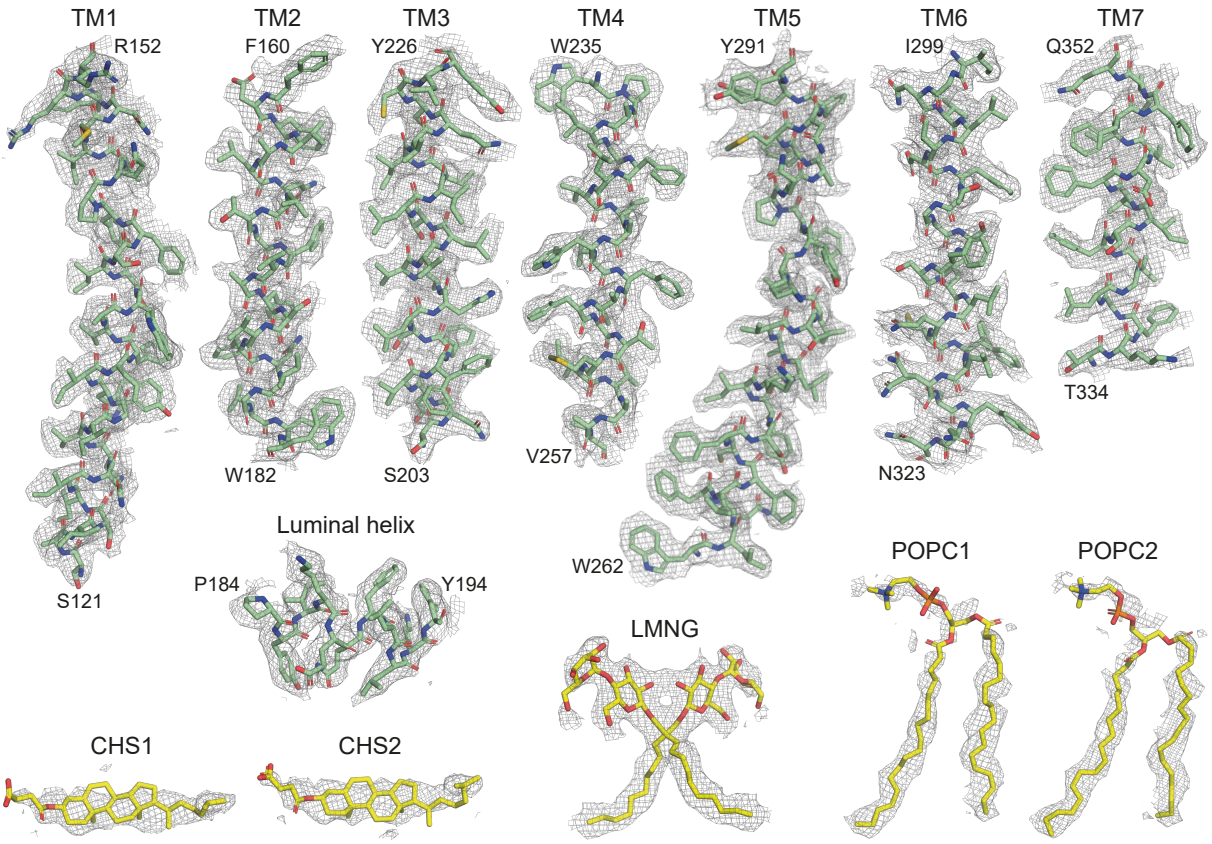


Fig. 8.18 – Essential Cryo-EM Maps of *Hs*CTNS(Δ2–115)-Syb30-Myc-His at 2.45 Å.
Density fit around all *Hs*CTNS(Δ2–115) helices and lipid and detergent molecules bound within the dimer interface. All maps are contoured at 6 σ and carved around the atoms at 2 Å.

collection. Following the workflow illustrated in **Fig. 8.17 a** and described in **Sec. 2.5.7**, almost 15 million particles were processed, leading to a number of 2D classes that clearly show a dimeric protein engulfed in a detergent micelle and bound to two additional molecules. Through *ab initio* reconstruction and various refinement and polishing steps, we ultimately obtained cryo-EM maps at 2.60 and 2.45 Å under *C*1 and *C*2 symmetry, respectively (**Fig. 8.17 b**).

8.4.3 Model Building, Refinement & Validation

Initial efforts to build the model for the *Hs*CTNS(Δ2–115) dimer using one of the *At*CTNS crystal structures shown in **Ch. 7** or the AlphaFold prediction of *Hs*CTNS (UniProt: O60931) proved difficult as only a few TM helices could be fitted simultaneously. This was caused by the fact that the transporter was captured in the outward-open (binding site facing the lysosomal lumen) rather than the inward-open conformation, which was observed for the structures mentioned above. A cryo-EM structure of *Hs*CTNS in the outward-open conformation was released around the time we published the crystal structures of the plant transporter (Guo *et al.*, 2022). That model (PDB: 8DKI) was easily fitted into the observed density, and upon manual refinement in Coot, followed by real-space refinement in PHENIX (Afonine *et al.*,

Table 8.2 – Cryo-EM Structure Refinement & Validation Statistics.
Ramachandran plots, generated by MolProbity (Lovell *et al.*, 2003), are illustrated in **Fig. A.10**.

	***Hs*CTNS(Δ2–115)-Syb30-Myc-His (apo)**
Refinement program	PHENIX real space refinement
Initial model used (PDB code)	8DKI
Model resolution (Å)	2.45
FSC threshold	0.143
Number of atoms:	
Protein	3,973
Ligand	317
Water	0
Average *B*-factor (Å^2):	
Protein	47.1
POPC lipid	72.8
LMNG detergent	44.4
CHS detergent	63.8
Water	N/A
RMSDs:	
Bond lengths (Å)	0.004
Bond angles (°)	0.708
MolProbity score	2.05
Clashscore, all atoms	8.03
MolProbity counts/frequencies (%):	
Poor rotamers	25/5.85
Favoured rotamers	309/72.4
Ramachandran outliers	0/0
Ramachandran favoured	463/97.9
Bad bonds	0/0
Bad angles	1/0.02

2018), yielded the fits illustrated in **Fig. 8.18**. The RMSD between the initial and final model was calculated at 0.878 Å over all 237 C_α atoms. Voluminous densities were observed within the dimer interface and modelled as two molecules of CHS, one LMNG detergent, and two 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) lipid molecules, although other phospholipids could have also been used (density fits shown in **Fig. 8.18**; chemical structures shown in **Tables 3.1** and **3.2**; POPC is POPE with a -N$(CH_3)^+$ instead of a -NH_3^+ group). The two copies of Syb30-Myc-His (see **Fig. 8.17 b**) were not fitted into the density. Refinement and validation statistics for the model are shown in **Table 8.2**, revealing adequate average *B*-factors for protein (47.1 Å^2) and ligand atoms (44.4–72.8 Å^2), as well as a good MolProbity score of 2.05, resulting from a low number of poor rotamers (25 or 5.85%) and no Ramachandran outliers.

8.4.4 *Hs*CTNS(Δ2–115) Dimer Structure

Like the *At*CTNS crystal dimer, the *Hs*CTNS(Δ2–115) dimer is also arranged symmetrically around a twofold rotational axis located in the center of the dimer interface (**Fig. 8.19 a**). However, strikingly, the dimer interface is composed of TMs 2, 3, and 4, rotated ~90° relative to

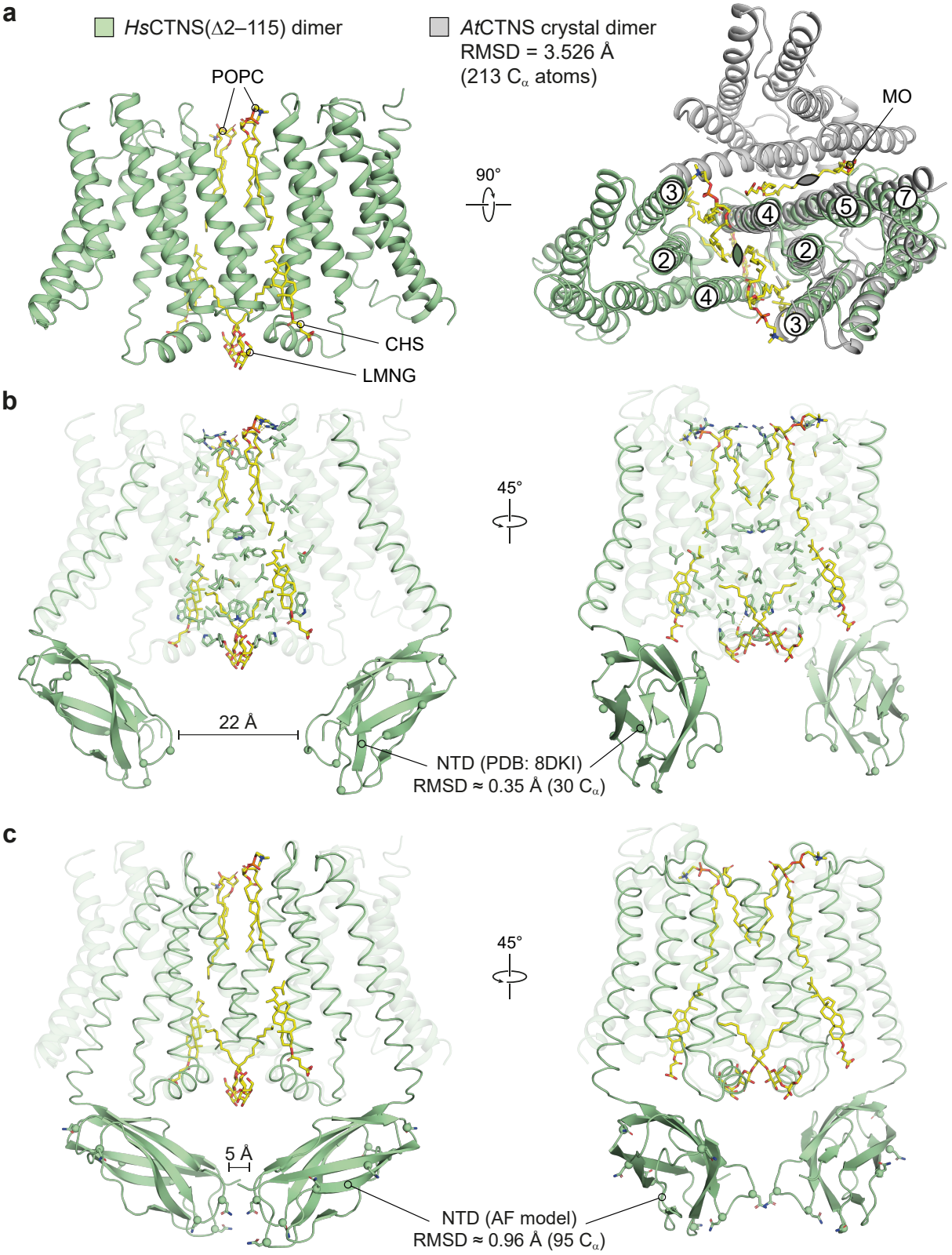


Fig. 8.19 – *Hs*CTNS(Δ2–115) Dimer Structure & Interface.
a, Cartoon representation of the *Hs*CTNS(Δ2–115) dimer bound to lipid and detergent molecules shown as sticks. The dimer arrangement is compared to the *At*CTNS crystal dimer from **Fig. 7.13**. **b**, Overview of the dimer interface, highlighting side chain residues within 4 Å distance as sticks. The position of the NTD results from superimposition along TM1 using the *Hs*CTNS model from PDB: 8DKI (Guo *et al.*, 2022). Asparagine-linked glycosylation sites are indicated as spheres. **c**, As in **b** but using the inward-open AlphaFold model (UniProt: O60931) superimposed along TMs 2–4.

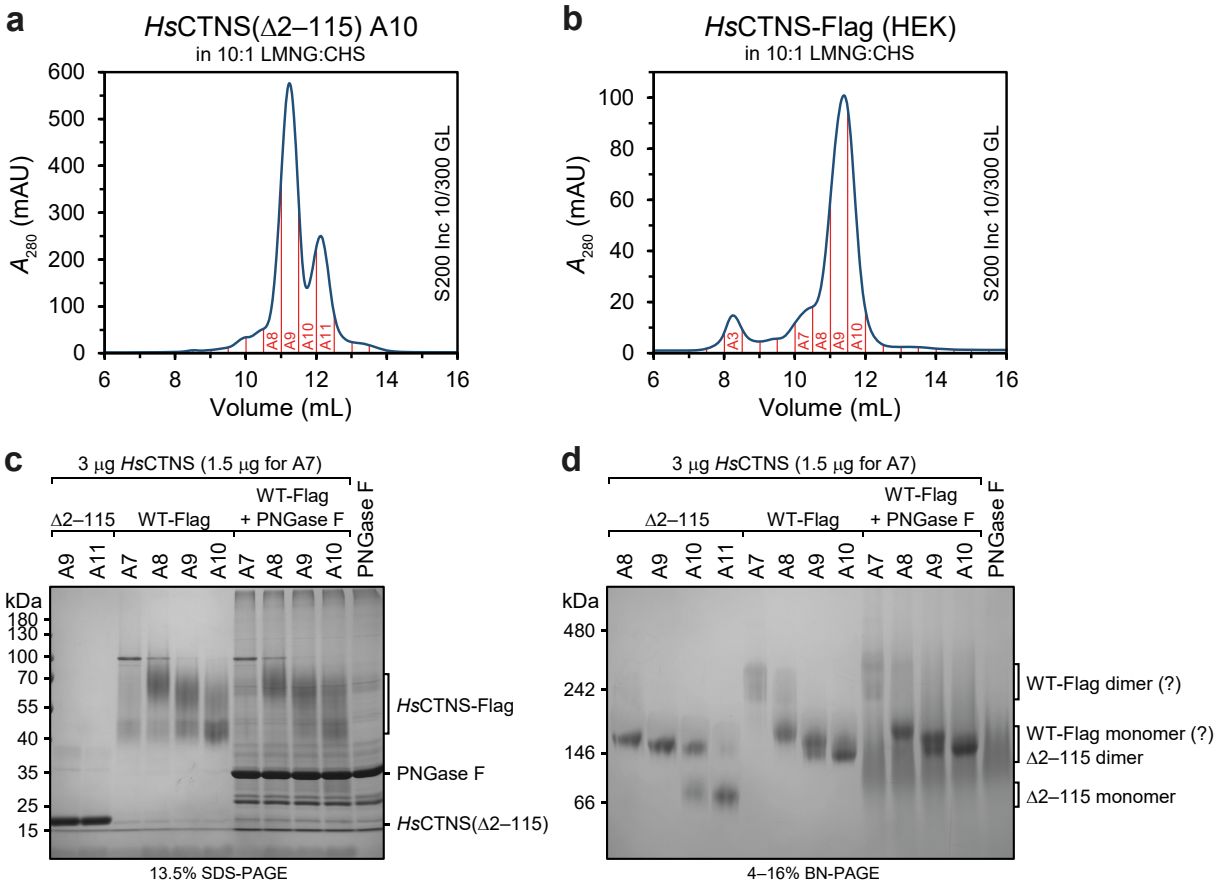


Fig. 8.20 – Initial Dimerization Analysis of Full-Length *Hs*CTNS.
a and **b**, SEC profiles of *Hs*CTNS(Δ2–115) expressed in yeast cells and full-length *Hs*CTNS-Flag expressed in mammalian HEK293F cells. **c** and **d**, SDS-PAGE and BN-PAGE analysis using the fractions shown in **a** and **b**.

the *At*CTNS crystal dimer interface. The lipid and detergent molecules bound in between the individual monomers form vital hydrophobic contacts that likely keep the dimer together. From each monomer, 24 predominantly aliphatic and aromatic side chain residues are positioned within a 4 Å distance from either the other monomer or the bound non-polymeric molecules (**Fig. 8.19 b**), generating a contact area that stretches over ~2,830 Å^2. Interestingly, only a small number of polar contacts are formed: Arginines 228 of either monomer form salt bridge interactions with the phosphate group of the POPC lipids, and W262 interacts with a hydroxyl group of the LMNG detergent molecule *via* a hydrogen bond. However, no water molecules were added to the model, rendering the identification of indirect hydrogen bonds impossible.

8.4.5 Potential Full-Length *Hs*CTNS Dimer

Superimposition of the *Hs*CTNS(Δ2–115) dimer using two full-length *Hs*CTNS monomers (PDB: 8DKI) indicates that the NTDs would not clash. In fact, a gap of ~22 Å is observed, which likely leaves enough room to allow the full-length protein to dimerize even when all seven asparagines are glycosylated, at least in the presented outward-open conformation (**Fig. 8.19 b**).

When predicting the position of the NTDs in the inward-open conformation, by superimposing the human CTNS AlphaFold prediction (UniProt: O60931) along the dimer interface (TMs 2–4), the gap narrows significantly to ~5 Å, with asparagines 107 meeting at the most narrow point (**Fig. 8.19 c**). Still, the formation of a full-length *Hs*CTNS dimer remains plausible, albeit less likely, due to the potential steric hindrance of the glycans. To address the oligomeric state of full-length *Hs*CTNS in solution, we started larger-scale expression trials in mammalian HEK293F cells (1 L volume). Preliminary analysis of the SEC fractions indicated that only a small fraction of the full-length transporter formed a dimer (**Fig. 8.20**), consistent with the fact that crystal and cryo-EM structures obtained by others depict *Hs*CTNS as a monomer (Guo *et al.*, 2022). Further analysis could not be included within the scope of this thesis. Thus, the molecular and functional relevance of a potential *Hs*CTNS dimer remains to be elucidated.

CHAPTER 9

Structural & Functional Analysis of CTNS

The crystal structures of *At*CTNS in the inward-open state (**Ch. 7**), as well as the cryo-EM structure of *Hs*CTNS(Δ2–115) in the outward-open state (**Ch. 8**), enabled us to perform a detailed analysis of the CTNS transporter. In this chapter, I will describe the general architecture of CTNS and compare the fold to related and otherwise prominent 7TM proteins. I will highlight common features shared between CTNS and other PQ-loop proteins, such as the family-defining PQ-motifs that introduce kinks in the TM helices, as well as novel features that have not been observed before, including two salt bridges that stabilize the transporter in the inward-open state. Following the structural analysis, I will briefly describe a number of *in vitro* and *in vivo* activity assays that were developed to characterize the transport function of *At*CTNS and *Hs*CTNS, respectively, followed by the functional data generated for a selection of relevant variants of these transporters, as identified based on the structures. Using this data, we established a mechanism that describes how CTNS recognizes its ligand and how it uses protons to couple transport across the vacuolar or lysosomal membrane, respectively. Finally, I will show that we can use this knowledge to explain the impact of cystinosis-causing mutations by mapping them onto the structures.

9.1 Structure of CTNS

As predicted (**Fig. 3.1**), the *At*CTNS and *Hs*CTNS(Δ2–115) transporters, obtained in **Chs. 7** and **8**, consist of seven TM α-helices, adopting a compact V-shaped structure in both the inward-open (binding site facing the cytoplasm) and outward-open (vacuolar or lysosomal luminal) conformations (**Figs. 9.1** and **9.2 a**, respectively). The TMs are organized into two three-helix bundles (THBs) (TMs 1–3 and TMs 5–7), each exhibiting a 1-3-2 topology, where the third helix is positioned in between the first two (**Fig. 9.1 b**). TMs 1 and 6, as well as 2 and 5, pack together to join the THBs in a pseudo-symmetric way, framing a central binding cavity, as shown later. TM4, the inversion linker helix, connects the THBs and resides in close proximity to both TMs 2 and 5. The THBs superimpose well in both conformational states, generating an RMSD of 2.784 Å over all C_α atoms of TM helical residues for the inward-open state of *At*CTNS (**Fig. 9.1 c**) and 1.265 Å for the outward-open conformation of *Hs*CTNS(Δ2–115)

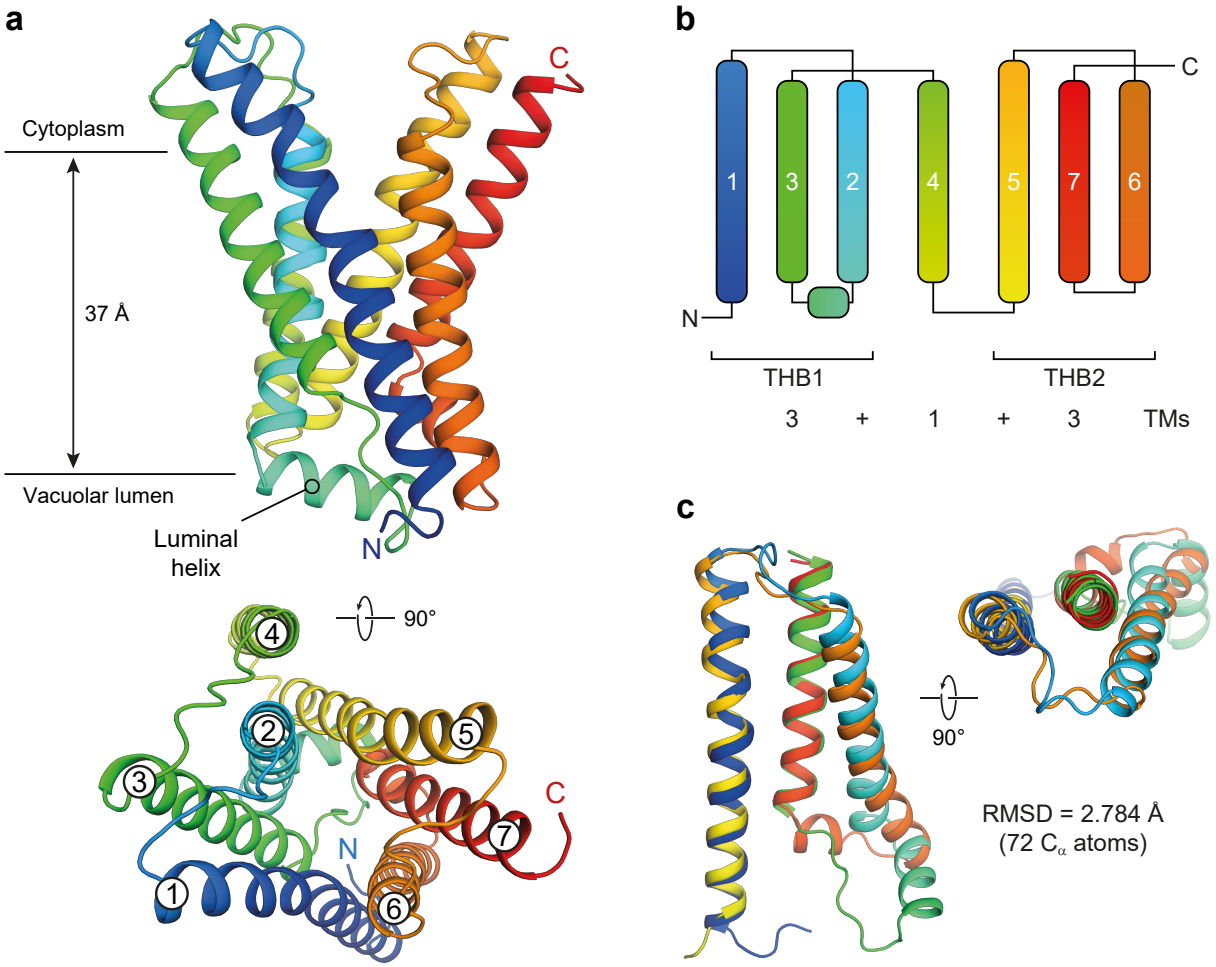


Fig. 9.1 – *At*CTNS Structure in the Inward-Open Conformation.
a, Cartoon representation of *At*CTNS from **Fig. 7.6** (PDB: 7ZK1), coloured by spectrum from blue (N-terminus) to red (C-terminus). The seven TM helices are numbered in the top view. **b**, Schematic representation of the TM topology, organized in two THBs. **c**, Superimposition of both THBs using TM helical residues only.

(**Fig. 9.2 b**). Whilst showing similar intra-THB connections on the cytoplasmic side (TM 1–2 and 5–6 loops), the main structural differences between the THBs are observed on the luminal side of the transporter, where TMs 2 and 3 are connected *via* a longer linker that contains a three-turn helix (termed luminal helix), whilst the TM 6–7 connection is much shorter. The different conformational states of the CTNS transporter superimpose to an RMSD of 4.293 Å when considering all C_α atoms (**Fig. 9.2 c**), highlighting the movement of all six core TM helices to facilitate the conformational transition. However, interestingly, the individual THBs superimpose to a lower 2.279 and 2.497 Å (using all C_α atoms), respectively, with the main divergence stemming from TMs 1 and 5 bending around their respective PQ-motif, thus demonstrating that the conformational transition results from both the helical movement within each THB and the precise interplay of both THBs.

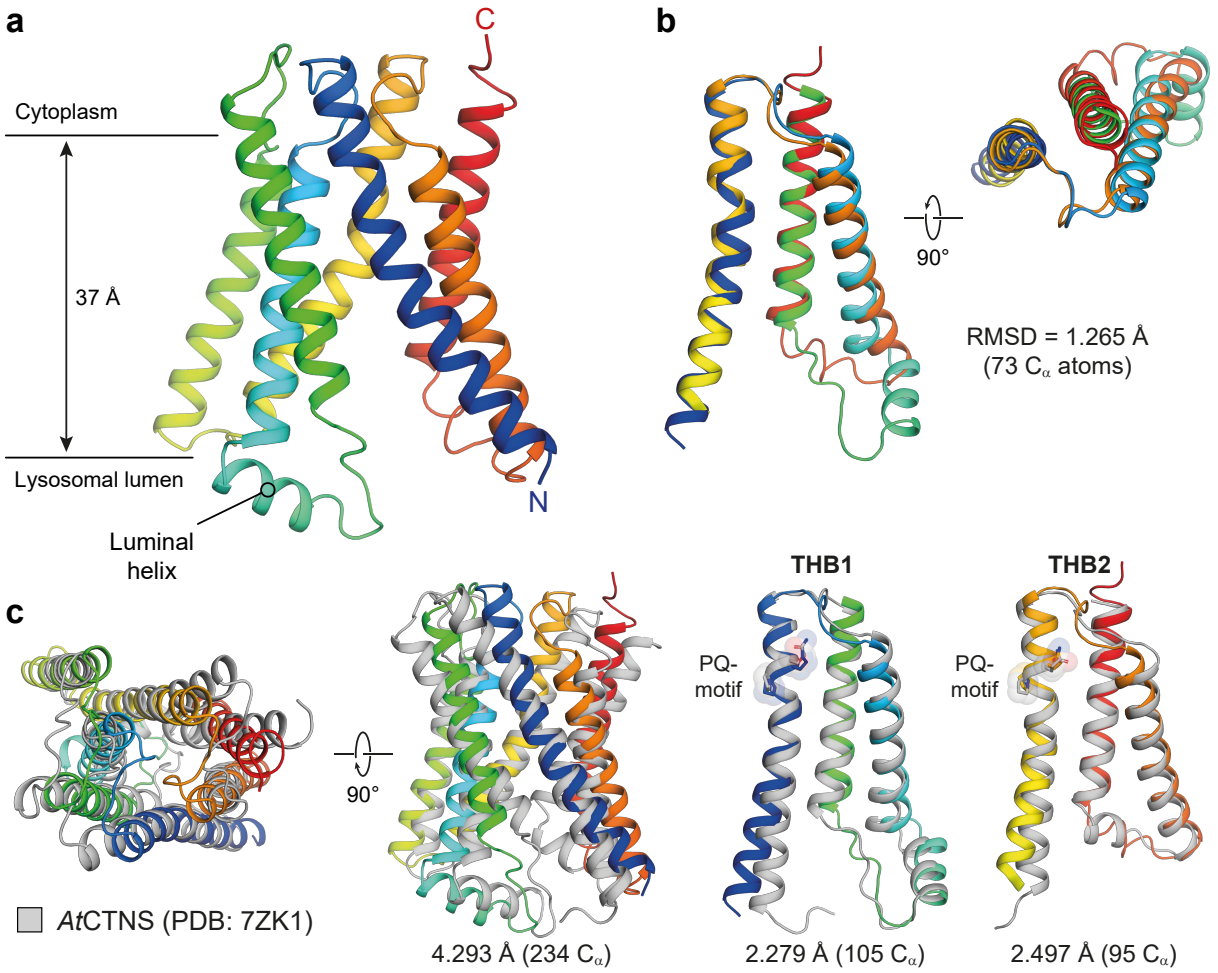


Fig. 9.2 – *Hs*CTNS(Δ2–115) Structure in the Outward-Open Conformation.
a, Cartoon representation of chain A of the *Hs*CTNS(Δ2–115) dimer from **Fig. 8.19**, coloured by spectrum from blue (N-terminus) to red (C-terminus). **b**, Superimposition of both THBs using TM helical residues only. **c**, Superimposition with *At*CTNS in the inward-open state (from **Fig. 9.1**) using all residues.

9.1.1 Architectural Comparison with Related Proteins

The CTNS transporter is a member of the so-called PQ-loop family, a group of 7TM proteins with conserved proline-glutamine repeats but diverse functions (Saudek, 2012). Other members of this family include amino acid transporters such as PQLC2 (Jézégou *et al.*, 2012) and Ypq2 (Kawano-Kawada *et al.*, 2019), the SWEET sugar transporters, the KDEL trafficking receptor and the lipid flippase MPDU1 (Schenk *et al.*, 2001). To date, the only structures of eukaryotic PQ-loop proteins available in the PDB are those of *Os*SWEET2b (Tao *et al.*, 2015) and *At*SWEET13 (Han *et al.*, 2017), both in the inward-open conformation, as well as several *Gg*KDELR2 structures (Bräuer *et al.*, 2019; Gerondopoulos *et al.*, 2021; Wu & Rapoport, 2021), bound to different peptides or binders, each in the outward-open state. Extending the selection to prokaryotic relatives yields a number of SemiSWEETs, the bacterial homologues of the SWEETs, which are constructed from a dimer of identical THBs, omitting the linker helix but retaining a highly similar core architecture. These have been studied in more detail, and multiple representatives

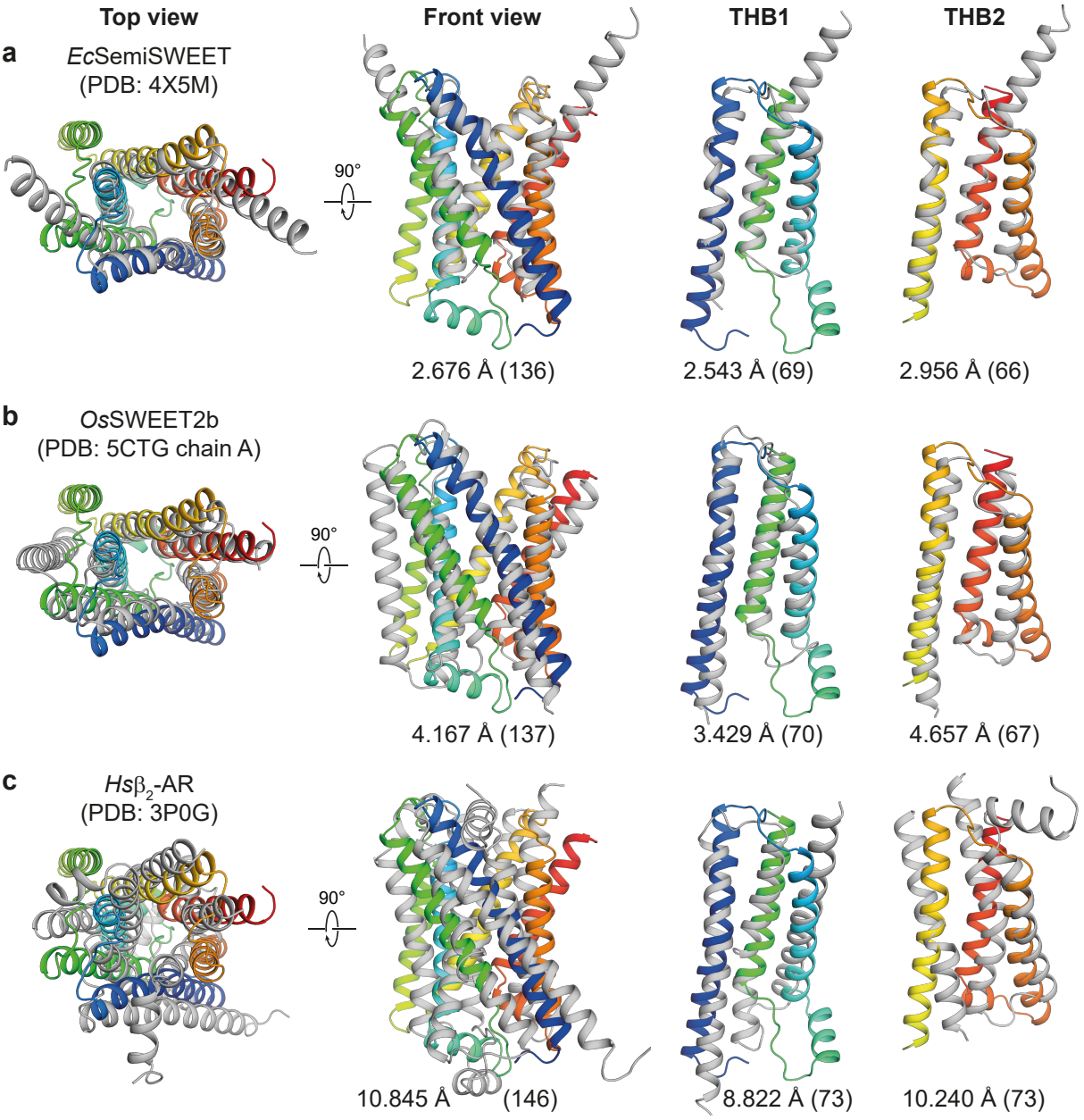


Fig. 9.3 – Structural Comparison Using *At*CTNS in the Inward-Open State.
Superimpositions of *At*CTNS from **Fig. 7.6** (PDB: 7ZK1; coloured by spectrum) with **a**, *Ec*SemiSWEET (Lee *et al.*, 2015), **b**, *Os*SWEET2b (Tao *et al.*, 2015), and **c**, *Hs*β_2-AR (Rasmussen *et al.*, 2011), each coloured in grey. RMSDs are given based on the alignment over all TM helical C_α atoms of the THBs (count in parentheses).

have been deposited in the PDB featuring inward- (Lee *et al.*, 2015) and outward-open (Latorraca *et al.*, 2017) as well as occluded states (Xu *et al.*, 2014).

Superimposition of the THB TMs of our *At*CTNS model with the available structures in the inward-open conformation (**Fig. 9.3**) and those of the *Hs*CTNS(Δ2–115) model with the structures in the outward-open conformation (**Fig. 9.4**), respectively, illustrates a significant structural similarity between the PQ-loop family members despite low sequence identities

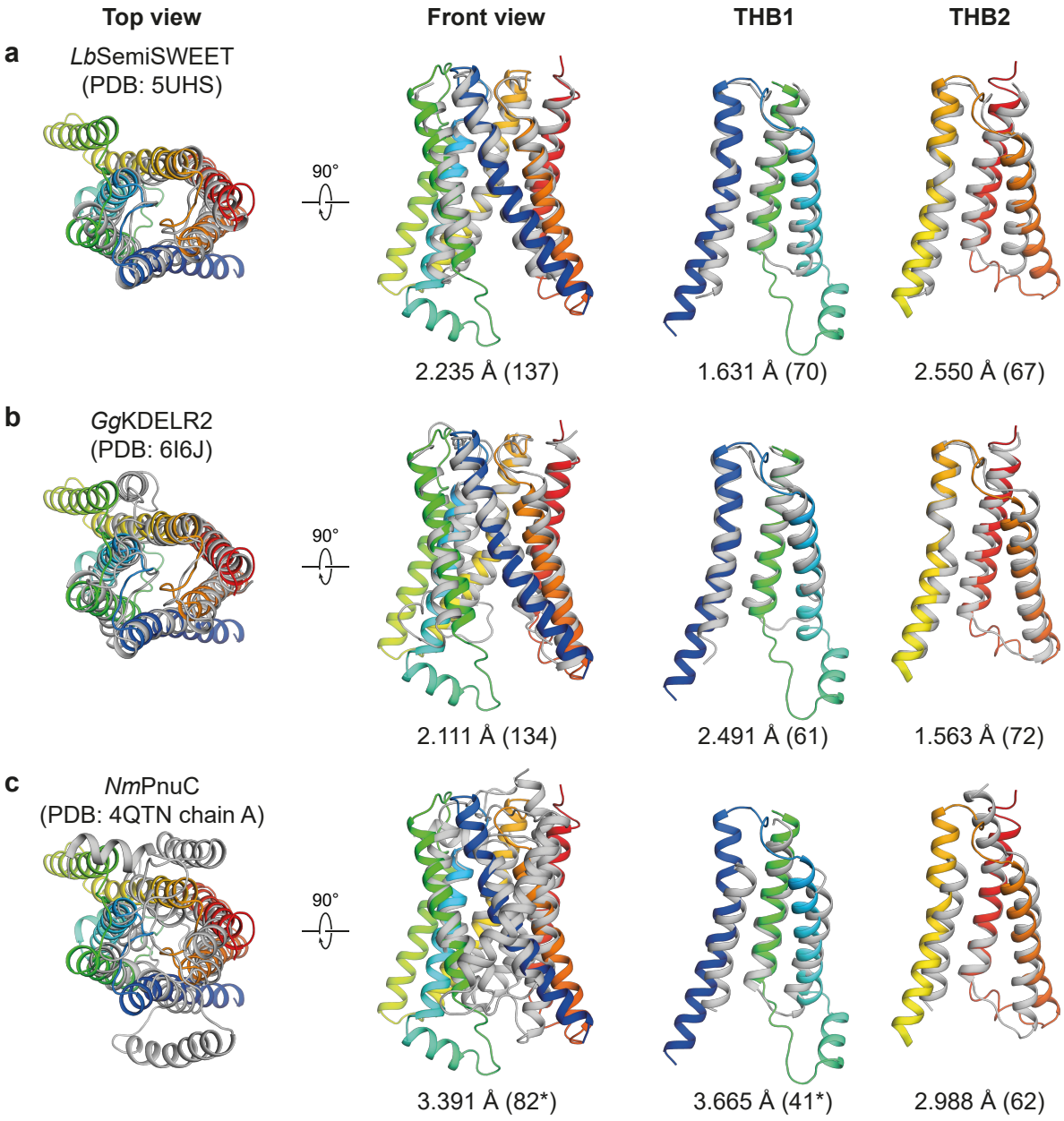


Fig. 9.4 – Structural Comparison Using *Hs*CTNS(Δ2–115) in the Outward-Open State. Superimposition of chain A of the *Hs*CTNS(Δ2–115) dimer from **Fig. 8.19** (coloured by spectrum) with **a**, *Gg*KDELR2 (Bräuer *et al.*, 2019), **b**, *Lb*SemiSWEET (Latorraca *et al.*, 2017), and **c**, *Nm*PnuC (Jaehme *et al.*, 2014), each coloured in grey. RMSDs are given based on the alignment over all TM helical C_α atoms of the THBs (count in parentheses; *indicates that some atoms were ignored).

of only ~15–21% (**Table 9.1**). CTNS aligns with the SemiSWEETs from different bacteria with RMSDs of ~2.2–2.7 Å over all superimposed TM helical C_α atoms and ~1.6–3.0 Å for the individual THBs (**Figs. 9.3 a** and **9.4 a**). Interestingly, TM3 of the *E. coli* homologue extends past TMs 3 and 7 of CTNS by about three α-helical turns, however, this is not observed for the homologue from *Leptospira biflexa* (*Lb*). *At*CTNS superimposes with the eukaryotic SWEET transporter to 4.167 Å over both THBs and ~3.4–4.7 Å for each individually (**Fig. 9.3 b**). The linker helix of *Os*SWEET2b is positioned closer toward the first THB, forming a less spherical

Table 9.1 – Sequence Identities among PQ-Loop & Related Proteins.

	Sequence identity (%)	1	2	3	4	5	6	7	8	9
1	*Os*SWEET2b (PDB: 5CTG)		38.3	18.9	19.9	18.4	19.9	18.6	12.6	18.8
2	*At*SWEET13 (PDB: 5XPD)	38.3		18.1	21.6	14.1	17.3	17.8	19.3	17.1
3	*Lb*SemiSWEET (PDB: 5UHS)	18.9	18.1		40.7	21.5	16.0	14.9	18.0	10.5
4	*Ec*SemiSWEET (PDB: 4X5M)	19.9	21.6	40.7		19.0	20.7	17.0	15.9	12.8
5	*Hs*CTNS (UniProt: O60931)	18.4	14.1	21.5	19.0		44.8	16.4	18.5	18.8
6	*At*CTNS (UniProt: P57757)	19.9	17.3	16.0	20.7	44.8		14.7	18.8	16.4
7	*Gg*KDELR2 (PDB: 6I6J)	18.6	17.8	14.9	17.0	16.4	14.7		19.7	14.4
8	*Nm*PnuC (PDB: 4QTN)	12.6	19.3	18.0	15.9	18.5	18.8	19.7		17.3
9	*Hs*β_2-AR (PDB: 3P0G)	18.8	17.1	10.5	12.8	18.8	16.4	14.4	17.3	

molecule than CTNS. The *Hs*CTNS(Δ2–115) structure aligns best with the *Gg*KDELR2 (2.111 Å for both THBs combined, ~1.6–2.5 Å individually, **Fig. 9.4 b**). Here, the linker helix is positioned closer to the C-terminal THB compared to CTNS. Unlike the PQ-loop family transporters, the KDELR does not transition into an inward-open state (Bräuer *et al.*, 2019). However, upon peptide binding, the C-terminal half of KDELR's TM7 kinks out, increasing the superimposition RMSD with *Hs*CTNS(Δ2–115) in the outward-open conformation. The main difference based on the secondary structure of CTNS and all other PQ-loop proteins described here is the presence of the luminal helix between TMs 2 and 3 and, more generally, the length of the connections between TM helices on the luminal side. In contrast, the loops on the cytoplasmic side appear to align well in most PQ-loop protein superimpositions, a result of the shared gating mechanism on this side, as described in **Sec. 9.1.2**.

Based on the 3+1+3 helical arrangement, CTNS was also grouped with G protein-coupled receptors (GPCRs) and the PnuC vitamin transporters, among others, to form the transporter–opsin–GPCR (TOG) superfamily (Yee *et al.*, 2013). GPCRs comprise the largest group of 7TM proteins in humans (Jacoby *et al.*, 2006) and have been extensively studied due to their clinical importance, with many hundreds of structures solved (Congreve *et al.*, 2020). PnuCs are found exclusively in prokaryotes and contain an additional TM segment, resulting in a 1+3+1+3 topology (Jaehme *et al.*, 2014). The sequence identities among representatives of these groups and CTNS (~16–19%) do not differ significantly from those observed within the PQ-loop family (**Table 9.1**). However, their THBs are arranged in a sequential 1-2-3 topology instead of the 1-3-2 observed for PQ-loop members (Feng & Frommer, 2016). As a result, superimposition of *At*CTNS with a human GPCR yields RMSDs of ~8.8–10.2 Å for the individual THBs and 10.845 Å for both combined (**Fig. 9.3 c**), demonstrating significant structural differences. The THB TMs of *Hs*CTNS(Δ2–115) and PnuC from *Neisseria mucosa* (*Nm*) align better, however, due to the fact that *Nm*PnuC was captured in the occluded state, a number of C_α atoms need to be ignored to generate RMSDs of ~3.0–3.7 Å (**Fig. 9.4 c**), otherwise similar values to the GPCR representative are obtained. Furthermore, *Nm*PnuC aligns with the CTNS THBs in

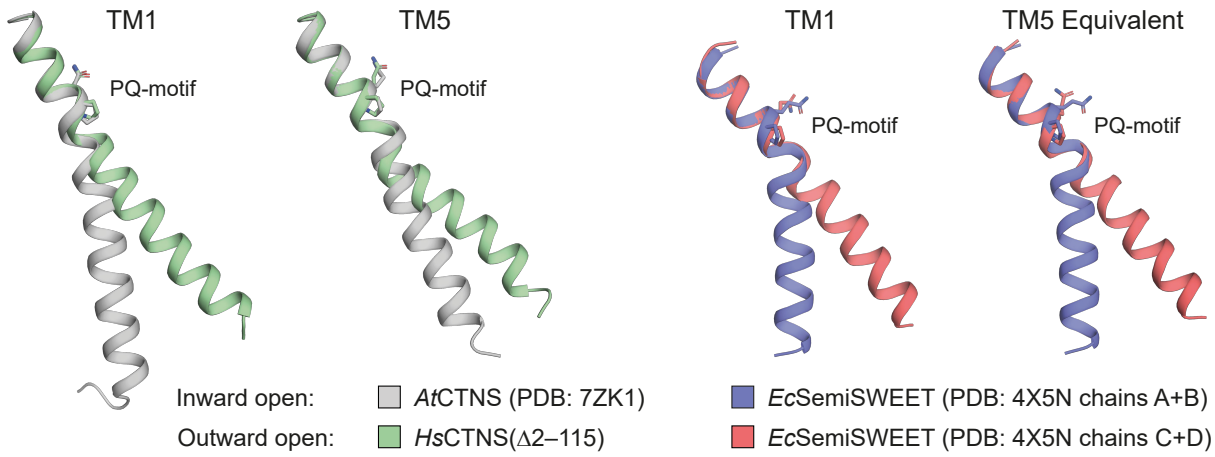


Fig. 9.5 – PQ-Motif Prolines as Molecular Hinges.
Cartoon representation of TMs 1 and 5 aligned from the top of the helix to the PQ-motif glutamine, illustrating the helical movement after the proline residue observed between the inward- and outward-open conformations.

a manner where TMs 1 and 5 are swapped, *i.e.* TMs 5, 2, and 3 of PnuC superimpose with TMs 1–3 of CTNS to form the N-terminal THB, and similarly for the C-terminal THB. This is consistent with earlier observations based on the comparison of PnuC with the SWEETs, which was described as the result of a possible domain swapping event, suggesting that PnuCs and members of the PQ-loop family have evolved from common ancestors (Jaehme *et al.*, 2015, 2016; Feng & Frommer, 2016).

9.1.2 Structural Characteristics of CTNS

The CTNS structure contains common features shared among the PQ-loop family members, as well as novel ones that have not been observed before. The defining characteristic of the family is the proline-glutamine (PQ) dipeptide, which resides within the most conserved region of these proteins: the end of two helices and their connecting loop (Ponting *et al.*, 2001). In a condensed format, this yielded the family name ‘‘PQ-loop’’ (Saudek, 2012), which is often mistaken for a location description of the dipeptide. Most PQ-loop transporters, such as CTNS and the SemiSWEETs, contain two PQ-motifs, one each on TM1 and TM5 (or its equivalent), whilst receptors like the KDELR only retain the latter (Newstead & Barr, 2020). The prolines were shown to function as molecular hinges, inducing a kink in the helix that allows the transporter to transition between inward- and outward-open conformations. This was first illustrated for the *Ec*SemiSWEET sucrose transporter (Lee *et al.*, 2015) and holds for CTNS, although the helices seem to bend to a lesser extent (**Fig. 9.5**). The prolines have been described as indispensable for the transport function of a number of PQ-loop transporters (Liu *et al.*, 2012; Xuan *et al.*, 2013; Lee *et al.*, 2015; Tao *et al.*, 2015; Kawano-Kawada *et al.*, 2019, 2021) and are further analysed regarding their importance for CTNS-mediated transport in **Sec. 9.3.2**.

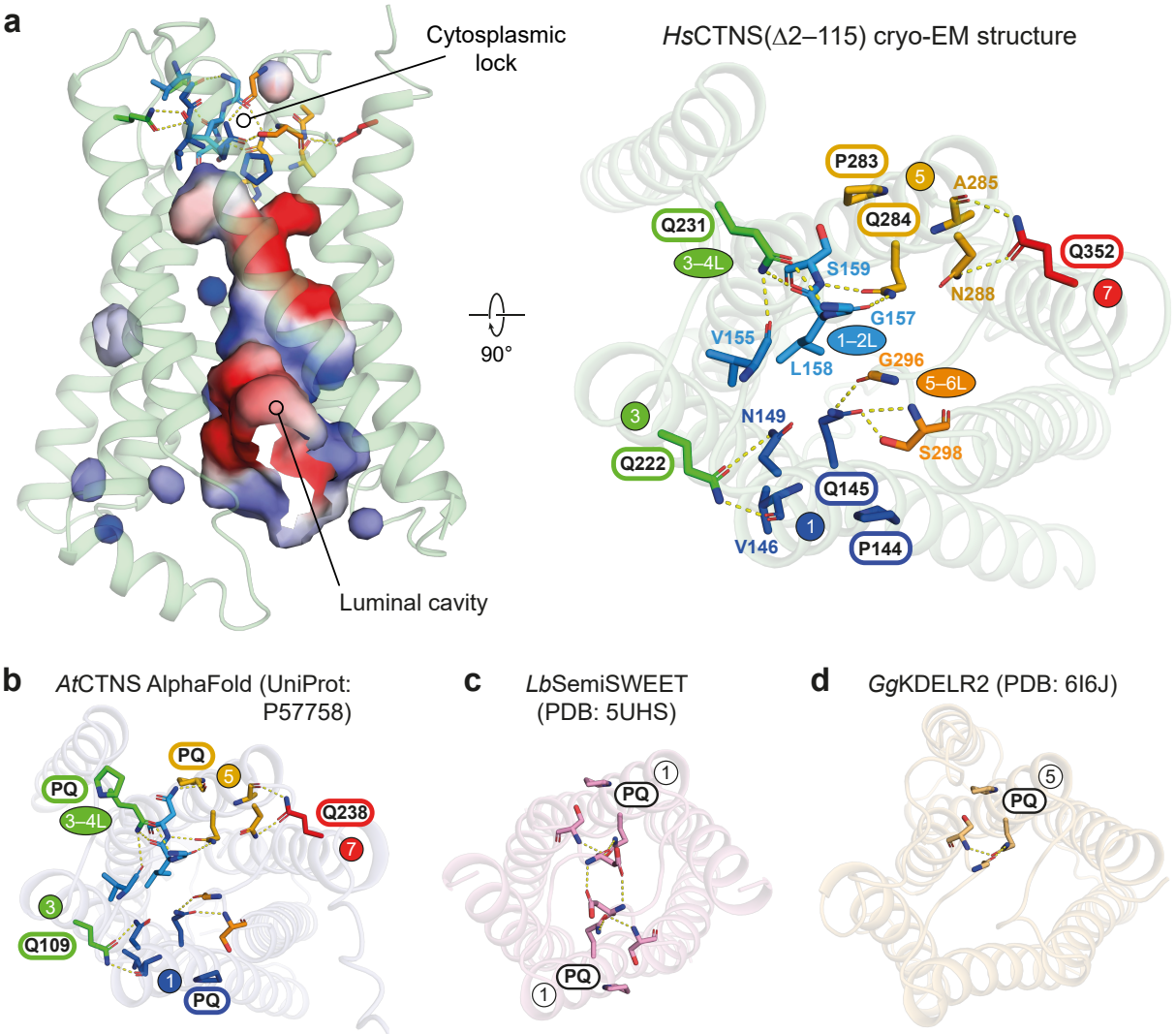


Fig. 9.6 – PQ-Motif Glutamines Seal CTNS on the Cytoplasmic Side.
a, View of *Hs*CTNS(Δ2–115) from **Fig. 8.19** in the outward-open state, highlighting the cytoplasmic lock and a cavity accessible from the lysosomal lumen. Conserved gating residues are shown as sticks, coloured based on their position, and hydrogen bonds are indicated using yellow dashed lines. **b**–**d**, Conserved gating residues present in other PQ-loop proteins.

The role of the PQ-motif glutamine was revealed in the first *Lb*SemiSWEET structure captured in an occluded state (Xu *et al.*, 2014). Here, the glutamine of one THB was found to coordinate the cytosol-facing loop of the other THB by forming hydrogen bonds with the backbone groups of two residues positioned on the loop, thus closing the transporter on that side. In *Hs*CTNS(Δ2–115), the cytoplasmic lock is formed through the PQ-motif glutamines 145 and 284 in a similar way, each interacting with a glycine (157 and 296) and serine (159 and 298) residue positioned on the TM1–2 and TM5–6 loops, respectively (**Fig. 9.6 a**). The remaining cytoplasmic loop connecting TMs 3–4 is linked to this structure *via* another glutamine (Q231), which, similar to Q284, binds the backbones of three residues (V155, G157, and L158) located on the TM1–2 loop. Interestingly, this Q231 is preceded by a proline in *At*CTNS (glycine in *Hs*CTNS), thus forming a third PQ-motif in the plant homologue, which, unlike the other

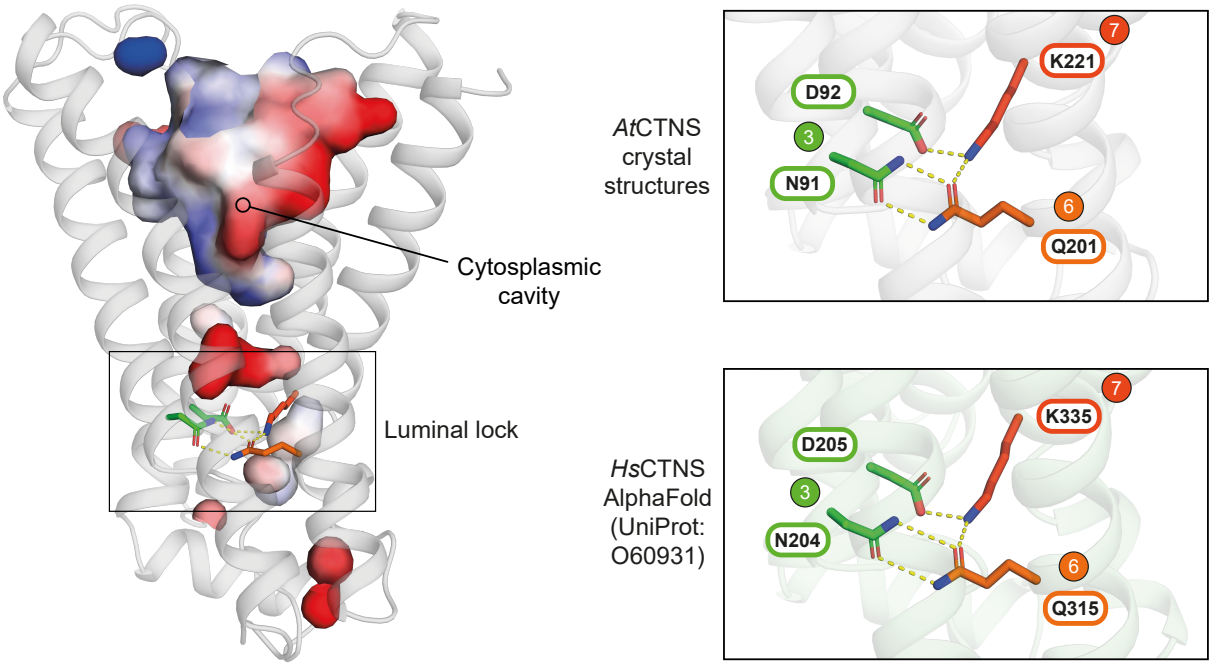


Fig. 9.7 – Unique Salt Bridge Seals CTNS on the Luminal Side.
View of *At*CTNS from **Fig. 7.6** (PDB: 7ZK1) in the inward-open state, highlighting the luminal lock and a cavity accessible from the cytoplasm. Conserved gating residues are shown as sticks, coloured based on their position, and hydrogen bonds and salt bridge interactions are indicated using yellow dashed lines.

two, resides in a loop region and remains bound to its interaction partners in the inward-open conformation, likely linking the TM1–2 and 3–4 loops throughout the transport cycle. Two more glutamines (222 and 352) form vital interactions near the cytoplasmic lock, linking TMs 3 and 7 to 1 and 5, respectively. All these glutamines are well conserved among other CTNS homologues, such as *At*CTNS (**Fig. 9.6 b**, using the occluded-state structure predicted by AlphaFold), demonstrating their functional importance in locking the transporter on its cytoplasmic side. Other members of the PQ-loop family, such as *Lb*SemiSWEET (Latorraca *et al.*, 2017), retain at least the interactions mediated by the PQ-motif glutamines whilst forming additional interactions that do not recur in CTNS (**Fig. 9.6 c**). The KDELR retains only the PQ-motif on the C-terminal THB (**Fig. 9.6 d**); its glutamine remains bound to the TM1–2 loop at all times. The absence of the N-terminal motif hereby allows the protein to transition asymmetrically and function as a trafficking receptor (Newstead & Barr, 2020). The eukaryotic SWEETs lack glutamines in both motifs, thus suggesting a distinct gating mechanism, however, no inward-occluded structures are available in the PDB.

In contrast to the cytoplasmic lock, gating on the luminal side of the transporter is mediated through a novel interaction that is not found in other PQ-loop family members. As highlighted for *At*CTNS and *Hs*CTNS (using the inward-open structure predicted by AlphaFold) in **Fig. 9.7**, a salt bridge is formed between an aspartate and a lysine, linking TMs 3 and 7 to close the

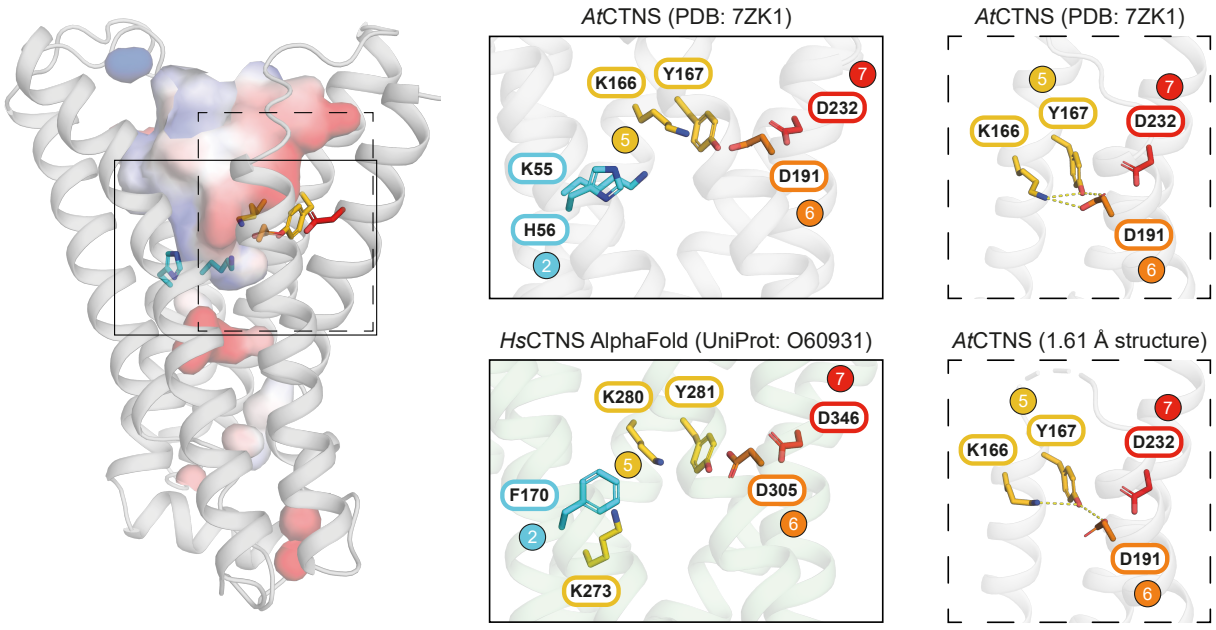


Fig. 9.8 – Charged Binding Site Residues of CTNS.
View of *At*CTNS from **Fig. 7.6** (PDB: 7ZK1) in the inward-open state, highlighting charged residues found within the binding site. The main differences regarding the binding site composition of the plant and human transporters are shown, respectively. Residues are shown as sticks, coloured based on their position, and hydrogen bonds and salt bridge interactions are indicated using yellow dashed lines.

transporters on the vacuolar and lysosomal luminal sides, respectively. The lysine residue is furthermore coordinated by yet another glutamine, located on TM6. This interaction network is strictly conserved within all CTNS homologues (see **Fig. 9.11**) and plays a crucial role in the transport mechanism, as outlined in **Sec. 9.3.3**. Additional hydrogen bonds are found that help seal the luminal-sided loops, however, the specific residues that form these interactions vary among different CTNS homologues.

Other PQ-loop transporters, such as the SWEETs and SemiSWEETs, also use salt bridges to facilitate gating, however, these interactions are formed by residues positioned within the inter-TM loops. Charged residues are generally not found within the protein core, including the substrate binding site. *Lb*SemiSWEET, for example, binds its ligand (glucose) through hydrophobic contacts and a small number of hydrogen bonds (Latorraca *et al.*, 2017). This is in contrast to other PQ-loop members like the KDELR or CTNS, where multiple charged residues are present within their respective binding pockets. *Gg*KDELR2 binds its peptide substrate using five salt bridges (Bräuer *et al.*, 2019). In CTNS, there are four charged residues found within the binding site: Two conserved aspartates and two lysines, one of which is located on a different TM helix in human CTNS compared to the plant homologue but ultimately, similar positions are observed for its ϵ-amino group (**Fig. 9.8**). *At*CTNS additionally contains a histidine on TM2; in *Hs*CTNS, this is a phenylalanine. Interestingly, most of these residues are positioned in the C-terminal THB. In some of our *At*CTNS crystal structures, we observe an

additional salt bridge formed between K166 and D191, and coordinated by Y167 (**Fig. 9.8**). We later propose that this salt bridge plays a crucial role during the transport cycle of L-cystine (see **Sec. 9.3.3**).

9.2 Assay Development to Address CTNS Function

CTNS uses the proton electrochemical gradient established in the lysosome to drive L-cystine transport across the lysosomal membrane (Kalatzis *et al.*, 2001), moving ligand molecules and protons at a 1:1 stoichiometry (Ruivo *et al.*, 2012). To characterize the function of CTNS, we developed a variety of complementary binding and transport assays (**Fig. 9.9**). As discussed in **Sec. 3.2**, *At*CTNS can be purified in a DDM and exchanged into a DM detergent micelle, enabling reconstitution into POPE:POPG vesicles by rapid dilution (**Sec. 2.6.2**), yielding proton-tight proteoliposomes (Parker *et al.*, 2014). We employed this reconstituted system to monitor CTNS-mediated proton flux using the pH-dependent fluorophore pyranine, as well as L-cystine transport using ^{14}C-radiolabelled ligand in separate *in vitro* assays. To address the importance of specific residues towards CTNS function, we generated a series of point mutants in the plant transporter (**Fig. 9.10**) and tested their function using these assays. Purification of *Hs*CTNS or its NTD-truncation *Hs*CTNS(Δ2–115) required the use of LMNG:CHS detergent (**Sec. 8.2**). Due to its low CMC, LMNG:CHS is thought to be difficult to remove during reconstitution into lipid vesicles, hindering the formation of proton-tight proteoliposomes. Thus, we opted to express *Hs*CTNS in *Xenopus laevis* oocytes and use electrophysiological measurements under two-electrode voltage clamp (TEVC) to complement the data generated for the plant homologue.

9.2.1 Cystine Binding Assay

Cystine binding was carried out using the nanoDSF-based tryptophan fluorescence quench assay as established in **Sec. 3.3.2**; methodical specifications are described in **Sec. 2.6.1**. For these experiments, we only used *At*CTNS and its variants solubilized in DDM, as this generally led to improved melting profiles with higher T_ms compared to the same proteins engulfed in a DM detergent micelle (see **Fig. 3.5 c** and **d**). The nanoDSF profiles for the utilized variants are illustrated in **Fig. A.11**.

9.2.2 Transport Assay Using Pyranine

Purified *At*CTNS and variants thereof were reconstituted into POPE:POPG liposomes (see **Fig. 9.10 b** and **Sec. 2.6.2**), and proton-coupled cystine transport was monitored using the pH-sensitive fluorophore pyranine (trisodium 8-hydroxypyrene-1,3,6-trisulfonate; Clement & Gould, 1981) in a similar way as reported previously (Parker *et al.*, 2014; Minhas *et al.*, 2018). In the presence of KCl, the pyranine dye was first entrapped within the proteoliposomes.

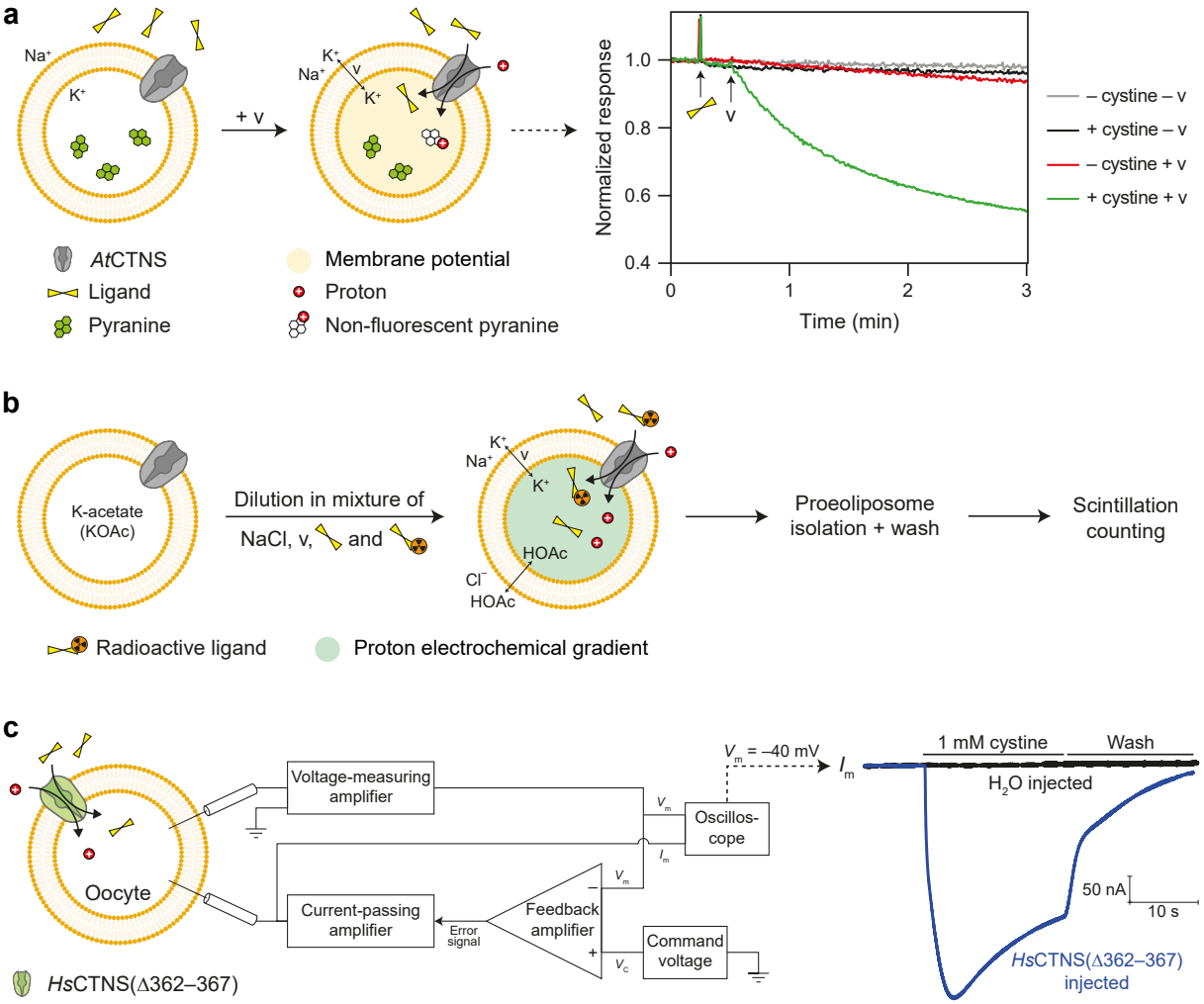


Fig. 9.9 – Assays Used to Study CTNS Function.
Schematic illustration and exemplary data for the transport assays used in this work: **a**, Proteoliposome based assay using pyranine to monitor proton movement, **b**, proteoliposome based assay using radioactive ligand, and **c**, TEVC assay using *Xenopus laevis* oocytes to measure currents elicited by L-cystine. The TEVC setup circuit was adapted from Schulien *et al.* (2019). v = valinomycin, a K^+-specific ionophore; V_m = membrane potential; I_m = membrane current; V_c = command voltage.

Then, after diluting the proteoliposomes in NaCl-containing buffer at the same molarity, the fluorescence was recorded. The substrate was added at the 15 s mark, and after 30 s, valinomycin (v), a K^+-specific ionophore, was added, equilibrating the potassium concentration across the membrane (Su *et al.*, 2019), thus creating a negative-inside membrane potential ($\Delta\Psi$) that drives CTNS-mediated proton-coupled cystine transport. Subsequent protonation of the pyranine reporter shifts its excitation spectrum, leading to a reduction in the measured fluorescence. The assay is illustrated schematically and annotated with typical results in **Fig. 9.9 a**; methodical specifications are described in **Sec. 2.6.3**.

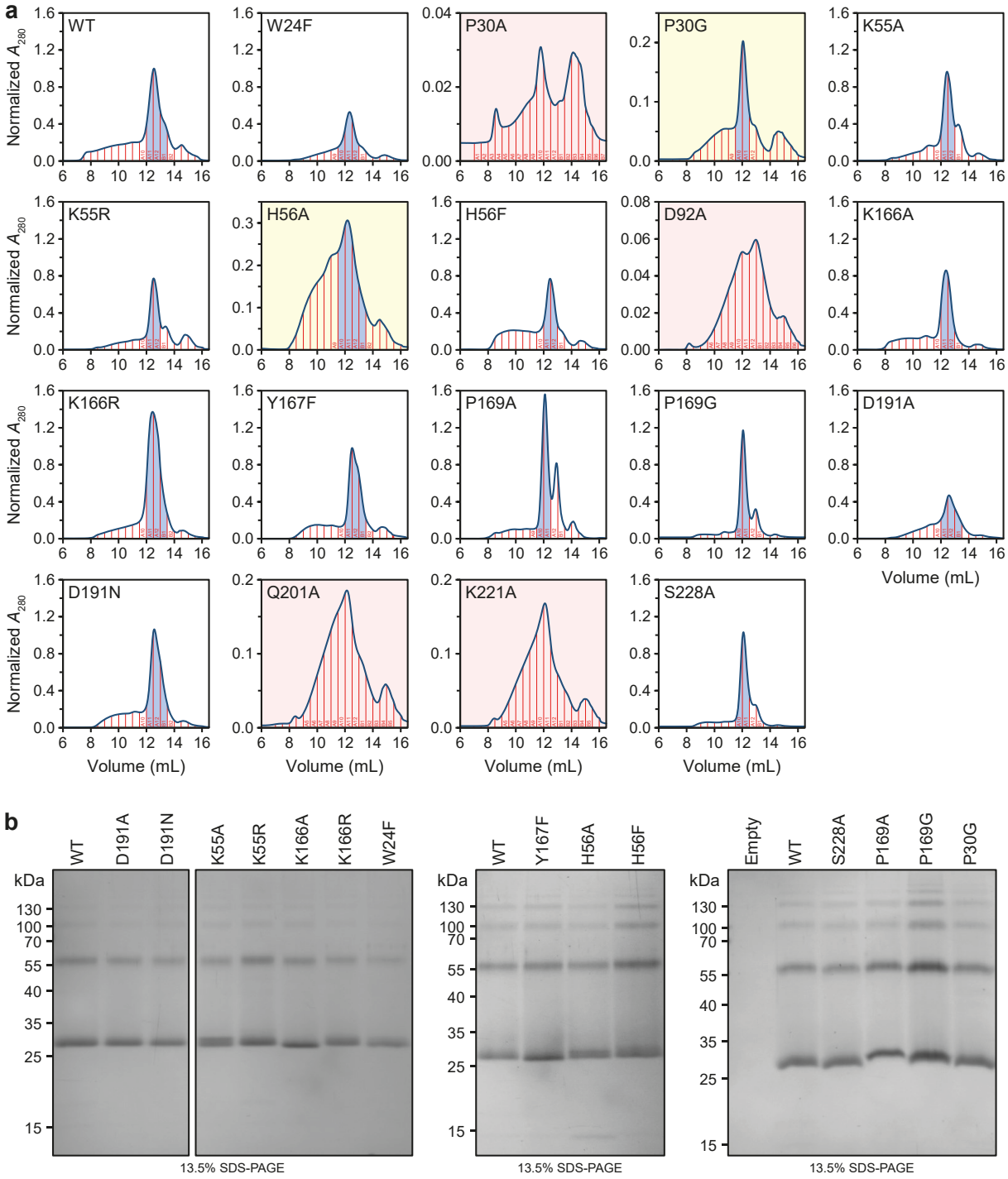


Fig. 9.10 – Purification & Reconstitution of *At*CTNS Variants.
a, Purification SEC traces (in DM detergent) normalized based on the WT transporter and accounting for differences in expression volume. Pooled fractions are shaded in blue. Reduced yields are highlighted in yellow, and poor-quality *At*CTNS variants in red. **b**, SDS-PAGE analysis of 2.5 µg re-solubilized protein (in 1% DDM) from each reconstituted sample, taken after a 30 min ultracentrifugation step at 200,000 × *g*.

9.2.3 Radioactive Transport Assay

Proteoliposome-based transport assays using radiolabelled L-cystine (^{14}C) were adapted from Parker *et al.* (2017). The liposomal lumen was loaded with K-acetate, and proteoliposomes were

subsequently diluted into a NaCl-containing buffer also supplemented with valinomycin (v) and a mixture of radioactive and conventional ligand. Acetic acid is membrane permeable; its equilibration across the membrane generates a chemical gradient (ΔpH), resulting in a negative potential inside the vesicles (Kaczorowski *et al.*, 1979). Coupled with the electric potential of the v-induced K^+ diffusion gradient ($\Delta\Psi$), these conditions create a proton electrochemical gradient ($\Delta\mu_{H^+}$) that drives ligand transport. To stop the reaction, vesicles were isolated on a filter and washed extensively. Scintillation counting was then used to analyse the ligand levels within the proteoliposomes, and a standard curve was used to calculate molar concentrations. The assay is illustrated schematically in **Fig. 9.9 b**; methodical specifications are described in **Sec. 2.6.4**.

9.2.4 Electrophysiological TEVC Assay

The TEVC assay was adapted from Ruivo *et al.* (2012). *Hs*CTNS without its C-terminal LTM (Δ362–367) was expressed at the cell surface of *Xenopus laevis* oocytes. Using TEVC, a negative-inside membrane potential of −40 mV was generated and continuously maintained. Specific currents elicited upon the addition of cystine were measured. The experiment was carried out by Sacha Salphati (Newstead and Tucker groups), a schematic illustration is presented in **Fig. 9.9 c**, however, methodical specifications are not provided here and will instead be found in his thesis or more briefly in Löbel *et al.* (2022).

9.3 Functional Characterization of CTNS

Much of this thesis focuses on the structure determination of *At*CTNS (**Chs. 4** to **7**), which was chosen as a model to study the human transporter. Despite sharing most structural features within the transmembrane domain (TMD), as described in **Sec. 9.1**, plant CTNS from *Arabidopsis thaliana* does not contain the N-terminal domain (NTD) found in the human and other eukaryotic transporters (**Fig. 9.11**). As the structure of *Hs*CTNS remains intact without the NTD, given that we were able to purify the truncated transporter and, furthermore, solve its structure using cryo-EM (**Ch. 8**), we sought to determine whether this domain holds any significant functional importance. Analysing *Hs*CTNS using our TEVC assay (**Fig. 9.9 c**), we found that the transport kinetics are not affected by the NTD, with both the full-length and truncated protein showing similar Michaelis-Menten constants (K_Ms) for L-cystine at 117 ± 2.4 and 117 ± 29 µM, respectively (**Fig. 9.12 a**). We then determined the K_M for L-cystine transport of the plant transporter using our *in vitro* assays (**Fig. 9.9 a** and **b**). We found that *At*CTNS, like its human homologue, transports cystine in a proton-coupled fashion with a K_M of 19 ± 2.9 µM when monitoring proton flux using pyranine (**Fig. 9.12 b**) and ~13 µM when measuring the transport of radiolabelled ligand directly (**Fig. 9.12 c**). These results suggest a slightly higher ligand affinity for *At*CTNS compared to the human transporter.

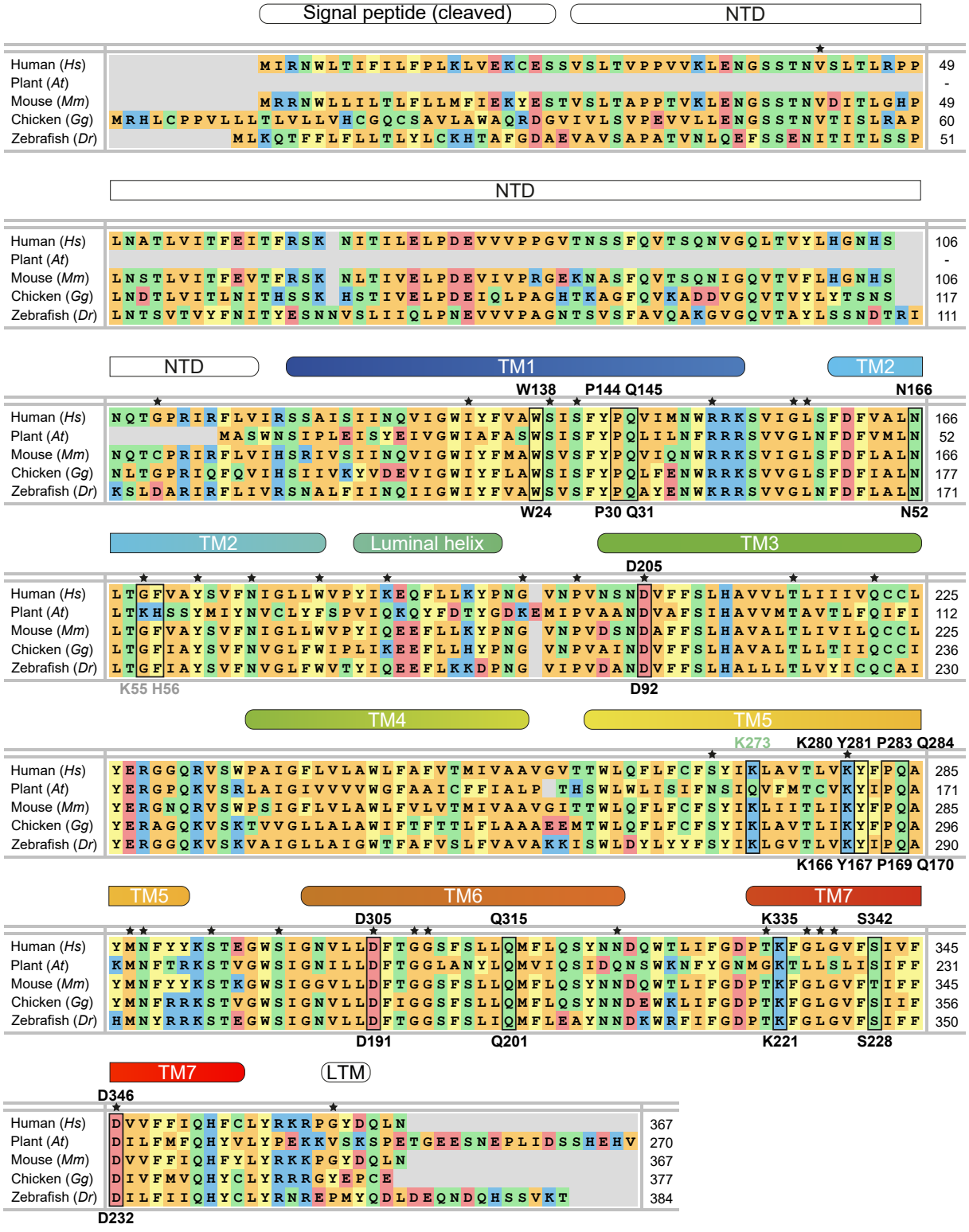


Fig. 9.11 – Annotated Sequence Alignment of CTNS Homologues.
a, Sequence alignment of CTNS homologues coloured *via* amino acid chemistry. TM helices 1–7 for human and plant CTNS as observed in **Figs. 9.1** and **9.2** are depicted above and below the alignment, respectively. Functionally important residues are labelled, and differences between the human and plant homologues are coloured in grey and green, respectively. Cystinosis-causing mutations, as listed in **Table 9.2**, are highlighted using stars.

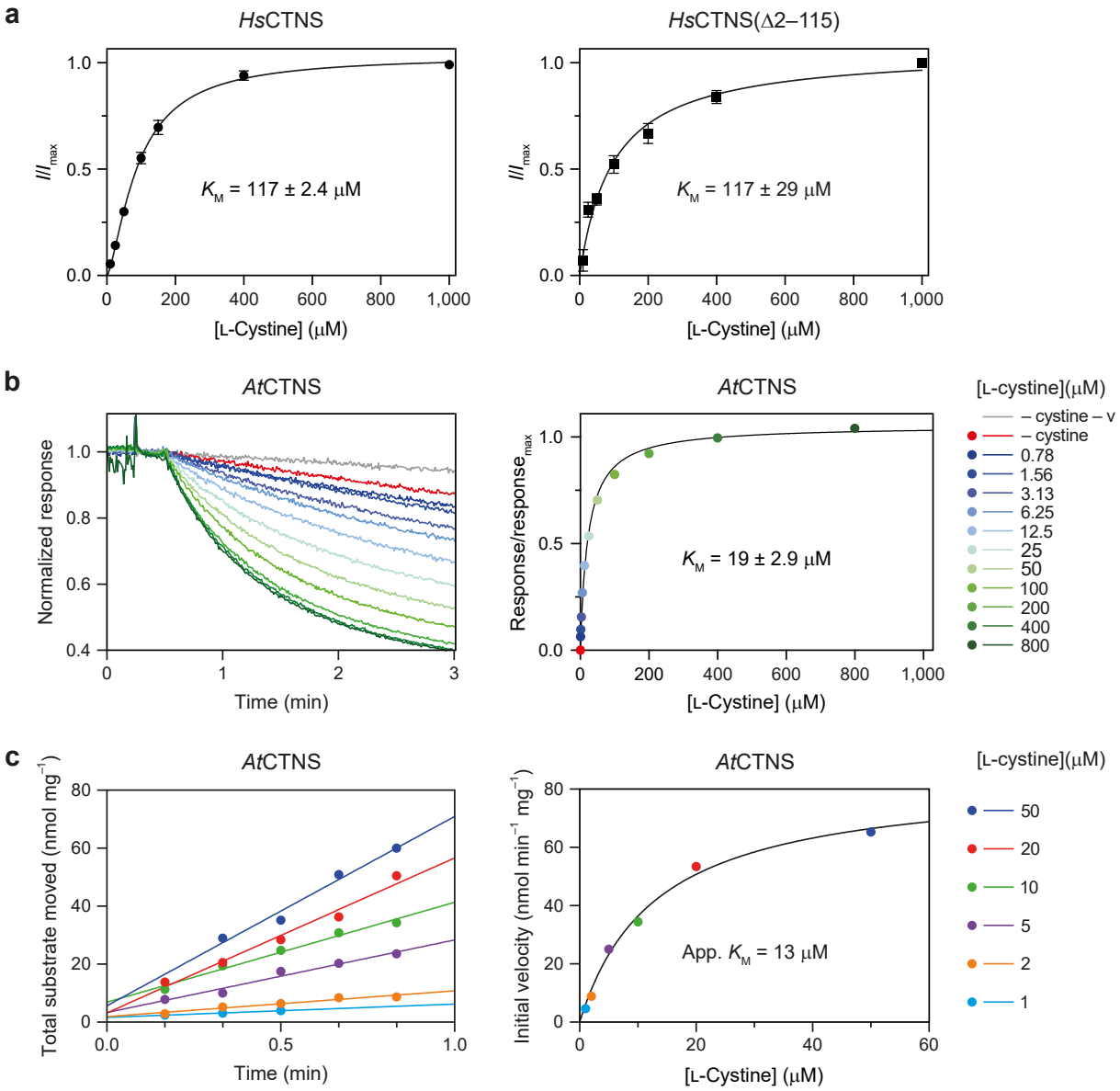


Fig. 9.12 – Functional Characterization of CTNS.
a, Representative K_M curves derived from TEVC for full-length *Hs*CTNS(Δ2–115) and a construct lacking the NTD (Δ2–115). Error bars SEM. The K_Ms were calculated from three independent experiments, error SD. **b**, Representative set of raw data for *At*CTNS from the pyranine assay and its kinetic analysis. The K_M was calculated from three independent experiments, error SD. **c**, Kinetic analysis of *At*CTNS using the radioactive transport assay.

9.3.1 Molecular Basis for Cystine Recognition

The cystine-bound co-crystal structure of *At*CTNS and Syb39 (**Sec. 7.4**) reveals detailed insights into the precise interactions formed between the ligand and the binding site of the transporter (**Fig. 9.8**). The use of Syb39 was crucial here as it enabled us to trap the ligand at the base of the polar pocket observed within the inward-open conformation of *At*CTNS, thus preventing it from dissociating (**Fig. 9.13 a**). Cystine adopts an extended conformation within the binding site and makes several notable interactions with conserved side chains (**Fig. 9.13 b** and **c**).

At physiological pH, cystine can exist in three different ionic forms: zwitterionic, negatively, or double negatively charged, depending on the protonation state of the two amino groups, which have different pK_a values of 7.48 and 9.02, respectively (Bannai & Kitamura, 1981). The ligand-bound complex was crystallized at pH 6.5, which favours the protonation of both amines to yield ammonium groups. In this zwitterionic state, cystine forms a salt bridge interaction between one of its ammonium groups and D191 on TM6, and two more between the two carboxylate groups and K55 (TM2) and K166 (TM5), respectively. Two additional hydrogen bonds are formed through interactions with N52 (TM2) and S228 (TM7). The disulfide of cystine is positioned within close proximity to H56 on TM2, which is a phenylalanine in the human transporter, representing the only major structural difference observed between their binding sites (**Fig. 9.13 b** and **f**).

To assess the importance of these interactions towards cystine binding and transport, we assayed the relevant *At*CTNS variants. We found that all three salt bridge interactions are critical for both cystine recognition and transport, with alanine variants resulting in inactive protein in both the binding and pyranine-based transport assays (**Fig. 9.13 d** and **e**). Only K55A retained a minor binding capacity for cystine at ~10%, similar to a more conservative mutation of D191 to asparagine. Both these variants are completely transport deficient, though, which is also true for conservative mutations of the lysine residues to arginine, highlighting the importance of the three salt bridge interactions for both cystine recognition and transport. Using our TEVC assay, we also tested the equivalent residues in the human homologue. K55 of the plant transporter is moved from TM2 to TM5 (K273 in *Hs*CTNS, **Fig. 9.13 f**), nevertheless, mutation to alanine eliminates ligand-elicited currents (**Fig. 9.13 g**). The same is true for K280A and D305A, the equivalent variants of K166A and D191A in *At*CTNS, thus suggesting a conserved transport mechanism between CTNS homologues that relies on three salt bridge interactions.

Another critical residue for transport is W24, which is positioned at the base of the binding pocket not far from the ammonium group of cystine that interacts with D191 (~3.4 Å). Exchanging W24 with a phenylalanine results in an alteration of the nanoDSF profile of the *At*CTNS variant compared to WT, transforming the blue-shift melting curve into a red-shift one whilst also reducing the T_m by ~10 °C (**Fig. A.11**). Nevertheless, cystine addition still quenches the initial tryptophan fluorescence signal, indicating binding is only reduced to ~60% of WT levels (**Fig. 9.13 d**). In contrast, cystine transport *via* the W24F variant is almost completely abolished, retaining only ~5% of WT transport levels (**Fig. 9.13 e**). Taken together, this suggests a possible cation-π interaction with the ammonium group of cystine and, as discussed, later demonstrates an important role for W24 in facilitating the conformational changes required for transport (**Sec. 9.3.3**).

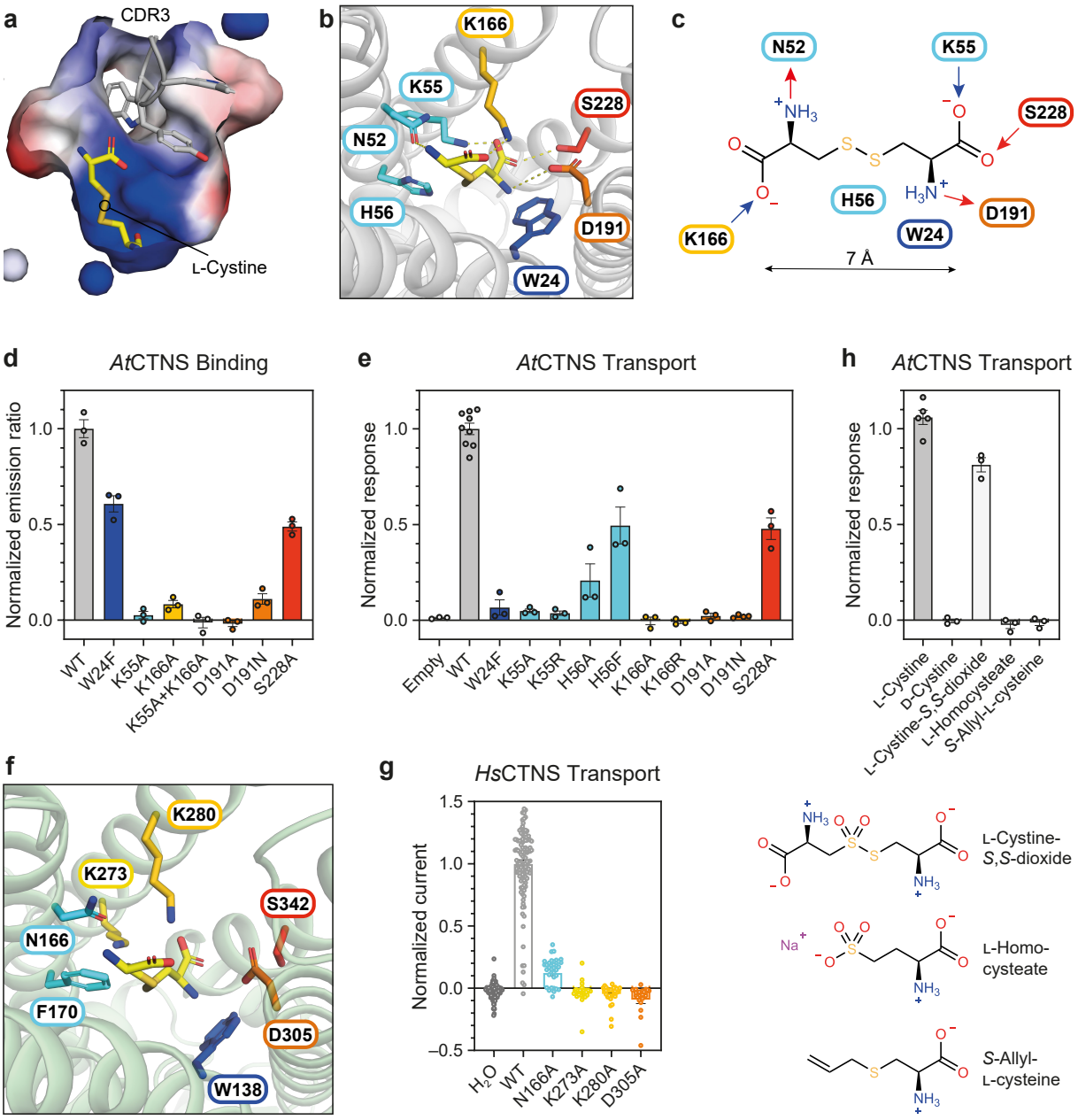


Fig. 9.13 – Functional Characterization of Binding Site Variants.
a, View of the cytosolic cavity of *At*CTNS in the inward-open and substrate-bound states (PDB: 7ZKW chain B), showing the position of L-cystine and the CDR3 loop of Syb39. **b**, Zoomed-in view of the binding site, showing interactions using yellow dashed lines. **c**, Schematic illustration of **b**. **d**, Tryptophan quench binding assay results for *At*CTNS variants at 3.2 mM L-cystine, error bars SEM. Typical nanoDSF traces are shown in **Fig. A.11**. **e**, Pyranine-based transport assay results for *At*CTNS variants at 200 µM L-cystine, error bars SEM. **f**, Ligand from **b** merged into the inward-open *Hs*CTNS structure predicted by AlphaFold (UniProt: O60931), showing a widely conserved binding site. **g**, TEVC assay results for *Hs*CTNS variants at 1 mM L-cystine, error bars SEM. **h**, Substrate specificity of *At*CTNS as measured using our pyranine-based assay with 500 µM substrate, error bars SEM.

Cystine also forms a hydrogen bond with S228 (**Fig. 9.13 b** and **c**). An alanine mutant retains about 50% activity in both binding and transport assays (**Fig. 9.13 d** and **e**), indicating a supporting role for this residue during the transport cycle. Mutating H56 to alanine reduces

transport to ~20%; exchanging it with a phenylalanine, as present in the human homologue, results in ~50% transport activity, suggesting the charge is not essential for cystine:proton symport at the experimental conditions. H56 may have evolved in *At*CTNS as a facilitator for protons to increase their availability within the binding site, considering that the proton gradient between the vacuole and cytoplasm of plant cells is less steep than that between the lysosome and cytoplasm of mammalian cells (Shen *et al.*, 2013; Ohkuma & Poole, 1978). A final hydrogen bond is formed with N52. We assayed its equivalent N166 mutated to alanine in *Hs*CTNS using our TEVC assay (**Fig. 9.13 g**) and found that it retains ~20% of its activity, suggesting a similar supporting role for function as S228 and H56 in the plant transporter.

As established in **Sec. 3.3.2** using our binding assay, *At*CTNS features a similarly narrow substrate preference compared to the human transporter (Ruivo *et al.*, 2012). Using our pyranine-based assay, we tested whether *At*CTNS can transport any of these alternative ligands. L-Cystine-*S*,*S*-dioxide showed a strong F_{Trp} quench (**Fig. 3.6 c**), indicative of binding, and interestingly, the two additional carbonyl groups do not affect transport as well, retaining ~80% of L-cystine levels (**Fig. 9.13 h**). The other two analogues are not transported, though, including *S*-allyl-L-cysteine, which retained similar binding levels compared to L-cystine, albeit at a threefold increase in concentration. These results emphasize the importance of the collective interaction network between the ligand and the binding site, suggesting that a subset of interactions may be formed but only if the charged groups on the substrate are at a precise distance (~7 Å for cystine or its sulfonated analogue) can the conformational changes be triggered and transport facilitated (**Sec. 9.3.3**). Finally, we also tested D-cystine, which is not transported, highlighting that *At*CTNS is stereospecific.

9.3.2 Molecular Basis for Cystine Transport

Secondary active transporters move their substrates *via* an alternating access mechanism (Jardetzky, 1966). In PQ-loop transporters, the required structural changes rely on proline residues positioned within the TM regions, which act as molecular hinges (**Fig. 9.5**), switching the conformation of the transporter from an outward-open into an inward-open state around the substrate in a symmetric fashion (Newstead & Barr, 2020). The functional importance of the PQ-motif prolines has been described for several family members. In SWEETs and SemiSWEETs, the prolines of both PQ-motifs are critical for sugar transport (Xuan *et al.*, 2013; Lee *et al.*, 2015; Tao *et al.*, 2015). More recently, this was also reported for the yeast PQ-loop protein Ypq2 (Kawano-Kawada *et al.*, 2019). For LAAT-1 in nematodes (Liu *et al.*, 2012) and its human homologue PQLC2 (Jézégou *et al.*, 2012), as well as yeast Stm1 (Kawano-Kawada *et al.*, 2021), the proline of the first PQ-motif was described as indispensable for their respective transport functions. To test the impact of both PQ-motif prolines on CTNS-mediated cystine transport, we assayed alanine variants using both our TEVC and pyranine-based assays. Mutating the proline of the second PQ-motif (on TM5) results in transport-deficient protein

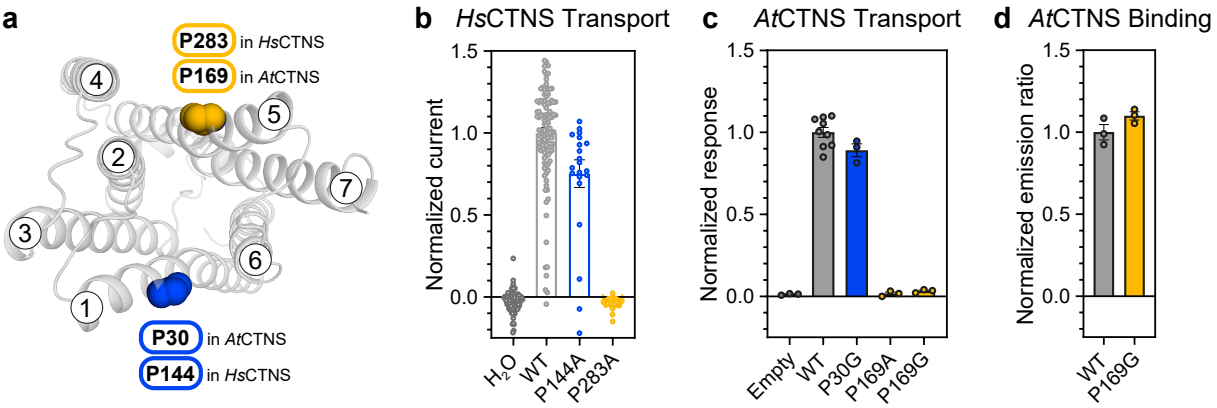


Fig. 9.14 – Functional Characterization of PQ-Motif Proline Variants.
a, Top view of *At*CTNS in the inward-open state (PDB: 7ZK1), highlighting the position of the PQ-motif prolines, illustrated as spheres. **b**, TEVC assay results for *Hs*CTNS variants at 1 mM L-cystine, error bars SEM. **c**, Pyranine-based transport assay results for *At*CTNS variants at 200 µM L-cystine, error bars SEM. **d**, Tryptophan quench binding assay results for *At*CTNS variants at 3.2 mM L-cystine, error bars SEM. Typical nanoDSF traces are shown in **Fig. A.11**.

(**Fig. 9.14 b** and **c**), however, L-cystine binding is not affected and even slightly increased to ~110% of WT levels (**Fig. 9.14 d**). Interestingly, the prolines of the first PQ-motif (on TM1) are dispensable as a *Hs*CTNS-P144A variant retains ~75% cystine transport levels compared to WT. An equivalent *At*CTNS variant could not be purified (**Fig. 9.10 a**), however, a glycine mutant could and retained ~90% of the transport function (**Fig. 9.14 c**). Taken together, this suggests that the second PQ-motif is responsible for facilitating the conformational changes required for transport. It also suggests that CTNS, unlike other PQ-loop transporters, may operate more asymmetrically (**Sec. 9.3.3**), resembling a close relationship to the KDELR, which has lost the PQ-motif on TM1 entirely (Newstead & Barr, 2020).

Cystine transport *via* CTNS is coupled to the translocation of protons (Kalatzis *et al.*, 2001), as demonstrated for *At*CTNS in **Fig. 9.12 b** and **c**. Scanning mutagenesis has previously identified D305 of the human protein as a potential site for proton binding (Ruivo *et al.*, 2012). When mutating the equivalent aspartate in the plant transporter (D191) to asparagine, we observed a slow proton leak despite the transporter's inability to move L-cystine (**Fig. 9.15 b** and **e**), suggesting this variant is uncoupled, thus confirming the previous observation. D191 is coordinated by Y167 through a hydrogen bond interaction. Interestingly, mutating Y167 to phenylalanine alters the transport kinetics, increasing the K_M to 73 ± 13 µM and the V_{max} to ~130% of WT levels (**Fig. 9.15 c**). Both L-cystine and protons are found to move at an increased rate as tested using our pyranine-based and radioactive transport assays (**Fig. 9.15 d** and **e**), suggesting that the hydroxyl group of Y167 is an important kinetic regulator of the transport cycle. D191 also forms a salt bridge with K166, however, this is only observed in our Syb39-bound structures (apo state in PDB: 7ZK1 and cystine-bound in PDB: 7ZKW) when the transporter is opened the furthest to the cytoplasm. Other structures of CTNS do not show

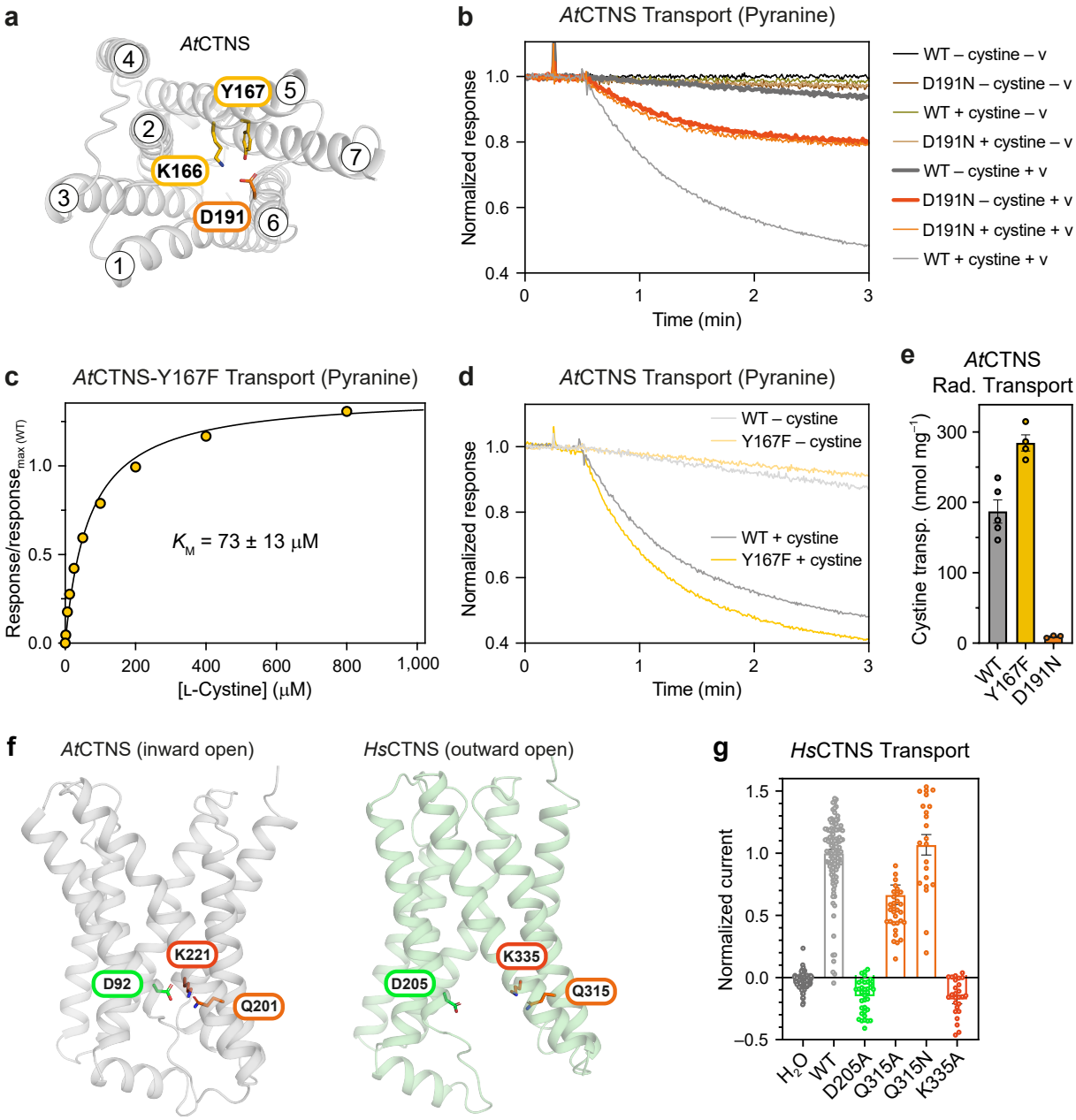


Fig. 9.15 – Functional Characterization of Salt Bridge Residue Variants.
a, Top view of *At*CTNS in the inward-open state (PDB: 7ZK1), highlighting the position of the cytosol-facing salt bridge. **b**, Pyranine-based transport assay results for WT *At*CTNS and the D191N variant, showing a slow proton leak. Substrate used at 200 µM. **c**, Representative fit for the kinetic analysis of *At*CTNS-Y167F using the pyranine assay. The K_M was calculated from three independent experiments, error SD. Typical traces for the data point at 200 µM L-cystine is illustrated in **d**. **e**, Radioactive transport assay results for WT *At*CTNS and the Y167F and D191N variants at 50 µM substrate after 4 min, error bars SEM. **f**, Front view of *At*CTNS and *Hs*CTNS(Δ2–115), showing the engaged and disengaged salt bridge on the luminal-facing side of the transporters, respectively. **g**, TEVC assay results for *Hs*CTNS variants at 1 mM L-cystine, error bars SEM.

this feature and instead link the two side chains *via* hydrogen bonds with Y167 (**Fig. 9.8**). The correct interplay of these three residues is undoubtedly crucial for the transport function of CTNS. The presence of a slow proton leak for the *At*CTNS-D191N variant also suggests a

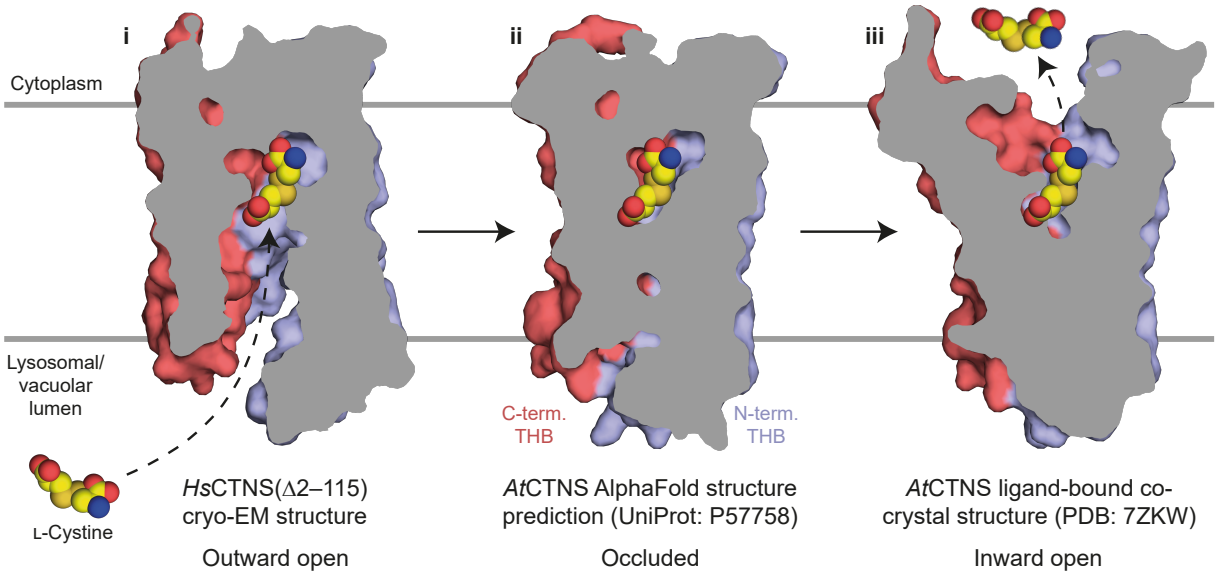


Fig. 9.16 – Alternating Access Cystine Transport *via* CTNS.
Slab representation of the major conformational states obtained using XRC, cryo-EM, and the AlphaFold structure prediction tool (Jumper *et al.*, 2021; Varadi *et al.*, 2022, 2024).

second proton binding site, as predicted before (Ruivo *et al.*, 2012). The salt bridge observed on the luminal side of the active site stands out as a likely candidate to serve this function (**Figs. 9.7** and **9.15 f**). Alanine variants of these residues could not be purified (**Fig. 9.10 a**), highlighting their structural importance. Instead, we used our TEVC assay to investigate their functional contribution. We found that mutating the coordinating glutamine was tolerated whilst the aspartate and lysine that form the salt bridge were indispensable for transport (**Fig. 9.15 g**), indicating that the luminal-sided salt bridge is also critical for transport.

9.3.3 Mechanism of Cystine Transport

To date, a detailed mechanism for alternating access transport of PQ-motif transporters has only been described for the bacterial SemiSWEETs (Lee *et al.*, 2015; Latorraca *et al.*, 2017). Transition from the outward-open to inward-open conformation was shown to occur spontaneously, even in the absence of substrate, and was driven by the formation of favourable interactions in the latter state (Latorraca *et al.*, 2017). Using the structures obtained in **Chs. 7** and **8**, as well as the AlphaFold prediction of *At*CTNS, which captures the transporter in an occluded state, we can illustrate the three main conformational states of CTNS, clearly showing alternating access to the vacuolar or lysosomal lumen and the cytoplasm (**Fig. 9.16**). Unlike the SemiSWEETs, CTNS couples substrate transport to proton translocation, thus introducing the necessity for tighter regulation. Our data suggest that proton-coupled transport is regulated through specific ligand-binding site interactions and the protonation/deprotonation of two conserved salt bridges, one on either side of the active site.

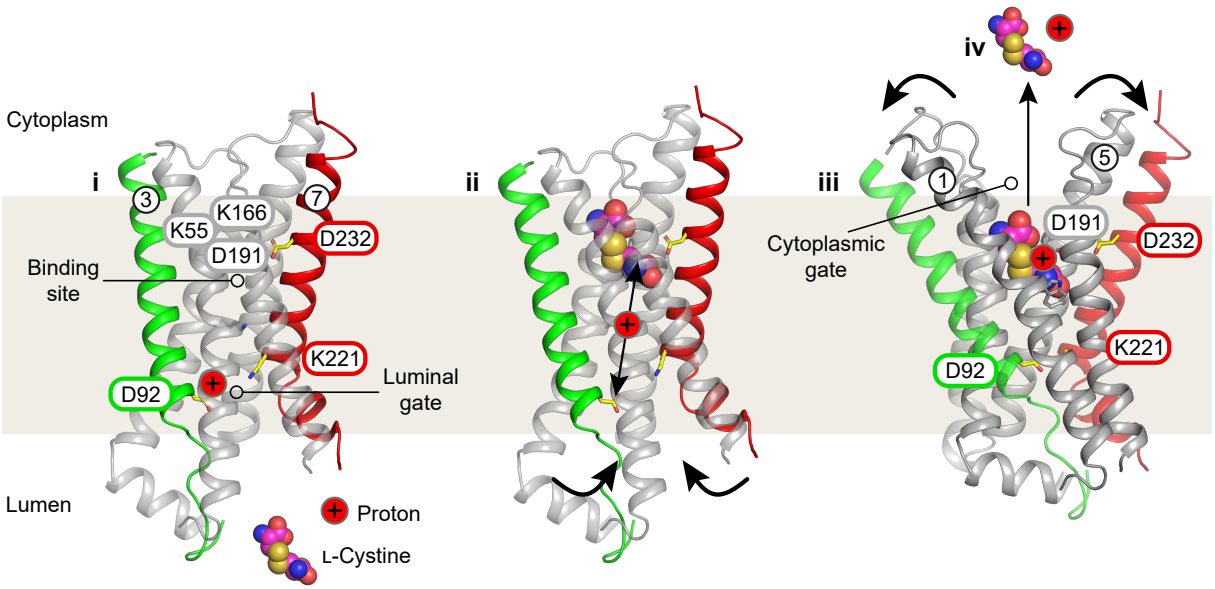


Fig. 9.17 – Proposed CTNS Transport Mechanism.
Residues are numbered based on the plant homologue of CTNS. For further information, refer to the main text. Adapted from Löbel *et al.* (2022).

As illustrated in **Fig. 9.17**, we propose the following mechanism: CTNS adopts the outward-open conformation (state **i**) upon protonation of D92, which weakens its interaction with K221, breaking the salt bridge that forms the luminal gate. In this state, the cytoplasmic gate is closed, aided by the PQ-motif glutamines, each of which coordinates the inter-TM loop of the opposite THB. TMs 1 and 6 and 2 and 5 are packed together, and the luminal ends of TMs 3 and 7 are separated to form a channel for the ligand to enter (see **Fig. 9.16**). Upon cystine binding, D92 likely deprotonates, and the luminal gate closes, yielding a substrate-bound occluded state (**ii**). In the binding pocket, cystine engages D191 and K166 on TMs 6 and 5, respectively, and induces the conformational changes necessary for the transporter to transition into the inward-open state (**iii**). In this conformation, the cytoplasmic ends of TMs 1 and 6 and 2 and 5 are spread apart, aided by the PQ-motif prolines acting as molecular hinges. This disengages the cytoplasmic lock, enabling cystine to exit the transporter (**iv**). Electrostatically, cystine release would be favoured by the protonation of D191, weakening the salt bridge interaction with one of the ammonium groups of the ligand. Upon exposure to the cytosol, D191 would then deprotonate, and a new transport cycle can commence.

Protonation and deprotonation of D92 and D191, respectively, are likely coordinated through ordered water molecules, as was previously shown for larger proton-coupled symporters (Parker *et al.*, 2017). In our 1.61 Å *At*CTNS crystal structure (from **Fig. 7.13**), we observe intermittent water networks that range from the cytoplasmic to the luminal side of the transporter (**Fig. 9.18**). Importantly, a continuous network is not found, otherwise, proton leakage through the transporter *via* the Grotthuss mechanism (Agmon, 1995) would be inevitable. In the inward-open state, D92 is connected to the vacuolar lumen *via* its salt bridge interaction

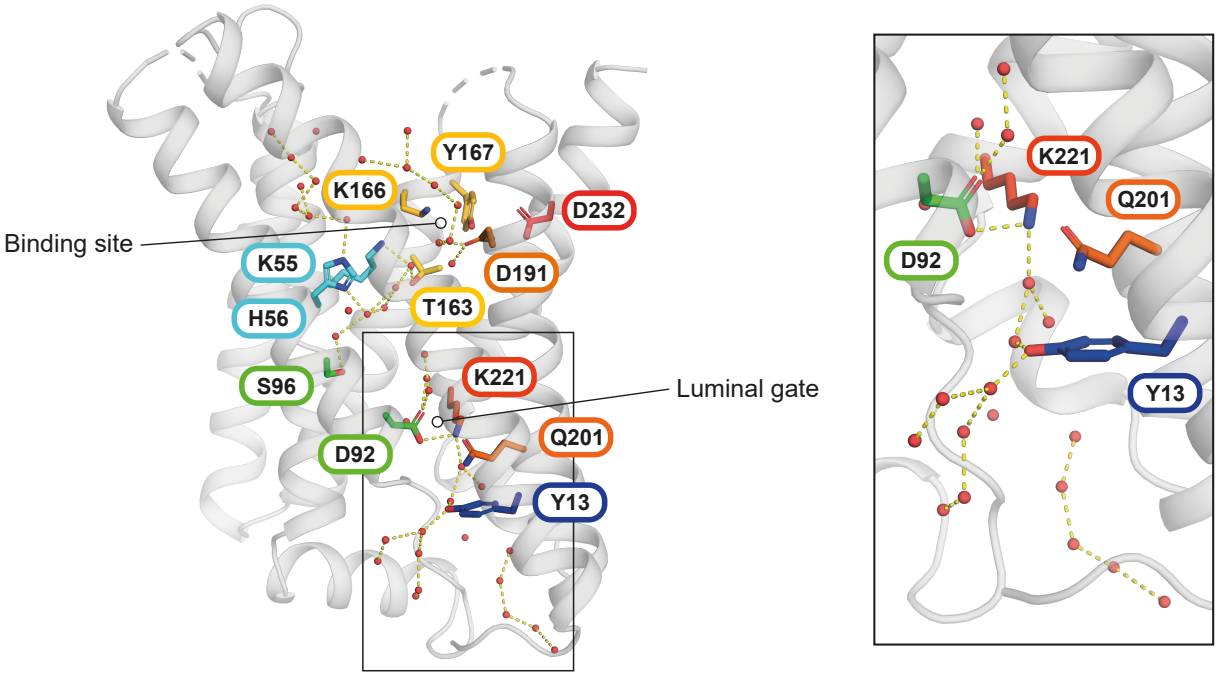


Fig. 9.18 – Water Networks in the 1.61 Å *At*CTNS-Syb1-Nb4 Co-Crystal Structure.
Cartoon representation of the structure from **Fig. 7.13**, showing intermittent water networks stretching from the cytoplasmic to the luminal side of the transporter. Water molecules are shown as spheres, and prominent side chain residues as sticks. Hydrogen bonds are illustrated using yellow dashed lines.

partner K221 and the hydroxyl group of Y13, which extends a network of coordinated water molecules, demonstrating a clear path that could facilitate the protonation of D92, with the proton ultimately coming from a hydronium ion present in the vacuolar lumen. Y13 is not a conserved residue, thus, the exact path for protonation of D92 may vary among different CTNS homologues. The carboxyl group of D191 is connected to a water network that ends in the cytoplasm, demonstrating a potential route for deprotonation. How the proton is moved from the luminal gate to the binding site cannot be answered using our structure. Perhaps a defined water network connecting D92 with D191 forms once the transporter has transitioned in the ligand-bound occluded state. Alternatively, the proton may be facilitated directly by cystine.

Assembled from a homodimer of THBs, the bacterial SemiSWEETs transition symmetrically around their substrates to facilitate transport (Lee *et al.*, 2015; Latorraca *et al.*, 2017). In contrast, the KDELR undergoes an asymmetric conformational cycle to facilitate its function as a trafficking receptor (Newstead & Barr, 2020). Following structural investigation and based on our functional data, we propose that cystine transport *via* CTNS is also mediated by an asymmetric transition from the outward-open to the inward-open state (**Fig. 9.19**). Unlike the KDELR, CTNS still contains both PQ-motifs, however, the proline on TM1 in the N-terminal THB is dispensable for function. When we superimpose the transporters from the outward-open and inward-open states, we do observe structural flexibility in TM1, however,

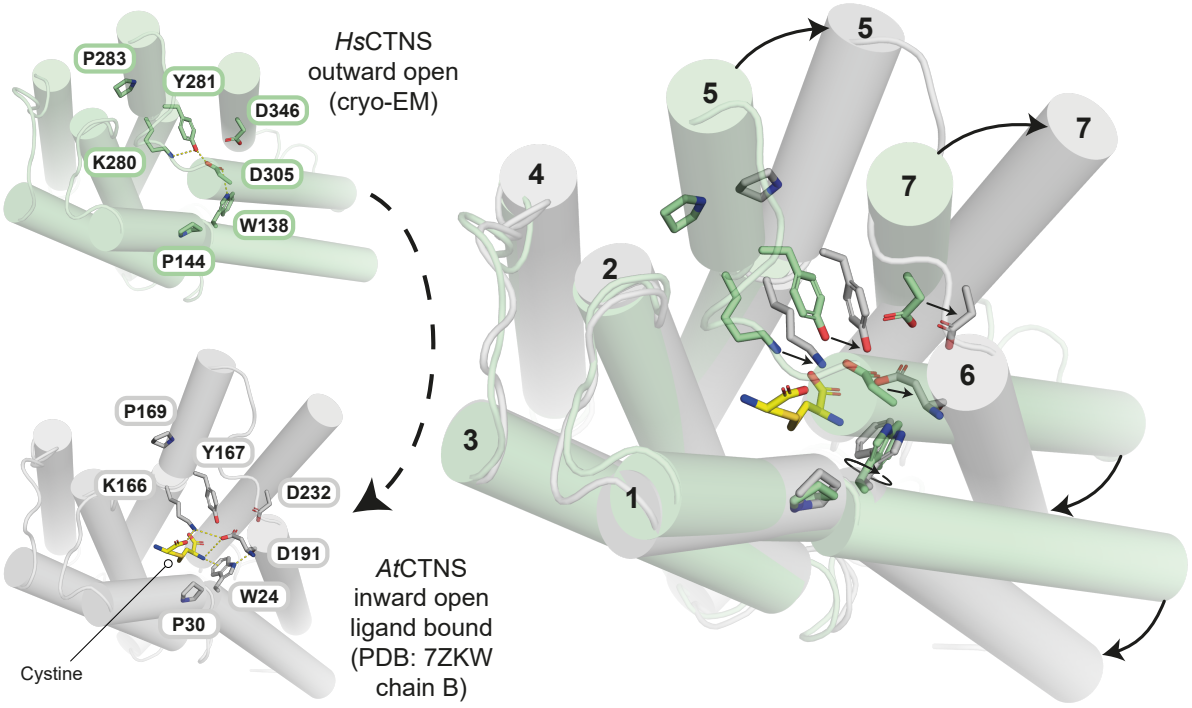


Fig. 9.19 – Model for the Asymmetric Transition Between Conformational States.
Cartoon representation of CTNS with cylinders as helices, showing the structural differences between the outward-open and inward-open conformations observed when the structures are superimposed based on N-terminal THB residue C_α atoms only.

the helix is bent much more homogeneously compared to TM5, the equivalent helix in the C-terminal THB, which is rather kinked than curved (see **Fig. 9.5**). This may suggest a division of labour within the transporter to facilitate the conformational changes. Structural analysis reveals that the N-terminal THB mainly composes hydrophobic and polar residues, whilst the C-terminal THB contains a number of charged elements, all positioned near the second PQ-motif proline. Upon cystine binding, W24 engages in a cation-π interaction with one of the ligand's ammonium groups, swivelling W24 around its axis (**Fig. 9.19**) and thus forming the base of the binding pocket. Flipping W24 involves the re-orientation of a hydrogen bond formed with the carboxylate of D191, which now interacts with its backbone carbonyl group instead. This likely induces a rotation in TM6 and the connected TM7, which would favour the formation of the luminal-sided salt bridge and, hence, the closure of the gate. However, additional rotational movement is required to open the cytoplasmic gate. Alignment of our outward-open and ligand-bound inward-open structures, using only C_α atoms belonging to the N-terminal THB, suggests that this is mediated through the interactions formed between cystine and the binding site residues (**Fig. 9.19**). It appears that cystine uses the N-terminal THB purely as an anchor point to reach the C-terminal THB and move the charged elements closer to one another, thus kinking the cytoplasmic half of TM5, aided by the PQ-motif proline, and further rotating TMs 6 and 7. As a result, the TM5–6 loop is moved away from the TM1–2 loop, and the transporter opens to the cytoplasm. Interestingly, this model of an asymmetric transition

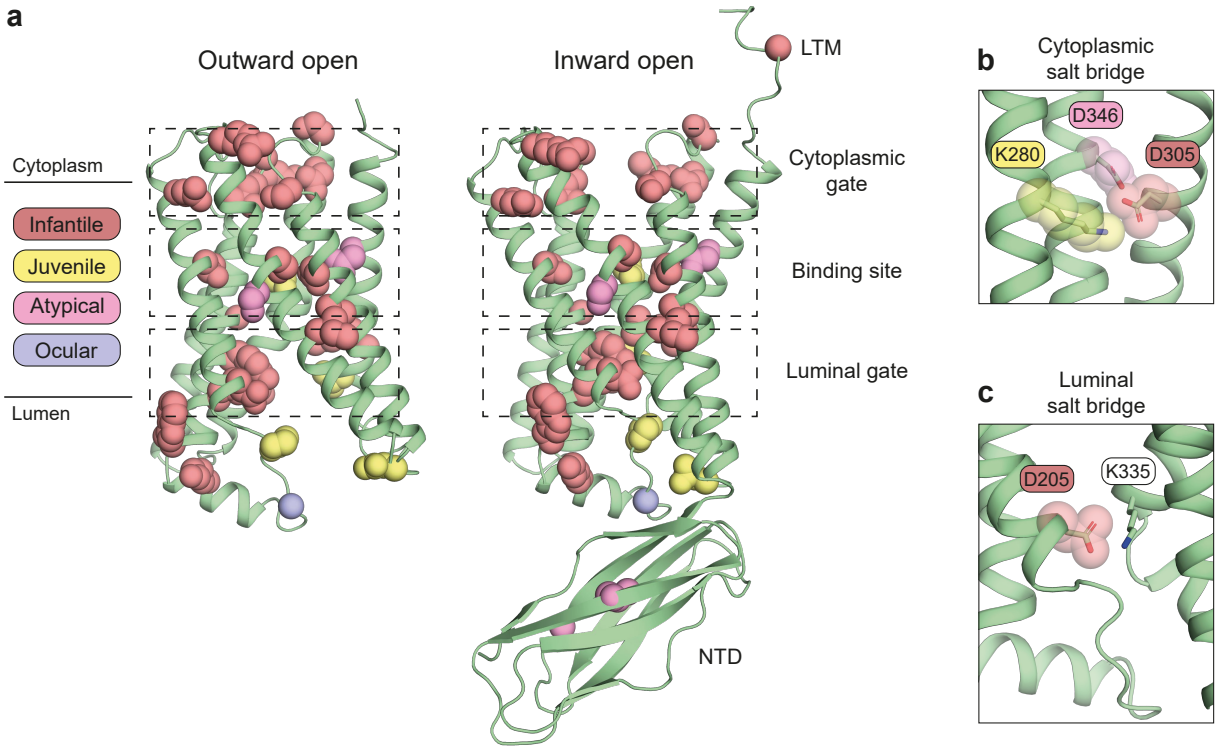


Fig. 9.20 – Structural Context of Cystinosis-Causing Mutations.
a, Position of the missense mutations listed in **Table 9.2** mapped onto our *Hs*CTNS(Δ2–115) structure in the outward-open conformation and the AlphaFold prediction (UniProt: O60931) in the inward-open conformation. Affected residues are shown as spheres and coloured based on the clinical phenotype. **b** and **c**, Zoomed-in view of the salt bridges as observed in the inward-open state.

leaves TMs 2–4 almost free of conformational changes, enabling potential dimerization along this interface without impairing the transport function. We observed the exact dimer interface in our cryo-EM structure of a *Hs*CTNS(Δ2–115) dimer (**Ch. 8**, specifically **Fig. 8.19**). The relevance of such a dimer remains elusive, however, it highlights another interesting topic of research.

9.3.4 Disease-Causing Mutation Mapping

Coupled with our functional analysis, the structures presented in this work provide a framework for understanding how specific mutations in the *CTNS* gene cause the different forms of cystinosis. To date, 42 missense mutations have been identified (listed in **Table 9.2**, with corresponding residues highlighted in **Fig. 9.11**) and linked to four different clinical phenotypes: infantile, juvenile, atypical, and ocular. Mapping the mutations onto the structure of the human transporter (**Fig. 9.20 a**) reveals that most of them result in the modification of a residue that is located within one of three regions critical for transport: the cytoplasmic and luminal gates as well as the binding site. In particular, disease-causing mutations are found to be present within the cytoplasmic-facing salt bridge, near the binding site, and the luminal salt bridge, which gates the transporter in the inward-open state (**Fig. 9.20 b** and **c**). Interestingly, mutations located

Table 9.2 – Annotated List of Cystinosis-Causing Missense Mutations.

Mutation	*At*CTNS equivalent	Region	Clinical phenotype
V42I	N/A	N-terminal domain	Atypical/juvenile
G110V	N/A	N-terminal domain	Atypical
I133F	I19	TM1	Infantile
S139F	S25	TM1	Atypical/juvenile
S141F	S27	TM1	Infantile
R151G	R37	TM1	Infantile
G157D	G43	TM1–2 loop	Infantile
L158P	L44	TM1–2 loop	Infantile
G169D	K55	TM2	Infantile
Y173H/C	Y59	TM2	Infantile
N177S/T	N63	TM2	Infantile/juvenile
W182R	F68	TM2	Infantile
K187R	Q73	Luminal helix (TM2–3 loop)	Infantile
G197R	K83	TM2–3 loop	Ocular
P200L	P87	TM2–3 loop	Juvenile
D205N	D92	TM3	Infantile
T216R	T103	TM3	Infantile
Q222R	Q109	TM3	Infantile
S270F/Y	N156	TM5	Infantile
K280R	K166	TM5	Juvenile
M287I	M173	TM5	Infantile
N288K	N174	TM5	Infantile
S293I	S179	TM5–6 loop	Infantile
S298N	S184	TM5–6 loop	Infantile
D305G/Y	D191	TM6	Infantile
G308R/E/V	G194	TM6	Infantile
G309D/V	G195	TM6	Infantile
N323K	Q209	TM6	Juvenile
T334N	G220	TM7	Juvenile
G337R	L223	TM7	Infantile
L338R/P	L224	TM7	Infantile
G339R	S225	TM7	Infantile
D346N	D232	TM7	Atypical/juvenile
G362R	V248	C-terminal tail	Infantile

close to either gate are almost always associated with the more severe infantile phenotype, whilst those closer to the binding site or at the luminal end of the transporter (including the NTD) may represent any of the phenotypes (**Fig. 9.20 a**). Mutations associated with infantile cystinosis include D205N and D305N, which we identified as transport-deficient (D205A in **Fig. 9.15 g** and D305A in **Fig. 9.13 g**). K280R was also found to be transport-deficient in both transporters (*At*CTNS(K166A/R) and *Hs*CTNS(K280A) in **Fig. 9.13 e** and **g**, respectively),

however, this variant is only linked to juvenile cystinosis. This finding is consistent with earlier observations (Kalatzis *et al.*, 2004) and may indicate an additional function of CTNS beyond cystine transport. Importantly, we also identified a minority of residues that map onto regions of the transporter that we identified to contain little functional contribution, such as V42 and G110 in the NTD. These may be folding mutations, however, further structural and functional analysis would be needed to understand the impact of these mutations and how they cause cystinosis.

CHAPTER 10

Conclusion

At the outset of this thesis, several questions were formulated regarding the structure and function of CTNS. After determination of the crystal structures of *At*CTNS in the inward-open ligand-free, as well as inward-open ligand-bound states (**Ch. 7**), and the cryo-EM structure of *Hs*CTNS(Δ2–115) in the outward-open conformation (**Ch. 8**), followed by the in-depth analysis of both the structure and cystine transport function of CTNS (**Ch. 9**), we can now answer some of these questions:

- How does CTNS fit structurally in the diverse PQ-loop family? What characteristics are shared, and are there any novel features?

The structures of CTNS superimpose well with other representatives of the PQ-loop family, both in the inward- and outward-facing conformations. The main architectural difference between CTNS and the eukaryotic SWEET transporters, as well as the chicken KDEL receptor, lies in the positioning of the inversion helix TM4. CTNS features two PQ-motifs, positioned in TMs 1 and 5, identical to other members of this family. Novel structural characteristics include the presence of two salt bridges, which are positioned at the cytosolic and lysosomal luminal gates, respectively. Furthermore, CTNS features a highly charged ligand binding site to facilitate specific interactions with L-cystine. Only the KDEL receptor was previously known to contain a substantial number of charged residues within its core.

- What is the underlying mechanism for cystine transport by CTNS? How does CTNS couple cystine movement to the translocation of protons?

CTNS acts as an alternating access transporter, moving its helices around the central L-cystine binding site to facilitate ligand movement across the membrane. Similar to the SWEETs and SemiSWEETs, the PQ-motif prolines play a crucial role in mediating the necessary conformational changes that enable transport. However, unlike these sugar transporters, our data indicate that CTNS moves in a more asymmetric fashion, given that only the second PQ-motif proline is indispensable for transport. CTNS couples ligand movement to the translocation of protons *via* two novel salt bridge interactions that coordinate the opening and closing of the lysosomal luminal and cytosolic gates based on their protonation states.

- Do the reported missense mutations affect the transport function of CTNS? Are these effects sufficient to explain the progression of cystinosis?

When mapping the 42 missense mutations onto the structures obtained in this work, their impact on the transport function of CTNS is obvious. Many of the residues mutated in cystinosis are located near one of the three functionally important areas: the cytoplasmic and lysosomal luminal gates, as well as the binding site. Due to the compactness of the transporter, a significant mutation, even when located further away from the core of the protein, may cause serious malfunction of the transport function. Interestingly, some mutants, such as K280R, are not associated with the most severe phenotype of cystinosis despite completely inhibiting the transport function. This highlights potential additional functions of CTNS outside of L-cystine transport.

- How does CTNS interact with and regulate other lysosomal structures?

Using the data provided in this thesis, this question remains more difficult to answer. The data clearly indicate that both *At*CTNS and *Hs*CTNS(Δ2–115) form dimers *in vitro*, similar to other members of the PQ-loop family, some of which have been described to function as multimers. However, there is no evidence that the transporters exist as dimers *in vivo*, whether functional or not. In case the interactions between CTNS and other lysosomal structures, such as the Ragulator-RagA/C component of mTORC1 (Andrzejewska *et al.*, 2016; Guo *et al.*, 2022), prove true, it would be conceivable that CTNS executes this function as a dimer. A potential dual role of CTNS to function as both a cystine transporter and a receptor would not be out of the ordinary: SLC38A9 functions as a transceptor at the lysosomal membrane, transporting arginine and recruiting mTORC1 (Rebsamen *et al.*, 2015; Wang *et al.*, 2015; Lei *et al.*, 2018). Furthermore, a similar dual function has also been proposed for human PQLC2 (Talaia *et al.*, 2021). Further biochemical and cell biology studies are needed to tackle this question.

Appendix

Construct Design

Table A.1 – Gene Strings.
This table continues on the next page.

Sequence 5'-3'
HR-*At*CTNS-HR
ACCCCGGATTCTAGAACTAGTGGATCCCCCATGGCTTCCTGGAATTCTATCCCATTGGAAATCTCTTACGAAATCGTTGGTTGGATTGCTTTTGCCTCTTGGTCTATTTCTTTCTACCCA CAATTGATCTTGAACTTCAGAAGAAGATCCGTTGTCGGTTTGAACTTCGATTTCGTTATGCTGAACTTGACCAAGCACTCTTCTTACATGATCTACAACGTTTGCTTGTACTTCTCCCCA GTTATTCAAAAGCAGTACTTCGATACTTACGGCGACAAAGAAATGATTCCAGTTGCTGCTAATGATGTCGCCTTTTCTATTCATGCTGTTGTTATGACTGCTGTCACCTTGTTCCAAATC TTCATATACGAAAGAGGTCCCCAGAAGGTTTCTAGATTGGCTATTGGTATAGTTGTTGTTGTTTGGGGTTTCGCTGCTATCTGTTTTTTCATTGCTTTGCCAACTCATTCCTGGTTGTGG TTGATTTCCATTTTCAACTCCATCCAGGTTTTCATGACCTGCGTTAAGTATATTCCACAAGCCAAGATGAACTTCACCAGAAAATCTACTGTTGGTTGGTCCATTGGTAACATCTTGTTG GATTTTACTGGTGGTTTGGCCAACTACTTGCAAATGGTTATTCAATCCATCGACCAGAACTCTTGGAAGAATTTTTACGGTAACATGGGCAAGACCCTGTTGTCTTTGATTTCTATCTTC TTCGACATCCTGTTCATGTTCCAACATTATGTCTTGTACCCAGAGAAGAAGGTGTCTAAATCTCCAGAAACTGGTGAAGAATCTAACGAACCATTGATCGATTCCTCTCATGAACACGTT GGGGAAAATTTATATTTTCAAGGTCAATTT
HR-*Dr*CTNS-HR
ACCCCGGATTCTAGAACTAGTGGATCCCCCATGTTGAAGCAGACCTTCTTCTTGTTCTTGTTGACCTTGTACTTGTGTAAGCACACTGCTTTTGGTGATGCTGAAGTTGCTGTTTCTGCT CCAGCTACTGTTAACTTGCAAGAATTCTCCTCTGAAAACATCACCATCACTTTGTCTAGTCCATTGAACACCTCTGTTACCGTTTACTTCAACATTACCTACGAGTCCAACAACGTGTCC TTGATTATTCAATTGCCAAACGAAGTTGTTGTTCCAGCTGGTAATACTTCTGTTTCTTTTGCTGTTCAAGCCAAAGGTGTTGGTCAAGTTACTGCTTACTTATCCTCTAACGATACCAGG ATCAAATCCTTGGATGCCAGAATTAGATTCTTGATCGTCAGATCTAACGCCCTGTTCATCATCAATCAAATTATCGGTTGGATCTACTTCGTTGCCTGGTCTGTTTCATTTTATCCACAA GCTTACGAGAACTGGAAGAGAAGATCTGTTGTTGGTTTGAACTTCGATTTCTTGGCCTTGAATTTGACCGGTTTCATTGCTTACTCTGTTTTCAACGTTGGTTTGTTCTGGGTTACCTAC ATCCAAGAAGAGTTCTTGAAAAAGGATCCAAACGGTGTTATTCCAGTTGATGCTAACGATGTGTTCTTCTCATTGCATGCTTTGTTGTTGACTTTGGTCTACATTTGTCAATGCGCCATA TACGAAAGAGGTGGTCAAAAAGTTTCCAAGGTTGCTATTGGTTTATTGGCCATTGGTTGGACTTTTGCCTTTGTTTCTTTGTTCGTTGCTGTTGCCAAAAAGATCTCTTGGTTAGACTAC CTGTACTACTTCTCCTACATTAAGTTGGGTGTTACCCTGGTTAAGTACATTCCACAAGCACATATGAACTACAGACGTAAGTCTACTGAAGGTTGGTCTATTGGTAACGTCTTGTTGGAT TTTACTGGTGGCTCTTTCTCCTTGATCCAAATGTTTTTGGAAGCCTACAACAACGATAAGTGGCGTTTTATCTTTGGTGACCCAACTAAGTTTGGTTTGGGTGTTTTCTCCATCTTCTTC GACATCTTGTTCATTATCCAACACTACTGCTTGTACAGAAACAGAGAACCTATGTACCAAGACTTGGATGAACAAAACGATCAACACTCTTCTGTCAAGACCGGGGAAAATTTATATTTT CAAGGTCAATTT
HR-*Gg*CTNS-HR
ACCCCGGATTCTAGAACTAGTGGATCCCCCATGAGACATTTGTGTCCACCAGTTTTGTTGTTGACCTTGGTTTTGTTGGTTCACTGTGATGGTGTTATCGTTTTGTCTGTTCCAGAAGTT GTCTTGTTGGAAAACGGTTCTTCTACCAACGTTACCATTTCTTTGAGAGCCCCATTGAACGATACTTTGGTTATTACCTTGAACATCACCCACTCTTCTAAGCACTCTACTATAGTTGAA TTGCCAGACGAAATTCAATTGCCAGCTGGTCATACAAAAGCTGGTTTTCAAGTTAAGGCTGATGATGTTGGTCAAGTTACCGTTTACTTGTACACCTCTAACTCTAATTTGACCGGTCCA AGAATCCAATTCCAAGTTATCCATTCCATCATCGTTAAGTACGTTGATGAAGTTATCGGCTGGATCTATTTTTTGGCCTGGTCTATTTCATTCTACCCACAGTTGTTTGAAAACTGGCGT AGAAAATCTGTTGTCGGTTTGTCCTTTGATTTCATTGCTTTGAACTTGACCGGTTTCATTGCCTACTCTGTTTTCAATGTTGGTTTGTTCTGGATCCCCTTGATCAAAGAAGAATTCCTG TTGCATTACCCAAACGGTGTTAATCCAGTTGCTATCAACGATGTGTTCTTCTCATTGCATGCTGTTGCTTTGACTTTGTTGACCATTATTCAATGCTGCATCTACGAAAGAGCTGGTCAA AAAGTTTCTAAGACTGTTGTTGGTTTATTGGCCTTGGCTTGGATTTTCACTTTCACTACTTTGTTTTTGGCTGCCGCTGAAGAAATGACTTGGTTGCAATTTTTGTTCTGCTTCTCCTAC ATTAAGTTGGCCGTTACCTTGATTAAGTACTTCCCACAAGCTTACATGAACTTCAGAAGAAAGTCTACTGTTGGTTGGTCTATCGGTAACGTGTTGTTGGATTTTATCGGTGGTTCTTTC TCCCTGTTGCAAATGTTCTTGCAATCCTACAACAATGACGAGTGGAAGTTGATTTTTGGTGACCCAACTAAGTTTGGTTTGGGCGTTTTTTCCATTATCTTCGACATCGTTTTCATGGTC CAACACTATTGCTTGTATAGACGTAGAGGTTACGAACCATGTGAGGGGGAAAATTTATATTTTCAAGGTCAATTT
HR-*Hs*CTNS-HR
ACCCCGGATTCTAGAACTAGTGGATCCCCCATGATCAGAAACTGGTTGACCATCTTCATCTTGTTCCCATTGAAGTTGGTTGAGAAGTGCGAATCTTCTGTTTCTTTGACTGTTCCACCA GTTGTCAAGTTGGAAAATGGTTCTTCTACCAATGTCTCTTTGACCTTAAGACCACCATTGAACGCTACTTTGGTTATCACTTTCGAAATCACCTTCAGGTCCAAGAACATTACCATTTTG GAATTGCCAGATGAAGTTGTTGTTCCACCTGGTGTTACTAATTCCTCATTCCAAGTTACCTCTCAAAACGTTGGTCAGTTGACTGTTTACTTGCATGGTAACCATTCTAATCAAACCGGT CCAAGGATTAGGTTCTTGGTGATTAGATCTTCCGCCATCTCCATCATCAATCAAGTTATTGGTTGGATCTACTTCGTTGCCTGGTCTATTTCATTTTACCCACAGGTTATCATGAACTGG CGTAGAAAATCTGTTATCGGTTTGTCCTTTGATTTCGTCGCTTTGAATTTGACTGGTTTTGTTGCCTACTCCGTTTTCAACATTGGTTTGTTGTGGGTCCCTTACATCAAAGAACAATTC TTGTTGAAGTACCCAAACGGTGTTAACCCAGTTAATTCCAACGATGTGTTCTTCTCATTGCATGCTGTTGTTTTGACCTTGATCATCATCGTTCAATGCTGCTTGTACGAAAGAGGTGGA CAAAGAGTTTCTTGGCCAGCTATTGGTTTTTTGGTTTTGGCTTGGTTGTTCGCTTTCGTTACTATGATAGTTGCTGCTGTTGGTGTTACAACTTGGTTGCAATTTTTGTTCTGCTTCTCC TACATTAAGTTGGCTGTTACCTTGGTTAAGTACTTCCCACAAGCTTACATGAACTTCTACTACAAGTCTACTGAAGGTTGGTCCATTGGTAACGTTTTGTTGGATTTTACTGGTGGCTCT TTCTCCTTGTTGCAAATGTTCTTACAGTCCTACAACAACGATCAGTGGACTTTGATTTTTGGTGACCCAACTAAGTTTGGTTTGGGCGTTTTTTCTATCGTTTTCGACGTTGTTTTCTTC ATCCAGCACTTTTGCCTGTACAGAAAAAGACCAGGTTACGACCAATTGAACGGGGAAAATTTATATTTTCAAGGTCAATTT
HR-*Hs*CTNS(1-121)-BglII
GAAAAAACCCCGGATTCTAGAACTAGTGGATCCCCCATGATCAGAAACTGGTTGACCATCTTCATCTTGTTCCCATTGAAGTTGGTTGAGAAGTGCGAATCTTCTGTTTCTTTGACTGTT CCACCAGTTGTCAAGTTGGAAgctGGTTCTTCTACCgctGTCTCTTTGACCTTAAGACCACCATTGgccGCTACTTTGGTTATCACTTTCGAAATCACCTTCAGGTCCAAGgccATTACC ATTTTGGAATTGCCAGATGAAGTTGTTGTTCCACCTGGTGTTACTgctTCCTCATTCCAAGTTACCTCTCAAAACGTTGGTCAGTTGACTGTTTACTTGCATGGTgccCATTCTgctCAA ACCGGTCCAAGGATTAGGTTCTTGGTGATTAGATCTTCC

Table A.1 (Continued)

Sequence 5'-3'
HR-*Mm*CTNS-HR
ACCCCGGATTCTAGAACTAGTGGATCCCCCATGCGTAGAAACTGGTTGTTGATCTTGACCTTGTTCCTGTTGATGTTCATCGAAAAGTACGAGTCTACTGTTTCTTTGACTGCTCCACCA
ACTGTTAAGTTGGAAAATGGTTCTTCTACCAACGTTGATATCACTTTGGGTCATCCATTGAACTCCACTTTGGTTATTACTTTCGAAGTCACCTTCAGGTCTAAGAACTTGACTATAGTT
GAATTGCCAGACGAAGTTATCGTTCCAAGAGGTGAAAAGAATGCCTCATTCCAAGTTACCTCTCAAAACATTGGTCAAGTCACCGTTTTCTTGCATGGTAACCATTCTAATCAAACCTGT
CCAAGAATCAGGTTCTTGGTTATTCACTCCAGAATCGTGTCCATTATCAACCAGGTTATTGGTTGGATCTACTTCATGGCTTGGTCTGTTTCATTTTACCCACAAGTCATCCAGAATTGG
AGAAGAAAGTCTGTTATCGGTTTGTCCTTTGATTTCTTGGCTTTGAATTTGACCGGTTTCGTTGCTTACTCCGTTTTCAATATTGGTTTGTTGTGGGTCCCTTACATCCAAGAAGAATTC
CTATTGAAGTACCCAAACGGTGTTAACCCAGTTGATTCTAATGATGCCTTCTTCTCATTGCATGCTGTTGCTTTAACTCTGATCGTTATCTTGCAATGCTGCTTGTACGAAAGAGGTAAT
CAAAGAGTTTCCTGGCCATCTATTGGTTTCTTGGTTTTAGCTTGGTTGTTCGTTTTGGTCACTATGATAGTTGCTGCTGTTGGTATTACTACCTGGTTGCAATTTTTGTTCTGCTTCTCC
TACATCAAGCTGATCATCACATTGATCAAGTACTTCCCACAAGCCTACATGAACTTCTACTACAAATCTACCAAAGGTTGGTCCATTGGTGGTGTTTTGTTGGATTTTACTGGTGGCTCT
TTCTCCTTGTTGCAAATGTTCTTGCAATCCTACAACAACGATCAATGGACCTTGATTTTTGGTGACCCAACTAAGTTTGGTTTGGGTGTTTTCACCATTTTCTTCGACGTTGTGTTCTTC
ATCCAACACTTCTACCTGTACAGAAAAAAGCCAGGTTACGACCAATTGAACGGGGAAAATTTATATTTTCAAGGTCAATTT
SacI-*Hs*LAMP2A-XhoI
GAGCTCGCCACCATGTGTCTGTCTCCAGTGAAGGGCGCCAAGCTGATCCTGATCTTCCTGTTTCTGGGCGCCGTGCAGAGCAACGCCCTGATCGTGAATCTGACCGACAGCAAGGGCACC
TGTCTGTACGCCGAGTGGGAGATGAACTTCACCATCACCTACGAAACCACCAACCAGACCAACAAGACCATCACAATCGCCGTGCCTGACAAGGCCACACACGATGGAAGCAGCTGCGGC
GACGATAGAAACAGCGCCAAGATCATGATTCAGTTCGGCTTCGCCGTGTCCTGGGCCGTGAATTTCACAAAAGAGGCCAGCCACTACAGCATCCACGACATCGTGCTGAGCTACAACACC
AGCGACAGCACCGTGTTTCCAGGCGCTGTGGCTAAAGGCGTGCACACCGTGAAGAACCCCGAGAACTTCAAGGTGCCCCTGGACGTGATCTTCAAGTGCAACAGCGTGCTGACGTACAAT
CTGACCCCTGTGGTGCAGAAGTACTGGGGCATCCATCTGCAGGCCTTCGTGCAGAATGGCACCGTGTCCAAGAACGAGCAAGTGTGCGAAGAGGATCAGACCCCTACCACAGTGGCCCCT
ATCATCCACACAACAGCCCCTAGCACCACCACCACACTGACCCCTACAAGCACCCCTACACCTACACCAACTCCAACACCTACCGTGGGCAACTACTCCATCCGGAACGGCAACACCACA
TGCCTGCTGGCCACAATGGGCCTGCAGCTGAACATCACCGAGGAAAAAGTGCCCTTCATCTTCAACATCAACCCCGCCACCACCAACTTCACCGGCTCTTGTCAGCCTCAGAGCGCCCAG
CTGAGACTGAACAACAGCCAGATCAAGTACCTGGATTTCATCTTCGCTGTGAAGAACGAGAAGCGGTTCTACCTGAAAGAAGTGAACGTCTACATGTACCTGGCCAACGGCAGCGCCTTC
AATATCAGCAACAAGAACCTGAGCTTCTGGGACGCCCCTCTGGGCAGCAGCTACATGTGCAACAAAGAACAGGTGCTGAGCGTGTCCCGGGCCTTCCAGATCAACACCTTCAACCTGAAG
GTGCAGCCCTTCAACGTGACCAAGGGCCAGTACAGCACAGCCCAGGATTGCAGCGCCGACGAGGACAATTTCCTGGTGCCTATTGCCGTGGGAGCCGCTCTTGGCGGAGTGCTGATTCTG
GTGCTGCTGGCCTACTTCATCGGCCTGAAGAGACACCACACCGGCTACGAGCAGTTCCTCGAG

Table A.2 – Primers Used for Sequencing & Cloning.
Newstead lab oligo collection identifiers are annotated for each primer. This table continues on the next two pages.

Identifier	Name	Sequence 5'-3'	Sites	Use
206	pDDGFP-Leu2d_SF2	CTTCTTATTCAAATGTAA	-	Sequencing
207	pDDGFP-Leu2d_SR2	AGAAAATTTGTGACCATTAAC	-	Sequencing
660	pLEXm_fw	GCTGGTTATTGTGCTGTCTCATC	-	Sequencing
661	pLEXm_rv	CACCAGCCACCACCTTCTGATAG	-	Sequencing
1632	yeGFP_rv	ATCACCATCTAATTCAACCAA	-	Sequencing
539	HR_fw	ACCCCGGATTCTAGAACTAGTGGATC	HR	Cloning (pLeu2d)
540	HR_rv	AAATTGACCTTGAAAATATAAATTTTC	HR	Cloning (pLeu2d)
1734	lac/M13_fw	AGCGGATAACAATTTCACACAGG	-	Cloning (pLeu2d)
2443	Avi-HR_rv	AAATTGACCTTGAAAATATAAATTTTCGTTAACCTCATGCCATTC GATTTTC	HR	Cloning (pLeu2d)
2505	*At*CTNS-Avi_rv	CTCATGCCATTCGATTTTCTGTGCCTCGAAGATATCGTTCAGCCC AACGTGTTCATGAGAGG	Avi-tag	Cloning (pLeu2d)
2655	Avi-*At*CTNS_fw	ATGGGGCTGAACGATATCTTCGAGGCACAGAAAATCGAATGGCAT GAGGCTTCCTGGAATTCTATCCC	Avi-tag	Cloning (pLeu2d)
2883	*At*CTNS(W24A)_fw	GTTGGTTGGATTGCTTTTGCCTCTGCTTCTATTTCTTTCTACCCA CAATTG	-	Mutagenesis
2884	*At*CTNS(W24A)_rv	CAATTGTGGGTAGAAAGAAATAGAAGCAGAGGCAAAAGCAATCCA ACCAAC	-	Mutagenesis
2885	*At*CTNS(W24F)_fw	GGTTGGATTGCTTTTGCCTCTTTCTCTATTTCTTTCTACCCACAA TTG	-	Mutagenesis
2886	*At*CTNS(W24F)_rv	CAATTGTGGGTAGAAAGAAATAGAGAAAGAGGCAAAAGCAATCCA AC	-	Mutagenesis
3070	*At*CTNS(S27A)_fw	CCTCTTGGTCTATTGCTTTCTACCCACAATTG	-	Mutagenesis
3071	*At*CTNS(S27A)_rv	CAATTGTGGGTAGAAAGCAATAGACCAAGAGGC	-	Mutagenesis
3072	*At*CTNS(S27C)_fw	CCTCTTGGTCTATTTGTTTCTACCCACAATTG	-	Mutagenesis
3073	*At*CTNS(S27C)_rv	CAATTGTGGGTAGAAACAAATAGACCAAGAGGC	-	Mutagenesis
3116	*At*CTNS(P30A)_fw	GTCTATTTCTTTCTACGCCCAATTGATCTTGAACTTC	-	Mutagenesis
3117	*At*CTNS(P30A)_rv	GAAGTTCAAGATCAATTGGGCGTAGAAAGAAATAGACC	-	Mutagenesis
3118	*At*CTNS(P30G)_fw	GGTCTATTTCTTTCTACGGTCAATTGATCTTGAACTTCAG	-	Mutagenesis
3119	*At*CTNS(P30G)_rv	CTGAAGTTCAAGATCAATTGACCGTAGAAAGAAATAGACC	-	Mutagenesis
2887	*At*CTNS(D47A)_fw	GTTGTCGGTTTGAACTTCGCTTTCGTTATGCTGAACTTG	-	Mutagenesis
2888	*At*CTNS(D47A)_rv	GTTCAGCATAACGAAAGCGAAGTTCAAACCGAC	-	Mutagenesis
2889	*At*CTNS(D47N)_fw	CCGTTGTCGGTTTGAACTTCAACTTCGTTATGCTGAACTTGAC	-	Mutagenesis
2890	*At*CTNS(D47N)_rv	CAAGTTCAGCATAACGAAGTTGAAGTTCAAACCGACAACG	-	Mutagenesis

Table A.2 (Continued)

Identifier	Name	Sequence 5'-3'	Sites	Use
3074	*At*CTNS(N52A)_fw	CGATTTCGTTATGCTGGCTTTGACCAAGCACTCTTC	-	Mutagenesis
3075	*At*CTNS(N52A)_rv	GAGTGCTTGGTCAAAGCCAGCATAACGAAATCGAAG	-	Mutagenesis
3076	*At*CTNS(N52Q)_fw	CGATTTCGTTATGCTGCAATTGACCAAGCACTC	-	Mutagenesis
3077	*At*CTNS(N52Q)_rv	GAGTGCTTGGTCAATTGCAGCATAACGAAATCG	-	Mutagenesis
2963	*At*CTNS(K55A)_fw	GCTGAACTTGACCGCTCACTCTTCTTACATG	-	Mutagenesis
2964	*At*CTNS(K55A)_rv	CATGTAAGAAGAGTGAGCGGTCAAGTTCAGC	-	Mutagenesis
2965	*At*CTNS(K55Q)_fw	GCTGAACTTGACCCAACACTCTTCTTACATG	-	Mutagenesis
2966	*At*CTNS(K55Q)_rv	GTAAGAAGAGTGTTGGGTCAAGTTCAGC	-	Mutagenesis
2967	*At*CTNS(K55R)_fw	GCTGAACTTGACCAGACACTCTTCTTAC	-	Mutagenesis
2968	*At*CTNS(K55R)_rv	CATGTAAGAAGAGTGTCTGGTCAAGTTCAG	-	Mutagenesis
2891	*At*CTNS(H56A)_fw	CGTTATGCTGAACTTGACCAAGGCTTCTTCTTACATGATCTACAACGTTTG	-	Mutagenesis
2892	*At*CTNS(H56A)_rv	GCAAACGTTGTAGATCATGTAAGAAGAAGCCTTGGTCAAGTTCAGCATAACG	-	Mutagenesis
2893	*At*CTNS(H56F)_fw	GTTATGCTGAACTTGACCAAGTTCTCTTCTTACATGATCTACAACGTTTG	-	Mutagenesis
2894	*At*CTNS(H56F)_rv	CAAACGTTGTAGATCATGTAAGAAGAGAACTTGGTCAAGTTCAGCATAAC	-	Mutagenesis
2895	*At*CTNS(P87A)_fw	CGGCGACAAAGAAATGATTGCTGTTGCTGCTAATGATG	-	Mutagenesis
2896	*At*CTNS(P87A)_rv	GACATCATTAGCAGCAACAGCAATCATTTCTTTGTCGCCG	-	Mutagenesis
2897	*At*CTNS(P87G)_fw	CGGCGACAAAGAAATGATTGGTGTTGCTGCTAATGATGTCGC	-	Mutagenesis
2898	*At*CTNS(P87G)_rv	GCGACATCATTAGCAGCAACACCAATCATTTCTTTGTCGCCG	-	Mutagenesis
2899	*At*CTNS(N91A)_fw	CCAGTTGCTGCTGCTGATGTCGCCTTTTCTATTCATGC	-	Mutagenesis
2900	*At*CTNS(N91A)_rv	GCATGAATAGAAAAGGCGACATCAGCAGCAGCAACTGG	-	Mutagenesis
2901	*At*CTNS(N91Q)_fw	GATTCCAGTTGCTGCTCAAGATGTCGCCTTTTCTATTCATG	-	Mutagenesis
2902	*At*CTNS(N91Q)_rv	CATGAATAGAAAAGGCGACATCTTGAGCAGCAACTGGAATC	-	Mutagenesis
2903	*At*CTNS(D92A)_fw	CCAGTTGCTGCTAATGCTGTCGCCTTTTCTATTC	-	Mutagenesis
2904	*At*CTNS(D92A)_rv	GAATAGAAAAGGCGACAGCATTAGCAGCAACTG	-	Mutagenesis
2905	*At*CTNS(D92N)_fw	GATTCCAGTTGCTGCTAATAACGTCGCCTTTTCTATTCATGCTG	-	Mutagenesis
2906	*At*CTNS(D92N)_rv	CATGAATAGAAAAGGCGACGTTATTAGCAGCAACTGGAATC	-	Mutagenesis
2907	*At*CTNS(K166A)_fw	CCAGGTTTTCATGACCTGCGTTGCTTATATTCCACAAGCCAAG	-	Mutagenesis
2908	*At*CTNS(K166A)_rv	CTTGGCTTGTGGAATATAAGCAACGCAGGTCATGAAAACCTGG	-	Mutagenesis
2909	*At*CTNS(K166Q)_fw	CCAGGTTTTCATGACCTGCGTTCAATATATTCCACAAGCCAAG	-	Mutagenesis
2910	*At*CTNS(K166Q)_rv	CTTGGCTTGTGGAATATATTGAACGCAGGTCATGAAAACC	-	Mutagenesis
2911	*At*CTNS(K166R)_fw	CAGGTTTTCATGACCTGCGTTAGATATATTCCACAAGCCAAGATG	-	Mutagenesis
2912	*At*CTNS(K166R)_rv	CATCTTGGCTTGTGGAATATATCTAACGCAGGTCATGAAAACCTG	-	Mutagenesis
3078	*At*CTNS(Y167A)_fw	CATGACCTGCGTTAAGGCTATTCCACAAGCCAAG	-	Mutagenesis
3079	*At*CTNS(Y167A)_rv	CTTGGCTTGTGGAATAGCCTTAACGCAGGTCATG	-	Mutagenesis
2913	*At*CTNS(Y167F)_fw	GTTTTCATGACCTGCGTTAAGTTCATTCCACAAGCCAAGATGAAC	-	Mutagenesis
2914	*At*CTNS(Y167F)_rv	CATCTTGGCTTGTGGAATGAACTTAACGCAGGTCATGAAAAC	-	Mutagenesis
3168	*At*CTNS(P169A)_fw	GACCTGCGTTAAGTATATTGCCCAAGCCAAGATGAAC	-	Mutagenesis
3169	*At*CTNS(P169A)_rv	GTTCATCTTGGCTTGGGCAATATACTTAACGCAGG	-	Mutagenesis
3170	*At*CTNS(P169G)_fw	GACCTGCGTTAAGTATATTGGTCAAGCCAAGATGAACTTCAC	-	Mutagenesis
3171	*At*CTNS(P169G)_rv	GTGAAGTTCATCTTGGCTTGACCAATATACTTAACGCAGGTC	-	Mutagenesis
2915	*At*CTNS(D191A)_fw	CATTGGTAACATCTTGTTGGCTTTTACTGGTGGTTTGGC	-	Mutagenesis
2916	*At*CTNS(D191A)_rv	CAAACCACCAGTAAAAGCCAACAAGATGTTACCAATG	-	Mutagenesis
3166	*At*CTNS(D191G)_fw	GGTAACATCTTGTTGGGTTTTACTGGTGGTTTG	-	Mutagenesis
3167	*At*CTNS(D191G)_rv	CAAACCACCAGTAAAACCCAACAAGATGTTACCAATG	-	Mutagenesis
2917	*At*CTNS(D191N)_fw	GTCCATTGGTAACATCTTGTTGAACTTTACTGGTGGTTTGGCCAAC	-	Mutagenesis
2918	*At*CTNS(D191N)_rv	GGCCAAACCACCAGTAAAGTTCAACAAGATGTTACCAATG	-	Mutagenesis
3164	*At*CTNS(D191Y)_fw	CCATTGGTAACATCTTGTTGTATTTTACTGGTGGTTTG	-	Mutagenesis
3165	*At*CTNS(D191Y)_rv	CAAACCACCAGTAAAATACAACAAGATGTTACCAATG	-	Mutagenesis
2919	*At*CTNS(Q201A)_fw	GGTGGTTTGGCCAACTACTTGGCTATGGTTATTCAATCCATCGAC	-	Mutagenesis
2920	*At*CTNS(Q201A)_rv	GGTCGATGGATTGAATAACCATAGCCAAGTAGTTGGCCAAACCAC	-	Mutagenesis
2921	*At*CTNS(Q201N)_fw	GTGGTTTGGCCAACTACTTGAACATGGTTATTCAATCCATCG	-	Mutagenesis
2922	*At*CTNS(Q201N)_rv	CGATGGATTGAATAACCATGTTCAAGTAGTTGGCCAAACCAC	-	Mutagenesis

Table A.2 (Continued)

Identifier	Name	Sequence 5'-3'	Sites	Use
2923	*At*CTNS(K221A)_fw	GAAGAATTTTTACGGTAACATGGGCGCTACCCTGTTGTCTTTGATTTCTATC	-	Mutagenesis
2924	*At*CTNS(K221A)_rv	GAAATCAAAGACAACAGGGTAGCGCCCATGTTACCGTAAAAATTCTTCC	-	Mutagenesis
2925	*At*CTNS(K221Q)_fw	GGAAGAATTTTTACGGTAACATGGGCCAAACCCTGTTGTCTTTGATTTC	-	Mutagenesis
2926	*At*CTNS(K221Q)_rv	CAAAGACAACAGGGTTTGGCCCATGTTACCGTAAAAATTC	-	Mutagenesis
2927	*At*CTNS(K221R)_fw	GAATTTTTACGGTAACATGGGCAGAACCCTGTTGTCTTTGATTTCTATC	-	Mutagenesis
2928	*At*CTNS(K221R)_rv	GAAATCAAAGACAACAGGGTTCTGCCCATGTTACCGTAAAAATTC	-	Mutagenesis
3180	*At*CTNS(S228A)_fw	CCTGTTGTCTTTGATTGCTATCTTCTTCGACATC	-	Mutagenesis
3181	*At*CTNS(S228A)_rv	GGATGTCGAAGAAGATAGCAATCAAAGACAACAGG	-	Mutagenesis
2929	*At*CTNS(D232A)_fw	GTTGTCTTTGATTTCTATCTTCTTCGCTATCCTGTTCATGTTCCAAC	-	Mutagenesis
2930	*At*CTNS(D232A)_rv	GACATAATGTTGGAACATGAACAGGATAGCGAAGAAGATAGAAATCAAAGACAAC	-	Mutagenesis
2931	*At*CTNS(D232N)_fw	GTCTTTGATTTCTATCTTCTTCAACATCCTGTTCATGTTCCAAC	-	Mutagenesis
2932	*At*CTNS(D232N)_rv	GGAACATGAACAGGATGTTGAAGAAGATAGAAATCAAAGAC	-	Mutagenesis
3069	*At*CTNS(Δ246–270)_rv	AAATTGACCTTGAAAATATAAATTTTCCTCTGGGTACAAGACATAATGTTGGAACATG	HR	Cloning (pLeu2d)
2593	*Dr*CTNS(Δ2–120)_fw	ACCCCGGATTCTAGAACTAGTGGATCCCCCATGTTGATCGTCAGATCTAACG	HR	Cloning (pLeu2d)
2508	*Gg*CTNS-Avi_rv	CTCATGCCATTCGATTTTCTGTGCCTCGAAGATATCGTTCAGCCCCTCACATGGTTCGTAACC	Avi-tag	Cloning (pLeu2d)
2590	*Gg*CTNS(Δ2–114)_fw	ACCCCGGATTCTAGAACTAGTGGATCCCCCATGCAAGTTATCCATTCCATCATCG	HR	Cloning (pLeu2d)
3825	*Hs*CTNS(N301A)_fw	GGTTGGTCCATTGGTGCCGTTTTGTTGG	-	Mutagenesis
3826	*Hs*CTNS(N301A)_rv	CCAACAAAACGGCACCAATGGACCAACC	-	Mutagenesis
N/A	*Hs*CTNS(D346N)_fw	CGTTTTTTCTATCGTTTTCAACGTTGTTTTCTTCATC	-	Mutagenesis
N/A	*Hs*CTNS(D346N)_rv	GATGAAGAAAACAACGTTGAAAACGATAGAAAAAACG	-	Mutagenesis
3615	*Hs*CTNS(Δ2–23)_fw	ACCCCGGATTCTAGAACTAGTGGATCCCCCATGGTTTCTTTGACTGTTCCACCAG	HR	Cloning (pLeu2d)
2591	*Hs*CTNS(Δ2–115)_fw	ACCCCGGATTCTAGAACTAGTGGATCCCCCATGTTGGTGATTAGATCTTCC	HR	Cloning (pLeu2d)
3044	*Hs*CTNS(Δ2–118)_fw	ACCCCGGATTCTAGAACTAGTGGATCCCCCATGAGATCTTCCGCCATCTCCATCATC	HR	Cloning (pLeu2d)
3045	*Hs*CTNS(Δ360–367)_rv	AAATTGACCTTGAAAATATAAATTTTCTTTTCTGTACAGGCAAAAGTGC	HR	Cloning (pLeu2d)
3333	*Hs*CTNS(NTD)_rv	GGAAGATCTAATCACCAAGAACCTAATCC	BglII	Cloning (pLeu2d)
2969	*Hs*CTNS-Avi_rv	CTCATGCCATTCGATTTTCTGTGCCTCGAAGATATCGTTCAGCCCGTTCAATTGGTCGTAACC	Avi-tag	Cloning (pLeu2d)
2970	Avi-*Hs*CTNS(Δ2–115)_fw	ATGGGGCTGAACGATATCTTCGAGGCACAGAAAATCGAATGGCATGAGATCAGAAACTGGTTGACC	Avi-tag	Cloning (pLeu2d)
3026	*Hs*CTNS_iso1_to_2_rv1	GCCCAATCTTGTCTCAATCTAGAACCAGAACCAGTTCTAGCAGCTTGCAAACCTGGTCTTTTTCTGTACAGG	-	Cloning (pLeu2d)
3027	*Hs*CTNS_iso1_to_2_rv2	GCAGAAACAGAAGTAGTTTGTGGCAAAGCTTTTGGTTGCAAAGATGGAGCCCAATCTTGTCTCAATCTAG	-	Cloning (pLeu2d)
3028	*Hs*CTNS_iso1_to_2_rv3	AAATTGACCTTGAAAATATAAATTTTCACCTTTCAAAGAAGAAGCAGAAACAGAAGTAGTTTGTG	HR	Cloning (pLeu2d)
2971	*Hs*CTNS_pLEXm-GFP_fw	CGGCGGGAGCTCGCCACCATGATCAGAAACTGGTTGAC	SacI	Cloning (pLEXm-GFP)
2972	*Hs*CTNS_pLEXm-GFP_rv	CCGCCGCTCGAGGTTCAATTGGTCGTAACCTGG	XhoI	Cloning (pLEXm-GFP)
3744	*Hs*CTNS(Δ2–115)_pLEXm_fw	CTGAGAGCTCGCCACCATGTTGGTGATTAGATCTTCCGCCATC	SacI	Cloning (pLEXm)
3745	*Hs*CTNS(Δ2–23)_pLEXm_fw	CTGAGAGCTCGCCACCATGGTTTCTTTGACTGTTCCACCAG	SacI	Cloning (pLEXm)
3746	*Hs*CTNS(Δ363–367)_pLEXm_rv	CTGACTCGAGACCTGGTCTTTTTCTGTACAGG	XhoI	Cloning (pLEXm)
3747	*Hs*CTNS2_pLEXm_rv	CTGACTCGAGACCTTTCAAAGAAGAAGCAGAAAC	XhoI	Cloning (pLEXm)
3970	*Hs*CTNS-StrepII_rv	CTGACTCGAGCTATTTCTCAAACTGTGGGTGCGACCAAGCGGAGTTCAATTGGTCGTAACCTGG	XhoI	Cloning (pLEXm)
3971	*Hs*CTNS2-StrepII_rv	CTGACTCGAGCTATTTCTCAAACTGTGGGTGCGACCAAGCGGAACCTTTCAAAGAAGAAGCAGAAAC	XhoI	Cloning (pLEXm)
2592	*Mm*CTNS(Δ2–115)_fw	ACCCCGGATTCTAGAACTAGTGGATCCCCCATGTTGGTTATTCACTCCAGAATCG	HR	Cloning (pLeu2d)
2531	Phagemid_Seq_fw	GCTTCCGGCTCGTATGTTG	-	Sequencing (Nb)
2631	Nb-BspQI_fw	ATATGCTCTTCTAGTCAGGTGCAGCTGGTCGAGTC	BspQI	Cloning (pBX vectors)
2632	Nb-BspQI_rv	TATAGCTCTTCATGCTGAGGAGACGGTGACCTG	BspQI	Cloning (pBX vectors)

Table A.3 – Construct Sequences.
Amino acid sequences of constructs used in this work are listed. For sybody sequences, X denotes a randomized position present within CDRs 1, 2, and 3. For the exact composition of these regions, see **Tables A.5** and **A.11** for α-*At*CTNS and α-*Hs*CTNS(Δ2–115) sybodies, respectively. This table continues on the next page.

Construct	Source	Length	Sequence
*Hs*CTNS isoform 1	Uniprot O60931-1	367	MIRNWLTIFILFPLKLVEKCESSVSLTVPPVVKLENGSSTNVSLTLRPPL NATLVITFEITFRSKNITILELPDEVVVPPGVTNSSFQVTSQNVGQLTVY LHGNHSNQTGPRIRFLVIRSSAISIINQVIGWIYFVAWSISFYPQVIMNW RRKSVIGLSFDFVALNLTGFVAYSVFNIGLLWVPYIKEQFLLKYPNGVNP VNSNDVFFSLHAVVLTLIIIVQCCLYERGGQRVSWPAIGFLVLAWLFAFV TMIVAAVGVTTWLQFLFCFSYIKLAVTLVKYFPQAYMNFYYKSTEGWSIG NVLLDFTGGSFSLLQMFLQSYNNDQWTLIFGDPTKFGLGVFSIVFDVVFF IQHFCLYRKRPGYDQLN
*Hs*CTNS isoform 2 (*Hs*CTNS2)	Uniprot O60931-2	400	MIRNWLTIFILFPLKLVEKCESSVSLTVPPVVKLENGSSTNVSLTLRPPL NATLVITFEITFRSKNITILELPDEVVVPPGVTNSSFQVTSQNVGQLTVY LHGNHSNQTGPRIRFLVIRSSAISIINQVIGWIYFVAWSISFYPQVIMNW RRKSVIGLSFDFVALNLTGFVAYSVFNIGLLWVPYIKEQFLLKYPNGVNP VNSNDVFFSLHAVVLTLIIIVQCCLYERGGQRVSWPAIGFLVLAWLFAFV TMIVAAVGVTTWLQFLFCFSYIKLAVTLVKYFPQAYMNFYYKSTEGWSIG NVLLDFTGGSFSLLQMFLQSYNNDQWTLIFGDPTKFGLGVFSIVFDVVFF IQHFCLYRKRPGLQAARTGSGSRLRQDWAPSLQPKALPQTTSVSASSLKG
*At*CTNS	Uniprot P57758	270	MASWNSIPLEISYEIVGWIAFASWSISFYPQLILNFRRRSVVGLNFDFVM LNLTKHSSYMIYNVCLYFSPVIQKQYFDTYGDKEMIPVAANDVAFSIHAV VMTAVTLFQIFIYERGPQKVSRLAIGIVVVVWGFAAICFFIALPTHSWLW LISIFNSIQVFMTCVKYIPQAKMNFTRKSTVGWSIGNILLDFTGGLANYL QMVIQSIDQNSWKNFYGNMGKTLLSLISIFFDILFMFQHYVLYPEKKVSK SPETGEESNEPLIDSSHEHV
*Mm*CTNS	Uniprot P57757	367	MRRNWLLILTLFLLMFIEKYESTVSLTAPPTVKLENGSSTNVDITLGHPL NSTLVITFEVTFRSKNLTIVELPDEVIVPRGEKNASFQVTSQNIGQVTVF LHGNHSNQTCPRIRFLVIHSRIVSIINQVIGWIYFMAWSVSFYPQVIQNW RRKSVIGLSFDFLALNLTGFVAYSVFNIGLLWVPYIQEEFLLKYPNGVNP VDSNDAFFSLHAVALTLIVILQCCLYERGNQRVSWPSIGFLVLAWLFVLV TMIVAAVGITTWLQFLFCFSYIKLIITLIKYFPQAYMNFYYKSTKGWSIG GVLLDFTGGSFSLLQMFLQSYNNDQWTLIFGDPTKFGLGVFTIFFDVVFF IQHFYLYRKKPGYDQLN
*Gg*CTNS	Uniprot E1BXK0	377	MRHLCPPVLLLTLVLLVHCGQCSAVLAWAQRDGVIVLSVPEVVLLENGSS TNVTISLRAPLNDTLVITLNITHSSKHSTIVELPDEIQLPAGHTKAGFQV KADDVGQVTVYLYTSNSNLTGPRIQFQVIHSIIVKYVDEVIGWIYFLAWS ISFYPQLFENWRRKSVVGLSFDFIALNLTGFIAYSVFNVGLFWIPLIKEE FLLHYPNGVNPVAINDVFFSLHAVALTLLTIIQCCIYERAGQKVSKTVVG LLALAWIFTFTTLFLAAAEEMTWLQFLFCFSYIKLAVTLIKYFPQAYMNF RRKSTVGWSIGNVLLDFIGGSFSLLQMFLQSYNNDEWKLIFGDPTKFGLG VFSIIFDIVFMVQHYCLYRRRGYEPCE
*Dr*CTNS	Uniprot F1QM07	384	MLKQTFFLFLLTLYLCKHTAFGDAEVAVSAPATVNLQEFSSENITITLSS PLNTSVTVYFNITYESNNVSLIIQLPNEVVVPAGNTSVSFAVQAKGVGQV TAYLSSNDTRIKSLDARIRFLIVRSNALFIINQIIGWIYFVAWSVSFYPQ AYENWKRRSVVGLNFDFLALNLTGFIAYSVFNVGLFWVTYIQEEFLKKDP NGVIPVDANDVFFSLHALLLTLVYICQCAIYERGGQKVSKVAIGLLAIGW TFAFVSLFVAVAKKISWLDYLYYFSYIKLGVTLVKYIPQAHMNYRRKSTE GWSIGNVLLDFTGGSFSLIQMFLEAYNNDKWRFIFGDPTKFGLGVFSIFF DILFIIQHYCLYRNREPMYQDLDEQNDQHSSVKT

Table A.3 (Continued)

Construct	Source	Length	Sequence
Concave (G) Syb	Zimmermann *et al*. (2020)	114	QVQLVESGGGLVQAGGSLRLSCAASGFPVXXXXMXWYRQAPGKEREWVAA IXSXGXXTXYADSVKGRFTISRDNAKNTVYLQMNSLKPEDTAVYYCXVXV GXXYXGQGTQVTVS
Loop (B) Syb	Zimmermann *et al*. (2020)	120	QVQLVESGGGLVQAGGSLRLSCAASGFPVXXXXMXWYRQAPGKEREWVAA IXSXGXXTXYADSVKGRFTISRDNAKNTVYLQMNSLKPEDTAVYYCNVKD XGXXXXXYDYWGQGTQVTVS
Convex (R) Syb	Zimmermann *et al*. (2020)	124	QVQLVESGGGSVQAGGSLRLSCAASGXIXXIXYLGWFRQAPGKEREGVAA LXTXXGXTYYADSVKGRFTVSLDNAKNTVYLQMNSLKPEDTALYYCAAAX XGXXXPLXXXXYXYWGQGTQVTVS
α-*At*CTNS Nb1	Löbel *et al*. (2022)	121	QVQLVESGGGSAQPGGSLRLSCAVSGSVSELNTMGWFRQAPGKQRELVAR ITATSDATNYADSVKGRFTISRDNGWNTVYLQMNSLKPEDSAVYYCNVEG APSWFSGIRSYWGQGTQVTIS
α-*At*CTNS Nb2	This thesis	124	QVQLVESGGGFVQDGGSLRLSCTASGRTFRRYAMAWFRQAPGKEREFVAA LSSSGLITRYTDSVQGRFTVSRDNTKNTVYLQMNSLKPEDSAVYYCAASD NYFVDPKGTTAYDYWGQGTQVTVS
α-*At*CTNS Nb3	This thesis	126	QVQLVESGGGLVQAGGSLRLSCAASGRTITPISTYVMGWFRQDPGKEREF VASISWNGANTYYADSVKGRFTISRDNAKNTVYLQMNSLKPEDTAVYYCA ADPESHVRLRLGVGAYWGRGTQLTVS
α-*At*CTNS Nb4	Löbel *et al*. (2022)	126	QVQLVEAGGGLVQAGGSLRLSCAASGRTITPISTYVMGWFRQDPGKEREF VASISWNGANTYYADSVKGRFTISRDNAKNTVYLQMNSLKPEDTAVYYCA ADPESHVRLRLGVGAYWGRGTQVTVS
α-*At*CTNS Nb5	This thesis	123	QVQLVESGGGSVQAGGSLRLSCTVTGRTLTNYNMGWFRQAPGKEREFVAA TSWIGGTPYADSVKGRFTISRDNAKNTVLLQMDSLRPEDTAVYYCAAGAT YRTNLRTTGTYPYWGQGTQVTVS
α-*At*CTNS Nb6	This thesis	123	QVQLVESGGGSVQAGGSLRLSCAVSGRTLSNVNMGWFRQAPGKEREFVAA TSWMGGTPYADSVKGRFTISRDNAKNTVLLQMNSLKPEDTAVYYCAAGAT YLTNLRTEGTYPYWGQGTQVTVS
α-*At*CTNS Nb7	This thesis	123	QVQLVESGGGSVQTGGSLRLSCAVSGRTLTNYNMGWFRQAPGKEREFVAA TSALGGTPYADSVKGRFTISRDNAKNTVLLQMNSLKPEDTAVYYCAGGAP WRTNLRTIRTWPYWGQGTQVTVS
α-*At*CTNS Nb8	This thesis	123	QVQLVESGGGSVRTGGSLRLSCAASGRTLNYYTLGWFRQAPGKEREFVAA TSWMGGTPYADSVKGRFTISRDNAKNTVLLQMNSLKPEDTAVYYCAAGAP WRTNLRAEGTWPTWGQGTQVTVS
*Hs*LAMP2A	Uniprot P13473-1	410	MVCFRLFPVPGSGLVLVCLVLGAVRSYALELNLTDSENATCLYAKWQMNF TVRYETTNKTYKTVTISDHGTVTYNGSICGDDQNGPKIAVQFGPGFSWIA NFTKAASTYSIDSVSFSYNTGDNTTFPDAEDKGILTVDELLAIRIPLNDL FRCNSLSTLEKNDVVQHYWDVLVQAFVQNGTVSTNEFLCDKDKTSTVAPT IHTTVPSPTTTPTPKEKPEAGTYSVNNGNDTCLLATMGLQLNITQDKVAS VININPNTTHSTGSCRSHTALLRLNSSTIKYLDFVFAVKNENRFYLKEVN ISMYLVNGSVFSIANNNLSYWDAPLGSSYMCNKEQTVSVSGAFQINTFDL RVQPFNVTQGKYSTAQDCSADDDNFLVPIAVGAALAGVLILVLLAYFIGL KHHHAGYEQF
*Mm*LAMP2A	Uniprot P17047-1	415	MCLSPVKGAKLILIFLFLGAVQSNALIVNLTDSKGTCLYAEWEMNFTITY ETTNQTNKTITIAVPDKATHDGSSCGDDRNSAKIMIQFGFAVSWAVNFTK EASHYSIHDIVLSYNTSDSTVFPGAVAKGVHTVKNPENFKVPLDVIFKCN SVLTYNLTPVVQKYWGIHLQAFVQNGTVSKNEQVCEEDQTPTTVAPIIHT TAPSTTTTLTPTSTPTPTPTPTPTVGNYSIRNGNTTCLLATMGLQLNITE EKVPFIFNINPATTNFTGSCQPQSAQLRLNNSQIKYLDFIFAVKNEKRFY LKEVNVYMYLANGSAFNISNKNLSFWDAPLGSSYMCNKEQVLSVSRAFQI NTFNLKVQPFNVTKGQYSTAQDCSADEDNFLVPIAVGAALGGVLILVLLA YFIGLKRHHTGYEQF

Table A.4 – Additional Constructs for Bacterial Cell Culture.
All constructs contain a pelB leader sequence, causing the binder to be secreted into the periplasm during expression. Newstead lab plasmid collection identifiers are annotated for each construct. This table continues on the next page.

a, α-*At*CTNS sybodies cloned in vector pSb_init (CAM resistance, L-(+)-arabinose induction).

Identifier	α-*At*CTNS
2943	Syb2-Myc-His
2944	Syb5-Myc-His aka G3a
2945	Syb6-Myc-His
2946	Syb8-Myc-His
2947	Syb9-Myc-His aka G2a
2948	Syb11-Myc-His
2949	Syb13-Myc-His
2950	Syb14-Myc-His
2951	Syb15-Myc-His aka G1a
2952	Syb16-Myc-His aka G3b
2953	Syb18-Myc-His aka G2b
2954	Syb19-Myc-His
2955	Syb20-Myc-His
2956	Syb21-Myc-His aka B5c
2957	Syb22-Myc-His
2958	Syb23-Myc-His
2959	Syb24-Myc-His aka B4a
2960	Syb26-Myc-His aka B3a
2961	Syb27-Myc-His
2962	Syb28-Myc-His aka B1b
2963	Syb29-Myc-His
2964	Syb30-Myc-His
2965	Syb31-Myc-His aka B1a
2966	Syb32-Myc-His
2967	Syb33-Myc-His
2968	Syb34-Myc-His aka B2a
2969	Syb35-Myc-His
2970	Syb37-Myc-His aka B4b
2971	Syb40-Myc-His
2972	Syb42-Myc-His
2973	Syb44-Myc-His aka B5a
2974	Syb45-Myc-His aka B3b
2975	Syb46-Myc-His
2976	Syb47-Myc-His aka B2b
2977	Syb48-Myc-His
2978	Syb49-Myc-His
2979	Syb50-Myc-His
2980	Syb51-Myc-His
2981	Syb52-Myc-His
2982	Syb53-Myc-His
2983	Syb54-Myc-His
2984	Syb55-Myc-His
2985	Syb57-Myc-His
2986	Syb58-Myc-His
2987	Syb59-Myc-His
2988	Syb60-Myc-His
2989	Syb61-Myc-His
2990	Syb62-Myc-His
2991	Syb63-Myc-His

b, α-*At*CTNS nanobodies cloned in vector pADL-23c (Amp resistance, IPTG induction).

Identifier	α-*At*CTNS
2992	Nb2-His-Myc
2993	Nb3-His-Myc
2994	Nb5-His-Myc
2995	Nb6-His-Myc
2996	Nb7-His-Myc
2997	Nb8-His-Myc

Table A.4 (Continued)

c, α-*Hs*CTNS(Δ2–115) sybodies cloned in vector pSb_init (CAM resistance, L-(+)-arabinose induction).

Identifier	α-*Hs*CTNS(Δ2–115)	Identifier	α-*Hs*CTNS(Δ2–115)
2998	Syb1-Myc-His	3021	Syb26-Myc-His
2999	Syb2-Myc-His	3022	Syb27-Myc-His
3000	Syb3-Myc-His	3023	Syb28-Myc-His
3001	Syb4-Myc-His	3024	Syb29-Myc-His
3002	Syb5-Myc-His	3025	Syb31-Myc-His
3003	Syb6-Myc-His	3026	Syb33-Myc-His
3004	Syb7-Myc-His	3027	Syb34-Myc-His
3005	Syb8-Myc-His	3028	Syb36-Myc-His
3006	Syb9-Myc-His	3029	Syb37-Myc-His
3007	Syb10-Myc-His	3030	Syb38-Myc-His
3008	Syb11-Myc-His	3031	Syb39-Myc-His
3009	Syb12-Myc-His	3032	Syb40-Myc-His
3010	Syb13-Myc-His	3033	Syb42-Myc-His
3011	Syb14-Myc-His	3034	Syb43-Myc-His
3012	Syb15-Myc-His	3035	Syb44-Myc-His
3013	Syb16-Myc-His	3036	Syb45-Myc-His
3014	Syb17-Myc-His	3037	Syb46-Myc-His
3015	Syb18-Myc-His	3038	Syb47-Myc-His
3016	Syb19-Myc-His	3039	Syb48-Myc-His
3017	Syb21-Myc-His	3040	Syb49-Myc-His
3018	Syb22-Myc-His	3041	Syb50-Myc-His
3019	Syb23-Myc-His	3042	Syb51-Myc-His
3020	Syb24-Myc-His	3043	Syb52-Myc-His

Generation of *At*CTNS-Specific Sybodies

Table A.5 – ELISA Results & Analysis of *At*CTNS-Specific Sybodies.
ELISA scores of concave (G) and loop (B) sybodies bound to *At*CTNS-Avi-biotin (+) and $PepT_{So}$-Avi-biotin (NC, -) are listed. Orange data bars illustrate ELISA ratios. CDR sequences are annotated, with green sequences indicating multiple appearances. Sybs are numbered and classified based on their position during phylogenetic analysis. This table continues on the next page.

Location	-	+	+/-	CDR1	CDR2	CDR3	Comment	Syb No	Equiv.	Analysis Class
G-C12	0.05	0.76	14.74	SAEGME	DSWGQSTL	EVWVGRWYT		1		G4b
G-A11	0.05	0.18	3.37	KENGME	DSYGHWTW	EVWVGYWYT		2		
G-E07	0.05	0.28	5.33	EQSGME	ESSGHNTE	EVWVGYWYT	K87R mutation	3	CDR ≡ 5	
G-E08	0.05	0.12	2.19	EQSGME	ESSGHNTE	EVWVGYWYT	P28L mutation	4	CDR ≡ 5	
G-A09	0.05	0.30	5.62	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-A12	0.05	0.30	5.49	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-B02	0.06	0.41	7.26	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-B04	0.06	0.48	8.60	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-C01	0.05	0.37	6.76	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-C03	0.06	0.43	7.63	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-C08	0.05	0.36	6.74	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-C11	0.05	0.32	5.89	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-D06	0.06	0.41	6.58	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-D07	0.05	0.32	6.08	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-D11	0.05	0.37	7.03	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-E02	0.09	0.40	4.58	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-E09	0.06	0.32	4.99	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-E10	0.05	0.31	5.96	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-E12	0.05	0.31	6.31	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-F03	0.05	0.40	7.45	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-F04	0.06	0.22	3.67	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-F07	0.05	0.30	5.68	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-G03	0.06	0.44	7.96	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-G09	0.05	0.30	5.79	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-G12	0.05	0.28	5.56	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-H04	0.06	0.37	6.39	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-H07	0.05	0.28	5.18	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-H08	0.05	0.31	6.10	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-H09	0.05	0.29	5.95	EQSGME	ESSGHNTE	EVWVGYWYT		5		G3a
G-A04	0.06	0.12	1.99	YANGME	SSGGYETL	EVWVGRWYA		6		
G-A05	0.06	0.13	2.18	RSSGME	SSGGYETL	EVWVGRWYA		7		G1b
G-C07	0.05	0.09	1.77	RSSGME	SSGGYETL	EVWVGRWYA		7		G1b
G-C06	0.06	0.14	2.36	SQAGME	VSEGQQTL	EVWVGVWYY		8		
G-C09	0.05	0.28	5.16	YAAGME	ESQGQQTE	EVWVGGWYY	L11M mutation	9		G2a
G-D01	0.07	0.30	4.64	YAAGME	ESQGQQTE	EVWVGGWYY	L11M mutation	9		G2a
G-G10	0.05	0.19	3.78	YAAGME	ESQGQQTE	EVWVGGWYY	L11M mutation	9		G2a
G-D05	0.15	0.33	2.14	YAAGME	ESQGQQTE	EVWVGGWYY	N74D mutation	10	CDR ≡ 11	
G-B10	0.05	0.27	5.19	YAAGME	ESQGQQTE	EVWVGGWYY		11		
G-A01	0.06	0.83	13.06	ELAGME	ESWGEQTE	EVWVGGWYY		12		G4a
G-A10	0.06	0.33	5.65	ELAGME	ESWGEQTE	EVWVGGWYY		12		G4a
G-B06	0.12	0.85	7.23	ELAGME	ESWGEQTE	EVWVGGWYY		12		G4a
G-F02	0.06	0.14	2.49	AKAGME	ESNGSFTE	EVWVGGWYY		13		
G-A07	0.05	0.08	1.62	AEASME	NSDGSETI	EVWVGRWYW		14		
G-E04	0.06	0.11	1.84	ATAGME	SSNGTQTE	EVWVGYWYH	G42C mutation	15		G1a
G-H05	0.06	0.09	1.50	ATAGME	SSNGTQTE	EVWVGYWYH	G42C mutation	15		G1a
G-E05	0.06	0.39	6.68	SWHME	SSLGWWTY	YVNVGEEYY	Δ30, T91S mutation	16		G3b
G-F05	0.06	0.35	6.13	SWHME	SSLGWWTY	YVNVGEEYY	Δ30, T91S mutation	16		G3b
G-H02	0.05	0.20	3.77	ESTHMW	HSNGISTW	FVKVGSGYY	G9D mutation	17	CDR ≡ 18	
G-E01	0.05	0.24	4.44	ESTHMW	HSNGISTW	FVKVGSGYY		18		G2b
G-F01	0.06	0.30	4.89	ESTHMW	HSNGISTW	FVKVGSGYY		18		G2b
B-A10	0.06	0.25	4.37	WHGHMY	DSTGWETY	WGWTWHA		19		
B-B06	0.06	0.13	2.25	YHEVMY	DSQGGHTY	YGFNMAA		20		

Table A.5 (Continued)

Location	-	+	+/-	CDR1	CDR2	CDR3	Comment	Syb No	Equiv.	Analysis Class
B-F04	0.06	1.12	20.04	HKDMMY	ASYGQSTH	FGWDYWS		21		B5c
B-F03	0.06	0.23	4.11	MEHYMY	RSYGHNTY	HGWDAWE		22		
B-H01	0.05	0.21	3.92	ARAEMY	DSIGHQTW	YGFWAWE	D73G mutation	23		
B-G11	0.05	0.46	9.17	AQAEMY	DSIGHQTW	YGFWAWE		24		B4a
B-H08	0.05	0.48	9.68	AQAEMY	DSIGHQTW	YGFWAWE		24		B4a
B-E12	0.05	0.14	2.73	YWSTMF	ESNGHNTW	DGWNAWE	N84S mutation	25	CDR ≡ 26	
B-A05	0.06	0.42	7.10	YWSTMF	ESNGHNTW	DGWNAWE		26		B3a
B-D10	0.06	0.32	5.64	YWSTMF	ESNGHNTW	DGWNAWE		26		B3a
B-G02	0.05	0.42	7.99	YWSTMF	ESNGHNTW	DGWNAWE		26		B3a
B-G09	0.05	0.30	5.80	YWSTMF	ESNGHNTW	DGWNAWE		26		B3a
B-H02	0.05	0.46	8.84	YWSTMF	ESNGHNTW	DGWNAWE		26		B3a
B-A01	0.06	0.13	2.00	VSEQMH	ESGGYRTY	YGWWWEA	D73N mutation	27		
B-D06	0.06	0.17	2.81	ATRQMY	QSNGWETW	YGWWWAH		28		B1b
B-H03	0.06	0.15	2.62	ATRQMY	QSNGWETW	YGWWWAH		28		B1b
B-E05	0.07	0.11	1.70	RSQTMY	ASKGAQTH	NGWWWAA		29		
B-G12	0.05	0.86	16.96	FVEHMY	ESKGLETH	HGWLYAA		30		
B-C09	0.05	0.13	2.51	NNDTMY	ESEGVYTH	AGWLYAT		31		B1a
B-D12	0.05	0.12	2.31	NNDTMY	ESEGVYTH	AGWLYAT		31		B1a
B-B01	0.06	0.14	2.42	DHNTMY	NSEGWETT	DGWLWGS		32		
B-C10	0.06	0.09	1.51	WQEVMY	ESVGYSTT	EGWMWSW		33		
B-A04	0.06	0.26	4.48	AENSMY	HSAGFETW	EGWLYSA		34		B2a
B-A08	0.05	0.20	3.69	AENSMY	HSAGFETW	EGWLYSA		34		B2a
B-B09	0.06	0.20	3.45	AENSMY	HSAGFETW	EGWLYSA		34		B2a
B-D05	0.07	0.24	3.60	AENSMY	HSAGFETW	EGWLYSA		34		B2a
B-F02	0.06	0.25	4.45	AENSMY	HSAGFETW	EGWLYSA		34		B2a
B-F12	0.05	0.34	6.24	NNEVMH	SSSGYSTW	DGWAWMA		35		
B-G06	0.06	0.68	11.67	DQDTMH	SSEGYGTY	DGWWLQS	Q113R mutation	36	CDR ≡ 37	
B-B08	0.05	0.54	10.31	DQDTMH	SSEGYGTY	DGWWLQS		37		B4b
B-G03	0.05	0.09	1.69	DQDTMH	SSEGYGTY	DGWWLQS		37		B4b
B-G04	0.06	0.70	11.72	DQDTMH	SSEGYGTY	DGWWLQS		37		B4b
B-G01	0.05	0.19	3.57	YRNRMH	ESAGQETH	EGWYWQT	P41L mutation	38	CDR ≡ 39	
B-D04	0.06	0.20	3.15	YRNRMH	ESAGQETH	EGWYWQT		39		B1c
B-F11	0.06	0.14	2.49	YRNRMH	ESAGQETH	EGWYWQT		39		B1c
B-A06	0.06	0.20	3.34	YSNTMH	ESHGHNTV	EGWYWAS		40		
B-B03	0.06	1.05	18.65	YHNTMH	ESWGASTY	EGWMYQS		41		B5b
B-B02	0.06	0.29	4.69	EREWMW	ISNGERTH	DGWWNNW	G42S mutation	42		
B-H06	0.06	0.76	13.76	KHETMY	ESHGYETH	NGLWNNY	S21N mutation	43	CDR ≡ 44	
B-A09	0.05	0.92	16.74	KHETMY	ESHGYETH	NGLWNNY		44		B5a
B-A12	0.05	0.91	17.15	KHETMY	ESHGYETH	NGLWNNY		44		B5a
B-C04	0.06	0.95	16.94	KHETMY	ESHGYETH	NGLWNNY		44		B5a
B-D02	0.06	1.01	17.25	KHETMY	ESHGYETH	NGLWNNY		44		B5a
B-H05	0.06	1.01	17.49	KHETMY	ESHGYETH	NGLWNNY		44		B5a
B-E01	0.07	0.54	8.22	AQEVMY	QSHGQETY	DGWWWYA		45		B3b
B-D03	0.06	0.29	5.17	EYEHMY	NSYGTTTY	EGWMYES		46		
B-A11	0.05	0.26	4.88	NIEHMY	ASYGEGTY	EGWWYSN		47		B2b
B-C07	0.05	0.29	5.49	NIEHMY	ASYGEGTY	EGWWYSN		47		B2b
B-D07	0.05	0.24	4.53	NIEHMY	ASYGEGTY	EGWWYSN		47		B2b
B-G10	0.05	0.20	3.77	NIEHMY	ASYGEGTY	EGWWYSN		47		B2b
B-H11	0.05	0.24	4.83	NIEHMY	ASYGEGTY	EGWWYSN		47		B2b
B-F07	0.05	0.17	3.25	ETEYMH	KSWGKKTW	EGFIWYA		48		
B-F05	0.06	0.60	10.11	NNNYMY	ESYGQETW	AGWMWEA		49		
B-D11	0.06	0.18	3.19	QENYMH	NSYGRIETW	YGWWYET	Insertion before T58	50		
B-A02	0.07	0.16	2.50	EANEMH	GSTGNETH	EGWWYIS		51		
B-H04	0.06	1.00	15.71	KEESMY	FSYGSETH	EGWLWMA		52		
G-B12	0.05	0.31	6.05				Sequencing failed			
G-F06	0.06	0.25	4.07				Sequencing failed			
G-F10	0.05	0.33	6.54				Sequencing failed			
G-G07	0.05	0.12	2.30				Sequencing failed			
B-D01	0.06	0.84	14.34				CDRs not unique			

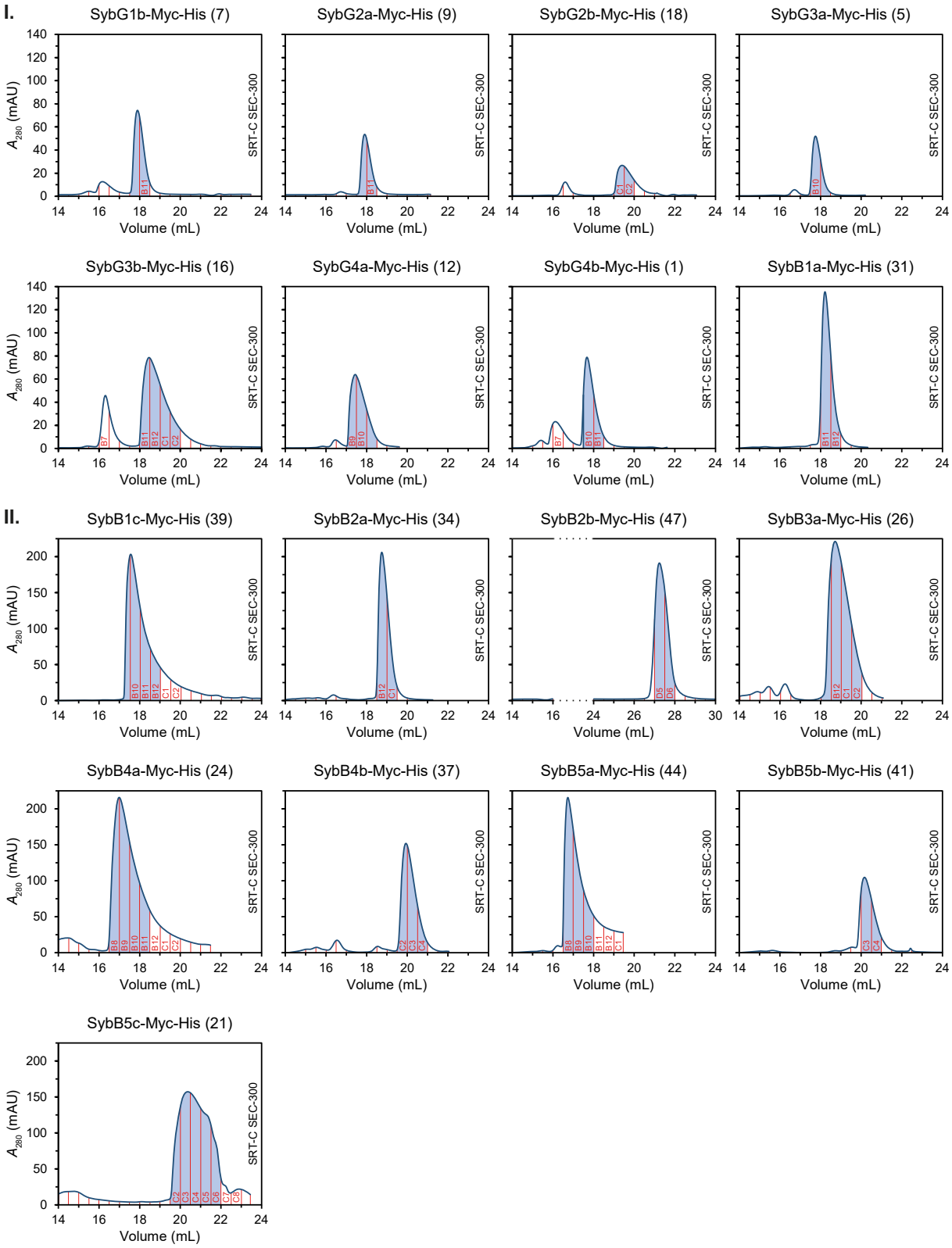


Fig. A.1 – SEC Traces of Myc-His-Tagged α-*At*CTNS Sybodies.
Small-scale expression and purification of sybodies from 50 mL cultures, performed in two batches (**I.** and **II.**). SEC was performed at 4 °C using a Sepax SRT-C SEC-300 column connected to an ÄKTA pure chromatography system (Cytiva). Samples were injected manually. Y-axes are fixed per batch to simplify the comparison of individual traces. Pooled and combined fractions are shaded in blue. Syb numbers are shown in parentheses.

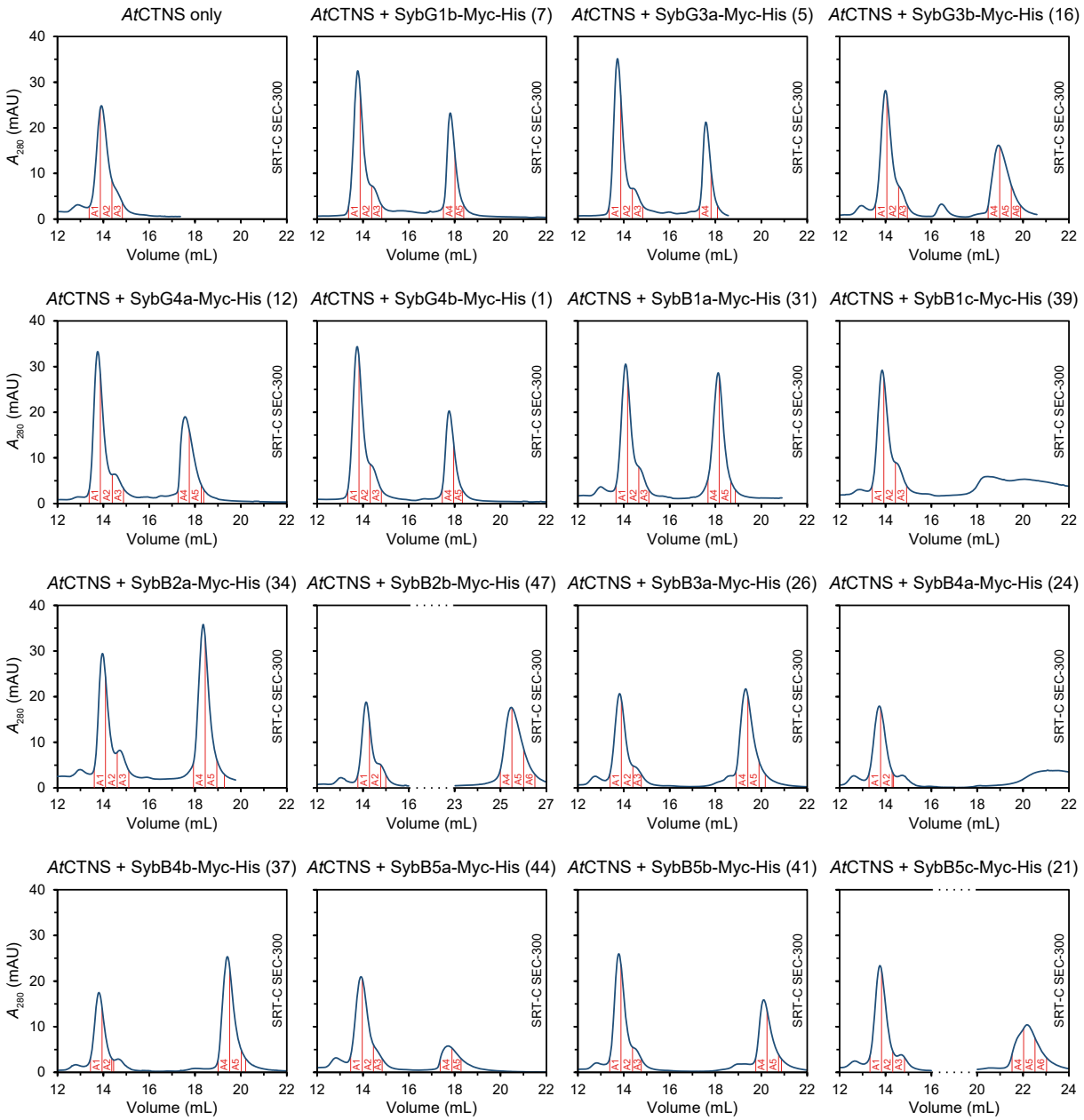


Fig. A.2 – SEC Co-Elutions of *At*CTNS & Myc-His-Tagged Sybodies.
SEC co-elutions were performed at 4 °C using a Sepax SRT-C SEC-300 column connected to an ÄKTA pure chromatography system (Cytiva). Samples were injected manually. Y-axes are fixed to simplify the comparison of individual traces. Syb numbers are shown in parentheses.

Generation of *At*CTNS-Specific Nanobodies

Table A.6 – ELISA Results & Analysis of *At*CTNS-Specific Nanobodies.
ELISA scores of nanobodies bound to *At*CTNS-Avi-biotin (+) and PepT$_{So}$-Avi-biotin (NC, -) are listed. Orange data bars illustrate ELISA ratios. CDR sequences are annotated, with green sequences indicating multiple appearances. Nbs are numbered based on their position during phylogenetic analysis.

Location	-	+	+/-	CDR1	CDR2	CDR3	Comment	Nb No	Equiv.
1-D01	0.05	0.25	4.83	GSVSELNTMGW	ARITATSDATNY	NVEGAPSWFSGIRSYW		1	
1-G10	0.05	0.08	1.57	GRTFRRYAMAW	AALSSSGLITRY	AASDNYFVDPKGTTAYDYW		2	
1-H02	0.05	0.12	2.23	GRTFRRYAMAW	AALSSSGLITRY	AASDNYFVDPKGTTAYDYW		2	
2-C07	0.05	0.10	1.88	GRTFRRYAMAW	AALSSSGLITRY	AASDNYFVDPKGTTAYDYW		2	
2-E09	0.06	0.10	1.83	GRTFRRYAMAW	AALSSSGLITRY	AASDNYFVDPKGTTAYDYW		2	
2-F05	0.05	0.11	2.17	GRTFRRYAMAW	AALSSSGLITRY	AASDNYFVDPKGTTAYDYW		2	
1-A12	0.06	0.40	6.78	GRTITPISTYVMGW	ASISWNGANTYY	AADPESHVRLRLGVGAYW		3	CDR ≡ 4
2-G02	0.05	0.38	6.93	GRTITPISTYVMGW	ASISWNGANTYY	AADPESHVRLRLGVGAYW		4	CDR ≡ 3
2-G10	0.05	0.12	2.38	GRTLTNYNMGW	AATSWIGGTPY	AAGATYRTNLRTTGTYPYW		5	
1-H06	0.06	0.07	1.35	GRTLSNVNMGW	AATSWMGGTPY	AAGATYLTNLRTEGTYPYW		6	
1-E05	0.05	0.07	1.42	GRTLTNYNMGW	AATSALGGTPY	AGGAPWRTNLRTIRTWPYW	Mixed CDRs/bad read	7	
2-G11	0.05	0.07	1.44	GRTLNYYTLGW	AATSWMGGTPY	AAGAPWRTNLRAEGTWPTW	Mixed CDRs/bad read	8	
1-A01	0.06	0.10	1.79				Syb39, not sequenced		

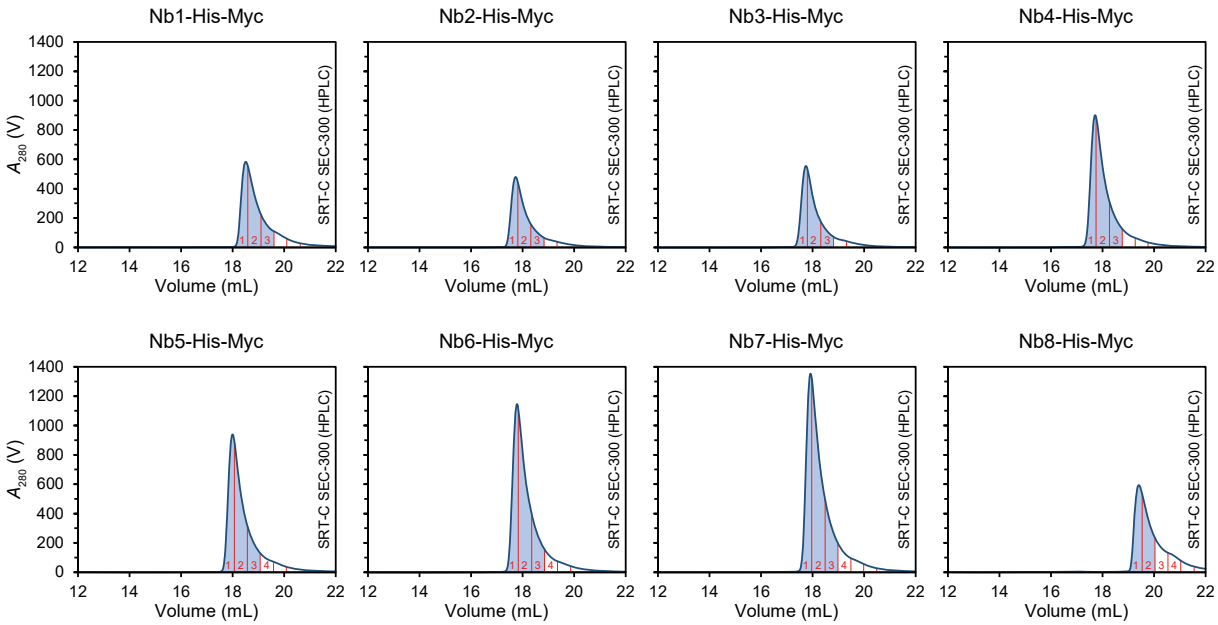


Fig. A.3 – SEC Traces of His-Myc-Tagged α-*At*CTNS Nanobodies.
Small-scale expression and purification of nanobodies from 50 mL cultures. SEC was performed at 4 °C using a Sepax SRT-C SEC-300 column connected to an HPLC system (Shimadzu). Samples were auto-injected. Y-axes are fixed to simplify the comparison of individual traces. Pooled and combined fractions are shaded in blue.

Crystallography Using Sybody- & Nanobody-Bound *At*CTNS Complexes

Table A.7 – X-Ray Data Collection & Processing Statistics of *At*CTNS-Syb39-Nb4 (S-SAD).

a, Relevant sweeps of data collected at DLS (proposal number mx23459-135).

Identifier	Beamline	Wavelength (Å)	Beamsize (μm)	Transmission (%)	Exposure (s)	Oscillation (°)	Total images	Resolution max. (Å)	Detector
6870_1	DLS - I23	2.7552	50 × 50	100	0.2000	0.10	3,600	2.00	Pilatus2 12M
6870_2	DLS - I23	2.7552	50 × 50	100	0.1000	0.10	3,600	2.00	Pilatus2 12M
6871_1	DLS - I23	2.7552	50 × 50	100	0.1500	0.10	3,600	2.00	Pilatus2 12M
6871_2	DLS - I23	2.7552	50 × 50	100	0.1500	0.10	3,600	2.00	Pilatus2 12M

b, Data processing and merging.

Identifier	Pipeline	Images used	High-resolution data cut	Space group	Cell dimensions *a*, *b*, *c* (Å)	Cell angles α, β, γ (°)	PDB
6870_1	XDS	1–3,600	3.00	$P2_12_12$	62.64, 319.71, 45.64	90.00, 90.00, 90.00	----
6870_2	XDS	1–3,600	3.00	$P2_12_12$	62.41, 318.37, 45.41	90.00, 90.00, 90.00	----
6871_1	XDS	1–3,600	3.00	$P2_12_12$	62.59, 319.44, 45.46	90.00, 90.00, 90.00	----
6871_2	XDS	1–3,600	3.00	$P2_12_12$	62.69, 320.00, 45.56	90.00, 90.00, 90.00	----
Merge	**XSCALE**	**All**	**3.20**	**$P2_12_12$**	**62.60, 319.49, 45.48**	**90.00, 90.00, 90.00**	**----**

c, Merging statistics by resolution bin and overall. *Significant at $p = 0.01$.

Resolution (Å)	Observations	Unique reflections	Completeness (%)	Mean $I/\sigma(I)$	R_{merge} (%)	R_{meas} (%)	$CC_{1/2}$ (%)
40.00–14.31	7,475	294	88.0	47.7	6.3	6.4	100*
14.31–10.12	16,140	594	100	48.4	6.5	6.6	99.9*
10.12–8.26	20,773	781	100	41.8	7.7	7.9	99.9*
8.26–7.16	25,287	900	100	35.0	10.2	10.4	99.8*
7.16–6.40	27,218	1,008	100	23.3	15.9	16.3	99.8*
6.40–5.84	31,876	1,163	100	19.0	21.2	21.6	99.6*
5.84–5.41	35,239	1,269	100	16.5	25.3	25.7	99.7*
5.41–5.06	34,676	1,302	100	16.3	25.2	25.7	99.7*
5.06–4.77	38,929	1,441	100	15.0	29.3	29.9	99.2*
4.77–4.53	38,042	1,470	100	12.5	35.4	36.1	99.6*
4.53–4.32	43,312	1,620	100	10.9	44.3	45.2	99.5*
4.32–4.13	41,264	1,614	100	8.5	57.5	58.7	99.4*
4.13–3.97	43,718	1,757	99.7	6.2	83.7	85.5	98.9*
3.97–3.82	44,686	1,770	99.9	4.6	119.5	122.0	97.8*
3.82–3.70	42,423	1,832	100	3.3	165.1	168.9	94.8*
3.70–3.58	46,437	1,964	99.9	2.3	231.3	236.6	91.0*
3.58–3.47	43,314	1,936	99.8	1.6	323.7	331.6	85.2*
3.47–3.37	41,987	2,046	99.6	1.1	438.4	449.9	74.7*
3.37–3.28	44,077	2,091	99.2	0.8	617.0	632.7	52.9*
3.28–3.20	43,436	2,141	98.5	0.5	877.4	900.3	33.9*
40.00–3.20	**710,309**	**28,993**	**99.6**	**10.5**	**32.4**	**33.1**	**99.8**

Table A.8 – X-Ray Data Collection & Processing Statistics of *At*CTNS-Syb39-Nb4 (Hg-SAD).

a, Relevant sweep of data collected at DLS (proposal number mx23459-140).

Identifier	Beamline	Wavelength (Å)	Beamsize (μm)	Transmission (%)	Exposure (s)	Oscillation (°)	Total images	Resolution max. (Å)	Detector
6883	DLS - I24	1.0080	7 × 7	10.41	0.0500	0.25	720	2.50	Pilatus3 6M

b, Data processing.

Pipeline	Images used	High-resolution data cut	Space group	Cell dimensions *a*, *b*, *c* (Å)	Cell angles α, β, γ (°)	PDB
ISPyB - xia2 3dii	1–720	Automatic	$P2_12_12$	63.30, 320.14, 46.72	90.00, 90.00, 90.00	----

c, Merging statistics by resolution bin and overall. *Significant at $p = 0.01$.

Resolution (Å)	Observations	Unique reflections	Multiplicity	Completeness (%)	Mean $I/\sigma(I)$	R_{merge} (%)	R_{meas} (%)	R_{pim} (%)	$CC_{1/2}$ (%)
58.87–9.22	3,996	803	4.98	100	23.1	4.7	5.4	2.5	99.9*
9.22–7.32	4,240	745	5.69	100	17.1	7.6	8.4	3.5	99.7*
7.32–6.40	4,311	718	6.00	100	9.9	14.4	15.8	6.4	99.7*
6.40–5.81	4,289	698	6.14	100	7.0	23.1	25.3	10.2	98.2*
5.81–5.40	4,441	713	6.23	100	5.5	28.8	31.4	12.4	98.8*
5.40–5.08	4,410	706	6.25	100	5.6	26.8	29.3	11.6	98.7*
5.08–4.82	4,305	686	6.28	100	5.0	29.7	32.6	13.0	98.8*
4.82–4.61	4,364	685	6.37	100	4.6	33.4	36.5	14.4	98.8*
4.61–4.44	4,446	702	6.33	99.6	3.6	44.6	48.6	19.1	96.8*
4.44–4.28	4,200	661	6.35	100	2.9	58.8	64.1	25.2	94.9*
4.28–4.15	4,192	697	6.01	100	2.5	62.9	69.0	27.8	96.0*
4.15–4.03	4,339	688	6.31	99.9	2.1	82.2	89.6	35.4	94.9*
4.03–3.93	4,119	653	6.31	99.9	1.5	116.6	127.2	50.2	86.7*
3.93–3.83	4,489	707	6.35	100	1.1	163.0	177.8	70.1	71.2*
3.83–3.74	4,208	656	6.41	99.9	1.0	190.0	206.9	81.0	65.6*
3.74–3.66	4,424	694	6.37	100	0.7	261.3	284.6	111.8	59.1*
3.66–3.59	4,006	653	6.13	100	0.8	224.0	245.0	98.2	62.5*
3.59–3.52	4,298	689	6.24	100	0.6	281.3	307.1	121.8	60.6*
3.52–3.46	4,340	689	6.30	99.6	0.4	372.1	406.0	160.4	30.6*
3.46–3.40	4,193	665	6.31	99.7	0.2	636.0	693.8	274.3	27.6*
58.86–3.40	**85,610**	**13,908**	**6.2**	**100**	**5.0**	**22.5**	**24.7**	**10.0**	**99.9**

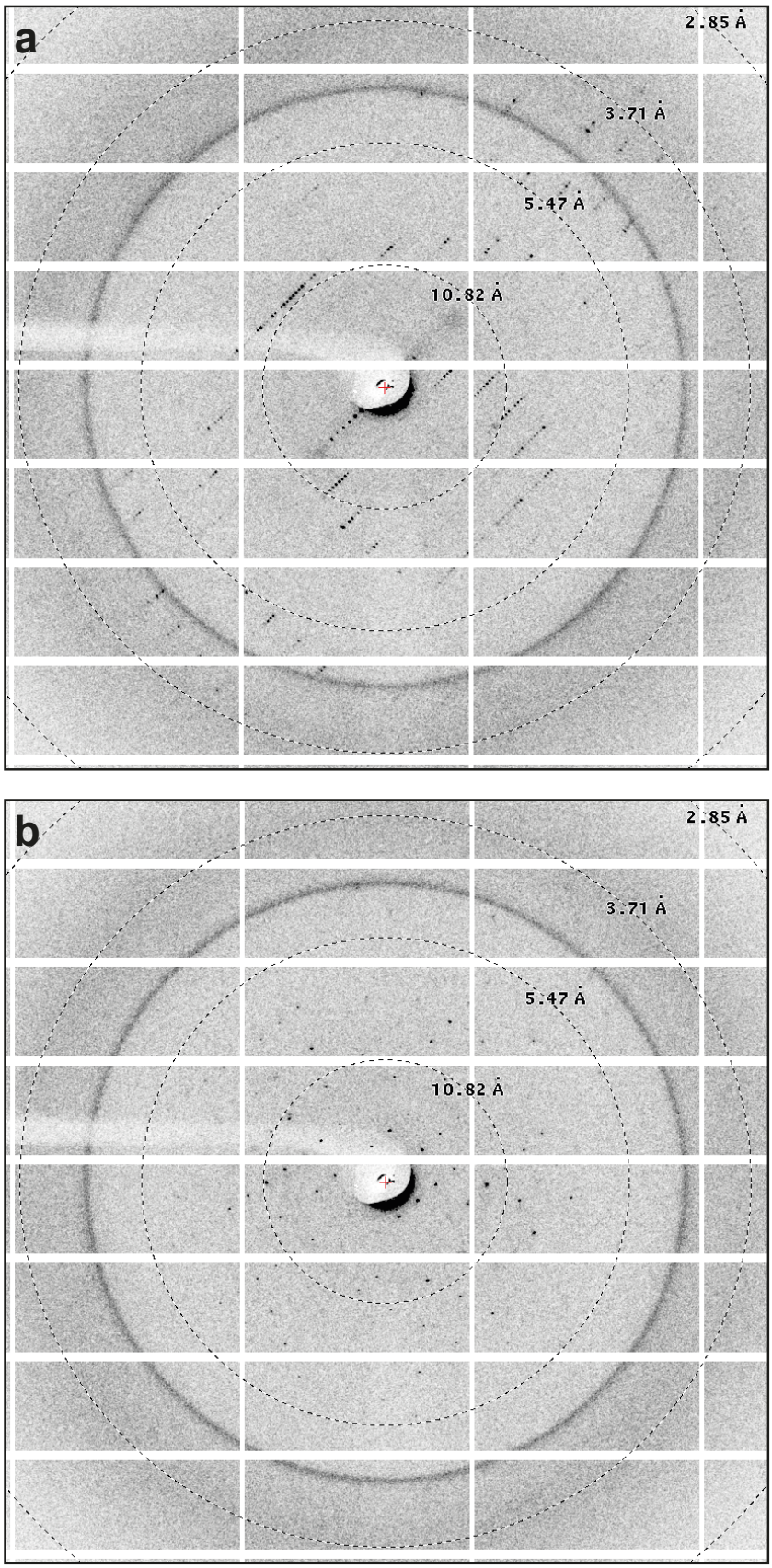


Fig. A.4 – Best Diffraction Patterns of *At*CTNS-Syb39-Nb4 (Hg-SAD).
Diffraction patterns obtained at $\varphi = 0°$ (**a**) and $\varphi = 90°$ (**b**), visualized using Adxv (Arvai, 2021).

Table A.9 – X-Ray Data Collection & Processing Statistics of *At*CTNS-Syb1-Nb4 (S-SAD).

a, Relevant sweep of data collected at DLS (proposal number mx23459-217).

Identifier	Beamline	Wavelength (Å)	Beamsize (µm)	Transmission (%)	Exposure (s)	Oscillation (°)	Total images	Resolution max. (Å)	Detector
7658_1	DLS - I23	2.00	100 × 140	100	2.7552	0.10	3,600	2.00	Pilatus2 12M

b, Data processing.

Pipeline	Images used	High-resolution data cut	Space group	Cell dimensions *a*, *b*, *c* (Å)	Cell angles α, β, γ (°)	PDB
ISPyB - autoPROC	1–3,600	Automatic	$P2_12_12$	65.67, 91.46, 102.43	90.00, 90.00, 90.00	----

c, Merging statistics by resolution bin and overall. *Significant at $p = 0.01$.

Resolution (Å)	Observations	Unique reflections	Multiplicity	Completeness (%)	Mean $I/\sigma(I)$	R_{merge} (%)	R_{meas} (%)	R_{pim} (%)	$CC_{1/2}$ (%)
68.22–5.32	28,383	2,448	11.59	100	57.7	3.4	3.5	1.0	99.8*
5.32–4.22	27,357	2,326	11.76	100	52.4	4.0	4.2	1.2	100*
4.22–3.69	26,111	2,293	11.39	100	39.7	4.8	5.0	1.4	99.9*
3.69–3.35	23,526	2,279	10.33	99.8	27.1	6.5	6.8	2.1	99.9*
3.35–3.11	22,670	2,229	10.18	98.1	19.2	9.4	9.9	3.0	99.7*
3.11–2.93	20,554	2,156	9.54	95.7	14.1	12.5	13.2	4.2	99.5*
2.93–2.78	19,585	2,111	9.29	93.5	10.4	16.2	17.1	5.4	99.5*
2.78–2.66	18,581	2,068	9.00	92.0	7.8	21.2	22.5	7.3	98.8*
2.66–2.56	17,886	2,036	8.80	90.9	6.2	25.9	27.5	9.0	98.3*
2.56–2.47	17,647	1,991	8.87	89.0	4.9	33.5	35.6	11.6	96.9*
2.47–2.39	16,875	1,984	8.52	88.3	3.7	43.3	46.2	15.5	95.3*
2.39–2.32	16,892	1,950	8.67	86.8	3.1	52.0	55.3	18.4	93.2*
2.32–2.26	15,348	1,915	8.03	86.8	2.3	69.6	74.4	25.6	86.7*
2.26–2.21	16,522	1,894	8.74	84.6	2.3	75.9	80.8	26.8	86.0*
2.21–2.16	14,628	1,897	7.72	84.4	1.5	107.9	115.9	41.1	72.0*
2.16–2.11	15,140	1,857	8.17	84.1	1.4	111.3	118.9	40.7	78.1*
2.11–2.07	14,151	1,843	7.69	82.4	1.0	153.3	164.6	58.2	60.3*
2.07–2.03	15,377	1,813	8.49	82.4	1.0	175.5	187.3	63.7	52.7*
2.03–1.99	12,822	1,826	7.04	81.8	0.6	236.0	255.6	94.7	36.1*
1.99–1.96	15,065	1,789	8.42	80.4	0.6	279.4	298.5	102.1	38.0*
68.22–1.96	**375,120**	**40,705**	**9.2**	**90.2**	**14.3**	**7.7**	**8.2**	**2.6**	**99.9**

Table A.10 – X-Ray Data Collection & Processing Statistics of *At*CTNS-Syb39-Cystine.

a, Relevant sweeps of data collected at DLS (proposal number mx23459-94).

Identifier	Beamline	Wavelength (Å)	Beamsize (μm)	Transmission (%)	Exposure (s)	Oscillation (°)	Total images	Resolution max. (Å)	Detector
6396_1	DLS - I24	0.9762	7 × 7	40.42	0.0500	0.50	360	2.50	PILATUS3 6M
6396_2	DLS - I24	0.9762	7 × 7	19.26	0.0500	0.50	360	2.50	PILATUS3 6M

b, Data processing.

Pipeline	Images used	High-resolution data cut	Space group	Cell dimensions *a*, *b*, *c* (Å)	Cell angles α, β, γ (°)	PDB
xia2 DIALS	All of sweeps 1–2	Automatic	*C*2	284.62, 64.43, 55.71	90.00, 99.86, 90.00	7ZKW

c, Merging statistics by resolution bin and overall. *Significant at $p = 0.01$.

Resolution (Å)	Observations	Unique reflections	Multiplicity	Completeness (%)	Mean $I/\sigma(I)$	R_{merge} (%)	R_{meas} (%)	R_{pim} (%)	$CC_{1/2}$ (%)
70.12–9.40	3,981	699	5.70	100	9.8	23.0	26.1	11.9	99.5*
9.40–7.46	3,975	686	5.79	99.9	5.0	40.7	45.4	19.7	69.9*
7.46–6.52	3,997	659	6.07	100	2.3	86.8	95.3	38.8	80.8*
6.52–5.92	4,161	672	6.19	100	1.7	100.6	110.0	44.0	73.3*
5.92–5.50	4,117	661	6.23	100	1.5	155.8	170.2	67.4	86.4*
5.50–5.18	4,115	660	6.23	100	1.9	125.8	138.3	56.4	78.9*
5.18–4.92	4,162	655	6.35	100	2.0	100.1	109.4	43.5	85.5*
4.92–4.70	3,982	639	6.23	100	2.0	141.8	155.2	62.2	44.8*
4.70–4.52	4,274	681	6.28	100	1.8	197.7	216.3	86.8	74.7*
4.52–4.37	4,097	651	6.29	100	2.4	81.1	88.7	35.6	70.0*
4.37–4.23	3,907	646	6.05	99.7	1.4	125.0	137.5	56.6	51.1*
4.23–4.11	3,842	661	5.81	99.7	1.3	397.7	442.4	191.3	56.7*
4.11–4.00	3,760	628	5.99	100	1.2	126.1	140.3	60.1	8.7
4.00–3.90	4,091	664	6.16	100	1.0	184.0	201.6	81.5	31.1*
3.90–3.81	4,047	665	6.09	100	1.0	144.5	158.6	64.7	66.3*
3.81–3.73	4,010	635	6.31	99.8	0.8	213.0	234.6	96.2	6.5
3.73–3.66	4,043	663	6.10	100	0.7	242.9	266.5	108.6	28.0*
3.66–3.59	3,735	643	5.81	100	0.6	289.8	319.0	132.1	9.5*
3.59–3.52	3,913	663	5.90	100	0.5	288.3	317.4	130.8	14.0*
3.52–3.47	3,643	613	5.94	95.3	0.6	266.4	292.9	120.5	10.0*
70.10–3.47	**79,852**	**13,144**	**6.1**	**100**	**2.0**	**109.4**	**120.4**	**49.7**	**79.7**

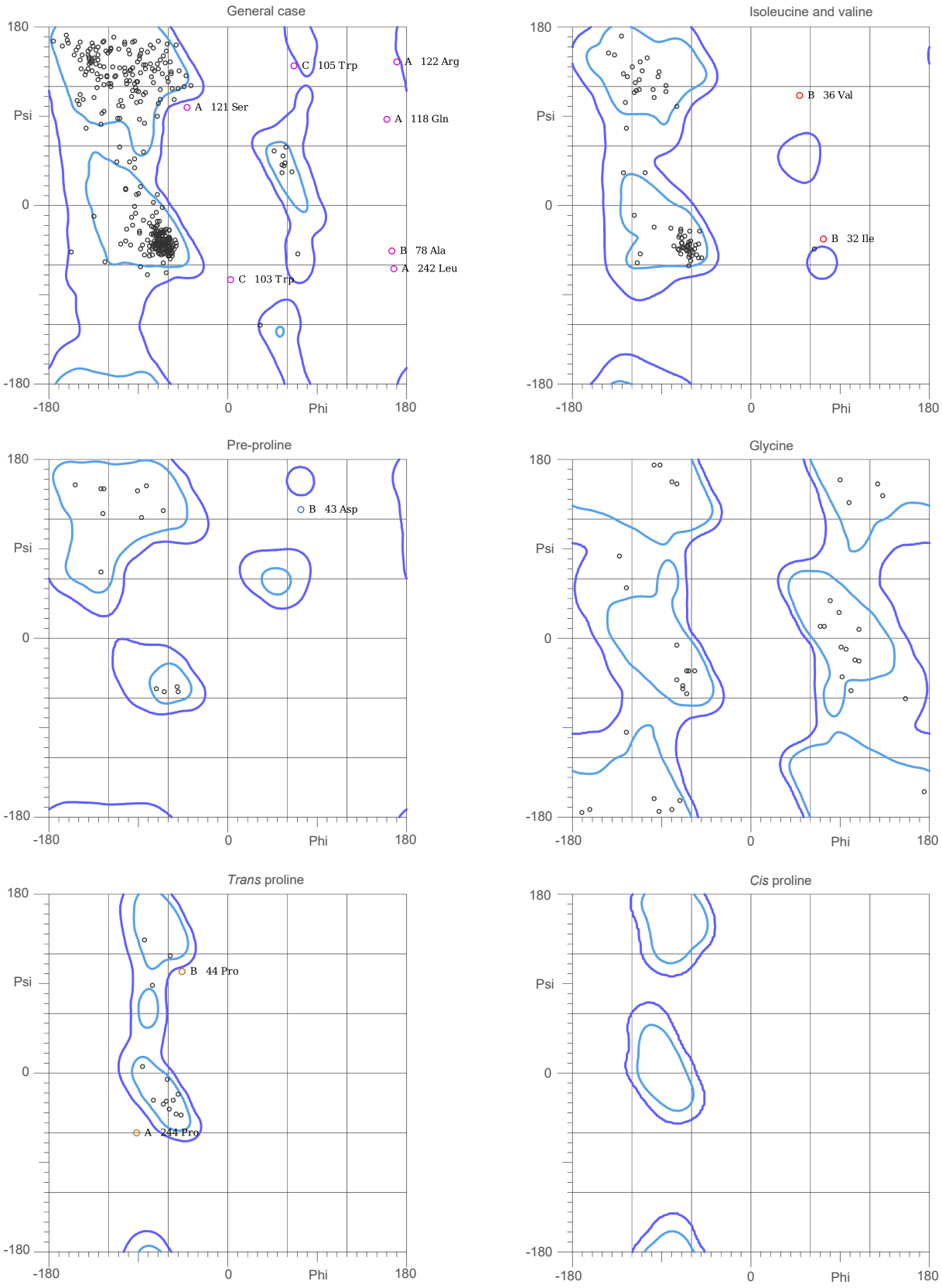


Fig. A.5 – Ramachandran Plots: *At*CTNS-Syb39-Nb4 Model.
Generated by MolProbity (Lovell *et al.*, 2003). The protein content of this model consists of chain A: *At*CTNS, B: Nb4, and C: Syb39.

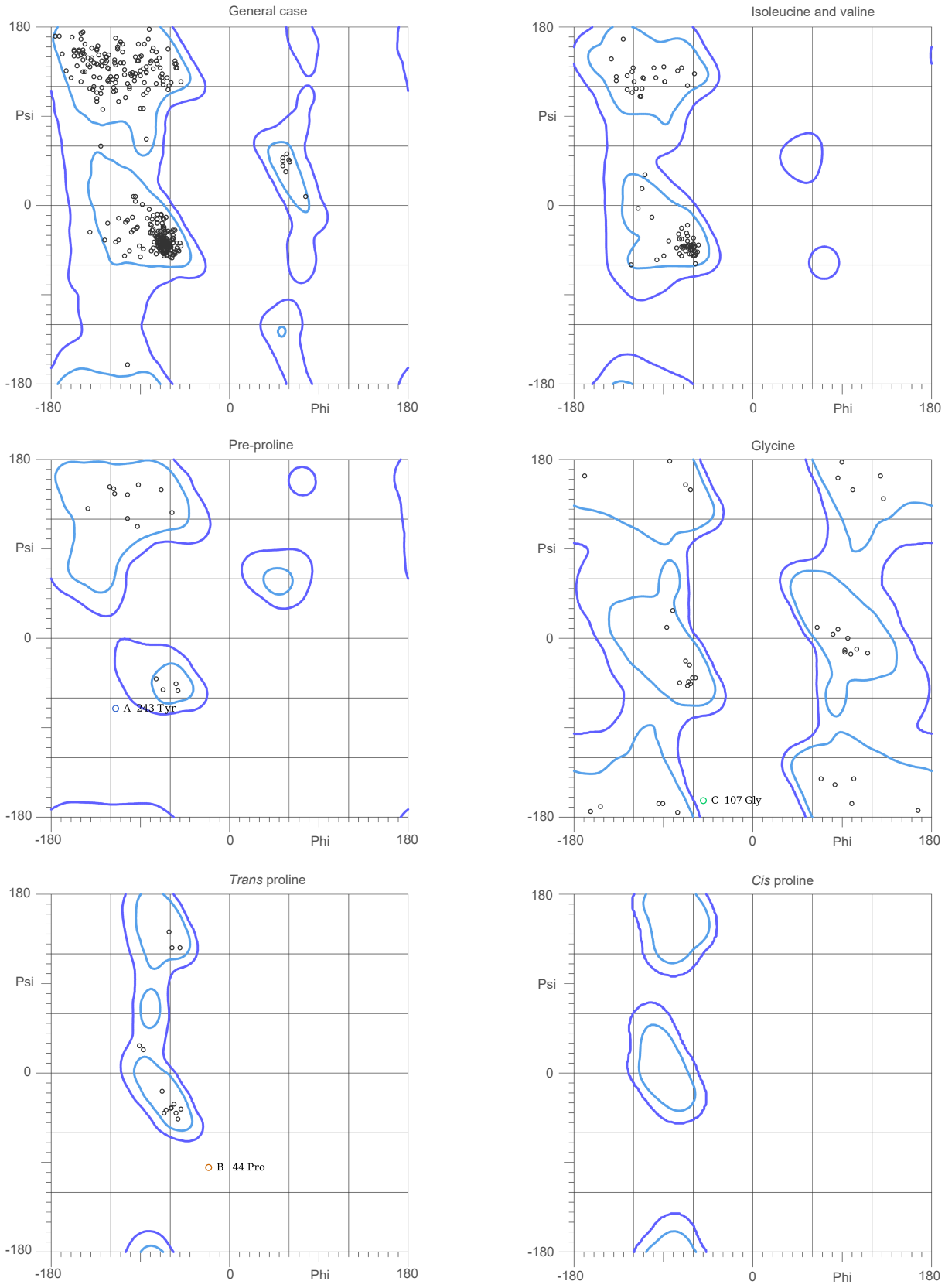


Fig. A.6 – Ramachandran Plots: *At*CTNS-Nb1-Nb4 Model.
Generated by MolProbity (Lovell *et al.*, 2003). The protein content of this model consists of chain A: *At*CTNS, B: Nb4, and C: Nb1.

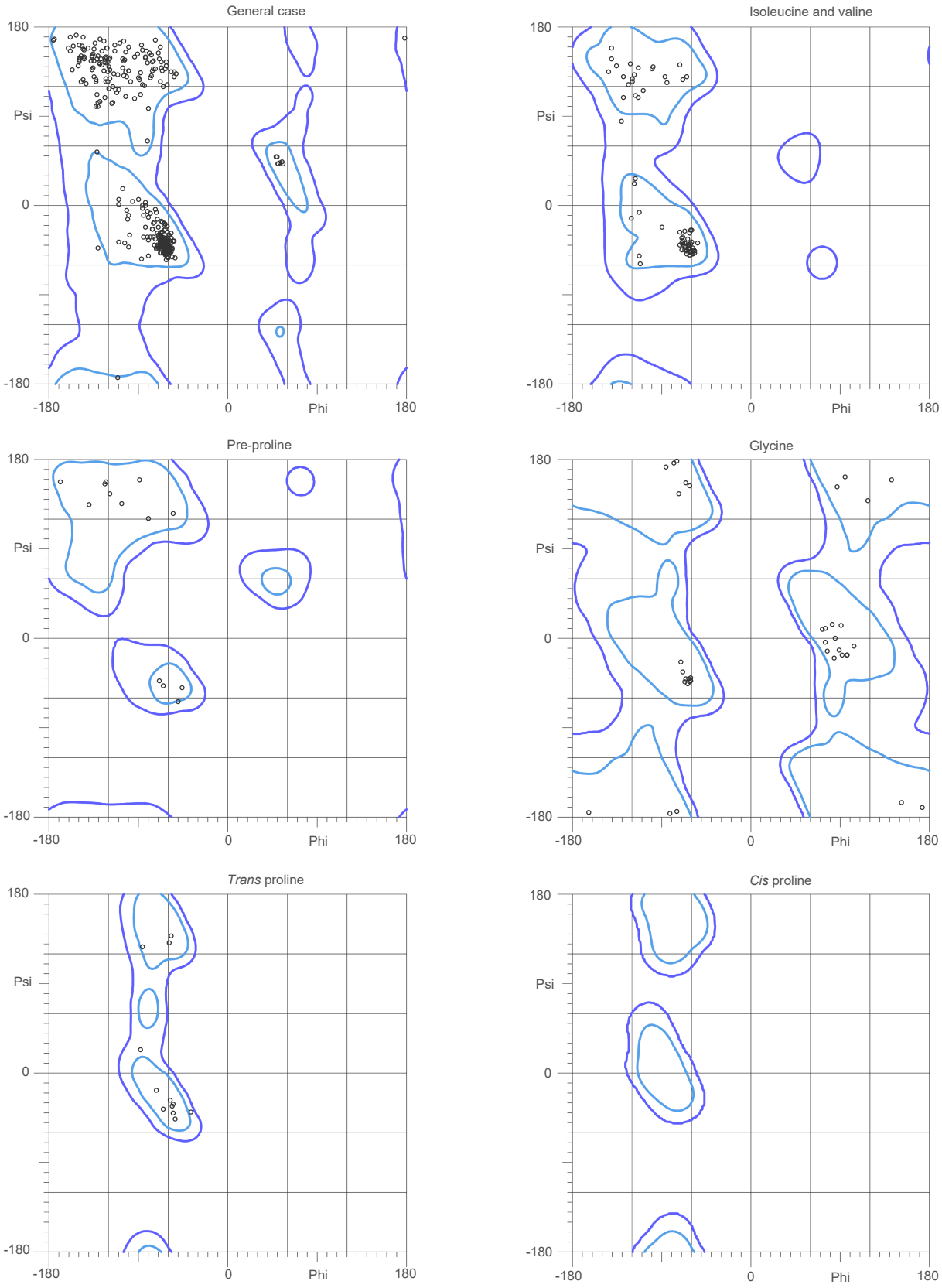


Fig. A.7 – Ramachandran Plots: *At*CTNS-Syb1-Nb4 Model.
Generated by MolProbity (Lovell *et al.*, 2003). The protein content of this model consists of chain A: *At*CTNS, B: Nb4, and C: Syb1.

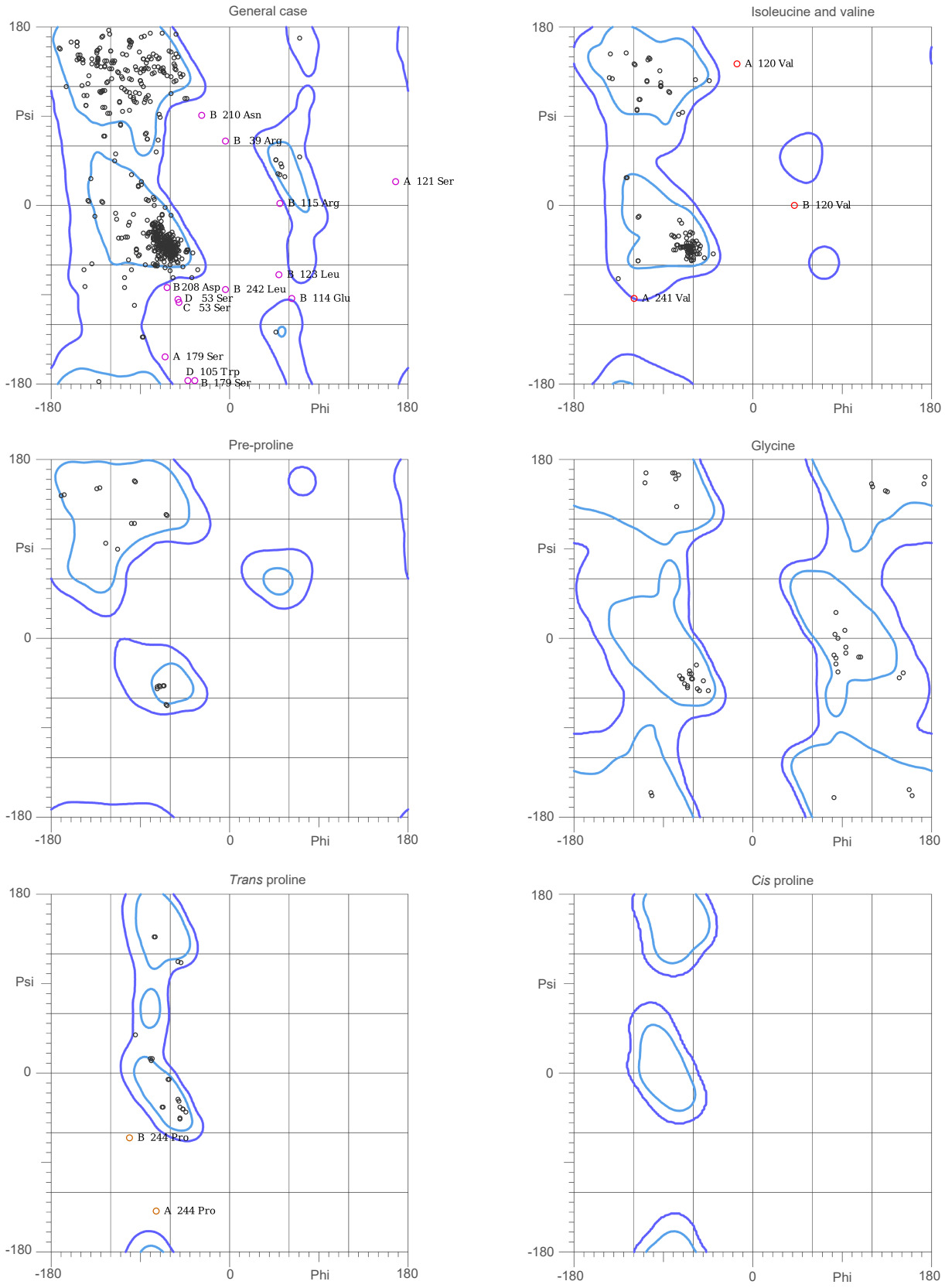


Fig. A.8 – Ramachandran Plots: *At*CTNS-Syb39-Cystine Model.
Generated by MolProbity (Lovell *et al.*, 2003). The protein content of this model consists of chains A and B: *At*CTNS, as well as C and D: Syb39.

Generation of *Hs*CTNS(Δ2–115)-Specific Sybodies

Table A.11 – ELISA Results & Analysis of *Hs*CTNS(Δ2–115)-Specific Sybodies.
ELISA scores of concave (G1 and G2 for monomer and dimer screening, respectively), loop (B1 and B2), and convex (R1 and R2) sybodies bound to the monomeric species of *Hs*CTNS(Δ2–115)-Avi-biotin (.), the dimeric species (:), and PepT_{So}-Avi-biotin (NC, -) are listed. ELISA ratios of *Hs*CTNS(Δ2–115) compared to the NC (./- and :/-), as well as the ratio between the dimeric and monomeric species, are shown. Data bars illustrate ELISA ratios (orange) and whether the Syb prefers binding to the monomeric (red) or dimeric (green) species of *Hs*CTNS(Δ2–115). CDR sequences are annotated, with green sequences indicating multiple appearances. Sybs are numbered based on their position during phylogenetic analysis. This table continues on the next page.

Location	-	.	:	./-	:/-	./:	:/.	CDR1	CDR2	CDR3	Comment	Syb No
B2-A11	0.06	0.08	0.18	1.29	2.83		2.20	SWSSMY	TSEGAWTH	YGSYGSY		1
G1-G09	0.06	0.27	0.71	4.28	11.19		2.61	AFYNMY	SSWGWSTE	YVYVGDGYK		2
G1-G03	0.07	0.22	0.51	3.25	7.46		2.29	THTRMY	WSYGTGTE	HASVGWGYH	A40V, V98A (bad read at end)	3
G2-D02	0.09	0.17	0.37	1.83	3.94		2.15	DHQRMY	HSTGSSTE	HVSVGWGYH		4
G2-G02	0.09	0.45	0.88	5.26	10.30		1.96	NFWEMA	YSSGHTH	AVQVGWSYH	Shorter CDR2 (Δ56 or 57)	5
G1-G01	0.06	0.10	0.15	1.68	2.53		1.50	EIDKMA	KSNGSMTY	AVKVGYYYE		6
G1-A06	0.07	0.23	0.56	3.55	8.44		2.38	WSEHMW	YSTGSKTE	HVSVGWGYH		7
G1-D05	0.06	0.21	0.61	3.18	9.43		2.97	WSEHMW	YSTGSKTE	HVSVGWGYH		7
G1-F08	0.06	0.24	0.54	3.96	8.73		2.20	WSEHMW	YSTGSKTE	HVSVGWGYH		7
G1-G11	0.07	0.21	0.56	3.16	8.46		2.68	WSEHMW	YSTGSKTE	HVSVGWGYH		7
G2-E09	0.07	0.16	0.52	2.31	7.36		3.18	ATEHMW	YSEGNYTE	HVSVGWGYH		8
G2-H12	0.06	0.14	0.56	2.29	9.23		4.03	ATEHMW	YSEGNYTE	HVSVGWGYH		8
G2-C05	0.09	0.13	0.19	1.51	2.11		1.40	HAQHMW	YSRGQTTE	YVSVGWGYH		9
G1-B06	0.06	0.13	0.41	2.04	6.53		3.19	YSEHMW	YSSGNMTE	HVSVGWGYA		10
G1-E03	0.07	0.11	0.43	1.61	6.35		3.93	YSEHMW	YSSGNMTE	HVSVGWGYA		10
G1-H12	0.07	0.12	0.51	1.77	7.21		4.08	YSEHMW	YSSGNMTE	HVSVGWGYA		10
G1-D12	0.07	0.07	0.11	1.06	1.70		1.60	MNGHMY	YSEGTTTE	HVSVGWGYT		11
G1-F02	0.07	0.10	0.34	1.38	4.88		3.54	ASEHMW	YSRGTETE	HVSVGWGYA		12
G1-D04	0.07	0.12	0.34	1.75	5.14		2.93	ASEHMW	YSKGTETE	HVSVGWGYA		13
G1-H02	0.06	0.09	0.35	1.61	6.10		3.80	ASEHMW	YSKGTETE	HVSVGWGYA		13
G2-G08	0.09	0.09	0.16	1.06	1.80		1.70	ASARMY	HSTGEHTE	HVSVGWGYH		14
G2-A09	0.08	0.09	0.32	1.12	4.21		3.75	DTARMY	WSKGKQTE	HVSVGWGYH		15
G2-C01	0.09	0.11	0.54	1.21	5.81		4.80	DTARMY	WSKGKQTE	HVSVGWGYH		15
G2-E04	0.07	0.11	0.45	1.62	6.60		4.08	DTARMY	WSKGKQTE	HVSVGWGYH		15
G2-A02	0.08	0.15	0.42	1.84	5.30		2.87	EKERMY	HSHGTTTE	HVSVGWGYH		16
G2-B05	0.09	0.17	0.41	1.91	4.64		2.43	EKERMY	HSHGTTTE	HVSVGWGYH		16
G2-D04	0.09	0.11	0.49	1.31	5.69		4.33	EKERMY	HSHGTTTE	HVSVGWGYH		16
G2-D05	0.09	0.23	0.47	2.69	5.45		2.02	EKERMY	HSHGTTTE	HVSVGWGYH		16
G2-D09	0.08	0.10	0.40	1.21	5.09		4.19	EKERMY	HSHGTTTE	HVSVGWGYH		16
G2-F07	0.07	0.15	0.51	2.05	6.86		3.34	EKERMY	HSHGTTTE	HVSVGWGYH		16
G2-A03	0.10	0.16	0.57	1.68	5.88		3.50	VWYNMY	QSYGWGTE	FVKVGRGYT		17
G2-E01	0.09	0.28	0.56	3.26	6.56		2.01	VWYNMY	QSYGWGTE	FVKVGRGYT		17
G2-F04	0.08	0.09	0.27	1.17	3.39		2.91	VWYNMY	QSYGWGTE	FVKVGRGYT	N84S mutation	17
G1-B03	0.07	0.08	0.12	1.15	1.75		1.52	DRSEMY	SSWGWFTH	AVGVGTQYY	Mixed CDRs	18
R1-F06(G)	0.06	0.19	0.06	3.23	1.05	3.07		DRSEMY	SSWGWFTH	AVGVGTQYY		18
B1-E10	0.07	0.11	0.26	1.59	3.84		2.41	YRAEMW	SSWGWFTH	YGQLATY	N84D mutation	19
B2-B11	0.09	0.12	0.78	1.33	8.83		6.66	HRAEMA	SSKGWFTH	YGAGSSA		20
B2-D08	0.09	0.13	0.65	1.50	7.36		4.90	HRAEMA	SSKGWFTH	YGAGSSA		20
B2-F04	0.07	0.13	0.52	1.82	7.11		3.91	HRAEMA	SSKGWFTH	YGAGSSA		20
B2-A05	0.08	0.13	0.42	1.64	5.52		3.37	KRAEMH	SSMGWWTH	YGAHYDV		21
B2-A04	0.07	0.14	0.10	2.07	1.56	1.33		ERAEMH	SSAGWWTH	YGYHAQN		22
B2-H06	0.10	0.07	0.16	0.71	1.61		2.28	DRAEME	SSLGWWTH	HGQDHRY		23

Table A.11 (Continued)

Location	-	.	:	./-	:/-	./:	:/.	CDR1	CDR2	CDR3	Comment	Syb No
B1-B05	0.08	0.08	0.24	1.02	3.00		2.92	ERSEMH	SSLGWWTH	HGYSNAY		24
B1-B10	0.07	0.08	0.21	1.11	2.98		2.68	ERSEMH	SSLGWWTH	HGYSNAY		24
B1-D12	0.07	0.08	0.20	1.10	2.71		2.46	SKAEMH	SSMGFWTH	YGYADSW		25
B1-H10	0.06	0.08	0.17	1.22	2.72		2.23	SKAEMH	SSMGFWTH	YGYADSW		25
B2-H08	0.07	0.59	0.85	8.52	12.21		1.43	GHAEMH	SSWGWWTH	YGASEWG	N84D mutation	26
B2-B10	0.08	0.07	0.12	0.83	1.52		1.84	SHAEMH	SSWGWFTH	YGYHQYW		27
B1-F10	0.09	0.10	0.24	1.10	2.77		2.52	FRSEMH	SSLGWWTH	YGNTGST		28
B1-G06	0.07	0.11	0.14	1.54	1.90		1.23	ERSEMH	SSLGWWTH	YGTTWWQ		29
B1-F09	0.06	0.09	0.25	1.34	3.92		2.93	HRAEMH	SSLGWMTH	YGSSIEV		30
B2-B09	0.08	0.13	0.37	1.63	4.83		2.96	NRAEMH	SSTGFFTH	YGRAADA	Mixed CDRs	31
B2-C01	0.07	0.15	0.45	2.09	6.08		2.90	NRSEMH	SSTGFFTH	YGRAVDA		32
B1-A05	0.07	0.11	0.09	1.52	1.17	1.30		HRAEMH	SSWGLWTH	YGEDNAS		33
B2-E08	0.07	0.10	0.12	1.47	1.73		1.18	WKSEMH	SSTGWWTH	YGDQTNA		34
B2-F10	0.09	0.29	0.75	3.28	8.51		2.59	KWTRMY	SSWGWRTW	TGSYQET		35
B2-C09	0.09	0.08	0.21	0.86	2.32		2.69	YHQRMY	SSEGAQTW	WGNITQW		36
B2-A03	0.07	0.29	0.85	4.35	12.87		2.96	MWSRMY	SSFGWHTW	WGEYEQY		37
B2-F11	0.08	0.30	0.93	3.98	12.23		3.07	MWSRMY	SSFGWHTW	WGEYEQY		37
B2-G02	0.08	0.10	0.29	1.16	3.55		3.05	AKSTMY	TSEGNHTE	FGAYYTY		38
R1-A08	0.06	0.06	0.16	0.97	2.61		2.68	SISSIT	ATHNGR	YHSHPLLYWKYW	Shorter CDR3 (Δ102–103), A98T	39
B1-H11	0.06	0.13	0.45	2.10	7.13		3.40	HIKTMH	TSEGRHTE	WGWYDSY		40
B1-D07	0.07	0.27	0.69	3.88	9.93		2.56	YRHSMY	TSEGRHTE	YGWYQAY		41
B1-D11	0.07	0.15	0.53	2.29	8.06		3.53	YRHSMY	TSEGRHTE	YGWYQAY		41
B1-B08	0.07	0.10	0.37	1.44	5.42		3.76	AHYSMH	TSEGHHTE	YGAFQSW		42
B1-F01	0.07	0.09	0.31	1.30	4.19		3.23	AHYSMH	TSEGHHTE	YGAFQSW		42
B1-G08	0.09	0.13	0.42	1.48	4.90		3.32	AHYSMH	TSEGHHTE	YGAFQSW		42
B2-C04	0.08	0.12	0.55	1.50	7.16		4.77	VWYTMY	TSEGHHTE	YGEYQSV		43
B2-A02	0.07	0.12	0.48	1.59	6.43		4.05	TWYTMY	TSDGHYTT	WGSYQVY		44
B1-C09	0.06	0.08	0.23	1.25	3.54		2.84	ERGSMY	TSEGYHTA	YGAYQSY		45
B1-D03	0.07	0.08	0.22	1.17	3.24		2.76	ERGSMY	TSEGYHTA	YGAYQSY		45
B1-G01	0.06	0.12	0.24	1.88	3.77		2.01	ERGSMY	TSEGYHTA	YGAYQSY		45
B1-H07	0.07	0.12	0.39	1.72	5.45		3.16	ERGSMY	TSEGYHTA	YGAYQSY	V2F mutation	45
R1-B11(B)	0.06	0.09	0.06	1.53	1.02	1.50		TNIYMH	LSHGYGTA	QGSGWEH		46
R1-C01	0.06	0.12	0.06	1.85	0.94	1.96		FISTIY	YTNMGN	RRGDSKPLDDSGYS		47
R1-B06	0.06	0.14	0.06	2.20	1.02	2.15		FISTIY	YTNMGN	RRGDSKPLDDSWYS		48
R1-E06	0.06	0.06	0.15	1.02	2.45		2.41	FISTIY	YTNMGN	RRGDSKPLDDSWYS	S11G mutation	48
R2-H08	0.06	0.09	0.25	1.39	4.00		2.89	TINWIE	NTWSGS	TVGHRYPLYSKSYW	A23V, N84D mutations	49
R2-H02	0.06	0.07	0.14	1.12	2.16		1.94	NIWSIE	NTWAGL	YIGKHSPLFHISYN	Q39L mutation	50
R2-A01	0.07	0.07	0.14	1.11	2.05		1.85	DIWYIE	NTWTGQ	YIGRVDPLEYEAYY		51
R2-A02	0.08	0.09	0.26	1.13	3.51		3.11	DIWYIE	NTWTGQ	YIGRVDPLEYEAYY		51
R2-A04	0.07	0.24	0.13	3.49	1.94	1.80		DIWYIE	NTWTGQ	YIGRVDPLEYEAYY		51
R2-A05	0.07	0.08	0.13	1.14	1.72		1.51	DIWYIE	NTWTGQ	YIGRVDPLEYEAYY		51
R2-A10	0.09	0.08	0.14	0.90	1.52		1.69	DIWYIE	NTWTGQ	YIGRVDPLEYEAYY		51
R2-B02	0.07	0.08	0.12	1.07	1.69		1.59	DIWYIE	NTWTGQ	YIGRVDPLEYEAYY	Bad read in CDR3	51
R2-B09	0.08	0.11	0.16	1.42	1.95		1.38	DIWYIE	NTWTGQ	YIGRVDPLEYEAYY		51
R2-C09	0.09	0.09	0.18	0.97	2.04		2.10	DIWYIE	NTWTGQ	YIGRVDPLEYEAYY		51
R2-F06	0.07	0.13	0.14	1.78	1.99		1.11	DIWYIE	NTWTGQ	YIGRVDPLEYEAYY		51
R2-G05	0.07	0.08	0.16	1.12	2.28		2.04	DIWYIE	NTWTGQ	YIGRVDPLEYEAYY		51
R2-G11	0.09	0.11	0.15	1.20	1.62		1.35	DIWYIE	NTWTGQ	YIGRVDPLEYEAYY		51
R2-H11	0.08	0.08	0.14	1.06	1.87		1.76	DIWYIE	NTWTGQ	YIGRVDPLEYEAYY		51
R2-D02	0.07	0.13	0.36	1.80	5.01		2.78	MIYLIE	NTWTGQ	YIGWVSPLYHQYYS		52
R2-F05	0.09	0.14	0.37	1.64	4.35		2.65	MIYLIE	NTWTGQ	YIGWVSPLYHQYYS		52
R2-G01	0.07	0.11	0.39	1.68	5.75		3.42	MIYLIE	NTWTGQ	YIGWVSPLYHQYYS		52
R2-H09	0.08	0.17	0.61	2.04	7.33		3.60	MIYLIE	NTWTGQ	YIGWVSPLYHQYYS		52
G1-A09	0.07	0.07	0.11	0.96	1.54		1.61				Sequencing failed twice	

Table A.12 – Further ELISA Results of Untested *Hs*CTNS(Δ2–115)-Specific Sybodies.
ELISA intensities and ratios are shown as in **Table A.11**. This table continues on the next page.

Location	-	.	:	./-	:/-	./:	:/.
B1-A01	0.31	0.09	0.26	0.28	0.83		
B1-A02	0.07	0.11	0.30	1.66	4.44		2.67
B1-A03	0.07	0.08	0.23	1.25	3.47		2.76
B1-A04	0.07	0.10	0.31	1.49	4.64		3.11
B1-A06	0.08	0.10	0.20	1.23	2.60		2.11
B1-A07	0.08	0.07	0.13	0.93	1.59		1.71
B1-A08	0.08	0.07	0.07	0.95	0.95		
B1-A09	0.09	0.08	0.09	0.94	1.00		
B1-A10	0.07	0.09	0.22	1.35	3.27		2.43
B1-A11	0.07	0.09	0.30	1.32	4.42		3.34
B1-A12	0.06	0.07	0.07	1.04	1.17		
B1-B01	0.06	0.09	0.22	1.43	3.45		2.42
B1-B02	0.07	0.08	0.19	1.22	2.82		2.31
B1-B03	0.08	0.09	0.23	1.15	3.10		2.68
B1-B04	0.06	0.08	0.22	1.28	3.59		2.81
B1-B06	0.07	0.08	0.08	1.11	1.07		
B1-B07	0.08	0.08	0.07	0.95	0.93		
B1-B09	0.07	0.07	0.07	1.06	1.12		
B1-B11	0.06	0.06	0.07	1.02	1.08		
B1-B12	0.06	0.06	0.06	1.07	1.02		
B1-C01	0.07	0.10	0.24	1.48	3.68		2.48
B1-C02	0.07	0.07	0.07	0.97	0.97		
B1-C03	0.25	0.22	0.21	0.91	0.84		
B1-C04	0.06	0.06	0.07	1.02	1.18		
B1-C05	0.07	0.11	0.25	1.58	3.80		2.40
B1-C06	0.29	0.42	0.26	1.43	0.91		
B1-C07	0.08	0.13	0.31	1.59	3.92		2.46
B1-C08	0.07	0.08	0.10	1.09	1.40		
B1-C10	0.07	0.09	0.26	1.27	3.89		3.06
B1-C11	0.07	0.06	0.08	0.94	1.15		
B1-C12	0.06	0.07	0.07	1.12	1.14		
B1-D01	0.07	0.07	0.17	1.06	2.53		2.38
B1-D02	0.07	0.08	0.16	1.13	2.28		2.02
B1-D04	0.06	0.07	0.07	1.12	1.11		
B1-D05	0.07	0.13	0.26	1.97	3.96		2.01
B1-D06	0.09	0.12	0.33	1.25	3.46		2.76
B1-D08	0.10	0.10	0.30	1.05	3.02		2.89
B1-D09	0.07	0.07	0.09	1.00	1.27		
B1-D10	0.07	0.08	0.21	1.17	3.05		2.62
B1-E01	0.06	0.08	0.20	1.22	3.13		2.58
B1-E02	0.07	0.08	0.19	1.24	2.96		2.39
B1-E03	0.07	0.08	0.19	1.16	2.85		2.45
B1-E04	0.06	0.07	0.08	1.09	1.27		
B1-E05	0.08	0.11	0.24	1.37	3.12		2.27
B1-E06	0.11	0.10	0.13	0.88	1.19		
B1-E07	0.09	0.10	0.09	1.17	1.08		
B1-E08	0.08	0.11	0.21	1.39	2.80		2.01
B1-E09	0.07	0.07	0.08	1.00	1.28		
B1-E11	0.14	0.12	0.13	0.90	0.93		
B1-E12	0.06	0.08	0.07	1.19	1.07		
B1-F02	0.06	0.07	0.06	1.10	1.01		
B1-F03	0.09	0.09	0.10	1.03	1.04		
B1-F04	0.06	0.07	0.07	1.11	1.11		
B1-F05	0.10	0.11	0.11	1.02	1.02		
B1-F06	0.12	0.15	0.19	1.29	1.67		1.29
B1-F07	0.09	0.10	0.10	1.15	1.15		
B1-F08	0.68	0.61	0.57	0.90	0.84		
B1-F11	0.06	0.08	0.21	1.23	3.35		2.72
B1-F12	0.06	0.09	0.24	1.44	3.95		2.74
B1-G02	0.06	0.09	0.29	1.37	4.54		3.31
B1-G03	0.07	0.09	0.24	1.31	3.71		2.83
B1-G04	0.06	0.08	0.20	1.33	3.52		2.64
B1-G05	0.09	0.12	0.11	1.43	1.29		
B1-G07	0.08	0.24	0.59	2.86	7.13		2.49
B1-G09	0.07	0.08	0.25	1.22	3.84		3.15
B1-G10	0.07	0.09	0.28	1.28	4.04		3.16
B1-G11	0.06	0.10	0.25	1.66	4.26		2.56
B1-G12	0.07	0.06	0.07	0.91	0.93		
B1-H01	0.07	0.07	0.07	1.03	1.13		
B1-H02	0.06	0.06	0.07	1.10	1.23		
B1-H03	0.06	0.07	0.19	1.13	3.06		2.70
B1-H04	0.06	0.06	0.06	1.04	1.02		
B1-H05	0.17	0.16	0.14	0.99	0.84		
B1-H06	0.08	0.11	0.25	1.37	3.22		2.36
B1-H08	0.08	0.11	0.33	1.38	4.19		3.04
B1-H09	0.06	0.06	0.06	1.03	1.02		
B1-H12	0.16	0.16	0.17	1.03	1.05		
B2-A01	0.07	0.08	0.17	1.07	2.26		2.11
B2-A06	0.08	0.10	0.23	1.27	2.81		2.22
B2-A07	0.06	0.08	0.09	1.20	1.48		

Location	-	.	:	./-	:/-	./:	:/.
B2-A08	0.06	0.09	0.30	1.40	4.80		3.43
B2-A09	0.07	0.10	0.12	1.37	1.69		1.24
B2-A10	0.06	0.07	0.09	1.15	1.35		
B2-A12	0.07	0.06	0.08	0.87	1.17		
B2-B01	0.07	0.07	0.10	1.06	1.51		1.42
B2-B02	0.08	0.10	0.22	1.22	2.70		2.21
B2-B03	0.07	0.09	0.30	1.22	4.17		3.42
B2-B04	0.07	0.07	0.08	0.93	1.09		
B2-B05	0.09	0.11	0.12	1.22	1.33		
B2-B06	0.42	0.10	0.15	0.23	0.36		
B2-B07	0.09	0.15	0.64	1.66	7.12		4.28
B2-B08	0.07	0.06	0.08	0.87	1.13		
B2-B12	0.08	0.07	0.14	0.89	1.69		1.91
B2-C02	0.07	0.09	0.28	1.28	3.81		2.98
B2-C03	0.07	0.35	0.79	5.42	12.12		2.24
B2-C05	0.12	0.08	0.13	0.69	1.08		
B2-C06	0.12	0.12	0.24	0.95	1.95		2.06
B2-C07	0.10	0.30	0.66	3.11	6.90		2.22
B2-C08	0.08	0.13	0.59	1.56	6.97		4.47
B2-C10	0.09	0.13	0.36	1.45	3.99		2.74
B2-C11	0.07	0.07	0.10	1.02	1.49		
B2-C12	0.07	0.08	0.08	1.15	1.20		
B2-D01	0.08	0.13	0.42	1.50	4.98		3.33
B2-D02	0.09	0.08	0.09	0.90	1.04		
B2-D03	0.07	0.08	0.16	1.13	2.38		2.11
B2-D04	0.07	0.08	0.21	1.05	2.85		2.70
B2-D05	0.19	0.09	0.34	0.50	1.80		3.58
B2-D06	0.12	0.10	0.10	0.84	0.89		
B2-D07	0.08	0.07	0.08	0.89	0.99		
B2-D09	0.10	0.08	0.17	0.75	1.62		2.15
B2-D10	0.22	0.08	0.12	0.34	0.53		
B2-D11	0.08	0.08	0.20	0.98	2.61		2.65
B2-D12	0.08	0.09	0.22	1.13	2.83		2.50
B2-E01	0.10	0.13	0.19	1.39	1.99		1.43
B2-E02	0.10	0.14	0.54	1.44	5.38		3.75
B2-E03	0.09	0.13	0.48	1.49	5.48		3.68
B2-E04	0.09	0.10	0.25	1.15	2.79		2.42
B2-E05	0.08	0.11	0.10	1.46	1.32		
B2-E06	0.10	0.09	0.12	0.95	1.26		
B2-E07	0.10	0.16	0.66	1.60	6.56		4.10
B2-E09	0.11	0.12	0.32	1.07	2.93		2.74
B2-E10	0.08	0.09	0.21	1.13	2.64		2.34
B2-E11	0.07	0.08	0.13	1.26	1.96		1.55
B2-E12	0.08	0.08	0.08	1.09	1.13		
B2-F01	0.10	0.14	0.50	1.50	5.25		3.50
B2-F02	0.09	0.09	0.28	1.08	3.20		2.96
B2-F03	0.08	0.09	0.21	1.10	2.60		2.37
B2-F05	0.15	0.17	0.59	1.12	3.97		3.53
B2-F06	0.10	0.25	0.19	2.52	1.90	1.32	
B2-F07	0.11	0.09	0.20	0.89	1.92		2.15
B2-F08	0.33	0.12	0.23	0.35	0.69		
B2-F09	0.09	0.08	0.21	0.87	2.30		2.66
B2-F12	0.06	0.08	0.08	1.21	1.29		
B2-G01	0.07	0.09	0.10	1.24	1.38		
B2-G03	0.07	0.09	0.08	1.30	1.19		
B2-G04	0.08	0.09	0.13	1.03	1.59		1.55
B2-G05	0.11	0.11	0.22	0.99	1.91		1.94
B2-G06	0.09	0.15	0.47	1.73	5.30		3.06
B2-G07	0.11	0.09	0.12	0.87	1.14		
B2-G08	0.06	0.07	0.11	1.16	1.71		1.47
B2-G09	0.09	0.10	0.12	1.11	1.35		
B2-G10	0.08	0.09	0.09	1.13	1.14		
B2-G11	0.11	0.09	0.12	0.79	1.06		
B2-G12	0.07	0.08	0.23	1.21	3.23		2.68
B2-H01	0.06	0.08	0.08	1.35	1.22		
B2-H02	0.07	0.13	0.55	1.79	7.63		4.26
B2-H03	0.07	0.08	0.10	1.03	1.34		
B2-H04	0.08	0.09	0.07	1.15	0.87		
B2-H05	0.08	0.10	0.08	1.21	0.97		
B2-H07	0.11	0.09	0.21	0.78	1.87		2.38
B2-H09	0.08	0.07	0.09	0.93	1.19		
B2-H10	0.08	0.08	0.14	1.06	1.73		1.63
B2-H11	0.07	0.09	0.08	1.30	1.11		
B2-H12	0.06	0.11	0.34	1.88	5.68		3.02
G1-A01	0.07	0.10	0.30	1.49	4.28		2.87
G1-A02	0.07	0.06	0.08	0.97	1.18		
G1-A03	0.06	0.09	0.30	1.49	4.88		3.27
G1-A04	0.06	0.07	0.07	1.13	1.11		
G1-A05	0.07	0.07	0.07	1.00	1.10		
G1-A07	0.07	0.08	0.17	1.25	2.67		2.13

Location	-	.	:	./-	:/-	./:	:/.
G1-A08	0.06	0.09	0.27	1.42	4.31		3.04
G1-A10	0.07	0.07	0.07	1.05	1.10		
G1-A11	0.07	0.07	0.08	1.00	1.08		
G1-A12	0.07	0.07	0.08	1.07	1.23		
G1-B01	0.06	0.07	0.08	1.11	1.22		
G1-B02	0.07	0.06	0.07	0.92	1.03		
G1-B04	0.06	0.06	0.07	1.02	1.19		
G1-B05	0.06	0.07	0.08	1.06	1.21		
G1-B07	0.06	0.06	0.07	1.01	1.16		
G1-B08	0.08	0.07	0.07	0.88	0.87		
G1-B09	0.06	0.06	0.08	1.05	1.29		
G1-B10	0.06	0.07	0.07	1.01	1.13		
G1-B11	0.06	0.06	0.07	1.03	1.11		
G1-B12	0.07	0.07	0.07	0.99	1.06		
G1-C01	0.08	0.07	0.08	0.95	1.00		
G1-C02	0.07	0.07	0.07	1.00	1.07		
G1-C03	0.06	0.07	0.07	1.05	1.10		
G1-C04	0.07	0.07	0.08	0.97	1.15		
G1-C05	0.07	0.07	0.07	1.02	1.10		
G1-C06	0.07	0.06	0.07	0.97	1.06		
G1-C07	0.06	0.06	0.07	0.97	1.14		
G1-C08	0.06	0.06	0.11	1.05	1.73		1.65
G1-C09	0.07	0.06	0.08	0.92	1.11		
G1-C10	0.07	0.07	0.07	0.98	1.09		
G1-C11	0.09	0.08	0.08	0.94	0.92		
G1-C12	0.06	0.07	0.08	1.07	1.26		
G1-D01	0.07	0.06	0.08	0.90	1.06		
G1-D02	0.06	0.06	0.06	0.99	1.00		
G1-D03	0.06	0.06	0.08	0.92	1.24		
G1-D06	0.06	0.06	0.07	1.01	1.04		
G1-D07	0.07	0.06	0.08	0.97	1.18		
G1-D08	0.06	0.07	0.07	1.12	1.10		
G1-D09	0.07	0.06	0.08	0.92	1.24		
G1-D10	0.07	0.06	0.06	1.00	0.98		
G1-D11	0.06	0.06	0.08	0.97	1.26		
G1-E01	0.07	0.13	0.36	2.00	5.52		2.77
G1-E02	0.06	0.15	0.38	2.45	6.23		2.54
G1-E04	0.07	0.07	0.08	0.96	1.25		
G1-E05	0.06	0.09	0.31	1.50	5.05		3.36
G1-E06	0.06	0.06	0.06	0.95	0.96		
G1-E07	0.06	0.06	0.07	1.06	1.21		
G1-E08	0.07	0.09	0.30	1.40	4.54		3.25
G1-E09	0.06	0.23	0.59	3.57	9.28		2.60
G1-E10	0.07	0.06	0.08	0.86	1.01		
G1-E11	0.06	0.07	0.08	1.01	1.17		
G1-E12	0.06	0.07	0.06	1.19	1.05		
G1-F01	0.07	0.07	0.08	1.04	1.10		
G1-F03	0.06	0.09	0.26	1.44	4.06		2.81
G1-F04	0.06	0.16	0.51	2.53	8.11		3.20
G1-F05	0.06	0.07	0.19	1.12	3.12		2.78
G1-F06	0.06	0.07	0.07	1.04	1.04		
G1-F07	0.06	0.07	0.17	1.25	2.92		2.34
G1-F09	0.06	0.06	0.07	1.05	1.25		
G1-F10	0.07	0.06	0.07	0.98	1.09		
G1-F11	0.06	0.07	0.08	1.10	1.24		
G1-F12	0.06	0.07	0.08	1.13	1.23		
G1-G02	0.07	0.07	0.08	1.06	1.15		
G1-G04	0.07	0.06	0.08	0.95	1.19		
G1-G05	0.06	0.06	0.07	1.01	1.12		
G1-G06	0.06	0.06	0.07	1.00	1.19		
G1-G07	0.06	0.07	0.06	1.11	0.98		
G1-G08	0.06	0.06	0.07	1.06	1.18		
G1-G10	0.06	0.06	0.07	1.02	1.18		
G1-G12	0.06	0.06	0.07	1.03	1.22		
G1-H01	0.06	0.06	0.07	0.98	1.17		
G1-H03	0.06	0.06	0.07	0.97	1.13		
G1-H04	0.06	0.06	0.07	1.00	1.17		
G1-H05	0.06	0.06	0.07	1.01	1.14		
G1-H06	0.06	0.06	0.06	1.01	1.06		
G1-H07	0.07	0.07	0.07	1.01	0.93		
G1-H08	0.05	0.06	0.07	1.04	1.33		
G1-H09	0.06	0.06	0.06	1.11	1.03		
G1-H10	0.07	0.06	0.07	0.95	1.05		
G1-H11	0.06	0.06	0.07	1.02	1.25		
G2-A01	0.10	0.12	0.42	1.30	4.46		3.42
G2-A04	0.08	0.11	0.36	1.39	4.72		3.40
G2-A05	0.07	0.14	0.43	2.03	6.13		3.03
G2-A06	0.09	0.43	0.82	5.05	9.52		1.88
G2-A07	0.08	0.10	0.28	1.19	3.38		2.84
G2-A08	0.07	0.11	0.44	1.56	6.22		3.98

Table A.12 (Continued)

Location	-	.	:	./-	:/-	./:	:/.
G2-A10	0.07	0.10	0.30	1.42	4.10		2.89
G2-A11	0.07	0.10	0.29	1.43	3.99		2.79
G2-A12	0.07	0.08	0.08	1.10	1.14		
G2-B01	0.08	0.15	0.46	1.88	5.95		3.17
G2-B02	0.08	0.12	0.11	1.43	1.25		
G2-B03	0.08	0.09	0.09	1.06	1.05		
G2-B04	0.09	0.11	0.37	1.23	4.26		3.46
G2-B06	0.09	0.09	0.11	0.96	1.17		
G2-B07	0.08	0.11	0.36	1.32	4.41		3.33
G2-B08	0.08	0.11	0.42	1.28	5.03		3.91
G2-B09	0.07	0.09	0.33	1.41	4.99		3.53
G2-B10	0.07	0.10	0.32	1.32	4.37		3.30
G2-B11	0.07	0.17	0.55	2.26	7.59		3.35
G2-B12	0.07	0.10	0.36	1.38	4.94		3.59
G2-C02	0.10	0.12	0.36	1.23	3.80		3.09
G2-C03	0.10	0.14	0.36	1.42	3.62		2.55
G2-C04	0.18	0.11	0.26	0.59	1.42		
G2-C06	0.09	0.11	0.34	1.29	3.94		3.04
G2-C07	0.08	0.09	0.13	1.05	1.58		1.51
G2-C08	0.15	0.11	0.44	0.74	2.89		3.88
G2-C09	0.09	0.13	0.29	1.51	3.41		2.25
G2-C10	0.07	0.08	0.07	1.12	0.95		
G2-C11	0.07	0.10	0.37	1.38	5.10		3.71
G2-C12	0.08	0.08	0.07	0.97	0.94		
G2-D01	0.11	0.07	0.08	0.68	0.75		
G2-D03	0.09	0.10	0.22	1.07	2.41		2.25
G2-D06	0.10	0.15	0.50	1.50	5.06		3.38
G2-D07	0.09	0.09	0.18	1.08	2.15		1.99
G2-D08	0.09	0.08	0.10	0.86	1.10		
G2-D10	0.07	0.11	0.41	1.57	5.67		3.61
G2-D11	0.07	0.09	0.29	1.23	3.98		3.24
G2-D12	0.07	0.10	0.28	1.41	3.88		2.76
G2-E02	0.08	0.12	0.36	1.49	4.37		2.94
G2-E03	0.07	0.14	0.51	2.03	7.72		3.79
G2-E05	0.11	0.13	0.44	1.14	3.89		3.43
G2-E06	0.07	0.09	0.13	1.28	1.81		1.42
G2-E07	0.08	0.10	0.37	1.24	4.70		3.79
G2-E08	0.09	0.12	0.44	1.33	4.90		3.70
G2-E10	0.07	0.09	0.30	1.31	4.22		3.23
G2-E11	0.07	0.11	0.36	1.44	4.86		3.38
G2-E12	0.07	0.11	0.37	1.50	5.21		3.47
G2-F01	0.12	0.19	0.33	1.60	2.80		1.75
G2-F02	0.09	0.11	0.31	1.28	3.58		2.79
G2-F03	0.08	0.11	0.34	1.33	4.26		3.20
G2-F05	0.11	0.10	0.12	0.97	1.11		
G2-F06	0.10	0.11	0.33	1.09	3.17		2.93
G2-F08	0.10	0.19	0.64	1.99	6.69		3.36
G2-F09	0.06	0.10	0.29	1.48	4.49		3.04
G2-F10	0.07	0.11	0.41	1.50	5.66		3.78
G2-F11	0.07	0.10	0.31	1.41	4.64		3.28
G2-F12	0.08	0.12	0.39	1.46	4.70		3.21
G2-G01	0.07	0.12	0.44	1.66	6.00		3.61
G2-G03	0.09	0.08	0.07	0.90	0.79		
G2-G04	0.08	0.10	0.38	1.24	4.69		3.79
G2-G05	0.07	0.21	0.65	2.85	8.66		3.04
G2-G06	0.08	0.09	0.11	1.02	1.32		
G2-G07	0.10	0.21	0.65	2.16	6.78		3.13
G2-G09	0.07	0.09	0.34	1.30	4.76		3.66
G2-G10	0.07	0.08	0.21	1.12	2.96		2.65
G2-G11	0.07	0.10	0.37	1.44	5.28		3.68
G2-G12	0.07	0.09	0.30	1.28	4.15		3.25
G2-H01	0.11	0.10	0.33	0.89	2.84		3.19
G2-H02	0.09	0.11	0.36	1.21	4.12		3.39
G2-H03	0.08	0.08	0.08	1.00	1.00		
G2-H04	0.07	0.08	0.10	1.18	1.39		
G2-H05	0.08	0.13	0.30	1.60	3.69		2.30
G2-H06	0.08	0.14	0.47	1.66	5.68		3.43
G2-H07	0.08	0.12	0.49	1.54	6.45		4.19
G2-H08	0.07	0.13	0.40	1.78	5.69		3.19
G2-H09	0.07	0.07	0.07	1.07	1.04		
G2-H10	0.07	0.09	0.25	1.32	3.83		2.89
G2-H11	0.07	0.09	0.31	1.31	4.72		3.59
R1-A01	0.06	0.06	0.06	0.97	0.93		
R1-A02	0.06	0.07	0.06	1.10	1.05		
R1-A03	0.06	0.06	0.06	1.05	1.07		
R1-A04	0.06	0.06	0.06	1.04	0.99		
R1-A05	0.07	0.07	0.06	1.06	0.92		
R1-A06	0.06	0.06	0.06	0.98	0.98		
R1-A07	0.06	0.06	0.06	1.00	0.98		
R1-A09	0.21	0.22	0.20	1.05	0.96		

Location	-	.	:	./-	:/-	./:	:/.
R1-A10	0.07	0.08	0.06	1.09	0.90		
R1-A11	0.07	0.06	0.07	0.87	0.92		
R1-A12	0.06	0.07	0.06	1.06	0.99		
R1-B01	0.06	0.06	0.06	1.06	1.03		
R1-B02	0.07	0.07	0.08	0.99	1.12		
R1-B03	0.19	0.14	0.13	0.75	0.68		
R1-B04	0.06	0.06	0.06	0.94	0.99		
R1-B05	0.06	0.06	0.06	0.97	1.00		
R1-B07	0.06	0.06	0.08	1.02	1.37		
R1-B08	0.06	0.06	0.06	0.97	1.03		
R1-B09	0.07	0.08	0.08	1.17	1.08		
R1-B10	0.06	0.07	0.06	1.03	0.93		
R1-B12	0.06	0.06	0.06	0.92	1.01		
R1-C02	0.06	0.06	0.07	1.04	1.11		
R1-C03	0.06	0.07	0.07	1.05	1.04		
R1-C04	0.06	0.06	0.06	1.02	1.03		
R1-C05	0.07	0.06	0.06	0.86	0.88		
R1-C06	0.06	0.06	0.06	1.05	1.03		
R1-C07	0.07	0.06	0.06	0.93	0.95		
R1-C08	0.06	0.07	0.06	1.07	1.02		
R1-C09	0.06	0.07	0.07	1.06	1.14		
R1-C10	0.07	0.06	0.07	0.99	1.01		
R1-C11	0.20	0.06	0.06	0.29	0.31		
R1-C12	0.06	0.07	0.07	1.10	1.10		
R1-D01	0.06	0.06	0.06	0.94	0.96		
R1-D02	0.06	0.06	0.06	1.00	0.96		
R1-D03	0.07	0.06	0.06	0.95	1.00		
R1-D04	0.06	0.06	0.06	1.07	1.06		
R1-D05	0.06	0.06	0.06	0.98	1.00		
R1-D06	0.06	0.06	0.06	0.98	0.94		
R1-D07	0.08	0.07	0.07	0.88	0.97		
R1-D08	0.06	0.06	0.06	1.08	1.01		
R1-D09	0.07	0.07	0.08	1.01	1.19		
R1-D10	0.06	0.07	0.07	1.01	1.03		
R1-D11	0.07	0.07	0.08	0.97	1.03		
R1-D12	0.07	0.07	0.07	1.02	0.97		
R1-E01	0.06	0.06	0.06	1.05	1.02		
R1-E02	0.06	0.06	0.07	0.99	1.08		
R1-E03	0.06	0.06	0.06	0.97	1.08		
R1-E04	0.06	0.07	0.06	1.11	1.00		
R1-E05	0.06	0.06	0.06	1.02	1.04		
R1-E07	0.06	0.06	0.07	1.00	1.06		
R1-E08	0.06	0.06	0.06	1.04	1.03		
R1-E09	0.07	0.07	0.07	1.03	1.03		
R1-E10	0.08	0.06	0.06	0.76	0.74		
R1-E11	0.06	0.08	0.07	1.16	1.12		
R1-E12	0.16	0.16	0.16	1.00	1.01		
R1-F01	0.06	0.07	0.06	1.13	1.11		
R1-F02	0.06	0.06	0.07	1.05	1.06		
R1-F03	0.06	0.06	0.06	1.04	1.00		
R1-F04	0.06	0.06	0.06	1.06	0.99		
R1-F05	0.07	0.06	0.07	0.84	0.92		
R1-F07	0.06	0.06	0.06	0.98	1.05		
R1-F08	0.06	0.06	0.06	1.02	1.01		
R1-F09	0.08	0.08	0.09	0.94	1.09		
R1-F10	0.07	0.07	0.07	0.99	1.03		
R1-F11	0.07	0.07	0.07	0.91	0.94		
R1-F12	0.06	0.06	0.07	0.99	1.20		
R1-G01	0.06	0.06	0.06	1.03	1.06		
R1-G02	0.06	0.07	0.07	1.24	1.13		
R1-G03	0.06	0.06	0.06	1.06	1.00		
R1-G04	0.06	0.06	0.07	1.06	1.15		
R1-G05	0.06	0.06	0.06	0.94	0.96		
R1-G06	0.06	0.06	0.06	1.00	1.00		
R1-G07	0.06	0.05	0.06	0.97	1.00		
R1-G08	0.08	0.07	0.08	0.85	1.00		
R1-G09	0.06	0.07	0.06	1.18	1.13		
R1-G10	0.19	0.19	0.19	0.96	0.98		
R1-G11	0.07	0.07	0.07	0.95	0.94		
R1-G12	0.06	0.07	0.07	1.11	1.07		
R1-H01	0.06	0.07	0.06	1.17	1.15		
R1-H02	0.06	0.06	0.06	0.94	0.95		
R1-H03	0.07	0.06	0.06	0.91	0.88		
R1-H04	0.06	0.06	0.06	1.06	1.02		
R1-H05	0.06	0.06	0.06	0.96	1.03		
R1-H06	0.05	0.06	0.05	1.12	1.01		
R1-H07	0.05	0.06	0.06	1.05	1.04		
R1-H08	0.05	0.06	0.07	1.06	1.26		
R1-H09	0.06	0.07	0.06	1.07	1.02		
R1-H10	0.07	0.08	0.07	1.03	0.98		

Location	-	.	:	./-	:/-	./:	:/.
R1-H11	0.07	0.06	0.06	0.98	0.93		
R1-H12	0.24	0.25	0.24	1.04	0.99		
R2-A03	0.07	0.09	0.13	1.35	1.96		1.46
R2-A06	0.07	0.08	0.13	1.05	1.79		1.70
R2-A07	0.08	0.07	0.07	0.98	0.89		
R2-A08	0.07	0.08	0.19	1.13	2.85		2.53
R2-A09	0.11	0.11	0.18	1.03	1.75		1.70
R2-A11	0.07	0.08	0.07	1.20	1.07		
R2-A12	0.08	0.07	0.13	0.88	1.61		1.83
R2-B01	0.07	0.07	0.07	1.00	1.04		
R2-B03	0.07	0.07	0.13	1.00	1.90		1.90
R2-B04	0.07	0.07	0.07	1.02	1.00		
R2-B05	0.07	0.09	0.13	1.29	1.86		1.44
R2-B06	0.07	0.09	0.10	1.30	1.43		
R2-B07	0.09	0.08	0.12	0.87	1.42		
R2-B08	0.08	0.08	0.15	0.94	1.90		2.03
R2-B10	0.10	0.11	0.12	1.11	1.18		
R2-B11	0.09	0.09	0.10	0.98	1.08		
R2-B12	0.08	0.08	0.17	1.02	2.11		2.07
R2-C01	0.07	0.07	0.13	1.01	1.79		1.77
R2-C02	0.08	0.08	0.13	0.97	1.56		1.60
R2-C03	0.07	0.07	0.12	1.13	1.81		1.60
R2-C04	0.07	0.07	0.13	1.03	1.91		1.86
R2-C05	0.08	0.09	0.16	1.20	2.15		1.80
R2-C06	0.09	0.07	0.08	0.81	0.86		
R2-C07	0.07	0.08	0.14	1.23	2.04		1.66
R2-C08	0.36	0.32	0.32	0.90	0.91		
R2-C10	0.10	0.10	0.16	1.02	1.67		1.64
R2-C11	0.08	0.07	0.09	0.87	1.11		
R2-C12	0.08	0.08	0.15	1.00	1.81		1.80
R2-D01	0.07	0.07	0.07	1.04	1.08		
R2-D03	0.07	0.07	0.13	1.09	2.02		1.85
R2-D04	0.08	0.07	0.07	0.87	0.86		
R2-D05	0.08	0.10	0.08	1.15	0.94		
R2-D06	0.09	0.07	0.09	0.81	0.93		
R2-D07	0.07	0.08	0.14	1.19	2.02		1.70
R2-D08	0.07	0.08	0.09	1.14	1.16		
R2-D09	0.11	0.07	0.13	0.64	1.21		
R2-D10	0.34	0.09	0.08	0.27	0.24		
R2-D11	0.08	0.09	0.15	1.04	1.87		1.79
R2-D12	0.07	0.08	0.15	1.11	1.98		1.78
R2-E01	0.07	0.07	0.13	1.11	1.95		1.76
R2-E02	0.07	0.07	0.12	1.13	1.88		1.66
R2-E03	0.07	0.07	0.13	1.13	1.95		1.73
R2-E04	0.08	0.08	0.07	1.04	0.84		
R2-E05	0.08	0.10	0.15	1.19	1.82		1.53
R2-E06	0.10	0.09	0.08	0.98	0.89		
R2-E07	0.13	0.12	0.11	0.91	0.80		
R2-E08	0.07	0.07	0.11	0.98	1.50		1.53
R2-E09	0.11	0.26	0.16	2.27	1.44	1.57	
R2-E10	0.08	0.09	0.13	1.10	1.68		1.54
R2-E11	0.09	0.09	0.15	0.95	1.56		1.64
R2-E12	0.07	0.08	0.16	1.13	2.33		2.06
R2-F01	0.09	0.09	0.08	1.00	0.94		
R2-F02	0.07	0.08	0.12	1.12	1.75		1.56
R2-F03	0.07	0.07	0.08	0.94	1.10		
R2-F04	0.07	0.07	0.07	1.05	1.09		
R2-F07	0.08	0.08	0.14	1.06	1.81		1.70
R2-F08	0.08	0.08	0.09	1.02	1.10		
R2-F09	0.10	0.07	0.08	0.71	0.76		
R2-F10	0.10	0.09	0.18	0.84	1.71		2.05
R2-F11	0.08	0.09	0.15	1.20	1.90		1.58
R2-F12	0.08	0.08	0.10	1.04	1.27		
R2-G02	0.35	0.32	0.27	0.91	0.76		
R2-G03	0.61	0.42	0.36	0.69	0.58		
R2-G04	0.07	0.07	0.14	1.11	2.09		1.88
R2-G06	0.09	0.10	0.08	1.12	0.99		
R2-G07	0.08	0.09	0.07	1.05	0.86		
R2-G08	0.07	0.09	0.19	1.17	2.67		2.28
R2-G09	0.10	0.11	0.21	1.06	2.08		1.97
R2-G10	0.08	0.08	0.17	0.98	2.05		2.10
R2-G12	0.07	0.08	0.15	1.11	2.26		2.04
R2-H01	0.20	0.30	0.23	1.48	1.13		
R2-H03	0.07	0.07	0.11	0.98	1.61		1.64
R2-H04	0.06	0.06	0.15	1.13	2.61		2.31
R2-H05	0.07	0.09	0.12	1.23	1.76		1.43
R2-H06	0.07	0.07	0.14	0.98	1.95		2.00
R2-H07	0.07	0.07	0.07	0.95	1.02		
R2-H10	0.12	0.08	0.14	0.66	1.15		
R2-H12	0.07	0.08	0.14	1.15	2.01		1.75

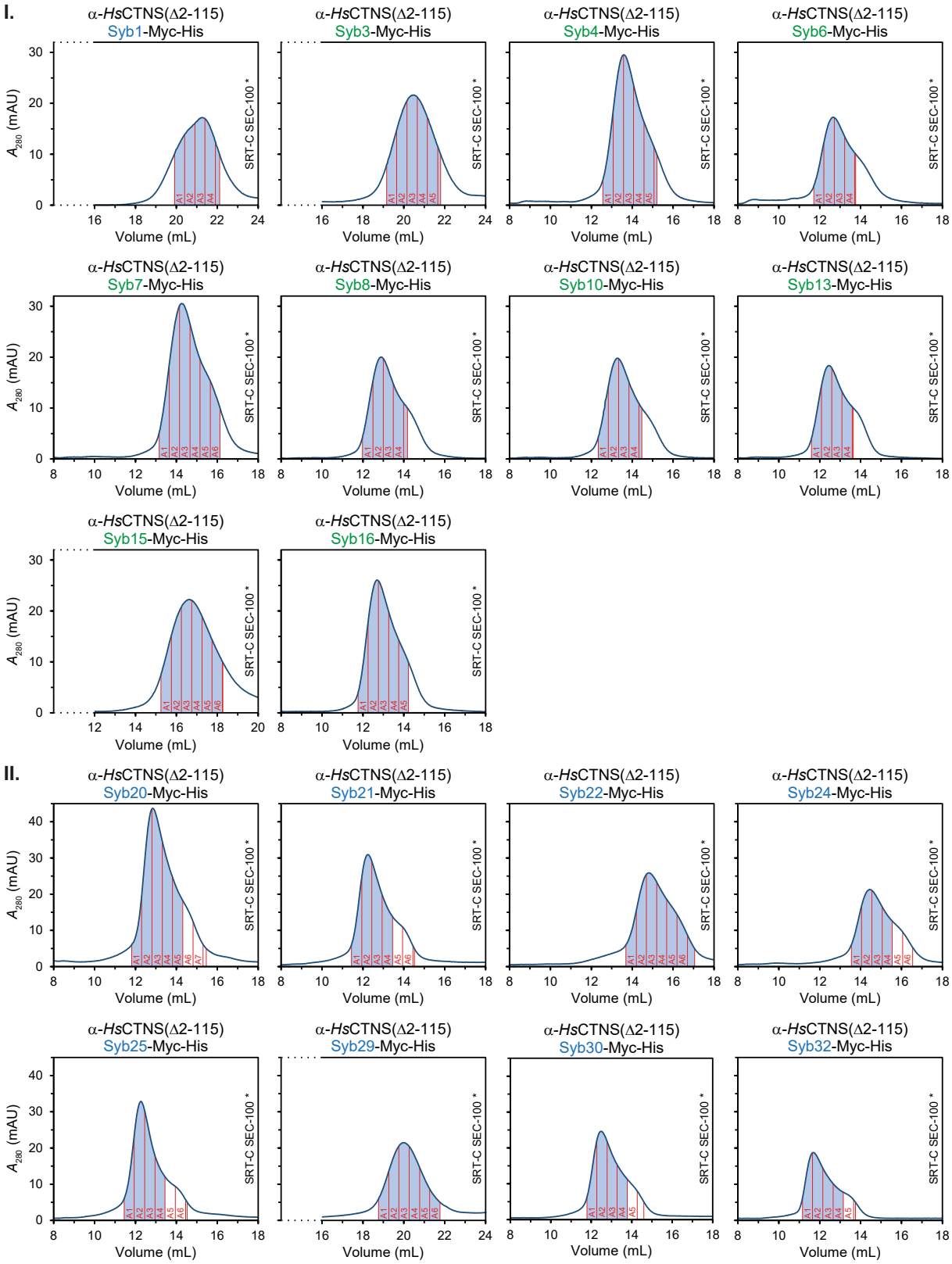
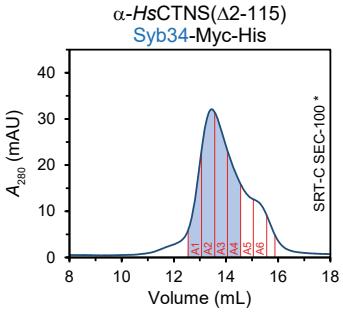


Fig. A.9 – SEC Traces of Myc-His-Tagged α-*Hs*CTNS(Δ2–115) Sybodies.
Small-scale expression and purification of sybodies from 50 mL cultures, performed in three batches (**I.**–**III.**). SEC was performed at RT (*highlights temperature) using a Sepax SRT-C SEC-100 column connected to an ÄKTA go chromatography system (Cytiva). Samples were injected automatically using an ALIAS autosampler (Cytiva). Y-axes are fixed per batch to simplify the comparison of individual traces. Pooled and combined fractions are shaded in blue. Sybs are colour-coded based on their origin from the concave (G), loop (B), or convex (R) library. This figure continues on the next page.

II. (continued)



III.

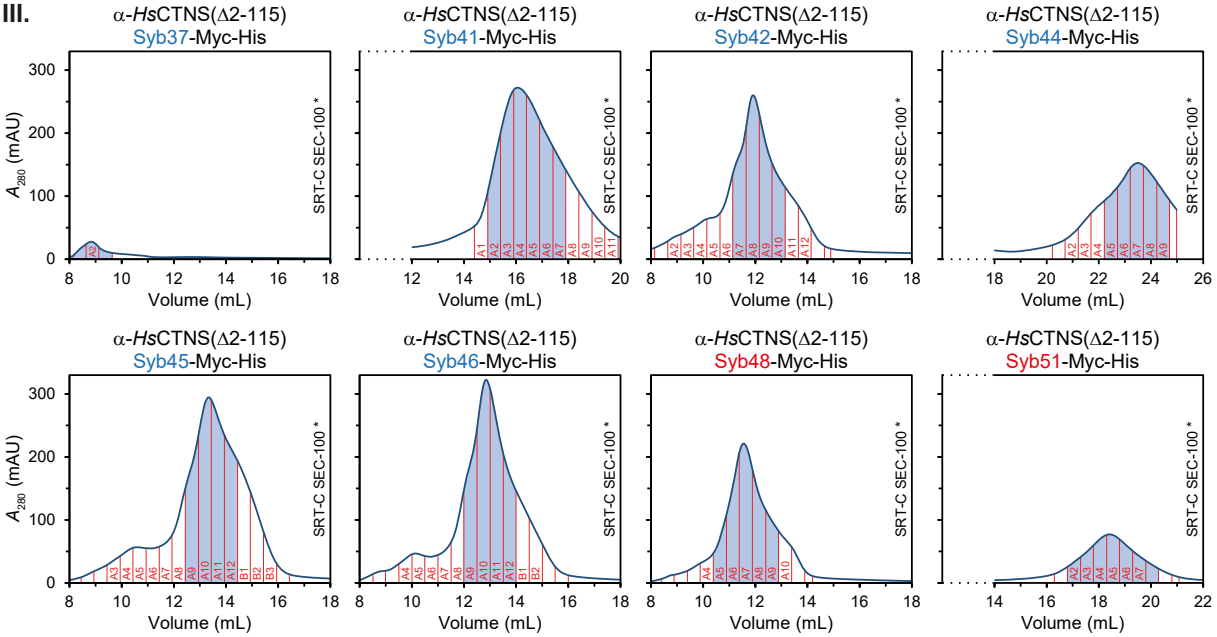


Fig. A.9 (Continued)

Table A.13 – *Hs*CTNS(Δ2–115)-Sybody Binding Analysis Overview.

Syb No	Yield (µg)	Ret. Vol. (mL)	Pulldown Dimer	Pulldown Monomer	BN-PAGE Dimer	BN-PAGE Monomer	App. K_D (nM) Dimer	k_{off} (min^{-1}) Dimer
1	39	21.3	✓✓✓	×				
2	×	×						
3	76	20.4	×					
4	90	13.6	×					
6	68	12.7	×					
7	87	14.3	×					
8	55	12.9	×					
10	55	13.3	✓	✓	(✓)	(✓)	N/A	N/A
13	48	12.4	✓	✓	(✓)	(✓)		
15	90	16.6	×	×	×	×	N/A	N/A
16	101	12.7	×					
17	×	×						
18	×	×						
20	146	12.8	✓✓✓✓✓	✓✓✓	✓✓	✓✓✓	94.6	0.46
21	78	12.2	✓✓✓✓	×	✓	✓✓		
22	83	14.8	✓					
24	54	14.4	✓✓					
25	42	12.3	✓✓✓✓✓	✓✓	✓✓✓✓✓	✓✓✓✓✓	457	0.32
27	×	×						
29	41	20.0	✓✓					
30	65	12.4	✓✓✓✓✓	✓✓✓	✓✓✓	✓✓✓	78.2	0.25
32	60	11.7	✓✓✓✓✓	✓✓	✓✓	✓✓✓	210	0.57
34	74	13.4	✓✓✓✓	✓	×	✓		
35	×	×						
37	11	8.9	×					
39	×	×						
41	1,242	16.1	✓✓✓✓✓	✓✓✓	×	×	281	0.49
42	748	11.9	✓✓✓					
44	511	23.5	✓✓✓					
45	1,350	13.3	✓✓✓✓	✓✓✓	×	✓		
46	1,219	12.9	×					
48	822	11.6	×					
49	×	×						
50	×	×						
51	321	18.4	✓✓✓	✓	✓✓	✓✓✓	N/A	N/A
52	×	×						

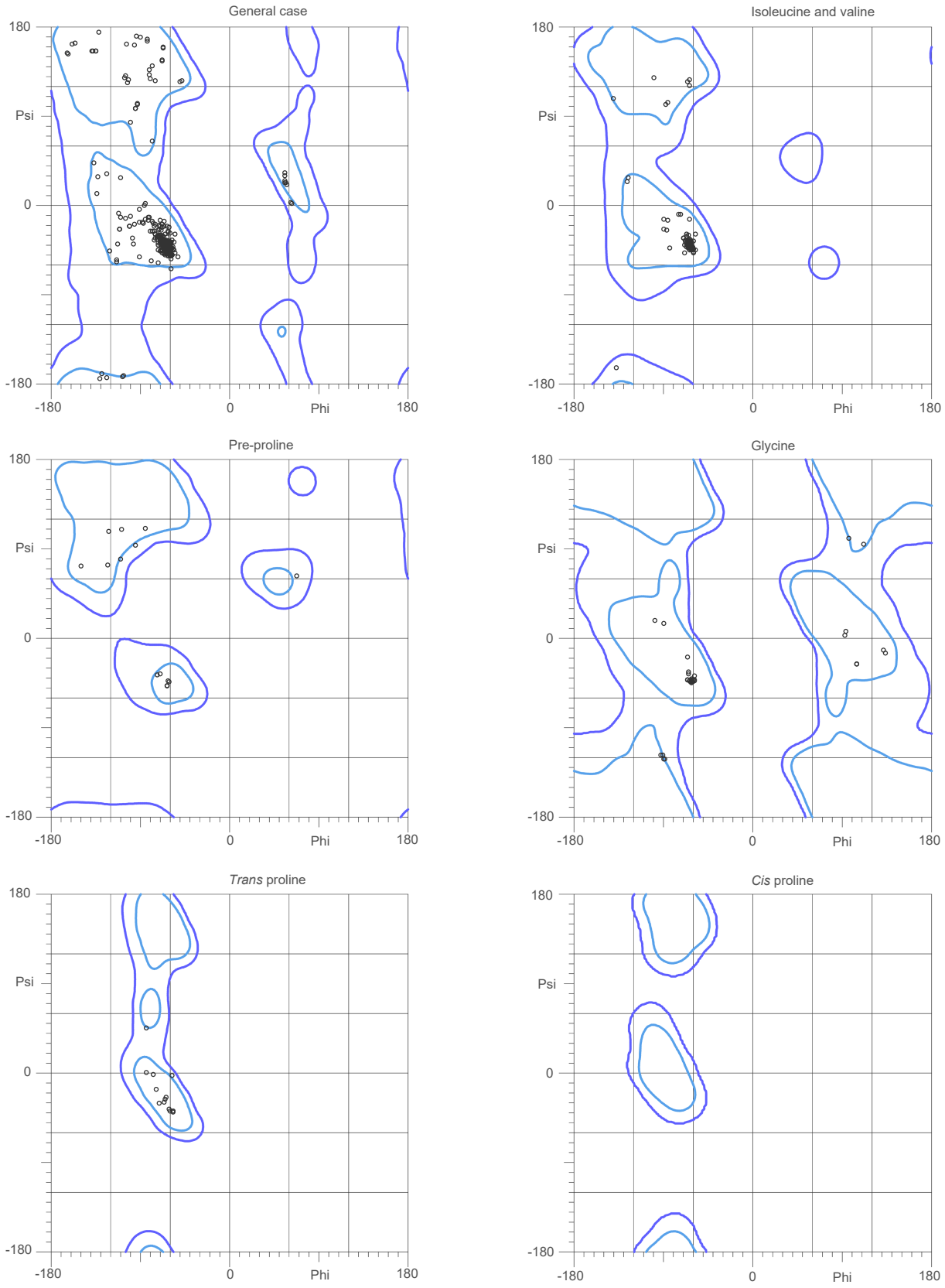


Fig. A.10 – Ramachandran Plots: *Hs*CTNS(Δ2–115)-Syb30-Myc-His Model.
Generated by MolProbity (Lovell *et al.*, 2003). The protein content of this model consists of chains A and B: *Hs*CTNS(Δ2–115), as well as chains C and D: Syb30-Myc-His.

Functional Analysis of CTNS

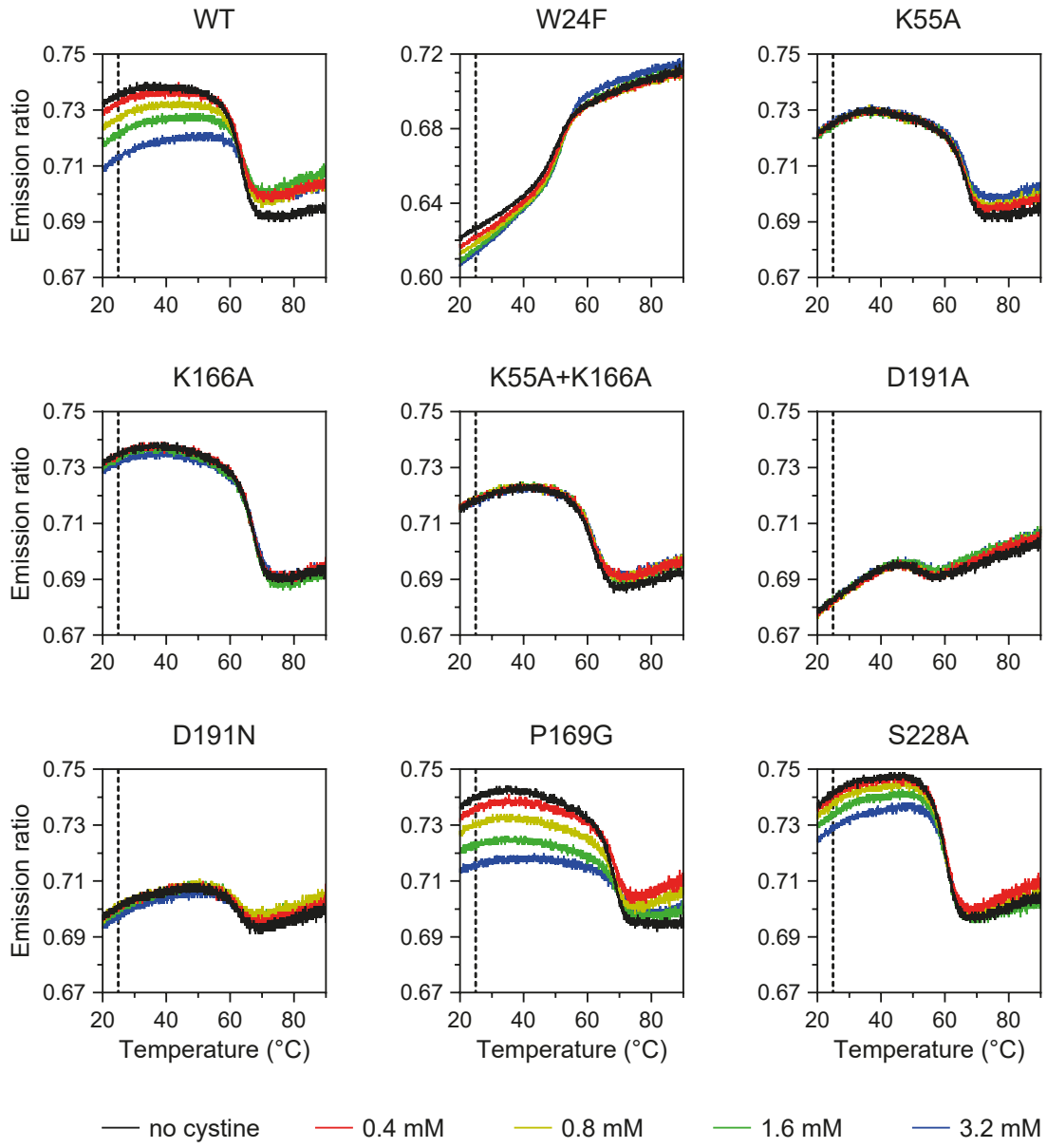


Fig. A.11 – *At*CTNS Mutants Cystine Binding NanoDSF Assay Data.
Typical nanoDSF profiles for WT and variant forms of *At*CTNS showing the tryptophan fluorescence quench elicited by the L-cystine ligand or the lack thereof.

Publications

At the time of submission, the research presented in this thesis resulted in the following publication:

Löbel, M., Salphati, S. P., El Omari, K., Wagner, A., Tucker, S. J., Parker, J. L. & Newstead, S. (2022). Structural basis for proton coupled cystine transport by cystinosin. *Nat. Commun.* **13**, 4845.

The full-text publication, including supplementary information, is presented hereafter.

nature communications

Article

https://doi.org/10.1038/s41467-022-32589-2

Structural basis for proton coupled cystine transport by cystinosin

Received: 20 April 2022

Accepted: 8 August 2022

Published online: 17 August 2022

Check for updates

Mark Löbel[1], Sacha P. Salphati[1,2], Kamel El Omari[3], Armin Wagner[3], Stephen J. Tucker[2,4], Joanne L. Parker[1] ✉ & Simon Newstead[1,4] ✉

Amino acid transporters play a key role controlling the flow of nutrients across the lysosomal membrane and regulating metabolism in the cell. Mutations in the gene encoding the transporter cystinosin result in cystinosis, an autosomal recessive metabolic disorder characterised by the accumulation of cystine crystals in the lysosome. Cystinosin is a member of the PQ-loop family of solute carrier (SLC) transporters and uses the proton gradient to drive cystine export into the cytoplasm. However, the molecular basis for cystinosin function remains elusive, hampering efforts to develop novel treatments for cystinosis and understand the mechanisms of ion driven transport in the PQ-loop family. To address these questions, we present the crystal structures of cystinosin from *Arabidopsis thaliana* in both apo and cystine bound states. Using a combination of in vitro and in vivo based assays, we establish a mechanism for cystine recognition and proton coupled transport. Mutational mapping and functional characterisation of human cystinosin further provide a framework for understanding the molecular impact of disease-causing mutations.

The lysosome is the main site of macromolecule degradation in mammalian cells and functions as a major signalling centre, integrating multiple inputs to regulate cellular metabolism via mTOR[1–3]. Central to lysosomal function and cell homoeostasis is the transport of amino acids across the lysosomal membrane via solute carrier (SLC) systems[4–6]. Solute carrier proteins are secondary active transporters that control the flow of nutrients between and within cells, often linked to ion gradients that function to concentrate nutrients and regulate levels in response to cellular energy demands[7–9]. Dysregulation of SLC function in the lysosome is linked to several lysosomal storage disorders, which are inherited metabolic diseases resulting from defective lysosomal function[10]. Mutations in the gene encoding cystinosin, *CTNS*, result in cystinosis, an autosomal recessive metabolic disease characterised by abnormal accumulation of cystine, the oxidised form of cysteine, inside lysosomes[11,12]. The *CTNS* gene was identified in 1998 and to date more than 140 mutations have been linked to nephropathic cystinosis in patients around the world[11,13]. Cystinosis often leads to renal Fanconi syndrome in paediatric patients, characterised by reduced kidney function and impaired growth and bone development due to mineral depletion[14]. Less severe forms of the disease result in blindness due to cystine crystallisation in the cornea, referred to as ocular cystinosis[15]. Defective lysosomal cystine transport also contributes to acute cellular oxidative stress, mitochondrial dysfunction, trafficking defects, and proteolysis defects in the cell[13,16,17]. However, a mechanistic link between cystinosis symptoms in patients and mutations in the *CTNS* gene is not fully established at the molecular level, hampering efforts to develop new treatments[18].

Cystinosin is a proton coupled lysosomal cystine transporter[19], which functions as the primary route for the export of cystine out of the lysosome (Fig. 1a). The transporter comprises seven transmembrane helices (TMs) and uses the outwardly directed proton electrochemical gradient, generated by the V-ATPase, to drive cystine export from the lysosome in a 1:1 stoichiometry[20]. Cystinosin is also a member of the so-called PQ-loop family[21,22], members of which function as trafficking receptors[23], lipid flippases[24], amino acid transporters[25] and sugar transporters in plants and mammals[26]. Eukaryotic PQ-loop

[1]Department of Biochemistry, University of Oxford, Oxford, UK. [2]Clarendon Laboratory, Department of Physics, University of Oxford, Parks Road, Oxford, UK. [3]Diamond Light Source, Harwell Science and Innovation Campus, Didcot, UK. [4]Kavli Institute for Nanoscience Discovery, University of Oxford, Oxford, UK. ✉e-mail: Joanne.parker@bioch.ox.ac.uk; Simon.newstead@bioch.ox.ac.uk

Article https://doi.org/10.1038/s41467-022-32589-2

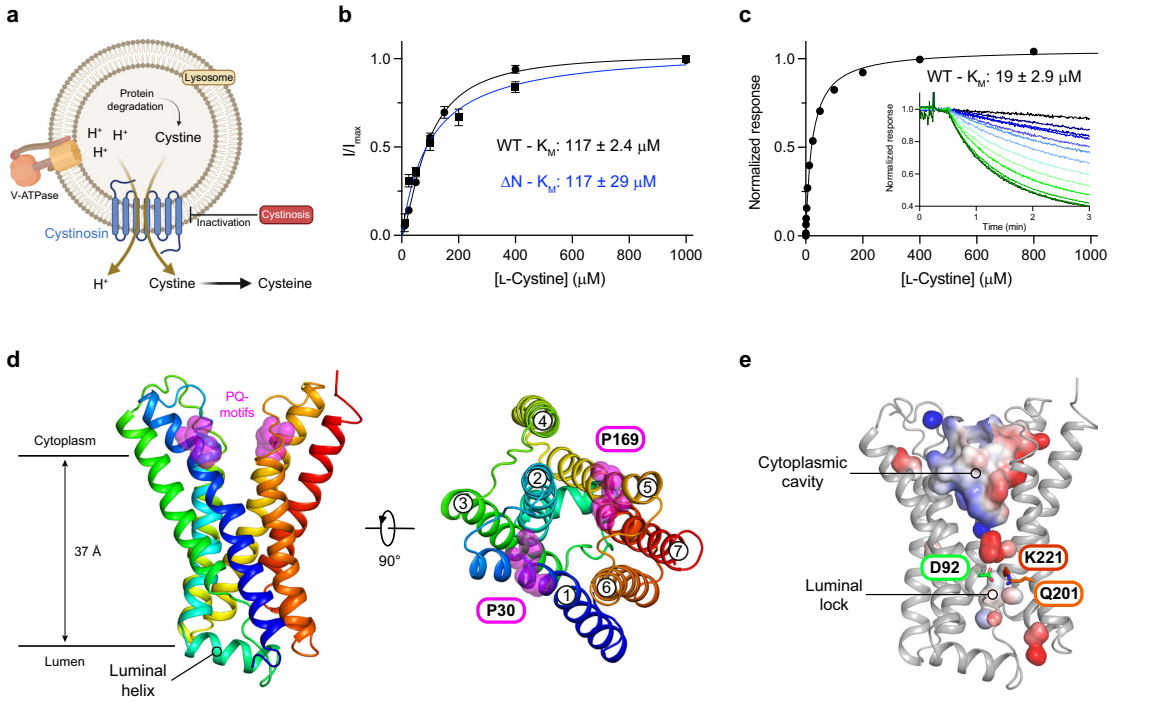


Fig. 1 | Functional characterisation and structure determination of cystinosin. **a** Cystinosin resides within the lysosomal membrane where it is responsible for the export of cystine from the lysosomal lumen. Transport is proton coupled, utilising the proton gradient across the membrane generated through the action of V-ATPase. Mutation of the CTNS gene, which encodes cystinosin, lead to cystinosis. **b** Representative K_M curve derived from TEVC for full length human cystinosin (WT – black) and a construct lacking N-terminal amino acids 2-115 (ΔN – blue). $n = 16$ independent experiments for every WT data point, $n = 10$ for ΔN, error bars SEM. K_M values were calculated from three biologically independent experiments, error SD. **c** Representative K_M analysis of plant cystinosin derived using a pyranine based in vitro assay. K_M was calculated from three independent experiments, error SD. Inset shows a typical set of raw data generated from the assay (lines are coloured according to cystine concentration dark green 800 via lighter colours to dark blue 0 µM). **d** Crystal structure of plant cystinosin with helices coloured blue to red from the N-terminus. Highlighted (magenta spheres) are the two PQ loop motifs which sit in the open cavity facing the cytoplasm in the conformation captured in the structure. **e** View of cystinosin highlighting the salt bridge network which forms the luminal lock and seals the binding site on the luminal side. Source data are provided as a Source Data file.

proteins contain two PQ-motifs, located on TM1 and TM5, which play important roles in facilitating the structural changes required for alternating access transport within these proteins[22,27].

Insights into the structural changes that occur in transporters belonging to the PQ-loop family were revealed through crystal structures of the bacterial PQ-loop transporters, which were captured in both the inward (cytoplasmic) and outward (extracellular) states of their transport cycles[27–29]. However, unlike eukaryotic PQ-loop family transporters, the bacterial homologues are obligate dimers, consisting of two 3-TM domains and function as facilitators, allowing sugar molecules to equilibrate across the membrane[30]. Although a plant homologue of the bacterial semi-SWEET transporters was also crystallised[31], revealing the architecture of the eukaryotic hepta-helical PQ-fold, we still have little insight into how PQ-loop transporters couple ligand binding and transport to secondary ion gradients. Previous work on human cystinosin, using two electrode voltage clamp (TEVC) assays, revealed the presence of several conserved aspartate residues, one of which undergoes reversible protonation during transport[20]. Cystine binding was proposed to alter the pK_a of this side chain, providing a mechanistic link for coupled transport. However, several key questions concerning the mechanism of cystinosin remain, including how cystine is recognised, how proton binding and release drive the full transport cycle and the rationalisation of patient mutations within the mechanism of transport.

To address these questions and understand the molecular basis for proton coupled cystine transport, we determined the structure of the plant cystinosin transporter from *Arabidopsis thaliana*. Using both Llama derived[32] and synthetic nanobodies[33], cystinosin was captured in both apo and ligand bound states, revealing a key role for conserved lysine side chains in cystine recognition. Combined with in vitro transport assays and TEVC comparisons on the human transporter we propose a mechanism for proton coupled cystine transport across the vacuolar and lysosomal membranes.

Results

Characterisation of *Arabidopsis thaliana* cystinosin

To identify homologues of cystinosin suitable for structural and biochemical analysis we screened several homologues from different organisms (Supplementary Fig. 1a). The homologue from *Arabidopsis thaliana* showed high expression levels in our yeast-based system[34] relative to other homologues and was therefore chosen for further analysis. Cystinosin from *A. thaliana*, however, does not contain the N-terminal lysosomal luminal domain found in human and other eukaryotic cystinosin proteins (Supplementary Fig. 1b), prompting us to determine the importance of this domain for transport function. Analysing the human transporter expressed in *Xenopus laevis* oocytes and measuring currents elicited by L-cystine under two electrode voltage clamp (TEVC), we found that the N-terminal domain of the human transporter does not impact the transport kinetics, with both having

Article https://doi.org/10.1038/s41467-022-32589-2

Table 1 | X-ray data collection and refinement statistics

		S-SAD data	Cystinosin apo (Syb39 & Nb4)	Cystinosin apo (Nb1 & Nb4)	Cystinosin cystine complex (Syb39)
Data collection	PDB	–	7ZK1	7ZKZ	7ZKW
	Space group	$P\,2_1\,2_1\,2$	$P\,2_1\,2_1\,2$	$C\,2$	$C\,2$
	Cell dimensions a, b, c (Å)	62.60, 319.49, 45.48	62.79, 319.96, 45.65	210.28, 77.46, 46.30	283.37, 64.07, 55.44
	Cell angles α, β, γ (°)	90, 90, 90	90, 90, 90	90, 93.25, 90	90, 99.95, 90
	Wavelength (Å)	2.7552	0.9999	0.9795	0.9762
	Resolution (Å) [a]	29.24–3.20 (3.28–3.20)	79.99–2.65 (2.69–2.65)	104.76–2.33 (2.37–2.33)	70.10–3.37 (3.52–3.37)
	CC1/2 (%) [a]	99.9 (67.1)	99.8 (35.7)	99.5 (42.7)	79.7 (10.0)
	R_{pim} (%) [a]	5 (99.2)	11.4 (227.6)	7.2 (91.7)	49.7 (120.5)
	I/σI [a]	13.7 (1.1)	5.9 (0.4)	7.9 (0.6)	2.0 (0.6)
	Completeness (%) [a]	99.7 (99.5)	100 (99.9)	100 (100)	99.7 (95.3)
	Multiplicity [a]	44.6 (38.6)	23.9 (13.3)	6.4 (6.7)	6.1 (5.9)
Refinement	Resolution (Å)		79.99–2.65	20.79–2.33	69.78–3.37
	Number of reflections		26170	31641	12426
	R_{work}/R_{free}		23.4/27.49	23.5/26.5	23.5/27.7
	F_o,F_c correlation		0.83	0.94	0.83
	Average B, all atoms (Å^2)		84.8	66.9	67.5
R.m.s deviations	Bond lengths (Å)		0.009	0.008	0.014
	Bong angles (°)		1.05	0.97	1.55
	Ramachandran statistics Favoured/outliers (%)		91.55/2.47	97.88/0.64	89.72/2.50
	Molprobity score		2.97	1.77	2.94

[a]Highest resolution shell shown in parentheses.

similar K_M values for cystine (117 ± 2.4 µM for WT compared to 117 ± 28.9 µM for the Δ2-115 variant) (Fig. 1b & Supplementary Fig. 2a). We then determined the K_M for cystine transport using a reconstituted system and employed pyranine to monitor proton movement[35]. These results establish *A. thaliana* cystinosin as a proton coupled cystine symporter with a K_M for cystine of 19 ± 2.9 µM, (Fig. 1c & Supplementary Fig. 2b) suggesting a slightly higher affinity for L-cystine than the human protein. Having established the plant transporter as a suitable functional homologue of human cystinosin, we sought to determine the structure using crystallisation in the lipidic cubic phase.

Structure of cystinosin

Single chain binders have been successfully used to obtain the structures of small, dynamic membrane proteins[36]. Using a similar strategy to our recent work on the KDEL receptor[23], we screened both a synthetic nanobody (sybody) library and a library generated via llama immunisation to identify specific binders to cystinosin (Supplementary Fig. 3). The structure was initially determined in complex with both a cytoplasmic binding sybody and luminal binding nanobody to 2.65 Å and phases calculated using long wavelength sulfur single wavelength anomalous dispersion (S-SAD) (Table 1 & Supplementary Fig. 4a). The transporter adopts a compact V-shaped structure, consisting of seven transmembrane (TM) alpha helices and adopts an inward (cytoplasmic) open state, with the CDR3 loop of the sybody (Syb39) extending into the cytoplasmic entrance of the binding site (Fig. 1d & Supplementary Fig. 4b). On the luminal side of the transporter the llama derived nanobody (Nb4) interacts with the N-terminus of TM1 and the luminal end of TM7 (Supplementary Fig. 4a, b). Within the crystal the binders act to bridge two of the transporters together, aiding crystal packing within the lipidic cubic phase. We also identified a second binder combination, consisting of two llama nanobodies, which diffracted to a slightly higher resolution of 2.33 Å (Table 1). The structures of the transporter in both crystal forms are very similar (root mean square deviation (RMSD) of 0.776 Å over 224 C_α atoms), with the only difference being loss of density for the cytoplasmic end of TM4 and the connecting loop to TM5 in the 2.33 Å structure (Supplementary Fig. 4c, d).

The structure has similarities to both the eukaryotic SWEET transporter, the bacterial semi-SWEET transporters and KDEL receptor[23,28,29,31], although with some notable differences. Cystinosin aligns with the previously determined SWEET transporter from *Oryza sativa* (rice) with an RMSD of 4.9 Å over 208 C_α atoms, which also adopts an inward open conformation[31]. However, the inversion helix (TM4) is positioned closer to TM5 forming a more spherical molecule. Another notable difference observed in cystinosin is the presence of a luminal helix connecting TM2 and TM3. This feature is likely to be present in all cystinosin homologues given the sequence conservation in this region (Supplementary Fig. 1b) and as discussed below, may play an important role in the transport mechanism by stabilising a shortunwound region of TM3. PQ-loop transporters contain a pair of three TM bundles, TMs 1-3 and TMs 5-7, which frame a central binding site[27]. In the structure of cystinosin, a polar cavity extends 23 Å into the centre of the transporter from the cytoplasmic side of the molecule, which contains a high density of positively charged side chains (Fig. 1e). The positive charges continue down into the binding site, which is dominated by the presence of two lysine side chains Lys55 (TM2) and Lys166 (TM5) and His56 (TM2), which as discussed below play an important role in cystine recognition. The luminal side of the transporter is sealed through the packing of TMs 3, 6 and 7, which are stabilised through interactions between Asp92 (TM3), Gln201 (TM6) and Lys221 (TM7) (Fig. 1e). This interaction network is strictly conserved within cystinosin homologues and as discussed below plays an important role in the transport mechanism.

Molecular basis for cystine recognition

To understand the structural basis of the recognition and transport of cystine we employed our previous binder trap methodology to stabilise the ligand bound state for structure determination[37]. Following

Article https://doi.org/10.1038/s41467-022-32589-2

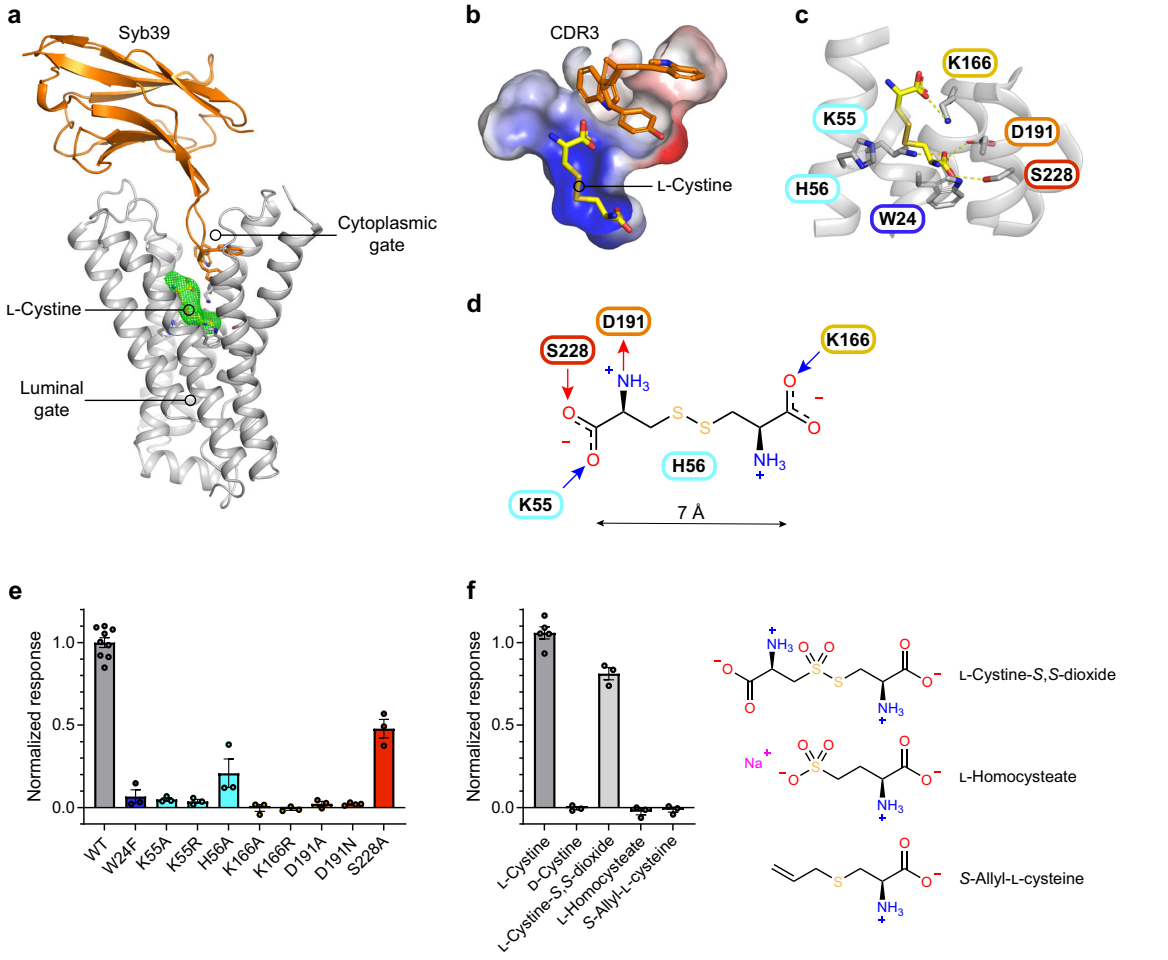


Fig. 2 | Structural basis for L-cystine recognition. a Crystal structure of cystinosin showing sybody 39, which aided in trapping the cystine ligand within the binding site. Cystine mFo-DFc difference electron density map (green mesh) contoured at 3σ. **b** The central binding site is positively charged complementing the negative charge on cystine. **c** Zoomed in view of the binding site showing the interaction network between cystinosin and cystine. **d** Schematic of cystine highlighting key interactions in the binding site. **e** Impact of binding site variants on cystine transport using a pyranine based in vitro assay. All variants were tested using 200 µM L-cystine. $n = 3$ independent experiments for variants ($n = 4$ for D191N) and $n = 9$ for WT, error bars SEM. **f** Substrate specificity of cystinosin analysed using a pyranine based in vitro assay system. All substrates were tested at 500 µM. $n = 5$ for L-cystine and $n = 3$ for all other substrates, error bars SEM. Source data are provided as a Source Data file.

several rounds of screening with different binder combinations we determined the structure of the ligand bound state in complex with Syb39 to 3.37 Å (Table 1). The ligand bound conformation of cystinosin is very similar to the apo state, with an RMSD of 0.48 Å over 240 C_α atoms. The cystine ligand was clearly resolved in the electron density map and sits at the base of the polar pocket observed in the apo state (Fig. 2a, b). Cystine adopts an extended conformation within the binding site and makes several notable interactions to conserved side chains (Fig. 2c, d). Cystine can exist in three different ionic forms depending on the pH of the environment[38]. Within physiological pH the carboxyl groups will be fully dissociated, however, the two amino groups have different pKa values of 7.48 and 9.02. Our crystallisation trials were conducted at pH 6.5, favouring the zwitterionic form of cystine, which we modelled into the structure. Indeed, we observe a salt bridge interaction between one of the amino groups of cystine and Asp191 (TM6). In addition, we observe two further salt bridge interactions between the two carboxylate groups and Lys55 (TM2) and Lys166 (TM5). A further hydrogen bond interaction is made to a conserved serine, Ser228 (TM7). The disulfide bond is sandwiched between Lys166 (TM5) and His56 (TM2), which would help to stabilise the delocalised negative charge, whereas the second amino group makes no specific interactions and points towards a hydrophobic cavity between TM2 and TM3.

To assess whether these observed interactions are important for L-cystine transport, we generated a series of variants in the plant transporter and tested their function using both a pyranine-based transport assay and a differential scanning fluorimetry (nanoDSF) binding assay. We found that alanine variants of the main interaction sites Lys55, Lys166 and Asp191, resulted in inactive protein in both assays (Fig. 2e and Supplementary Fig. 5). Furthermore, even conservative mutations of these side chains to arginine and asparagine respectively, could not recover transport activity, highlighting their importance to the mechanism. Interestingly the tryptophan, Trp24, which sits at the base of the binding site, is also important for L-cystine uptake, with even a

Article https://doi.org/10.1038/s41467-022-32589-2

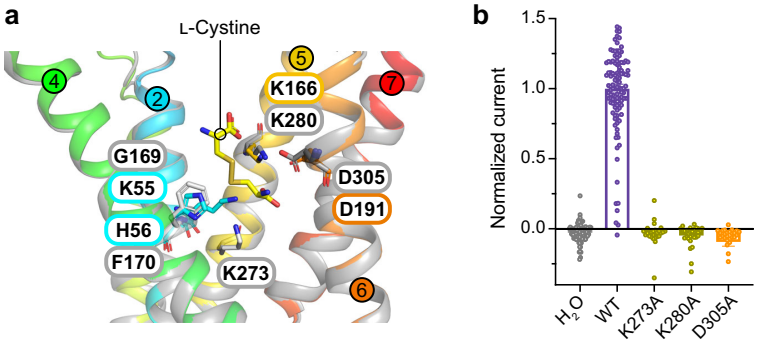


Fig. 3 | Differences in the binding site between plant and human cystinosin. **a** Structural superimposition of the plant crystal structure (coloured helices and numbers) and human AlphaFold2 model (grey) revealing the movement of key residues within the binding site. **b** TEVC data showing the effect of mutations within the binding site of human cystinosin. $n = 109$ independent experiments for H_2O, $n = 113$ for WT, $n = 28$ for K273A and D305A, $n = 30$ for K280A, error bars SEM. Source data are provided as a Source Data file.

conservative mutation to phenylalanine only exhibiting ~10% WT transport levels. In the binding assay this variant loses ~50% binding. Together these results suggest a possible cation-π interaction with the amino group that sits ~3.4 Å away in the binding site and demonstrates an essential role for Trp24 in the conformational changes required for transport. The His56Ala variant retained ~20% transport activity, indicating that this side chain is not essential for L-cystine recognition. L-Cystine also contacts Ser228 in the base of the binding site. An alanine variant of this side chain reduces both binding and transport by ~50%, indicating this side chain plays only a supporting role in transport (Fig. 2e and Supplementary Fig. 5).

A noticeable feature of the L-cystine binding pose are the interactions observed between the two carboxylate groups of the ligand and the Lys55 and Lys166 side chains. To assess the importance of these interactions we screened a library of structural analogues of L-cystine (Fig. 2f). We observed that only L-cystine-*S*,*S*-dioxide, which contains two carboxylates could be transported, at ~80% WT levels and consistent with previous assays conducted using the human transporter[20]. The uptake of L-cystine-*S*,*S*-dioxide also suggests that the position of the disulfide group is likely to be flexible to accommodate the additional two carbonyl groups which bind to one sulfur atom of the disulfide, consistent with our crystal structure. Interestingly, neither *S*-allyl-L-cysteine, which has an amino acid group replaced with an allyl group, nor L-homocysteate, which contains a sulfonate group in place of the second L-cysteine moiety, could be transported. These results show that the length of the ligand, in addition to the presence of the two carboxylate groups is important. Finally, D-cystine was also not transported, demonstrating that cystinosin is stereospecific. These results thus establish a preliminary binding model for plant cystinosin (Fig. 2d) emphasising the importance of the two salt bridges and an ideal distance of ~7 Å between the two carboxylate groups.

We next compared our structure to that predicted by AlphaFold2 for the human protein[39]. Many of the side chains identified in the binding site of plant cystinosin are strictly conserved within the human transporter, nevertheless, notable differences exist. In human cystinosin the equivalent side chain to His56 is phenylalanine (Phe170) and Lys55 is a glycine (Gly169) (Supplementary Fig. 1b). However, the AlphaFold2 model shows that in the human transporter a lysine, Lys273, extends from TM5 to sit in a similar position in the binding site (Fig. 3a). Using our oocyte transport assay, we tested the Lys273Ala variant and found this lysine is essential for function (Fig. 3b & Supplementary Fig. 6), similar to the structurally equivalent lysine in plant cystinosin. We then analysed the alanine variants of Lys280 and Asp305, which are the equivalent side chains to Lys166 and Asp191 in the plant transporter. These side chains are also essential for L-cystine transport, highlighting an evolutionary conserved mechanism of L-cystine recognition.

Mechanism of cystine transport

Secondary active transporters operate via an alternating access mechanism, whereby the transporter switches between outward and inward facing states to move ligands and ions across the membrane[40]. Within the PQ-loop family, alternating access transport is understood to occur following the symmetrical movement of helices within the two 3TM bundles[22]. Specifically, the PQ-motifs on TM1 and TM5 are required to enable the bending of these helices to open and close the intracellular gate in response to ligand and/or ion binding. In the plant SWEET transporter, the glutamine residues are absent from both PQ-motifs, with the transporter still functional in cells, however, the prolines were essential for function[31]. Therefore, to understand the role of the prolines within the transport mechanism of cystinosin, we assayed the proline to alanine variants in the human and plant transporters (Fig. 4a). In human cystinosin the proline of the first PQ-motif is dispensable, with transport levels of an alanine variant exhibiting ~75% WT levels. The equivalent alanine variant in plant cystinosin proved too unstable to purify and assay, however, a glycine variant was stable and surprisingly exhibited almost WT transport levels (~90%) (Fig. 4a & Supplementary Fig. 5c). We next analysed the Pro169Gly variant using our binding assay, and discovered that although this variant and the Pro169Ala variant are transport defective, it still retains WT levels of L-cystine binding (Supplementary Fig. 5a). Together, these results indicate the second PQ-motif proline facilitates the conformational changes during transport. The importance of the second PQ-motif proline is perhaps explained through its proximity to Lys166, which sits adjacent to Tyr167 and Asp191 (Supplementary Fig. 7a). Scanning mutagenesis previously identified the equivalent aspartate in the human transporter, Asp305, as being a potential site of proton binding[20].

To further explore the role of these side chains in cystinosin, we mutated Asp191 to asparagine and Tyr167 to phenylalanine. Interestingly, we observed increased proton flux through the Tyr167Phe variant, as measured using our pyranine-based assay (Supplementary Fig. 7b), indicating that either more protons are moving during the transport cycle or that the kinetics of transport have been altered. However, when we analysed the uptake of radioactive cystine we also observed an increase in the amount of substrate moved (Supplementary Fig. 7c). Further kinetic analysis of the Tyr167Phe variant demonstrated that Vmax increased ~30% relative to WT and the K_M increased from 19 µM to 73.3 ± 12.8 µM (Fig. 4b), confirming the role of this side chain in

 https://doi.org/10.1038/s41467-022-32589-2

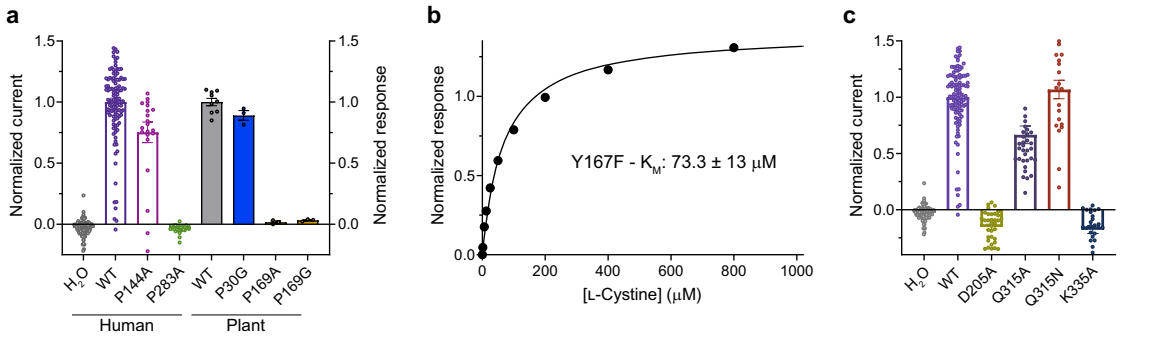


Fig. 4 | Role of PQ motifs in cystinosin. a Effect of mutating the PQ motifs in both human and plant (*A. thaliana*) cystinosin. The human protein was assayed using TEVC ($n = 109$ independent experiments for H_2O, $n = 113$ for WT, $n = 22$ for P144A and $n = 31$ for P283A) whereas the plant transporter was analysed in reconstituted liposomes ($n = 9$ independent experiments for WT and $n = 3$ for variants). Error bars SEM. **b** Representative K_M of plant cystinosin variant Tyr167Phe. K_M was calculated from three independent experiments, error SD. **c** Summary data showing the effect of mutations on the proposed luminal lock of human cystinosin expressed in Xenopus oocytes. $n = 109$ independent experiments for H_2O, $n = 113$ for WT, $n = 35$ for D205A, $n = 34$ for Q315A, $n = 22$ for Q315N, $n = 27$ for K335A, error bars SEM. Source data are provided as a Source Data file.

regulating transport kinetics. These results are consistent with the previous analysis of the human transporter and demonstrate that Tyr167 is important for modulating L-cystine recognition[20]. In contrast, the asparagine 191 variant displayed a marked, but slow proton leak and was unable to transport L-cystine, suggesting this variant is uncoupled (Supplementary Fig. 7d). The presence of a slow proton leak indicates a second site of proton binding is involved to gate the transporter, consistent with suggestions from the previous functional analysis using TEVC[20]. Our structure highlighted an interaction between Asp92 on TM3 with Lys221 on TM7 and Gln201 on TM6, which forms the only inter bundle salt bridge between the two 3TM repeats (Fig. 1e). Alanine variants of these side chains proved too unstable to purify (Supplementary Fig. 5c), highlighting their importance for structural stability. To address the role of these side chains in the transport mechanism of cystinosin, we made the equivalent mutations in the human transporter and assayed their function using our TEVC assay (Fig. 4c). Unlike the plant variants, the human proteins all expressed in the oocytes (Supplementary Fig. 6). Our results indicate that whilst the glutamine in the human transporter (Gln315) is not required for function, both the aspartate (Asp205) and lysine (Lys335) are essential and as discussed below, mutation of Asp205 to asparagine in the human transporter leads to infantile cystinosis[41].

As noted above, a major difference between human and plant cystinosin is the presence of His56 on TM2, which is located close to the disulfide bond of the ligand (Fig. 2b). In the human protein, His56 is a phenylalanine; however, a His56Phe plant cystinosin variant was still functional, displaying ~50% WT activity (Supplementary Fig. 7e). Although this side chain is not essential for transport, the proximity to Lys55 suggested a functional role, which combined with the slightly different pH environments found between the plant vacuole and mammalian lysosome[42,43], prompted us to study the effects of different pH conditions on the His56Phe variant in the plant transporter. The WT transporter shows transport activity across a wide pH range (6.5–8.2), whereas the His56Phe variant is unable to transport at pH 6.5 but remains functional at pH 7.2 and 8.2 (Supplementary Fig. 7e). As the pH gradient across the plant tonoplast membrane is less steep than observed in mammalian lysosomes, this adaption may have evolved to facilitate more efficient transport under the conditions found within plant cells.

Discussion

The crystal structure of the plant homologue provides a framework for understanding how disease-causing mutations in human cystinosin lead to different severities of cystinosis. Currently, 42 missense mutations have been identified (Supplementary Table 1), which are categorised into four broad groups and relating to age of onset: infantile, juvenile, atypical and ocular[13,15,18]. Omitting those mutations that occur in the N-terminal domain, we can map on the remaining mutations using the model of the human transporter (Fig. 5a and Supplementary Table 1). Many of these mutations map to the functionally important regions, and in particular cluster around the cytoplasmic gate, ligand binding region and luminal gate (Fig. 5b, c).

Previous structures of the bacterial PQ-motif transporters revealed that alternating access transport within this family is likely to occur following the movement of the two 3TM helical bundles around the central ligand, aided by the location of the two PQ-motifs in TM1 and 5[27,28]. However, unlike the bacterial transporters, cystinosin is proton coupled, which introduces the need to couple ligand binding to proton translocation. Our results suggest this occurred in cystinosin through the introduction of two conserved salt bridge interactions that exist on either side of the cystine binding site, which function to couple the open/closed state of the cytoplasmic and luminal gates to proton and cystine binding.

Taken together, our data enables us to propose a working model for proton coupled L-cystine transport (Fig. 6). The transporter adopts an outward open state in the lysosome (state i), with the cytoplasmic gate closed and TMs 1-2 and TMs 5-6 packed together. In this conformation, the luminal gate is open, with TMs 3 and 7 separated. In this state, Asp92 (Asp205; human numbering in parentheses) would be protonated, given the acidic luminal pH. The protonation of Asp92 would function to keep the luminal gate open, as it would weaken the interaction with Lys221 (Lys335). Following L-cystine binding, Asp92 (Asp205) likely deprotonates to engage Lys221 (Lys335), closing the luminal gate (state ii). L-Cystine engages both Lys166 (Lys280) and Asp191 (Asp305), which reside on TM5 and 6 respectively, triggering opening of the cytoplasmic gate, aided through interactions to Trp24 (Trp138) and Lys55 (Lys273). The transporter will transition to the inward (cytoplasmic) open state following the movement of the cytoplasmic ends of TMs 1 and 2 and TMs 5 and 6 away from one another (state iii). There are no salt bridge interactions coordinating the closure of the cytoplasmic gate (i.e., connecting helices of the first and second 3TM bundle), instead the loops connecting these helices contain polar and bulky hydrophobic side chains that pack together to seal the binding site from the cytoplasm. A similar feature is observed in the bacterial PQ-loop homologues[27]. Following opening of the cytoplasmic gate, L-cystine release could be triggered by protonation of Asp191

Article https://doi.org/10.1038/s41467-022-32589-2

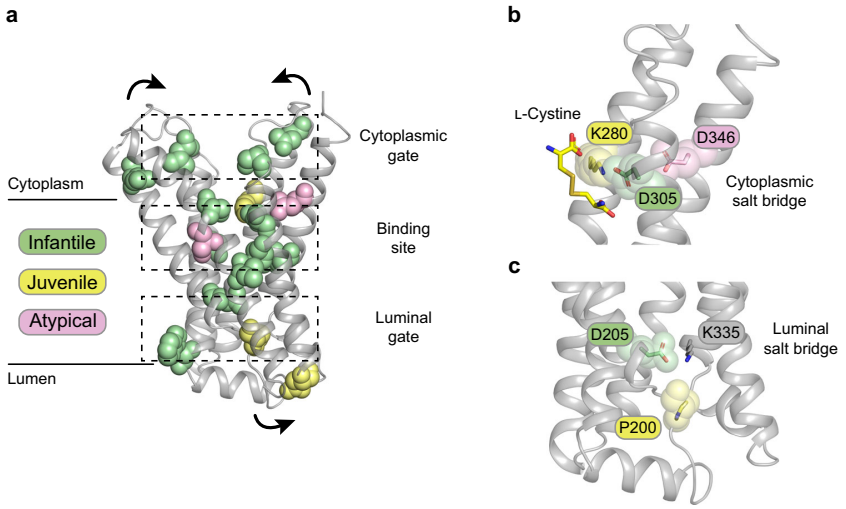


Fig. 5 | Structural context of cystinosis-causing mutations. a A subset of the missense mutations that localise to the cytoplasmic and luminal gates or binding site mapped onto the AlphaFold2 model of human cystinosin. Mutations are coloured according to the severity and onset of disease. Arrows indicate structural movements that occur during transport. **b** Zoomed in view of the cytoplasmic gate with the position of L-cystine superimposed. **c** Zoomed in view of the luminal gate with the inter bundle salt bridge residues highlighted.

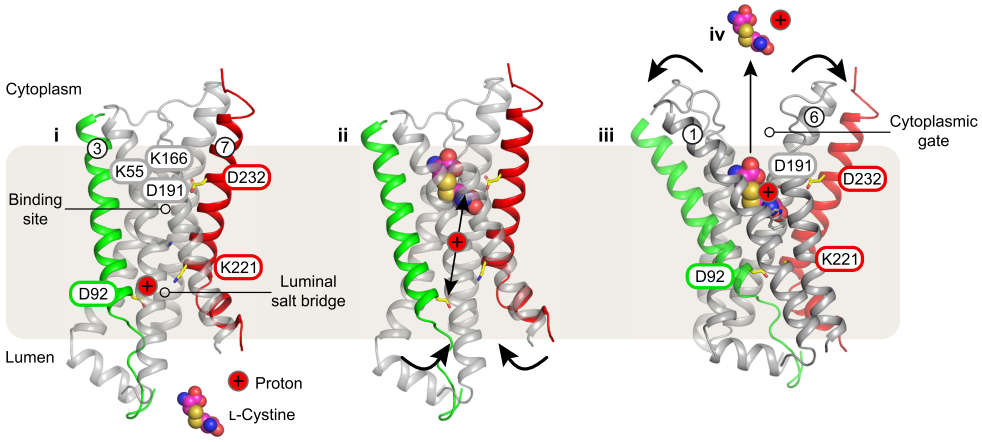


Fig. 6 | Conserved salt bridge interactions mediate L-cystine transport. Proposed mechanism for proton coupled cystine transport via cystinosin. **i** in the apo state the transporter adopts an outward open state due to protonation of Asp92 on TM3. **ii** cystine binding results in deprotonation of Asp92 and promotes opening of the cytoplasmic gate through interactions within the binding site, with Asp191, Lys55 and Lys166 playing important roles in ligand recognition. Closure of the luminal gate is facilitated through a salt-bridge interaction between Asp92 and Lys221. Movement of Lys221 towards Asp92 would reinforce the movements required to open the cytoplasmic gate as the helices pivot around the bound cystine ligand. **iii** transition to the inward open state occurs and Asp191 become protonated, weakening the interaction with cystine, and promoting release of the ligand and proton into the cytoplasm (**iv**). Deprotonation of Asp191 would create the conditions necessary for the return of the transporter to restart the cycle, facilitated by the proximity of Asp232.

(Asp305), as this would weaken the salt bridge interaction we observe with the amino terminus of the ligand (Fig. 2c, d). The movement of protons between Asp92 (Asp205) and Asp191 (Asp305) could occur through ordered water molecules, as shown for POT family peptide transporters[44] or may be facilitated directly by L-cystine. Following L-cystine release, Asp191 (Asp305) would be expected to deprotonate on exposure to the cytoplasm and further protonation of Asp92 (Asp205) would result in the transporter resetting for another cycle. The absence of a pH sensitive element within the cytoplasmic side of the helices suggests this gate is structurally coupled to the luminal gate, which contains the pH responsive element, Asp191. Mechanical coupling between the two gates is achieved by having Lys221 (Lys335) and Asp232 (Asp346) on TM7. Aspartate 232 (Asp346) interacts with Asp191 (Asp305) in the cystinosin structure (Fig. 6), providing a link between movement in the luminal gate to structural changes in the cytoplasmic region of the transporter and L-cystine binding.

Finally, our observation that the first PQ-motif is not required for cystinosin function suggests a division of labour within the transporter, with the second 3TM bundle coordinating a more sensitive interaction network. Certainly, more mutations map to the second 3TM

Article https://doi.org/10.1038/s41467-022-32589-2

bundle in cystinosis patients, supporting a more prominent role in the transport mechanism (Fig. 5a & Supplementary Table 1). Fusing the two 3TM bundles together clearly represented an important step in the evolution of function within the PQ-loop family and provided an ideal framework to evolve more complicated coupling mechanisms. It would also provide opportunities for oligomerisation and asymmetric assemblies in the membrane, for which we currently have little insight. It will be important to pursue this aspect of PQ-protein biology to fully understand their role in the cell. Nevertheless, the structures and functional data presented, combined with the functional annotation of key disease-causing mutations now provides a firm foundation for developing a molecular blueprint for proton coupled lysosomal cystine transport and cystinosis.

Methods

Cloning, expression, and purification of *Arabidopsis thaliana* cystinosin

The gene encoding *Arabidopsis thaliana* cystinosin (Uniprot code P57758) was codon optimised, synthesised (Geneart, ThermoFisher) and cloned into pDDGFP-Leu2d[34] for expression as a cleavable GFP^{His} fusion in *Saccharomyces cerevisiae* strain BJ5460[45]. Variant forms of plant cystinosin were created by site-directed mutagenesis using overlap PCR. Yeast cells transformed with their respective plasmids were cultivated in synthetic complete medium minus leucine (-leu) with 2% (w/v) D-glucose for 32 h before diluting ninefold into -leu containing 2% (v/v) DL-lactate; wild type protein was expressed using a 15 l fermentation vessel (Eppendorf BioFlo 415) whilst variants were grown in baffled flasks, using 6 l volumes. Following overnight culture, expression was induced at an optical density (OD_{600}) of 2.5 by addition of 1.5% (w/v) D-galactose, dissolved in -leu media. After 20–23 h the yeast cells were harvested by centrifugation and lysed using a high pressure (38 kpsi). Following a low-speed centrifuge spin to remove cell debris and unlysed cells (30,000 x *g*) membranes were isolated through two spins at 200,000 x *g* with a resuspension in between using 20 mM HEPES-NaOH, pH 7.5, 1 M K-acetate, and finally resuspended in phosphate-buffered saline (PBS). Wild type cystinosin and variants were solubilized in *n*-dodecyl-β-D-maltoside (DDM, Anatrace) and purified to homogeneity using standard immobilised metal affinity chromatography (IMAC) steps. Solubilisation was carried out for 1 h in 1% (w/v) DDM, 1 × PBS, 150 mM NaCl and 10% (v/v) glycerol. Following ultra-centrifugation at 200,000 x *g* for 1 h at 4 °C, the soluble fraction was supplemented with 25 mM imidazole and bound to Ni-NTA resin (Thermo) in batch for 2.5–3 h at 4 °C. The resin was transferred into a gravity flow column, washed with 25–30 CV of solubilisation buffer supplemented with 0.1% (w/v) DDM and 35 mM imidazole and subsequently eluted with 250 mM imidazole. After overnight dialysis (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.02% (w/v) DDM) at 4 °C in the presence of Tobacco Etch Virus (TEV) protease, a reverse IMAC step was performed using the same resin (1.5 h binding at a final concentration of 10 mM imidazole). The flow-through, containing the purified cystinosin protein, was concentrated using a 50 kDa MWCO centricon (Amicon, Sigma Aldrich) and applied to a Superdex 200 Increase 10/300 GL column (Cytiva) equilibrated in size exclusion buffer (20 mM MES-NaOH pH 6, 150 mM NaCl and 0.03 (w/v) DDM). For crystallisation the detergent concentration was reduced to 0.013% (w/v) DDM. C-terminally Avi-tagged protein was purified as above, and subsequent biotin modification was carried out using the BirA ligase according to the manufacturer's instructions (Avidity).

Nanobody selection and purification

To identify cystinosin-specific nanobodies a library was raised through immunisation of a llama using reconstituted protein and screened using C-terminally Avi-tagged and biotinylated protein (method modified from[32]). Specifically, nanobodies were raised in a llama following intramuscular immunisation with purified protein reconstituted into liposomes and using Gerbu LQ#3000 as the adjuvant. Immunisations and handling of the llama were performed under the authority of the project license PPL 70/8108. Blood (150 ml) was collected, and peripheral blood mononuclear cells were prepared using Ficoll-Paque PLUS according to the manufacturers protocol. Total RNA was extracted using TRIzol™ and VHH cDNAs were generated by reverse transcription-PCR using primer as detailed previously[46]. The pool of VHH encoding sequences were amplified by two rounds of nested PCR: firstly with 'CALL_001' and 'CALL_002', followed by 'VHH_For' and 'VHH_Rev_IgG2' and 'VHH_Rev_IgG3', and cloned into the SfiI sites of the phagemid vector pADL-23c. In this vector, the VHH encoding sequence is preceded by a pelB leader sequence followed by a linker, His_6- and cMyc-tag (GPGGQHHHHHHGAEQKLISEEDLS). Electro-competent *E. coli* TG1 cells were transformed with the recombinant pADL-23c vector resulting in a VHH library of about 2×10^8 independent transformants. The resulting TG1 library stock was then infected with M13K07 helper phage to obtain a library of VHH-presenting phages. Phages displaying VHHs specific for cystinosin were enriched after two rounds of bio-panning on 50 nM of biotinylated cystinosin, through capturing with Dynabeads™ MyOne™ Streptavidin T1 (Thermo Fisher). Enrichment after each round of panning was determined by plating the cell culture with 10-fold serial dilutions. After the second round of panning, 192 individual phagemid clones were picked, VHH displaying phages were recovered by infection with M13K07 helper phage and tested for binding to cystinosin by ELISA. ELISA positive clones were sequenced and unique nanobodies identified and grouped according to sequence identity. These were analysed by small scale pull-down using the His_6-tagged constructs as well as co-elution using a Sepax SRT-C SEC-300 size exclusion column (Chromex). The better binders were further characterised using biolayer interferometry using an Octet Red384 (Sartorius) and streptavidin biosensors loaded with biotinylated cystinosin at 20 nM in 20 mM MES-NaOH pH 6, 150 mM NaCl, 0.03% (w/v) DDM to compare k_{on} and k_{off}. To calculate the K_D a serial dilution of the nanobody was made and following a short baseline step the nanobody was allowed to associate for 10 min followed by a dissociation step of 10 min. Data were analysed in the Octet v9.0 software package and fit to a single binding site model in Prism (GraphPad). All raw data was baseline and reference subtracted, in-step corrected, *y*-axis aligned and filtered with a Savitzky–Golay filter.

Synthetic nanobody (Sybody) selection

Sybody selection was performed against C-terminally Avi-tagged and biotinylated cystinosin, prepared using the same procedure as above for the llama based nanobody screen. The protocols for sybody selection have been described elsewhere[33,47]. A high affinity sybody, Syb39, was identified from the loop library[47], which formed a stable complex with the transporter and co-eluted down a Sepax SRT-C SEC-300 column (Chromex). High affinity binders were identified using small scale pull-down assays using the His_6-tagged binders. Biolayer interferometry was performed on an Octet Red 384 (Sartorius) as described above to compare k_{on} and k_{off} and calculate K_D. The identified sybodies were cloned and expressed as tag-free binders using the protocols previously described[33].

Crystallisation

Crystallisation was performed using protein at 30–37.5 mg ml^{-1} final concentration, as determined using absorbance at 280 nm. L-Cystine was added to a final concentration of 0.5 mM and the protein was left on ice for at least 2 h prior to crystallisation using the lipidic cubic phase method[48]. Protein-laden mesophase was obtained by monoolein with protein in a 60:40 (w:w) ratio using a coupled syringe device (Art Robbins, USA). Initial crystals appeared at 20 °C after 1–3 days in

Article https://doi.org/10.1038/s41467-022-32589-2

27–28% PEG 500DME, 100 mM MES-NaOH, pH 5.50–5.75, 100–150 mM K-formate and matured after 5–7 days. Wells were opened using a tungsten carbide glasscutter and the crystals were harvested using 50 or 100 µm micromounts (MiTeGen). Crystals were cryo-cooled directly in liquid nitrogen and stored in unipucks.

Structure determination

X-ray diffraction data were collected at beamlines I24 and I04, Diamond Light Source, UK (Table 1). Indexing and integration were performed with XIA2 using either the DIALS[49–51] or autoPROC[52] pipeline, followed by scaling and merging with AIMLESS[53]. Initial phases were obtained by sulfur SAD phasing. Model building into the electron density map was performed in COOT[54], with structure refinement carried out using BUSTER[55]. The cystine bound crystal structure (PDB: 7ZKW) contained two molecules in the asymmetric unit and exhibited pseudo merohedral twinning in space group C2 with twin law -h-4/2 l,-k,l applied during refinement using Refmac[56]. The twin fraction was 0.204. The figures and interactions detailed in the study are based on chain B, which gave the strongest density in the refined maps. Geometry restraints for L-cystine were calculated using the grade server supplied by Global Phasing Ltd. Model validation was carried out using the Molprobity server[57]. Images were prepared using PyMol[58].

Sulfur-SAD phasing

X-ray diffraction images were collected at the Diamond Light Source I23 beamline[59] on a PILATUS 12 M detector (Dectris). A single dataset of 360° (rotation increment 0.1°, exposure 0.1 s) was collected at a temperature close to 60 K and a wavelength of 2.7552 Å. The dataset was automatically processed with DIALS[60] in the $P2_12_12$ space group to a resolution of 3.2 Å. The structure was solved by sulfur-SAD using the increased sulfur anomalous signal at longer wavelengths, Fast_EP[61], Anode[62] and Phenix.autobuild[63] as implemented in the Diamond automatic downstream processing pipelines.

Proteoliposome reconstitution

Purified wild type and mutated cystinosin were reconstituted into lipid vesicles consisting of a 3:1 (w:w) ratio of POPE:POPG by rapid dilution. To do so, the proteins were first exchanged into *n*-decyl-β-D-maltoside (DM, Glycon) by gel filtration using the same buffer as before but supplemented with 0.3% (w/v) DM, and subsequently mixed with the extruded vesicles at a 40:1 (w:w) lipid:protein ratio. The lipid protein mix was incubated at room temperature for 20 min before diluting it 27-fold into cold reconstitution buffer (50 mM K-phosphate, pH 7.0). Proteoliposomes were harvested by ultra-centrifugation at 200,000 × g for >2 h and dialysed extensively overnight against two batches of reconstitution buffer. Proteoliposomes were recovered and resuspended to a final concentration of 0.5 µg µl^{-1} as analyzed by SDS-PAGE using ImageJ, comparing all bands of reconstituted protein to a known concentration of detergent-solubilized protein. Proteoliposomes were subjected to two rounds of freeze-thaw using liquid nitrogen and stored at −80 °C. For llama immunisation to raise nanobodies, a total of 600 µg of cystinosin was reconstituted at a lipid to protein ratio of 30:1 and proteoliposomes were resuspended to a final concentration of 1 mg ml^{-1}.

Transport assay using pyranine

Proteoliposomes were pelleted by ultra-centrifugation at 108,000 × g and 4 °C for 25 min and subsequently resuspended in INSIDE buffer consisting of 5 mM HEPES-NaOH, 120 mM KCl, 2 mM $MgSO_4$ and 2 mM pyranine (trisodium 8-hydroxypyrene-1,3,6-trisulfonate) at pH 7.2. Proteoliposomes were subjected to eight rounds of freeze-thaw before being extruded through a 0.4 µm polycarbonate membrane. The uniquely formed vesicles were then harvested through ultra-centrifugation as before but at 18 °C and external pyranine was removed by applying the sample to a G-25 column (Cytiva) equilibrated in INSIDE buffer without pyranine. Vesicles were collected again and finally resuspended in INSIDE buffer without pyranine. Transport assays were performed at 25 °C in a cuvette equipped with a small magnetic flea using a Cary Eclipse Fluorescence Spectrophotometer (Agilent) set at dual excitation at 460 and 415 nm and emission at 510 nm. At the start of each assay, vesicles containing 5 µg of cystinosin were diluted into 545 µl of OUTSIDE buffer (5 mM HEPES-NaOH, 120 mM NaCl, 2 mM $MgSO_4$ at pH 7.2). After 15 seconds substrates were added per 100 µl from a freshly prepared stock in OUTSIDE buffer (200 µM final if not stated otherwise) and transport was initiated with a final concentration of 1 µM valinomycin (v) in OUTSIDE buffer after 30 seconds. The ratio of the resulting fluorescence profiles were calculated and plotted using Prism (GraphPad) or Excel (Microsoft), transport data from triplicate experiments were normalised to one to allow for comparison. For the pH experiments shown in Supplementary Fig. 6d, the buffer components were as stated above except 5 mM MES was used for pH 6.5 and 5 mM HEPES was used for pH 8.2.

L-Cystine binding assay

DDM-purified wild type and mutated cystinosin were diluted to 10 µM using a binding assay buffer consisting of 20 mM HEPES-NaOH, 150 mM NaCl and 0.03% (w/v) DDM, also containing up to 3.2 mM L-cystine, set to pH 7.2. After a 5 min incubation step at room temperature the solutions were loaded onto a Prometheus NT.48 (NanoTemper) and nanoDSF (differential scanning fluorimetry) experiments were run at 25% intensity using a 0.5 °C min^{-1} increment. We note that the addition of L-cystine did not significantly increase the melting temperature of the protein. Instead, we found that ligand addition led to a reduction in the tryptophan fluorescence emission ratio measured at 350 over 330 nm before the melting temperature was reached. Ten subsequent data points commencing from 25 °C were averaged and subtracted from a control melt without L-cystine. The fluorescence quench observed from each cystinosin variant with 3.2 mM L-cystine present compared to no ligand was finally normalised by dividing it by that of the average of a WT triplicate. All variants were completed in triplicate using a fresh L-cystine stock for each replicate.

Radioactive transport assay

Proteoliposomes were pelleted and resuspended in INSIDE buffer (10 mM MES-NaOH, 120 mM K-acetate, 2 mM $MgSO_4$ at pH 6.2). Four rounds of freeze-thaw were followed by extrusion through a 0.4 µm polycarbonate membrane. To reduce the volume, vesicles were harvested again at 18 °C and resuspended in INSIDE buffer.

Starting each individual assay, vesicles containing 2 µg protein were diluted into 250 µl OUTSIDE buffer (10 mM MES-NaOH, 120 mM NaCl, 2 mM $MgSO_4$ at pH 6.2) also containing 48 µM L-cystine, 2 µM radiocarbon-labelled ^{14}C-L-cystine (specific activity 0.2 Ci mmol^{-1}, Hartmann Analytic) as well as 10 µM valinomycin (Merck). Assays were performed at 20 °C. 45 µl fractions were removed at different timepoints, diluted into 2 ml of OUTSIDE buffer and vesicles were immediately isolated on a 0.22 µm cellulose filter (Merck) using a vacuum manifold. Filters were washed twice with 2 ml OUTSIDE buffer and subsequently transferred into Ultima Gold (PerkinElmer) scintillation liquid. Remaining radioactivity stemming from the inside of the vesicles was measured using a Wallac scintillation counter.

Electrophysiology

Human cystinosin without its C-terminal lysosomal targeting motif (GYDQLN) was subcloned with and without a C-terminal GFP into the pFAW vector for microscopy and electrophysiology experiments respectively. Variants of the protein were generated by site-directed mutagenesis and verified via sequencing. In vitro transcription of wild type and mutant *CTNS* was carried out using AmpliCap-Max T7 high

yield message marker kit (Cellscript). *Xenopus laevis* oocytes were defolliculated by collagenase treatment. They were then injected with 50 nl of nuclease-free water or 75 ng (50 nl of 1.5 µg µl^{-1} RNA) of human cystinosin wild type or mutant mRNA, and were incubated for 2 days at 19 °C in Barth's solution (54 mM NaCl, 30 mM KCl, 2.4 mM $NaHCO_3$, 0.8 mM $MgSO_4(H_2O)_7$, 0.4 mM $CaCl_2$, 0.3 mM $Ca(NO_3)_2(H_2O)_4$, 7.5 mM Tris-HCl, adjusted to pH 7.4) with gentamicin (50 µg ml^{-1})[20]. Two-electrode voltage clamp measurements were carried out on a Roboocyte system (Multi Channel Systems) using 3 M KCl back-filled borosilicate glass pipettes with a resistance of 200–1000 k Ω. Oocytes were voltage-clamped to −40 mV for the measurements. For each oocyte, an initial recording was made with ND-96 solution (96 mM NaCl, 2 mM KCl, 2 mM $MgCl_2$, 1.8 mM $CaCl_2$, 5 mM MES-NaOH, adjusted to pH 5) without cystine, followed by a second recording in ND-96 with 1 mM cystine (pH 5). For the dose-response curves, these measurement pairs were repeated on a single oocyte for each of the tested concentrations (0–1000 µM). Traces were analysed using the Roboocyte software (Multi Channel Systems). Cystine-evoked current values were obtained by subtracting the mean current after the addition of cystine from the mean current before the addition of cystine. Currents were normalised based on the cystine-evoked currents for the wild type human cystinosin injected oocytes of the corresponding oocyte batch. The data obtained for dose-response curves were normalised based on the maximum response measured (i.e., current at 400 µM or 1000 µM) for that particular oocyte and was fit using Prism (GraphPad). All experiments were repeated on at least three different oocyte batches.

Fluorescence Microscopy

Oocytes injected with either water, wild type or mutant human cystinosin-GFP mRNA were washed in 1× PBS, stained with 0.05 mg ml^{-1} of CF®633 wheat germ agglutinin (Biotium) in 1 × PBS for 5 min at RT, and washed with 1 × PBS. Surface expression was imaged using an LSM-780 confocal microscope (10 × lens) with an Argon (488 nm) and He-Ne laser (633 nm).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support this study are available from the corresponding authors upon reasonable request. The atomic coordinates have been in the RCSB Protein Data Bank (PDB) under accession code 7ZK1 (no substrate, Syb39 & Nb4), 7ZKZ (no substrate, Nb1 & Nb4), and 7ZKW (cystine bound, Syb39). Source data underlying Figs. 1b, c, 2e, f, 3b, 4a–c, and Supplementary Figs. 3b, 5a, b, 7c, e is provided as a Source Data file. Source data are provided with this paper.

References

1. Liu, G. Y. & Sabatini, D. M. mTOR at the nexus of nutrition, growth, ageing and disease. *Nat. Rev. Mol. Cell Biol.* **21**, 183–203 (2020).
2. Saftig, P. & Puertollano, R. How lysosomes sense, integrate, and cope with stress. *Trends Biochem Sci.* **46**, 97–112 (2021).
3. Hediger, M. A. et al. The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins. *Pflgers Arch. Eur. J. Physiol.* **447**, 465–468 (2004).
4. Wyant, G. A. et al. mTORC1 activator SLC38A9 is required to efflux essential amino acids from lysosomes and use protein as a nutrient. *Cell* **171**, 642–654.e612 (2017).
5. Talaia, G., Amick, J. & Ferguson, S. M. Receptor-like role for PQLC2 amino acid transporter in the lysosomal sensing of cationic amino acids. *Proc. Natl. Acad. Sci. USA* **118**, https://doi.org/10.1073/pnas.2014941118 (2021).
6. Ballabio, A. & Bonifacino, J. S. Lysosomes as dynamic regulators of cell and organismal homeostasis. *Nat. Rev. Mol. Cell Biol.* **21**, 101–118 (2020).
7. Cesar-Razquin, A. et al. A call for systematic research on solute carriers. *Cell* **162**, 478–487 (2015).
8. Drew, D. & Boudker, O. Shared molecular mechanisms of membrane transporters. *Annu Rev. Biochem* **85**, 543–572 (2016).
9. Zhang, Y., Zhang, Y., Sun, K., Meng, Z. & Chen, L. The SLC transporter in nutrient and metabolic sensing, regulation, and drug development. *J. Mol. Cell Biol.* **11**, 1–13 (2019).
10. Platt, F. M., d'Azzo, A., Davidson, B. L., Neufeld, E. F. & Tifft, C. J. Lysosomal storage diseases. *Nat. Rev. Dis. Prim.* **4**, 27 (2018).
11. Town, M. et al. A novel gene encoding an integral membrane protein is mutated in nephropathic cystinosis. *Nat. Genet* **18**, 319–324 (1998).
12. Attard, M. et al. Severity of phenotype in cystinosis varies with mutations in the CTNS gene: predicted effect on the model of cystinosin. *Hum. Mol. Genet* **8**, 2507–2514 (1999).
13. David, D. et al. Molecular basis of cystinosis: geographic distribution, functional consequences of mutations in the CTNS gene, and potential for repair. *Nephron* **141**, 133–146 (2019).
14. Cherqui, S. & Courtoy, P. J. The renal Fanconi syndrome in cystinosis: pathogenic insights and therapeutic perspectives. *Nat. Rev. Nephrol.* **13**, 115–131 (2017).
15. Castro-Balado, A. et al. Recent research in ocular cystinosis: drug delivery systems, cysteamine detection methods and future perspectives. *Pharmaceutics* **12**, https://doi.org/10.3390/pharmaceutics12121177 (2020).
16. Andrzejewska, Z. et al. Cystinosin is a component of the vacuolar H+-ATPase-Ragulator-Rag complex controlling mammalian target of rapamycin complex 1 signaling. *J. Am. Soc. Nephrol.* **27**, 1678–1688 (2016).
17. Johnson, J. L. et al. Upregulation of the Rab27a-dependent trafficking and secretory mechanisms improves lysosomal transport, alleviates endoplasmic reticulum stress, and reduces lysosome overload in cystinosis. *Mol. Cell Biol.* **33**, 2950–2962 (2013).
18. Jamalpoor, A., Othman, A., Levtchenko, E. N., Masereeuw, R. & Janssen, M. J. Molecular mechanisms and treatment options of nephropathic cystinosis. *Trends Mol. Med* **27**, 673–686 (2021).
19. Kalatzis, V., Cherqui, S., Antignac, C. & Gasnier, B. Cystinosin, the protein defective in cystinosis, is a H(+)-driven lysosomal cystine transporter. *EMBO J.* **20**, 5940–5949 (2001).
20. Ruivo, R. et al. Mechanism of proton/substrate coupling in the heptahelical lysosomal transporter cystinosin. *Proc. Natl Acad. Sci. USA* **109**, E210–E217 (2012).
21. Saudek, V. Cystinosin, MPDU1, SWEETs and KDELR belong to a well-defined protein family with putative function of cargo receptors involved in vesicle trafficking. *PLoS One* **7**, e30876 (2012).
22. Newstead, S. & Barr, F. Molecular basis for KDEL-mediated retrieval of escaped ER-resident proteins - SWEET talking the COPs. *J. Cell Sci.* **133**, https://doi.org/10.1242/jcs.250100 (2020).
23. Brauer, P. et al. Structural basis for pH-dependent retrieval of ER proteins from the Golgi by the KDEL receptor. *Science* **363**, 1103–1107 (2019).
24. Schenk, B. et al. MPDU1 mutations underlie a novel human congenital disorder of glycosylation, designated type If. *J. Clin. Invest* **108**, 1687–1695 (2001).
25. Jezegou, A. et al. Heptahelical protein PQLC2 is a lysosomal cationic amino acid exporter underlying the action of cysteamine in cystinosis therapy. *Proc. Natl Acad. Sci. USA* **109**, E3434–E3443 (2012).
26. Chen, L. Q. et al. Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* **468**, 527–532 (2010).
27. Latorraca, N. R. et al. Mechanism of substrate translocation in an alternating access transporter. *Cell* **169**, 96–107.e112 (2017).

28. Lee, Y., Nishizawa, T., Yamashita, K., Ishitani, R. & Nureki, O. Structural basis for the facilitative diffusion mechanism by SemiSWEET transporter. *Nat. Commun.* **6**, 6112 (2015).
29. Xu, Y. et al. Structures of bacterial homologues of SWEET transporters in two distinct conformations. *Nature* **515**, 448–452 (2014).
30. Feng, L. & Frommer, W. B. Structure and function of SemiSWEET and SWEET sugar transporters. *Trends Biochem Sci.* **40**, 480–486 (2015).
31. Tao, Y. et al. Structure of a eukaryotic SWEET transporter in a homotrimeric complex. *Nature* **527**, 259–263 (2015).
32. Pardon, E. et al. A general protocol for the generation of Nanobodies for structural biology. *Nat. Protoc.* **9**, 674–693 (2014).
33. Zimmermann, I. et al. Generation of synthetic nanobodies against delicate proteins. *Nat. Protoc.* https://doi.org/10.1038/s41596-020-0304-x (2020).
34. Parker, J. L. & Newstead, S. Method to increase the yield of eukaryotic membrane protein expression in Saccharomyces cerevisiae for structural and functional studies. *Protein Sci.* **23**, 1309–1314 (2014).
35. Parker, J. L., Mindell, J. A. & Newstead, S. Thermodynamic evidence for a dual transport mechanism in a POT peptide transporter. *Elife* **3**, https://doi.org/10.7554/eLife.04273 (2014).
36. Brunner, J. D. & Schenck, S. Production and application of nanobodies for membrane protein structural biology. *Methods Mol. Biol.* **2127**, 167–184 (2020).
37. Parker, J. L. et al. Structural basis of antifolate recognition and transport by PCFT. *Nature*, https://doi.org/10.1038/s41586-021-03579-z (2021).
38. Bannai, S. & Kitamura, E. Role of proton dissociation in the transport of cystine and glutamate in human diploid fibroblasts in culture. *J. Biol. Chem.* **256**, 5770–5772 (1981).
39. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
40. Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211**, 969–970 (1966).
41. Shotelersuk, V. et al. CTNS mutations in an American-based population of cystinosis patients. *Am. J. Hum. Genet* **63**, 1352–1362 (1998).
42. Shen, J. et al. Organelle pH in the Arabidopsis endomembrane system. *Mol. Plant* **6**, 1419–1437 (2013).
43. Ohkuma, S. & Poole, B. Fluorescence probe measurement of the intralysosomal pH in living cells and the perturbation of pH by various agents. *Proc. Natl Acad. Sci. USA* **75**, 3327–3331 (1978).
44. Parker, J. L. et al. Proton movement and coupling in the POT family of peptide transporters. *Proc. Natl Acad. Sci. USA* **114**, 13182–13187 (2017).
45. Drew, D. et al. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae. *Nat. Protoc.* **3**, 784–798 (2008).
46. Parker, J. L. et al. Cryo-EM structure of PepT2 reveals structural basis for proton-coupled peptide and prodrug transport in mammals. *Sci. Adv.* **7**, https://doi.org/10.1126/sciadv.abh3355 (2021).
47. Zimmermann, I. et al. Synthetic single domain antibodies for the conformational trapping of membrane proteins. *Elife* **7**, https://doi.org/10.7554/eLife.34317 (2018).
48. Caffrey, M. & Porter, C. Crystallizing membrane proteins for structure determination using lipidic mesophases. *J. Vis. Exp.* https://doi.org/10.3791/1712 (2010).
49. Winter, G., Lobley, C. M. C. & Prince, S. M. Decision making in xia2. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **69**, 1260–1273 (2013).
50. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2009).
51. Waterman, D. G. et al. Diffraction-geometry refinement in the DIALS framework. *Acta Crystallogr D. Struct. Biol.* **72**, 558–575 (2016).
52. Vonrhein, C. et al. Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. Sect. D., Biol. Crystallogr.* **67**, 293–302 (2011).
53. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **69**, 1204–1214 (2013).
54. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 486–501 (2010).
55. Blanc, E. et al. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **60**, 2210–2221 (2004).
56. Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **57**, 122–133 (2001).
57. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 12–21 (2010).
58. Schrodinger, L. L. C. *The PyMOL Molecular Graphics System, Version 1.3r1* (2010).
59. Wagner, A., Duman, R., Henderson, K. & Mykhaylyk, V. In-vacuum long-wavelength macromolecular crystallography. *Acta Crystallogr D. Struct. Biol.* **72**, 430–439 (2016).
60. Winter, G. et al. DIALS as a toolkit. *Protein Sci.* **31**, 232–250 (2022).
61. Sheldrick, G. M. Experimental phasing with SHELXC/ D/ E: combining chain tracing with density modification. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 479–485 (2010).
62. Thorn, A. & Sheldrick, G. M. ANODE: anomalous and heavy-atom density calculation. *J. Appl Crystallogr* **44**, 1285–1287 (2011).
63. Terwilliger, T. C. et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. Sect. D., Biol. Crystallogr.* **64**, 61–69 (2008).

Acknowledgements

The authors would like to thank Dr Alicia Churchill-Angus for help with crystal data analysis and Diamond Light Source for beamtime (proposal mx23459), and the staff of beamlines I24, I04 and I23 for assistance with crystal testing and data collection. Llama immunisation was carried out in collaboration with the Rosalind Franklin Institute under the supervision of Dr Ray Owens and Dr Jiangdong Huo. The synthetic nanobody (sybody) libraries were obtained from Professor Markus Seeger, University of Zurich. This research was supported by Wellcome awards to S.N. (215519;219531), and through BBSRC grants to S.J.T. (BB/T002018/1) and MRC grants to J.L.P. (MR/S021043/1). M.L. is a UKRI MRC DPhil. student and supported by a Clarendon Fund Scholarship in partnership with a Christ Church Clarendon award. S.S. is a UKRI BBSRC DPhil. student supported by a studentship from the Doctoral Training Partnership award (BB/T008784/1).

Author contributions

J.L.P. and S.N. conceived the project. M.L., J.L.P. and S.N. performed protein expression and purification work and analysed the purification and liposome assay data. LCP crystallisation, data collection and structure determination were undertaken by M.L., J.L.P., and supervised by S.N. J.L.P. designed and supervised the in vitro assays conducted and analysed by M.L. M.L. generated and screened the sybodies and identified the nanobodies. S.J.T. supervised and analysed the two electrode voltage experiments conducted and analysed by S.P.S. K.E.O. and A.W. undertook the long wavelength data collection, phasing, and initial model building, aided by M.L. and S.N. M.L., S.P.S., J.L.P., and S.N. wrote the manuscript and prepared the figures, with assistance from all the authors.

Competing interests

The authors declare no competing interests.

Article https://doi.org/10.1038/s41467-022-32589-2

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-32589-2.

Correspondence and requests for materials should be addressed to Joanne L. Parker or Simon Newstead.

Peer review information *Nature Communications* thanks Joseph Lyons and Manuel Palacin for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

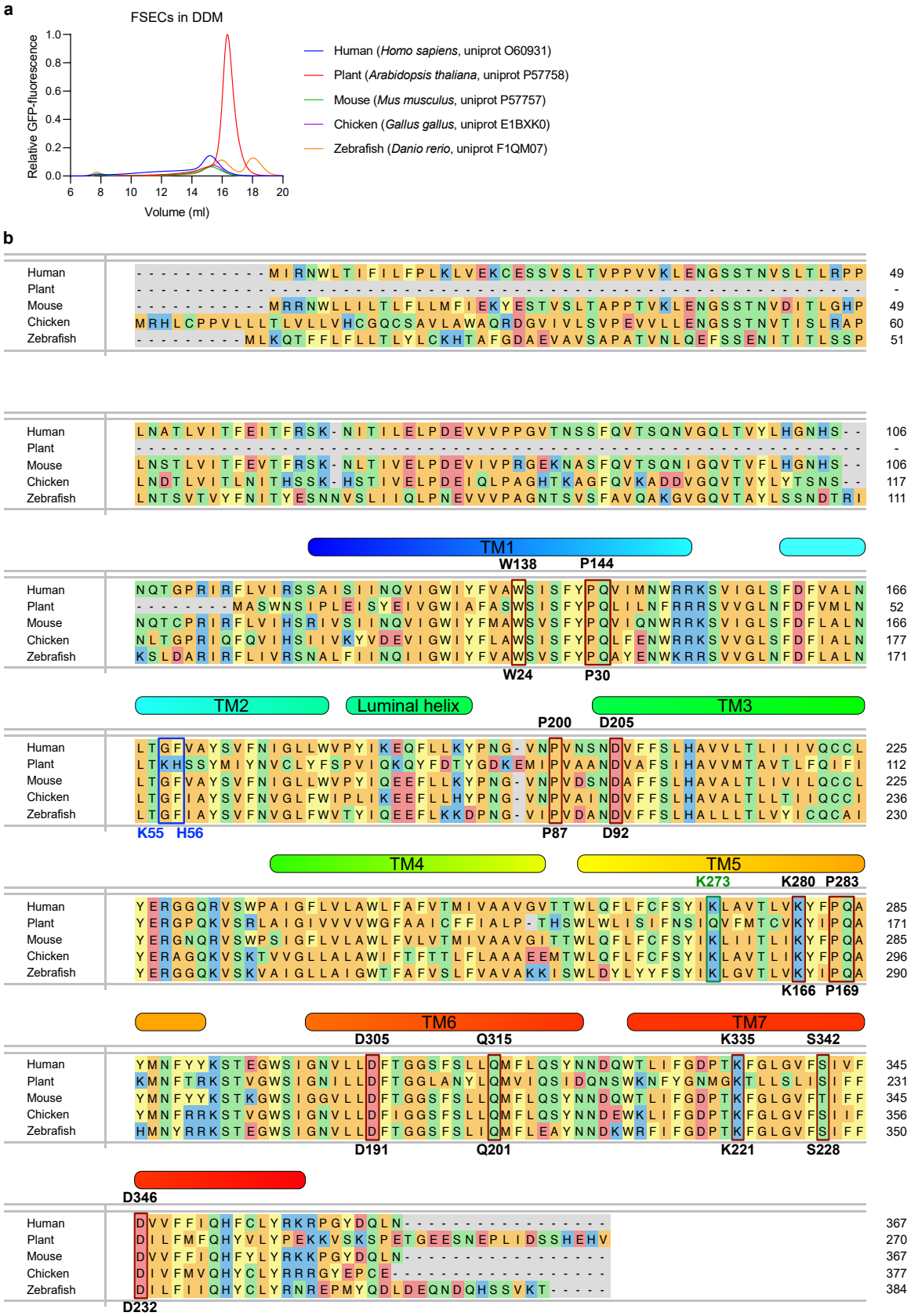
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022

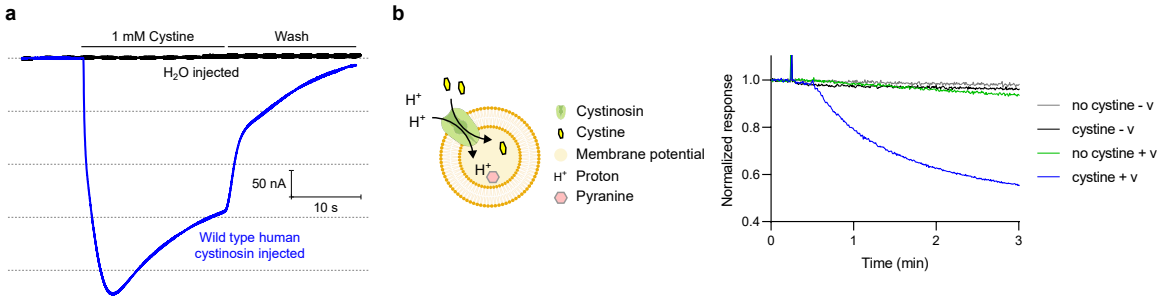
Supplementary Figures 1-7 & Supplementary Table 1

Structural basis for proton coupled cystine transport by cystinosin.

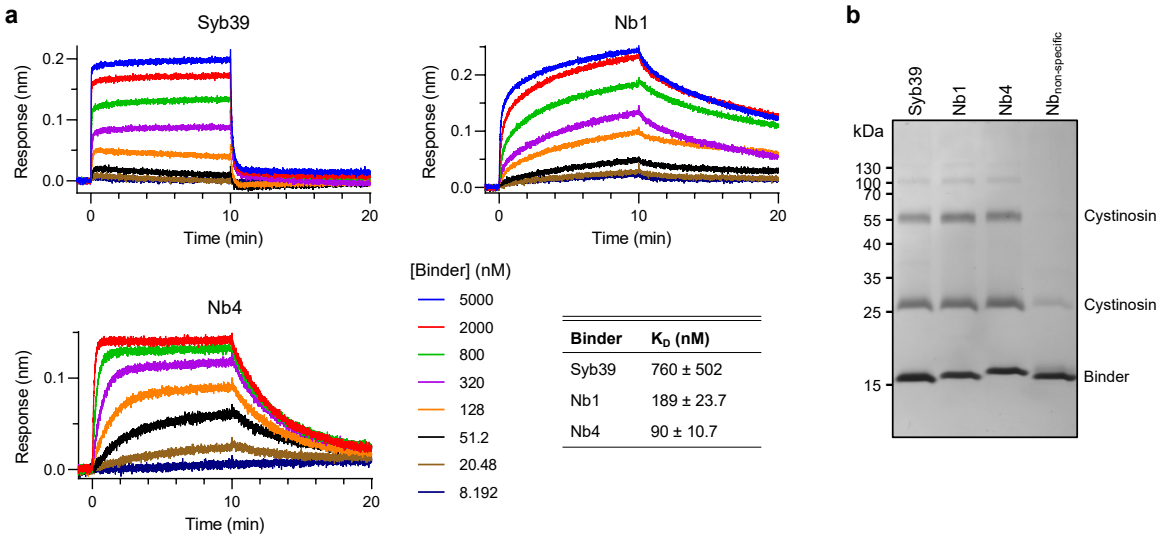
Mark Löbel[1], Sacha P Salphati[1], Kamel El Omari[2], Armin Wagner[2], Stephen Tucker[3,4], Joanne L Parker[1], Simon Newstead[1,4,]



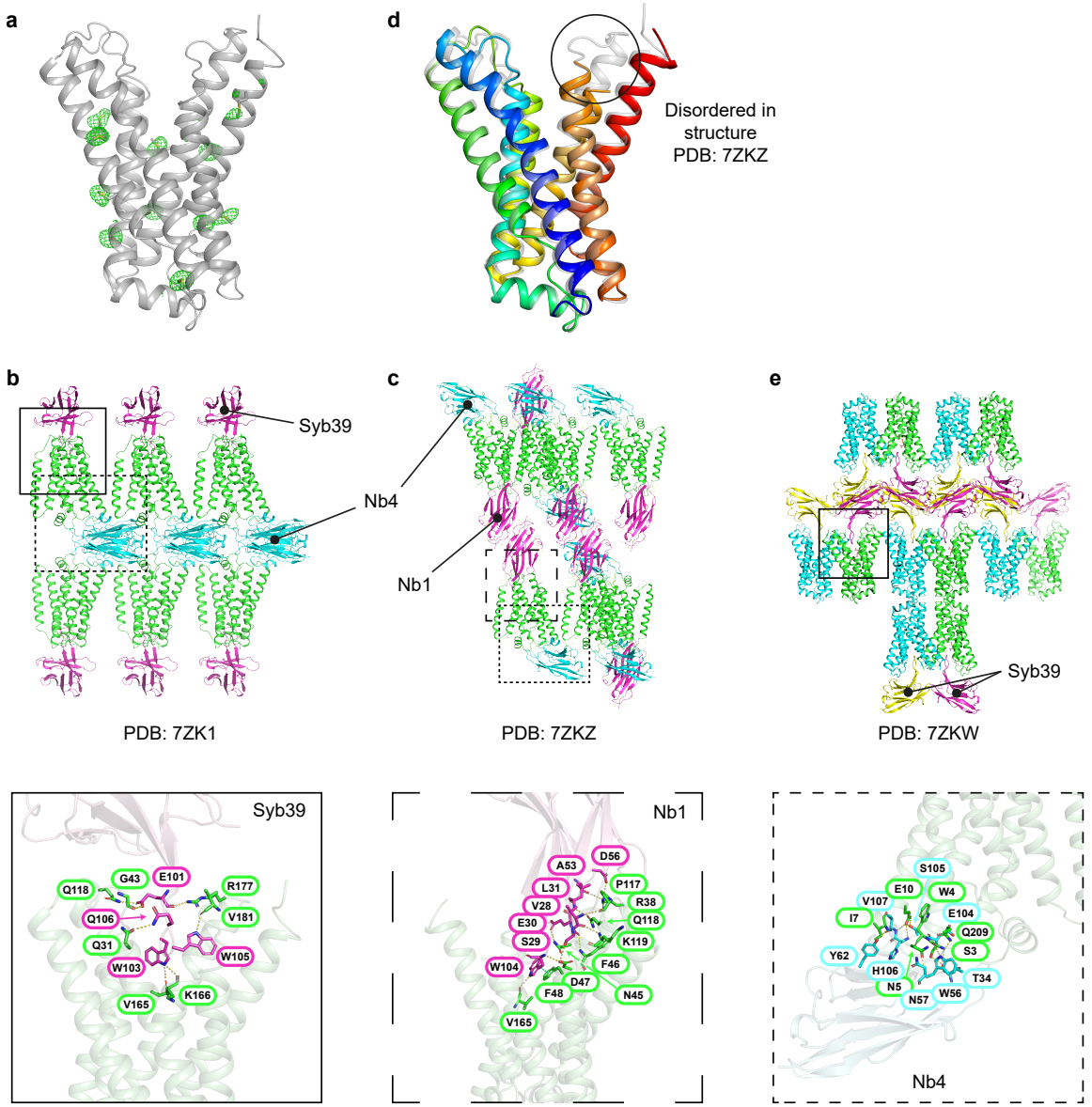
Supplementary Fig. 1: Analysis of cystinosin homologues. a Fluorescence size exclusion chromatography of five cystinosin homologues. Latin names are indicated along with the Uniprot code for each homologue. **b** Sequence alignment of cystinosin homologues (from a) coloured via amino acid chemistry. Residues with functional importance described in the manuscript are highlighted with the human numbering above the alignment and the plant (*Arabidopsis thaliana*) below. Highlighted in blue and green are the major differences between human and plant homologues found within the binding site of the protein.



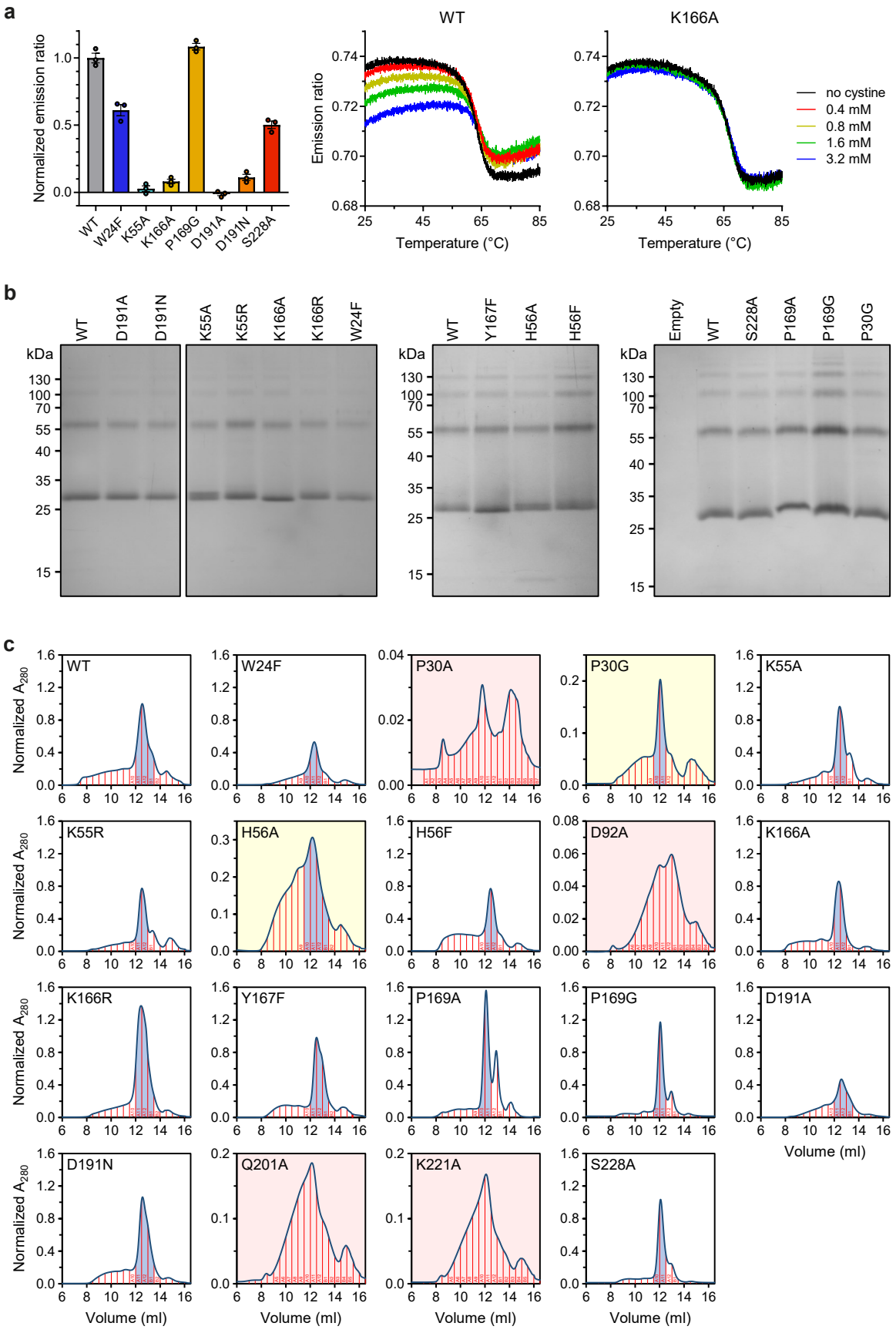
Supplementary Fig. 2: Exemplar data and controls for the assays used in the study. a A typical reading from TEVC on *Xenopus laevis* oocytes used to test human wild type cystinosin and mutant variants. Blue is the trace from oocytes injected with mRNA from WT cystinosin and black is an oocyte injected with water. **b** *Arabidopsis thaliana* cystinosin is a proton coupled cystine transporter. A reconstituted assay system was used to test plant cystinosin and mutant variants reading the change in fluorescence of an intra-liposomal fluorescent dye (pyranine) upon protonation. Cystine was added at 15 seconds and transport was initiated through the formation of a membrane potential through potassium diffusion gradient via the addition of valinomycin (v) at 30 seconds. While there is a very slight proton leak in the system in the presence of just a membrane potential (green - no cystine + v) robust protonation and hence transport is observed with the addition of cystine (blue).



Supplementary Fig. 3: Characterisation of the binders used in the study. a Representative curves for binding of cystinosin to the binders used in the study as determined using bilayer interferometry. The K_D for each binder was calculated from the mean of three independent experiments with the standard deviation shown. **b** Small-scale nickel-affinity pulldowns using His-tagged versions of the binders were used to screen for interactions between potential binders and cystinosin. Four independent repeats with similar results. Source data are available as a Source Data file.

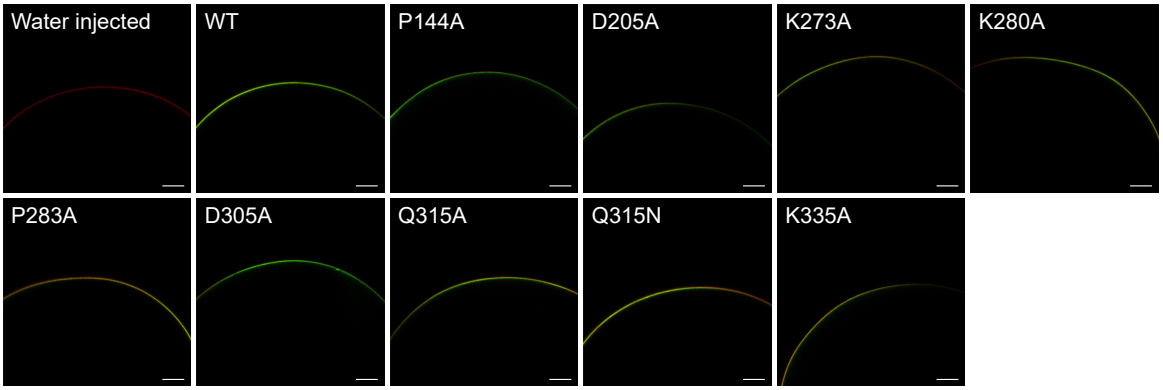


Supplementary Fig. 4: Anomalous difference maps, crystal packing and interactions for binder-cystinosin complexes. **a** Anomalous difference peaks calculated using Anode from data collected on I23 at Diamond Light Source (Table 1). Peaks corresponding to the location of the sulphur atoms are shown (green mesh) contoured at 3 σ mapped onto the final cystinosin model. **b** Crystal packing of structure PDB: 7KZ1, showing the contribution of the sybody and nanobody to the crystal contacts. **c** Crystal packing of structure PDB: 7ZKZ. **d** Overlay of the structures from PDB: 7ZK1 (grey) and 7ZKZ (rainbow) highlighting the region of disorder in the structure obtained from the two nanobody complex. **e** Crystal packing of structure PDB: 7ZKW. Interactions formed between cystinosin and Syb39, Nb1 and Nb4, respectively, are highlighted using yellow dashed lines, π-π stacking is indicated by cyan dashed lines.

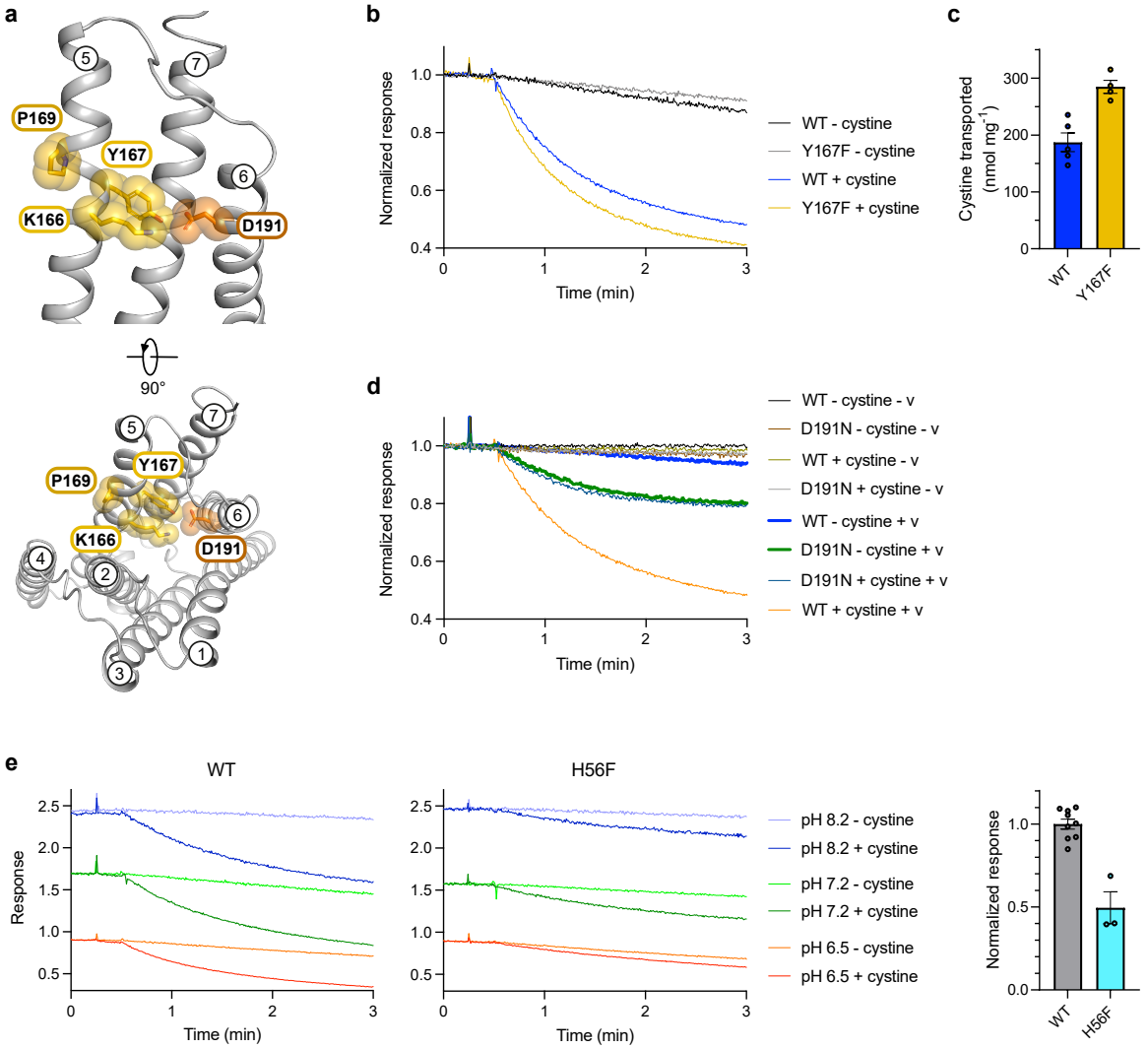


Supplementary Fig. 5: Analysis of the mutant variants. **a** Impact of cystinosin variants on cystine binding using a tryptophan fluorescence quench assay. n = 3 independent experiments for WT and variants, error bars SEM. Typical nano differential scanning fluorimetry (nanoDSF) traces are shown for wild type cystinosin and mutant Lys166. **b** Analysis of WT and variants of plant cystinosin studied in liposome

assays. SDS-PAGE gel of 2.5 μg of re-solubilised protein (in 1% DDM), taken after a 30-minute ultra-centrifugation step at 200,000*g*. Two independent repeats with similar results. **c** Size exclusion profiles of all wild type and mutant variants of cystinosin used in the study. The amount of protein obtained for some of the variants was considerably lower than WT in particular variants Pro30Ala, Asp92Ala, Gln201Ala, Lys221Ala (all coloured red) showed both a low amount of protein and poor quality as judged by the comparatively broad SEC profile. Mutants P30G and H56A (both coloured yellow) expressed lower than WT but yielded enough material for reconstitution. SEC fractions are labelled and pooled fractions are shaded blue-grey. Source data are available as a Source Data file.



Supplementary Fig. 6: Human cystinosin surface expression in oocytes. Cell surface expression of the human WT and mutant variants of cystinosin in the *Xenopus laevis* oocytes used for TEVC. A C-terminally GFP tagged construct was used for all variants tested and they all localised to the cell surface albeit to slightly different extents; with Pro283Ala, Gln315Ala and Lys335Ala showing lower expression compared to wild type. Scale bar 100 μm. Three independent repeats with similar results.



Supplementary Fig. 7: Analysis of Tyr167Phe, Asp191Asn and His56Phe variants of cystinosin. **a** Side and top view of the plant cystinosin crystal structure highlighting the position of Lys166, Tyr167, Pro169 and Asp191 side chains on transmembrane helices 5 and 6. **b** Tyr167Phe variant of cystinosin moves more protons in comparison to WT as determined using a pyranine based fluorescent *in vitro* assay. **c** In addition to the higher transport of protons shown in b, Tyr167Phe also transports more cystine per mg of protein. n = 5 independent experiments for WT and n = 4 for Y167F, error bars SEM. **d** Asp191Asn variant of cystinosin has a marked larger proton leak in the presence of a membrane potential (green line) in comparison to WT (blue line). In addition, this variant no longer responds to the presence of cystine with no additional fluorescence changed on addition of cystine (dark blue line in comparison to the WT, orange line). **e** WT cystinosin (left) functions over a wide pH range in particularly showing robust activity at pH 6.5. Whereas the His56Phe variant (right) shows a very low level of activity at pH 6.5. Representative traces from the pyranine based transport assay at pH 6.5, 7.2 and 8.2 in the presence (200 μM) or absence of cystine. Overall the activity of His56Phe is also reduced compared to WT (pH 7.2 shown). n = 9 independent experiments for WT and n = 3 for H56F, error bars SEM. Source data are available as a Source Data file.

Supplementary Table 1: List of cystinosis-causing mutations.

Mutation	Location	Clinical phenotype
V42I	N-terminal domain	Atypical/juvenile
G110V	N-terminal domain	Atypical
I133F	TM1	Infantile
S139F	TM1	Atypical/juvenile
S141F	TM1	Infantile
R151G	TM1	Infantile
G157D	TM1-2 loop	Infantile
L158P	TM1-2 loop	Infantile
G169D	TM2	Infantile
Y173H/C	TM2	Infantile
N177S/T	TM2	Infantile/juvenile
W182R	TM2	Infantile
K187R	Luminal helix (TM2-3 loop)	Infantile
G197R	TM2-3 loop	Ocular
P200L	TM2-3 loop	Juvenile
D205N	TM3	Infantile
T216R	TM3	Infantile
Q222R	TM3	Infantile
S270F/Y	TM5	Infantile
K280R	TM5	Juvenile
M287I	TM5	Infantile
N288K	TM5	Infantile
S293I	TM5-6 loop	Infantile
S298N	TM5-6 loop	Infantile
D305G/Y	TM6	Infantile
G308R/E/V	TM6	Infantile
G309D/V	TM6	Infantile
N323K	TM6	Juvenile
T334N	TM7	Juvenile
G337R	TM7	Infantile
L338R/P	TM7	Infantile
G339R	TM7	Infantile
D346N	TM7	Atypical/juvenile
G362R	C-terminal tail	Infantile

Bibliography

Abderhalden, E. (1903). Familiäre Cystindiathese. *Z. Physiol. Chem.* **38**, 557–561.

Abergel, C. (2013). Molecular replacement: Tricks and treats. *Acta Crystallogr. D* **69**, 2167–2173.

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W. *et al.* (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221.

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Echols, N., Headd, J. J., Hung, L.-W., Jain, S., Kapral, G. J., Kunstleve, R. W. G. *et al.* (2011). The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106.

Afonine, P. V., Poon, B. K., Read, R. J., Sobolev, O. V., Terwilliger, T. C., Urzhumtsev, A. & Adams, P. D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544.

Agmon, N. (1995). The grotthuss mechanism. *Chem. Phys. Lett.* **244**, 456–462.

Alcantara-Ortigoza, M. A., Belmont-Martinez, L., Vela-Amieva, M. & Gonzalez-Del Angel, A. (2008). Analysis of the CTNS gene in nephropathic cystinosis Mexican patients: Report of four novel mutations and identification of a false positive 57-kb deletion genotype with LDM-2/exon 4 multiplex PCR assay. *Genet. Test.* **12**, 409–414.

Aldahmesh, M. A., Humeidan, A., Almojalli, H. A., Khan, A. O., Rajab, M., Al-Abbad, A. A., F. Meyer, B. & Alkuraya, F. S. (2009). Characterization of CTNS mutations in Arab patients with cystinosis. *Ophthalmic Genet.* **30**, 185–189.

Andrzejewska, Z., Névo, N., Thomas, L., Bailleux, A., Chauvet, V., Benmerah, A. & Antignac, C. (2015). Lysosomal targeting of cystinosin requires AP-3. *Traffic* **16**, 712–726.

Andrzejewska, Z., Nevo, N., Thomas, L., Chhuon, C., Bailleux, A., Chauvet, V., Courtoy, P. J., Chol, M., Guerrera, I. C. & Antignac, C. (2016). Cystinosin is a component of the vacuolar H^+-ATPase-ragulator-rag complex controlling mammalian target of rapamycin complex 1 signaling. *J. Am. Soc. Nephrol.* **27**, 1678–1688.

Anikster, Y., Shotelersuk, V. & Gahl, W. A. (1999). CTNS mutations in patients with cystinosis. *Hum. Mutat.* **14**, 454–458.

Aricescu, A. R., Lu, W. & Jones, E. Y. (2006). A time-and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr. D* **62**, 1243–1250.

Arndt, U. W., Crowther, R. A. & Mallett, J. F. W. (1968). A computer-linked cathode-ray tube microdensitometer for X-ray crystallography. *J. Phys. E Sci. Instrum.* **1**, 510.

Arvai, A. (2021). Adxv (https://www.scripps.edu/tainer/arvai/adxv.html). San Diego, California, USA: The Scripps Research Institute.

Asarnow, D., Palovcak, E. & Cheng, Y. (2019). UCSF pyem v0.5. *Zenodo* (https://doi.org/10.5281/zenodo.3576630).

Attard, M., Jean, G., Forestier, L., Cherqui, S., van't Hoff, W., Broyer, M., Antignac, C. & Town, M. (1999). Severity of phenotype in cystinosis varies with mutations in the CTNS gene: Predicted effect on the model of cystinosin. *Hum. Mol. Genet.* **8**, 2507–2514.

Ballabio, A. & Bonifacino, J. S. (2020). Lysosomes as dynamic regulators of cell and organismal homeostasis. *Nat. Rev. Mol. Cell Biol.* **21**, 101–118.

Bannai, S. (1984). Transport of cystine and cysteine in mammalian cells. *Biochim. Biophys. Acta* **779**, 289–306.

Bannai, S. & Kitamura, E. (1981). Role of proton dissociation in the transport of cystine and glutamate in human diploid fibroblasts in culture. *J. Biol. Chem.* **256**, 5770–5772.

Beckett, D., Kovaleva, E. & Schatz, P. J. (1999). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* **8**, 921–929.

Beilsten-Edmands, J., Winter, G., Gildea, R., Parkhurst, J., Waterman, D. & Evans, G. (2020). Scaling diffraction data in the DIALS software package: Algorithms and new approaches for multi-crystal scaling. *Acta Crystallogr. D* **76**, 385–399.

Bellomo, F., Taranta, A., Petrini, S., Venditti, R., Rocchetti, M. T., Rega, L. R., Corallini, S., Gesualdo, L., De Matteis, M. A. & Emma, F. (2016). Carboxyl-terminal SSLKG motif of the human cystinosin-LKG plays an important role in plasma membrane sorting. *PLOS ONE* **11**, e0154805.

Better, M., Chang, C. P., Robinson, R. R. & Horwitz, A. H. (1988). Escherichia coli secretion of an active chimeric antibody fragment. *Science* **240**, 1041–1043.

Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S. & Bricogne, G. (2004). Refinement of severely incomplete structures with maximum likelihood in BUSTER–TNT. *Acta Crystallogr. D* **60**, 2210–2221.

Bonifacino, J. S. & Traub, L. M. (2003). Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu. Rev. Biochem.* **72**, 395–447.

Bos, E. S., van der Doelen, A. A., van Rooy, N. & Schuurs, A. H. W. M. (1981). 3,3',5,5'-Tetramethylbenzidine as an Ames test negative chromogen for horse-radish peroxidase in enzyme-immunoassay. *J. Immunoass. Immunochem.* **2**, 187–204.

Bragg, W. H. & Bragg, W. L. (1913). The reflection of X-rays by crystals. *Proc. Roy. Soc.* **88**, 428–438.

Bräuer, P., Parker, J. L., Gerondopoulos, A., Zimmermann, I., Seeger, M. A., Barr, F. A. & Newstead, S. (2019). Structural basis for pH-dependent retrieval of ER proteins from the Golgi by the KDEL receptor. *Science* **363**, 1103–1107.

Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O. S., Vonrhein, C. & Womack, T. O. (2017). BUSTER version 2.10.4. Cambridge, UK: Global Phasing Ltd.

Brünger, A. T. (1992). Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475.

Caesar, J., Reboul, C. F., Machello, C., Kiesewetter, S., Tang, M. L., Deme, J. C., Johnson, S., Elmlund, D., Lea, S. M. & Elmlund, H. (2020). SIMPLE 3.0. Stream single-particle cryo-EM analysis in real time. *J. Struct. Biol.: X* **4**, 100040.

Caffrey, M. & Cherezov, V. (2009). Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* **4**, 706–731.

Callis, J. (1995). Regulation of protein degradation. *Plant Cell* **7**, 845.

Casanas, A., Warshamanage, R., Finke, A. D., Panepucci, E., Olieric, V., Nöll, A., Tampé, R., Brandstetter, S., Förster, A., Mueller, M. *et al.* (2016). EIGER detector: Application in macromolecular crystallography. *Acta Crystallogr. D* **72**, 1036–1048.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21.

Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H.-J., Kuhn, P., Weis, W. I., Kobilka, B. K. *et al.* (2007). High-resolution crystal structure of an engineered human β_2-adrenergic G protein-coupled receptor. *Science* **318**, 1258–1265.

Cherqui, S. & Courtoy, P. J. (2017). The renal Fanconi syndrome in cystinosis: Pathogenic insights and therapeutic perspectives. *Nat. Rev. Nephrol.* **13**, 115–131.

Cherqui, S., Kalatzis, V., Trugnan, G. & Antignac, C. (2001). The targeting of cystinosin to the lysosomal membrane requires a tyrosine-based signal and a novel sorting motif. *J. Biol. Chem.* **276**, 13314–13321.

Chothia, C. & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5**, 823–826.

Clement, N. R. & Gould, J. M. (1981). Pyranine (8-hydroxy-1, 3, 6-pyrenetrisulfonate) as a probe of internal aqueous hydrogen ion concentration in phospholipid vesicles. *Biochemistry* **20**, 1534–1538.

Congreve, M., de Graaf, C., Swain, N. A. & Tate, C. G. (2020). Impact of GPCR structures on drug discovery. *Cell* **181**, 81–91.

Cormack, B. P., Bertram, G., Egerton, M., Gow, N. A., Falkow, S. & Brown, A. J. (1997). Yeast-enhanced green fluorescent protein (yEGFP): A reporter of gene expression in Candida albicans. *Microbiology* **143**, 303–311.

Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190.

David, D., Princiero Berlingerio, S., Elmonem, M. A., Oliveira Arcolino, F., Soliman, N., van den Heuvel, B., Gijsbers, R. & Levtchenko, E. (2019). Molecular basis of cystinosis: Geographic distribution, functional consequences of mutations in the CTNS gene, and potential for repair. *Nephron* **141**, 133–146.

De Duve, C. (1963). The lysosome. *Sci. Am.* **208**, 64–73.

De Duve, C., Pressman, B., Gianetto, R., Wattiaux, R. & Appelmans, F. (1955). Tissue fractionation studies. 6. Intracellular distribution patterns of enzymes in rat-liver tissue. *Biochem. J.* **60**, 604–617.

Delagenière, S., Brenchereau, P., Launer, L., Ashton, A. W., Leal, R., Veyrier, S., Gabadinho, J., Gordon, E. J., Jones, S. D., Levik, K. E. *et al.* (2011). ISPyB: An information management system for synchrotron macromolecular crystallography. *Bioinformatics* **27**, 3186–3192.

Di Guan, C., Li, P., Riggs, P. D. & Inouye, H. (1988). Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. *Gene* **67**, 21–30.

Diederichs, K. & Karplus, P. A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* **4**, 269–275.

Diederichs, K. & Karplus, P. A. (2013). Better models by discarding data? *Acta Crystallogr. D* **69**, 1215–1222.

Drew, D. & Boudker, O. (2016). Shared molecular mechanisms of membrane transporters. *Annu. Rev. Biochem.* **85**, 543–572.

Drew, D., Newstead, S., Sonoda, Y., Kim, H., von Heijne, G. & Iwata, S. (2008). GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae. *Nat. Protoc.* **3**, 784–798.

Dunbrack Jr, R. L. & Karplus, M. (1994). Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nat. Struct. Biol.* **1**, 334–340.

El Omari, K., Duman, R., Mykhaylyk, V., Orr, C. M., Latimer-Smith, M., Winter, G., Grama, V., Qu, F., Bountra, K., Kwong, H. S. *et al.* (2023). Experimental phasing opportunities for macromolecular crystallography at very long wavelengths. *Commun. Chem.* **6**, 219.

Emsley, P. & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta crystallogr. D* **60**, 2126–2132.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501.

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82.

Evans, P. & McCoy, A. (2008). An introduction to molecular replacement. *Acta Crystallogr. D* **64**, 1–10.

Evans, P. R. (2011). An introduction to data reduction: Space-group determination, scaling and intensity statistics. *Acta Crystallogr. D* **67**, 282–292.

Evans, P. R. & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214.

Fairhead, M. & Howarth, M. (2015). Site-specific biotinylation of purified proteins using BirA. *Methods Mol. Biol.* **1266**, 171–184.

Fantini, J. & Barrantes, F. J. (2013). How cholesterol interacts with membrane proteins: An exploration of cholesterol-binding sites including CRAC, CARC, and tilted domains. *Front. Physiol.* **4**, 31.

Fantus, D., Rogers, N. M., Grahammer, F., Huber, T. B. & Thomson, A. W. (2016). Roles of mTOR complexes in the kidney: Implications for renal disease and transplantation. *Nat. Rev. Nephrol.* **12**, 587–609.

Feng, L. & Frommer, W. B. (2016). Evolution of transporters: The relationship of SWEETs, PQ-loop, and PnuC transporters. *Trends Biochem. Sci.* **41**, 118–119.

Fisher, S., Levik, K., Williams, M., Ashton, A. & McAuley, K. (2015). SynchWeb: A modern interface for ISPyB. *J. Appl. Crystallogr.* **48**, 927–932.

Frost, L., Suryadevara, P., Cannell, S. J., Groundwater, P. W., Hambleton, P. A. & Anderson, R. J. (2016). Synthesis of diacylated γ-glutamyl-cysteamine prodrugs, and in vitro evaluation of their cytotoxicity and intracellular delivery of cysteamine. *Eur. J. Med. Chem.* **109**, 206–215.

Gahl, W. A., Bashan, N., Tietze, F., Bernardini, I. & Schulman, J. D. (1982). Cystine transport is defective in isolated leukocyte lysosomes from patients with cystinosis. *Science* **217**, 1263–1265.

Gahl, W. A., Reed, G. F., Thoene, J. G., Schulman, J. D., Rizzo, W. B., Jonas, A. J., Denman, D. W., Schlesselman, J. J., Corden, B. J. & Schneider, J. A. (1987). Cysteamine therapy for children with nephropathic cystinosis. *N. Engl. J. Med.* **316**, 971–977.

Gahl, W. A., Thoene, J. G. & Schneider, J. A. (2002). Cystinosis. *N. Engl. J. Med.* **347**, 111–121.

Gao, X.-D., Wang, J., Keppler-Ross, S. & Dean, N. (2005). ERS1 encodes a functional homologue of the human lysosomal cystine transporter. *FEBS J.* **272**, 2497–2511.

Geertsma, E. R. (2013). FX cloning: A versatile high-throughput cloning system for characterization of enzyme variants. *Methods Mol. Biol.* **978**, 133–148.

Gerondopoulos, A., Bräuer, P., Sobajima, T., Wu, Z., Parker, J. L., Biggin, P. C., Barr, F. A. & Newstead, S. (2021). A signal capture and proofreading mechanism for the KDEL-receptor explains selectivity and dynamic range in ER retrieval. *eLife* **10**, e68380.

Green, N. M. (1963). Avidin. 1. The use of [14C] biotin for kinetic studies and for assay. *Biochem. J.* **89**, 585–591.

Guo, X., Schmiege, P., Assafa, T. E., Wang, R., Xu, Y., Donnelly, L., Fine, M., Ni, X., Jiang, J., Millhauser, G. *et al.* (2022). Structure and mechanism of human cystine exporter cystinosin. *Cell* **185**, 3739–3752.

Han, J. & Burgess, K. (2010). Fluorescent indicators for intracellular pH. *Chem. Rev.* **110**, 2709–2728.

Han, L., Zhu, Y., Liu, M., Zhou, Y., Lu, G., Lan, L., Wang, X., Zhao, Y. & Zhang, X. C. (2017). Molecular mechanism of substrate recognition and transport by the AtSWEET13 sugar transporter. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 10089–10094.

Hattori, M., Hibbs, R. E. & Gouaux, E. (2012). A fluorescence-detection size-exclusion chromatography-based thermostability assay for membrane protein precrystallization screening. *Structure* **20**, 1293–1299.

Hill, J., Donald, K. A., Griffiths, D. E. & Donald, G. (1991). DMSO-enhanced whole cell yeast transformation. *Nucleic Acids Res.* **19**, 5791.

Hochuli, E., Döbeli, H. & Schacher, A. (1987). New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *J. Chromatogr. A* **411**, 177–184.

Hodgkin, D. C. (1965). The X-ray analysis of complicated molecules. *Science* **150**, 979–988.

Hopp, T. P., Prickett, K. S., Price, V. L., Libby, R. T., March, C. J., Pat Cerretti, D., Urdal, D. L. & Conlon, P. J. (1988). A short polypeptide marker sequence useful for recombinant protein identification and purification. *Nat. Biotechnol.* **6**, 1204–1210.

Huerta-Cepas, J., Serra, F. & Bork, P. (2016). ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638.

Ito, H., Fukuda, Y., Murata, K. & Kimura, A. (1983). Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.* **153**, 163–168.

Jacoby, E., Bouhelal, R., Gerspacher, M. & Seuwen, K. (2006). The 7 TM G-protein-coupled receptor target family. *ChemMedChem* **1**, 760–782.

Jaehme, M., Guskov, A. & Slotboom, D. J. (2014). Crystal structure of the vitamin B3 transporter PnuC, a full-length SWEET homolog. *Nat. Struct. Mol. Biol.* **21**, 1013–1015.

Jaehme, M., Guskov, A. & Slotboom, D. J. (2015). The twisted relation between Pnu and SWEET transporters. *Trends Biochem. Sci.* **40**, 183–188.

Jaehme, M., Guskov, A. & Slotboom, D. J. (2016). Pnu transporters: Ain't they SWEET? *Trends Biochem. Sci.* **41**, 117–118.

Jamalpoor, A., Othman, A., Levtchenko, E. N., Masereeuw, R. & Janssen, M. J. (2021). Molecular mechanisms and treatment options of nephropathic cystinosis. *Trends Mol. Med.* **27**, 673–686.

Jaquinod, M., Villiers, F., Kieffer-Jaquinod, S., Hugouvieux, V., Bruley, C., Garin, J. & Bourguignon, J. (2007). A proteomics dissection of Arabidopsis thaliana vacuoles isolated from cell culture. *Mol. Cell. Proteom.* **6**, 394–412.

Jardetzky, O. (1966). Simple allosteric model for membrane pumps. *Nature* **211**, 969–970.

Jézégou, A., Llinares, E., Anne, C., Kieffer-Jaquinod, S., O'Regan, S., Aupetit, J., Chabli, A., Sagné, C., Debacker, C., Chadefaux-Vekemans, B. *et al.* (2012). Heptahelical protein PQLC2 is a lysosomal cationic amino acid exporter underlying the action of cysteamine in cystinosis therapy. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E3434–E3443.

Jia, J., Zhu, L., Yue, X., Tang, S., Jing, S., Tan, C., Du, Y., Gao, J., Lee, I. & Qian, Y. (2024). Crosstalk between KDEL receptor and EGF receptor mediates cell proliferation and migration via STAT3 signaling. *Cell Commun. Signal.* **22**, 140.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A. *et al.* (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589.

Kabsch, W. (2010). xds. *Acta Crystallogr. D* **66**, 125–132.

Kaczorowski, G. J., Robertson, D. E. & Kaback, H. R. (1979). Mechanism of lactose translocation in membrane vesicles from Escherichia coli. 2. Effect of imposed $\Delta\Psi$, ΔpH, and $\Delta\mu_{H^+}$. *Biochemistry* **18**, 3697–3704.

Kaiser-Kupfer, M. I., Fujikawa, L., Kuwabara, T., Jain, S. & Gahl, W. A. (1987). Removal of corneal crystals by topical cysteamine in nephropathic cystinosis. *N. Engl. J. Med.* **316**, 775–779.

Kalatzis, V., Cherqui, S., Antignac, C. & Gasnier, B. (2001). Cystinosin, the protein defective in cystinosis, is a H^+-driven lysosomal cystine transporter. *EMBO J.* **20**, 5940–5949.

Kalatzis, V., Cohen-Solal, L., Cordier, B., Frishberg, Y., Kemper, M., Nuutinen, E. M., Legrand, E., Cochat, P. & Antignac, C. (2002). Identification of 14 novel CTNS mutations and characterization of seven splice site mutations associated with cystinosis. *Hum. Mutat.* **20**, 439–446.

Kalatzis, V., Nevo, N., Cherqui, S., Gasnier, B. & Antignac, C. (2004). Molecular pathogenesis of cystinosis: Effect of CTNS mutations on the transport activity and subcellular localization of cystinosin. *Hum. Mol. Genet.* **13**, 1361–1371.

Kantardjieff, K. A. & Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Sci.* **12**, 1865–1871.

Kapust, R. B. & Waugh, D. S. (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8**, 1668–1674.

Karplus, P. A. & Diederichs, K. (2012). Linking crystallographic model and data quality. *Science* **336**, 1030–1033.

Karplus, P. A. & Diederichs, K. (2015). Assessing and maximizing data quality in macromolecular crystallography. *Curr. Opin. Struct. Biol.* **34**, 60–68.

Kawano-Kawada, M., Manabe, K., Ichimura, H., Kimura, T., Harada, Y., Ikeda, K., Tanaka, S., Kakinuma, Y. & Sekito, T. (2019). A PQ-loop protein Ypq2 is involved in the exchange of arginine and histidine across the vacuolar membrane of Saccharomyces cerevisiae. *Sci. Rep.* **9**, 15018.

Kawano-Kawada, M., Ueda, T., Mori, H., Ichimura, H., Takegawa, K. & Sekito, T. (2021). Stm1 is a vacuolar PQ-loop protein involved in the transport of basic amino acids in Schizosaccharomyces pombe. *Biochim. Biophys. Acta – Biomembr.* **1863**, 183507.

Keenan, R. J., Freymann, D. M., Stroud, R. M. & Walter, P. (2001). The signal recognition particle. *Annu. Rev. Biochem.* **70**, 755–775.

Kiehntopf, M., Schickel, J., von der Gönne, B., Koch, H. G., Superti-Furga, A., Steinmann, B., Deufel, T. & Harms, E. (2002). Analysis of the CTNS gene in patients of German and Swiss origin with nephropathic cystinosis. *Hum. Mutat.* **20**, 237–237.

Kim, S. (2021). Exploring chemical information in PubChem. *Curr. Protoc.* **1**, e217.

Kraft, P., Bergamaschi, A., Broennimann, C., Dinapoli, R., Eikenberry, E., Henrich, B., Johnson, I., Mozzanica, A., Schlepütz, C., Willmott, P. *et al.* (2009). Performance of single-photon-counting PILATUS detector modules. *J. Synchrotron Radiat.* **16**, 368–375.

Krebs, M., Beyhl, D., Görlich, E., Al-Rasheid, K. A., Marten, I., Stierhof, Y.-D., Hedrich, R. & Schumacher, K. (2010). Arabidopsis v-ATPase activity at the tonoplast is required for efficient nutrient storage but not for sodium accumulation. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3251–3256.

Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**, 567–580.

Latorraca, N. R., Fastman, N. M., Venkatakrishnan, A., Frommer, W. B., Dror, R. O. & Feng, L. (2017). Mechanism of substrate translocation in an alternating access transporter. *Cell* **169**, 96–107.

Lawrence, R. E. & Zoncu, R. (2019). The lysosome as a cellular centre for signalling, metabolism and quality control. *Nat. Cell Biol.* **21**, 133–142.

Lee, Y., Nishizawa, T., Yamashita, K., Ishitani, R. & Nureki, O. (2015). Structural basis for the facilitative diffusion mechanism by SemiSWEET transporter. *Nat. Commun.* **6**, 6112.

Lei, H.-T., Ma, J., Sanchez Martinez, S. & Gonen, T. (2018). Crystal structure of arginine-bound lysosomal transporter SLC38A9 in the cytosol-open state. *Nat. Struct. Mol. Biol.* **25**, 522–527.

Lei, S.-P., Lin, H. C., Wang, S.-S., Callaway, J. & Wilcox, G. (1987). Characterization of the Erwinia carotovora pelB gene and its product pectate lyase. *J. Bacteriol.* **169**, 4379–4383.

Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C. & Adams, P. D. (2017). Polder maps: Improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. D* **73**, 148–157.

Liu, B., Du, H., Rutkowski, R., Gartner, A. & Wang, X. (2012). LAAT-1 is the lysosomal lysine/arginine transporter that maintains amino acid homeostasis. *Science* **337**, 351–354.

Löbel, M., Salphati, S. P., El Omari, K., Wagner, A., Tucker, S. J., Parker, J. L. & Newstead, S. (2022). Structural basis for proton coupled cystine transport by cystinosin. *Nat. Commun.* **13**, 4845.

Lovell, S. C., Davis, I. W., Arendall III, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins* **50**, 437–450.

Macías-Vidal, J., Rodés, M., Hernández-Pérez, J., Vilaseca, M. & Coll, M. (2009). Analysis of the CTNS gene in 32 cystinosis patients from Spain. *Clin. Genet.* **76**, 486–489.

Malekzadeh, M. H., Neustein, H. B., Schneider, J. A., Pennisi, A. J., Ettenger, R. B., Uittenbogaart, C. H., Kogut, M. D. & Fine, R. N. (1977). Cadaver renal transplantation in children with cystinosis. *Am. J. Med.* **63**, 525–533.

Manceau, A. & Nagy, K. L. (2008). Relationships between Hg (II)–S bond distance and Hg (II) coordination in thiolates. *Dalton Trans.* 1421–1425.

Manglik, A., Kobilka, B. K. & Steyaert, J. (2017). Nanobodies to study G protein-coupled receptor structure and function. *Annu. Rev. Pharmacol. Toxicol.* **57**, 19–37.

Mason, S., Pepe, G., Dall'Amico, R., Tartaglia, S., Casciani, S., Greco, M., Bencivenga, P., Murer, L., Rizzoni, G., Tenconi, R. *et al.* (2003). Mutational spectrum of the CTNS gene in Italy. *Eur. J. Hum. Genet.* **11**, 503–508.

McCoy, A. J. (2004). Liking likelihood. *Acta Crystallogr. D* **60**, 2169–2183.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.

McDowell, G. A., Gahl, W. A., Stephenson, L. A., Schneider, J. A., Weissenbach, J., Polymeropoulos, M. H., Town, M. M., van't Hoff, W., Farrall, M. & Mathew, C. G. (1995). Linkage of the gene for cystinosis to markers on the short arm of chromosome 17. *Nat. Genet.* **10**, 246–248.

McGowan-Jordan, J., Stoddard, K., Podolsky, L., Orrbine, E., McLaine, P., Town, M., Goodyer, P., MacKenzie, A. & Heick, H. (1999). Molecular analysis of cystinosis: Probable Irish origin of the most common French Canadian mutation. *Eur. J. Hum. Genet.* **7**, 671–678.

Minhas, G. S., Bawdon, D., Herman, R., Rudden, M., Stone, A. P., James, A. G., Thomas, G. H. & Newstead, S. (2018). Structural basis of malodour precursor transport in the human axilla. *eLife* **7**, e34995.

Mueller, M., Wang, M. & Schulze-Briese, C. (2012). Optimal fine φ-slicing for single-photon-counting pixel detectors. *Acta Crystallogr. D* **68**, 42–56.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367.

Muyldermans, S. (2013). Nanobodies: Natural single-domain antibodies. *Annu. Rev. Biochem.* **82**, 775–797.

Napolitano, G. & Ballabio, A. (2016). TFEB at a glance. *J. Cell Sci.* **129**, 2475–2481.

Nevo, N., Thomas, L., Chhuon, C., Andrzejewska, Z., Lipecka, J., Guillonneau, F., Bailleux, A., Edelman, A., Antignac, C. & Guerrera, I. C. (2017). Impact of cystinosin glycosylation on protein stability by differential dynamic stable isotope labeling by amino acids in cell culture (SILAC). *Mol. Cell. Proteom.* **16**, 457–468.

Newstead, S. & Barr, F. (2020). Molecular basis for KDEL-mediated retrieval of escaped ER-resident proteins – SWEET talking the COPs. *J. Cell Sci.* **133**, jcs250100.

Newstead, S., Kim, H., von Heijne, G., Iwata, S. & Drew, D. (2007). High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13936–13941.

Nilsson, I., Sääf, A., Whitley, P., Gafvelin, G., Waller, C. & von Heijne, G. (1998). Proline-induced disruption of a transmembrane α-helix in its natural environment. *J. Mol. Biol.* **284**, 1165–1175.

Oeffner, R. D., Afonine, P. V., Millán, C., Sammito, M., Usón, I., Read, R. J. & McCoy, A. J. (2018). On the application of the expected log-likelihood gain to decision making in molecular replacement. *Acta Crystallogr. D* **74**, 245–255.

Ohkuma, S. & Poole, B. (1978). Fluorescence probe measurement of the intralysosomal pH in living cells and the perturbation of pH by various agents. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3327–3331.

Padilla, J. E. & Yeates, T. O. (2003). A statistic for local intensity differences: Robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Crystallogr. D* **59**, 1124–1130.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Trong, I. L., Teller, D. C., Okada, T., Stenkamp, R. E. *et al.* (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739–745.

Pardon, E., Laeremans, T., Triest, S., Rasmussen, S. G., Wohlkönig, A., Ruf, A., Muyldermans, S., Hol, W. G., Kobilka, B. K. & Steyaert, J. (2014). A general protocol for the generation of Nanobodies for structural biology. *Nature Protoc.* **9**, 674–693.

Park, M., Helip-Wooley, A. & Thoene, J. (2002). Lysosomal cystine storage augments apoptosis in cultured human fibroblasts and renal tubular epithelial cells. *J. Am. Soc. Nephrol.* **13**, 2878–2887.

Parker, J. L., Corey, R. A., Stansfeld, P. J. & Newstead, S. (2019). Structural basis for substrate specificity and regulation of nucleotide sugar transporters in the lipid bilayer. *Nat. Commun.* **10**, 4657.

Parker, J. L., Deme, J. C., Kuteyi, G., Wu, Z., Huo, J., Goldman, I. D., Owens, R. J., Biggin, P. C., Lea, S. M. & Newstead, S. (2021a). Structural basis of antifolate recognition and transport by PCFT. *Nature* **595**, 130–134.

Parker, J. L., Deme, J. C., Wu, Z., Kuteyi, G., Huo, J., Owens, R. J., Biggin, P. C., Lea, S. M. & Newstead, S. (2021b). Cryo-EM structure of PepT2 reveals structural basis for proton-coupled peptide and prodrug transport in mammals. *Sci. Adv.* **7**, eabh3355.

Parker, J. L., Li, C., Brinth, A., Wang, Z., Vogeley, L., Solcan, N., Ledderboge-Vucinic, G., Swanson, J. M., Caffrey, M., Voth, G. A. *et al.* (2017). Proton movement and coupling in the POT family of peptide transporters. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 13182–13187.

Parker, J. L., Mindell, J. A. & Newstead, S. (2014). Thermodynamic evidence for a dual transport mechanism in a POT peptide transporter. *eLife* **3**, e04273.

Parker, J. L. & Newstead, S. (2013). Phasing statistics for alpha helical membrane protein structures. *Protein Sci.* **22**, 1664–1668.

Parker, J. L. & Newstead, S. (2014). Method to increase the yield of eukaryotic membrane protein expression in Saccharomyces cerevisiae for structural and functional studies. *Protein Sci.* **23**, 1309–1314.

Parsons, S. (2003). Introduction to twinning. *Acta Crystallogr. D* **59**, 1995–2003.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82.

Platt, F. M., D'Azzo, A., Davidson, B. L., Neufeld, E. F. & Tifft, C. J. (2018). Lysosomal storage diseases. *Nat. Rev. Dis. Primers* **4**, 27.

Ponting, C. P., Mott, R., Bork, P. & Copley, R. R. (2001). Novel protein domains and repeats in Drosophila melanogaster: Insights into structure, function, and evolution. *Genome Res.* **11**, 1996–2008.

Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. (2017). cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296.

Punjani, A., Zhang, H. & Fleet, D. J. (2020). Non-uniform refinement: Adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221.

Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99.

Rasmussen, S. G., Choi, H.-J., Fung, J. J., Pardon, E., Casarosa, P., Chae, P. S., DeVree, B. T., Rosenbaum, D. M., Thian, F. S., Kobilka, T. S. *et al.* (2011). Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. *Nature* **469**, 175–180.

Rebsamen, M., Pochini, L., Stasyk, T., de Araújo, M. E. G., Galluccio, M., Kandasamy, R. K., Snijder, B., Fauster, A., Rudashevskaya, E. L., Bruckner, M. *et al.* (2015). SLC38A9 is a component of the lysosomal amino acid sensing machinery that controls mTORC1. *Nature* **519**, 477–481.

Rhodes, G. (2006). *Crystallography made crystal clear: A guide for users of macromolecular models.* Academic Press, Elsevier, 3rd ed.

Rigaud, J.-L. & Lévy, D. (2003). Reconstitution of membrane proteins into liposomes. *Meth. Enzymol.* **372**, 65–86.

Rosenbaum, D. M., Cherezov, V., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H.-J., Yao, X.-J., Weis, W. I., Stevens, R. C. *et al.* (2007). GPCR engineering yields high-resolution structural insights into β_2-adrenergic receptor function. *Science* **318**, 1266–1273.

Rosenbaum, D. M., Rasmussen, S. G. & Kobilka, B. K. (2009). The structure and function of G-protein-coupled receptors. *Nature* **459**, 356–363.

Ruivo, R., Bellenchi, G. C., Chen, X., Zifarelli, G., Sagné, C., Debacker, C., Pusch, M., Supplisson, S. & Gasnier, B. (2012). Mechanism of proton/substrate coupling in the heptahelical lysosomal transporter cystinosin. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E210–E217.

Sadeghipour, F., Basiratnia, M., Derakhshan, A. & Fardaei, M. (2017). Mutation analysis of the CTNS gene in Iranian patients with infantile nephropathic cystinosis: Identification of two novel mutations. *Hum. Genom. Var.* **4**, 1–6.

Saudek, V. (2012). Cystinosin, MPDU1, SWEETs and KDELR belong to a well-defined protein family with putative function of cargo receptors involved in vesicle trafficking. *PLoS One* **7**, e30876.

Saxton, R. A. & Sabatini, D. M. (2017). mTOR signaling in growth, metabolism, and disease. *Cell* **168**, 960–976.

Schenk, B., Imbach, T., Frank, C. G., Grubenmann, C. E., Raymond, G. V., Hurvitz, H., Raas-Rotschild, A., Luder, A. S., Jaeken, J., Berger, E. G. *et al.* (2001). MPDU1 mutations underlie a novel human congenital disorder of glycosylation, designated type If. *J. Clin. Investig.* **108**, 1687–1695.

Schiestl, R. H. & Gietz, R. D. (1989). High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr. Genet.* **16**, 339–346.

Schmidt, T. G. M., Koepke, J., Frank, R. & Skerra, A. (1996). Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin. *J. Mol. Biol.* **255**, 753–766.

Schrödinger, LLC (2020). PyMOL version 2.4.0 (http://www.pymol.org/pymol).

Schulien, A., Justice, J. & Aizenman, E. (2019). Ion channels and action potential generation. In: Meriney, S. D. & Fanselow, E. E. (eds.). *Synaptic Transmission*, Academic Press, Elsevier, pp. 35–63.

Schulman, J. D., Bradley, K. H. & Seegmiller, J. E. (1969). Cystine: Compartmentalization within lysosomes in cystinotic leukocytes. *Science* **166**, 1152–1154.

Servais, A., Morinie, V., Gru, J.-P., Goujon, J.-M., Chadefaux-Vekemans, B., Antignac, C. *et al.* (2008). Late-onset nephropathic cystinosis: Clinical presentation, outcome, and genotyping. *Clin. J. Am. Soc. Nephrol.* **3**, 27–35.

Shahkarami, S., Galehdari, H., Ahmadzadeh, A., Babaahmadi, M., Pedram, M. *et al.* (2013). The first molecular genetics analysis of individuals suffering from nephropatic cystinosis in the Southwestern Iran. *Nefrología (English Ed.)* **33**, 308–315.

Sheldrick, G. M. (2008). A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122.

Sheldrick, G. M. (2010). Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallogr. D* **66**, 479–485.

Shen, J., Zeng, Y., Zhuang, X., Sun, L., Yao, X., Pimpl, P. & Jiang, L. (2013). Organelle pH in the Arabidopsis endomembrane system. *Mol. Plant* **6**, 1419–1437.

Shotelersuk, V., Larson, D., Anikster, Y., McDowell, G., Lemons, R., Bernardini, I., Guo, J., Thoene, J. & Gahl, W. A. (1998). CTNS mutations in an American-based population of cystinosis patients. *Am. J. Hum. Genet.* **63**, 1352–1362.

Smart, O. S., A, S., Holstein, J., Womack, T. O., Flensburg, C., Keller, P., Paciorek, W., Vonrhein, C. & G, B. (2020). Grade version 1.2.20 (http://grade.globalphasing.org). Cambridge, UK: Global Phasing Ltd.

Soliman, N. A., Elmonem, M. A., van den Heuvel, L., Abdel Hamid, R. H., Gamal, M., Bongaers, I., Marie, S. & Levtchenko, E. (2014). Mutational spectrum of the CTNS gene in Egyptian patients with nephropathic cystinosis. *JIMD Rep.* **14**, 87–97.

Stetsenko, A. & Guskov, A. (2017). An overview of the top ten detergents used for membrane protein crystallization. *Crystals* **7**, 197.

Strutz, W. (2016). Exploring protein stability by NanoDSF. *Biophys. J.* **110**, 393a.

Student (1908). The probable error of a mean. *Biometrika* **6**, 1–25.

Su, Z., Ran, X., Leitch, J. J., Schwan, A. L., Faragher, R. & Lipkowski, J. (2019). How valinomycin ionophores enter and transport K+ across model lipid bilayer membranes. *Langmuir* **35**, 16935–16943.

Sultana, A. & Lee, J. E. (2015). Measuring protein-protein and protein-nucleic acid interactions by biolayer interferometry. *Curr. Protoc. Protein Sci.* **79**, 19–25.

Talaia, G., Amick, J. & Ferguson, S. M. (2021). Receptor-like role for PQLC2 amino acid transporter in the lysosomal sensing of cationic amino acids. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2014941118.

Tao, Y., Cheung, L. S., Li, S., Eom, J.-S., Chen, L.-Q., Xu, Y., Perry, K., Frommer, W. B. & Feng, L. (2015). Structure of a eukaryotic SWEET transporter in a homotrimeric complex. *Nature* **527**, 259–263.

Taranta, A., Bellomo, F., Petrini, S., Polishchuk, E., De Leo, E., Rega, L. R., Pastore, A., Polishchuk, R., De Matteis, M. A. & Emma, F. (2017). Cystinosin-LKG rescues cystine accumulation and decreases apoptosis rate in cystinotic proximal tubular epithelial cells. *Pediatr. Res.* **81**, 113–119.

Taranta, A., Petrini, S., Citti, A., Boldrini, R., Corallini, S., Bellomo, F., Levtchenko, E. & Emma, F. (2012). Distribution of cystinosin-LKG in human tissues. *Histochem. Cell Biol.* **138**, 351–363.

Taranta, A., Petrini, S., Palma, A., Mannucci, L., Wilmer, M. J., De Luca, V., Diomedi-Camassei, F., Corallini, S., Bellomo, F., van den Heuvel, L. P. *et al.* (2008). Identification and subcellular localization of a new cystinosin isoform. *Am. J. Physiol. Renal Physiol.* **294**, F1101–F1108.

Taylor, G. (2003). The phase problem. *Acta Crystallogr. D* **59**, 1881–1890.

Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H. & Hung, L.-W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: The PHENIX AutoSol wizard. *Acta Crystallogr. D* **65**, 582–601.

Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D* **64**, 61–69.

Thoene, J., Lemons, R., Anikster, Y., Mullet, J., Paelicke, K., Lucero, C., Gahl, W., Schneider, J., Shu, S. & Campbell, H. T. (1999). Mutations of CTNS causing intermediate cystinosis. *Mol. Genet. Metab.* **67**, 283–293.

Thompson, A. A., Liu, J. J., Chun, E., Wacker, D., Wu, H., Cherezov, V. & Stevens, R. C. (2011). GPCR stabilization using the bicelle-like architecture of mixed sterol-detergent micelles. *Methods* **55**, 310–317.

Thorn, A. & Sheldrick, G. M. (2011). ANODE: Anomalous and heavy-atom density calculation. *J. Appl. Crystallogr.* **44**, 1285–1287.

Tickle, I. J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C. & Bricogne, G. (2016). STARANISO (http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi). Cambridge, UK: Global Phasing Ltd.

Topaloglu, R., Gulhan, B., İnözü, M., Canpolat, N., Yilmaz, A., Noyan, A., Dursun, I., Gökçe, İ., Gürgöze, M. K., Akinci, N. *et al.* (2017). The clinical and mutational spectrum of Turkish patients with cystinosis. *Clin. J. Am. Soc. Nephrol.* **12**, 1634–1641.

Topaloglu, R., Vilboux, T., Coskun, T., Ozaltin, F., Tinloy, B., Gunay-Aygun, M., Bakkaloglu, A., Besbas, N., van den Heuvel, L., Kleta, R. *et al.* (2012). Genetic basis of cystinosis in Turkish patients: A single-center experience. *Pediatr. Nephrol.* **27**, 115–121.

Town, M., Jean, G., Cherqui, S., Attard, M., Forestier, L., Whitmore, S. A., Callen, D. F., Gribouval, O., Broyer, M., Bates, G. P., van't Hoff, W. & Antignac, C. (1998). A novel gene encoding an integral membrane protein is mutated in nephropathic cystinosis. *Nat. Genet.* **18**, 319–324.

Uchański, T., Pardon, E. & Steyaert, J. (2020). Nanobodies to study protein conformational states. *Curr. Opin. Struct. Biol.* **60**, 117–123.

Urner, L. H., Liko, I., Yen, H.-Y., Hoi, K.-K., Bolla, J. R., Gault, J., Almeida, F. G., Schweder, M.-P., Shutin, D., Ehrmann, S. *et al.* (2020). Modular detergents tailor the purification and structural analysis of membrane proteins including G-protein coupled receptors. *Nat. Commun.* **11**, 564.

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A. *et al.* (2022). AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444.

Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., Tsenkov, M., Nair, S., Mirdita, M., Yeo, J. *et al.* (2024). AlphaFold Protein Structure Database in 2024: Providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* **52**, D368–D375.

Von Heijne, G. (1990). The signal peptide. *J. Membr. Biol.* **115**, 195–201.

Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T. & Bricogne, G. (2011). Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D* **67**, 293–302.

Wagner, A., Duman, R., Henderson, K. & Mykhaylyk, V. (2016). In-vacuum long-wavelength macromolecular crystallography. *Acta Crystallogr. D* **72**, 430–439.

Walden, H. (2010). Selenium incorporation using recombinant techniques. *Acta Crystallogr. D* **66**, 352–357.

Wang, J. & Wing, R. A. (2014). Diamonds in the rough: A strong case for the inclusion of weak-intensity X-ray diffraction data. *Acta Crystallogr. D* **70**, 1491–1497.

Wang, S., Tsun, Z.-Y., Wolfson, R. L., Shen, K., Wyant, G. A., Plovanich, M. E., Yuan, E. D., Jones, T. D., Chantranupong, L., Comb, W. *et al.* (2015). Lysosomal amino acid transporter SLC38A9 signals arginine sufficiency to mTORC1. *Science* **347**, 188–194.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L. *et al.* (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303.

Weiss, M. S. (2001). Global indicators of X-ray data quality. *J. Appl. Crystallogr.* **34**, 130–135.

Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B. *et al.* (2018). MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315.

Winn, M., Isupov, M. & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* **57**, 122–133.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A. *et al.* (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242.

Winter, G. (2010). xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190.

Winter, G., Beilsten-Edmands, J., Devenish, N., Gerstel, M., Gildea, R. J., McDonagh, D., Pascal, E., Waterman, D. G., Williams, B. H. & Evans, G. (2022). DIALS as a toolkit. *Protein Sci.* **31**, 232–250.

Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D. *et al.* (2018). DIALS: Implementation and evaluation of a new integration package. *Acta Crystallogr. D* **74**, 85–97.

Woody, R. W. (1995). [4] Circular dichroism. *Meth. Enzymol.* **246**, 34–71.

Wu, X. & Rapoport, T. A. (2021). Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies). *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2115001118.

Wu, Z., Newstead, S. & Biggin, P. C. (2022). State-dependent oligomerization and lipid saturation preferences of the KDEL receptor. *Biophys. J.* **121**, 433a.

Xiang, L., Etxeberria, E. & Van den Ende, W. (2013). Vacuolar protein sorting mechanisms in plants. *FEBS J.* **280**, 979–993.

Xu, Y., Tao, Y., Cheung, L. S., Fan, C., Chen, L.-Q., Xu, S., Perry, K., Frommer, W. B. & Feng, L. (2014). Structures of bacterial homologues of SWEET transporters in two distinct conformations. *Nature* **515**, 448–452.

Xuan, Y. H., Hu, Y. B., Chen, L.-Q., Sosso, D., Ducat, D. C., Hou, B.-H. & Frommer, W. B. (2013). Functional role of oligomerization for bacterial and plant SWEET sugar transporter family. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E3685–E3694.

Yang, D., Zhou, Q., Labroska, V., Qin, S., Darbalaei, S., Wu, Y., Yuliantie, E., Xie, L., Tao, H., Cheng, J. *et al.* (2021). G protein-coupled receptors: Structure- and function-based drug discovery. *Signal Transduct. Target Ther.* **6**, 7.

Yee, D. C., Shlykov, M. A., Västermark, Å., Reddy, V. S., Arora, S., Sun, E. I. & Saier Jr, M. H. (2013). The transporter–opsin–G protein-coupled receptor (TOG) superfamily. *FEBS J.* **280**, 5780–5800.

Yeetong, P., Tongkobpetch, S., Kingwatanakul, P., Deekajorndech, T., Bernardini, I. M., Suphapeetiporn, K., Gahl, W. A. & Shotelersuk, V. (2012). Two novel CTNS mutations in cystinosis patients in Thailand. *Gene* **499**, 323–325.

Zhang, J., Johnson, J. L., He, J., Napolitano, G., Ramadass, M., Rocca, C., Kiosses, W. B., Bucci, C., Xin, Q., Gavathiotis, E., Cuervo, A. M., Cherqui, S. & Catz, S. D. (2017). Cystinosin, the small GTPase Rab11, and the Rab7 effector RILP regulate intracellular trafficking of the chaperone-mediated autophagy receptor LAMP2A. *J. Biol. Chem.* **292**, 10328–10346.

Zimmermann, I., Egloff, P., Hutter, C. A., Arnold, F. M., Stohler, P., Bocquet, N., Hug, M. N., Huber, S., Siegrist, M., Hetemann, L. *et al.* (2018). Synthetic single domain antibodies for the conformational trapping of membrane proteins. *eLife* **7**, e34317.

Zimmermann, I., Egloff, P., Hutter, C. A., Kuhn, B. T., Bräuer, P., Newstead, S., Dawson, R. J., Geertsma, E. R. & Seeger, M. A. (2020). Generation of synthetic nanobodies against delicate proteins. *Nat. Protoc.* **15**, 1707–1741.

Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E. & Scheres, S. H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166.

Zivanov, J., Nakane, T. & Scheres, S. H. (2019). A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17.

UNIVERSITY OF
OXFORD

Clarendon

CHRIST
CHURCH
UNIVERSITY OF OXFORD

Medical
Research
Council

department of
B100HEMISTRY
1923-2023