

STRUCTURE OF THE PLANT-CONSERVED REGION OF CELLULOSE SYNTHASE AND ITS INTERACTIONS WITH THE CATALYTIC CORE

by

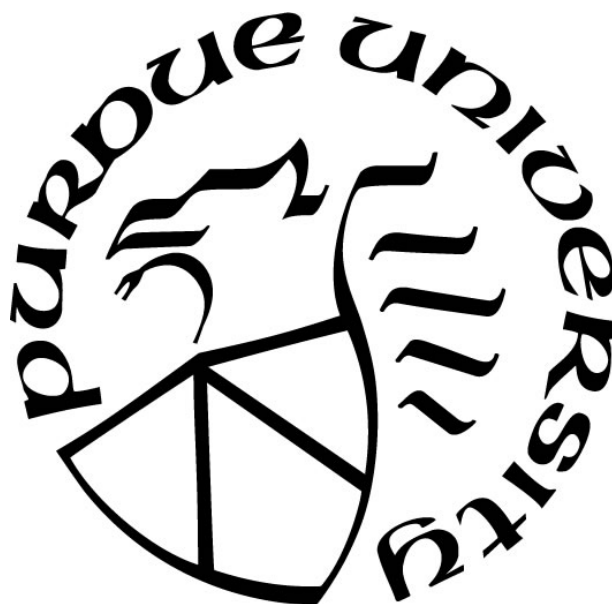
Phillip S. Rushton

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



Department of Biological Sciences

West Lafayette, Indiana

August 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Cynthia V. Stauffacher, Co-Chair

Department of Biological Sciences

Dr. Nicholas C. Carpita, Co-Chair

Department of Plant Pathology and Botany

Dr. Maureen C. McCann

Department of Biological Sciences

Dr. Daisuke Kihara

Department of Biological Sciences

Approved by:

Dr. Janice Evans

I would like to dedicate this thesis to my wife Mary Rushton for all her support.

ACKNOWLEDGMENTS

This work was supported as part of the Center for Direct Catalytic Conversion of Biomass to Biofuels, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award # DE-SC0000997. This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Crystallographic data were collected at GM/CA@APS, funded in whole or in part with Federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM- 12006). As well as, the Macromolecular Crystallography Facility supported by the Purdue Center for Cancer Research (PCCR) core grant P30 CA023168. SAXS data collection utilized the BioCAT resource supported by grant 9 P41 GM103622 from the National Institute of General Medical Sciences of the National Institutes of Health.

I am grateful for the love and support mother Deborah Rushton and father Scott Rushton, grandparents Phillip D. and JoAnn Rushton, Robert and Caroline Ward, brothers Matthew and Daniel Rushton, aunts Rebecca Patton for all of their support. I appreciate that I have been fortunate to have had excellent mentors in Fletcher Boyd, Kurt Hosteller, David Polley, Amanda Ingram, Eric Wetzel, David Krohne, Joel Sheets, Rob Cicchillo, Carla Yerkes, Gerald Watson, John Babcock, Tom Meade and Tom Sparks who helped me to achieve this goal. I want to thank my dogs Nikki, Rugby, Caesar, Akira and Apollo who always understood when I was too busy to play. I would also like to thank Cynthia Stauffacher, Nicholas Carpita, Maureen McCann, Daisuke Kihara, John Badger, Lee Makowski and all their lab members who have helped me, in particular, Anna Olek, Carolyn Zhou, Ziyun Ding, Nic Steussy, Tim Schmidt, Jay Mermoud, Matheus Bennatti, Haibing Yang, Ikenna Okekeogbu, Bryan Penning, Chun-liang Chen, Satchal Erramilli, Kaibo Zhang, Gillian Barth, Vatsal Purohit, Leang-Chung Choh (Chris) and Sanniv Ganguly.

TABLE OF CONTENTS

| | |
|---|----|
| LIST OF TABLES..... | 8 |
| LIST OF FIGURES | 9 |
| LIST OF ABBREVIATIONS..... | 13 |
| ABSTRACT..... | 16 |
| CHAPTER 1. LITERATURE REVIEW | 18 |
| 1.1 Introduction..... | 18 |
| 1.2 CESA Primary and Secondary Structure | 21 |
| 1.3 Plant-Specific Domains of CESA..... | 27 |
| 1.4 The <i>CESA</i> Gene Family | 30 |
| 1.4.1 Primary Cell Wall (PCW) <i>cesa</i> Mutants | 30 |
| 1.4.2 Secondary Cell Wall (SCW) <i>cesa</i> Mutants | 36 |
| 1.4.3 Post-Translational Modifications of CESAs and CESA-Interacting Proteins..... | 41 |
| 1.5 Synthesis, Trafficking and Life Cycle of CESAs and the CSC | 45 |
| 1.6 CESA Integration into a CSC | 49 |
| 1.7 Conclusion | 52 |
| CHAPTER 2. RICE CELLULOSE SYNTHASEA8 PLANT-CONSERVED REGION IS A COILED-COIL AT THE CATALYTIC CORE ENTRANCE | 55 |
| 2.1 Introduction..... | 55 |
| 2.2 Materials and Methods..... | 56 |
| 2.2.1 Cloning and Heterologous Protein Expression..... | 56 |
| 2.2.2 Protein Purification..... | 57 |
| 2.2.3 Circular Dichroism Analysis of P-CR Secondary Structure..... | 58 |
| 2.2.4 SAXS Data Collection and Processing..... | 59 |
| 2.2.5 Reconstruction of Molecular Envelopes..... | 59 |
| 2.2.6 Fitting of Crystal Structures to Solution Scattering Reconstructions..... | 60 |
| 2.2.7 Crystallization of P-CR..... | 60 |
| 2.2.8 X-ray Data Collection, Processing and Three-Dimensional Atomic Modeling..... | 61 |
| 2.3 Results..... | 62 |
| 2.3.1 P-CR Solution Structure and Molecular Envelope..... | 62 |

| | | |
|--|--|-----|
| 2.3.2 | P-CR Crystal Structure and Architecture | 64 |
| 2.3.3 | Comparison of the P-CR Crystallographic Structure with the SAXS Data..... | 69 |
| 2.3.4 | Molecular Docking of the P-CR with the Catalytic Core..... | 70 |
| 2.4 | Discussion | 72 |
| 2.4.1 | The P-CR Domain is a Unique Structure in Plant Cellulose Synthases | 73 |
| 2.4.2 | Data-Driven Modeling of the OsCESA8 Catalytic Domain..... | 73 |
| 2.4.3 | Potential Functions of the P-CR Domain | 75 |
| 2.4.4 | Uncertainty Remains About the Number of CESAs per Cellulose Synthase Complex | 77 |
| 2.5 | Supplemental Materials | 79 |
| CHAPTER 3. HYPOTHETICAL MODELS OF THE CELLULOSE SYNTHASE COMPLEX (CSC) | | 85 |
| 3.1 | Introduction..... | 85 |
| 3.2 | Tetrameric P-CR Models of the CSC Particle | 87 |
| 3.3 | Dimeric P-CR Models of the CSC Particle..... | 92 |
| 3.4 | Trimeric P-CR Models of the CSC Particle..... | 98 |
| 3.5 | Discussion | 102 |
| CHAPTER 4. STRUCTURAL COMPARISONS AND CHARACTERIZATIONS OF THE CELLULOSE SYNTHASEA8 PLANT-CONSERVED REGION REVEALS POTENTIALLY IMPORTANT RESIDUES FOR FUNCTION | | 104 |
| 4.1 | Introduction..... | 104 |
| 4.2 | Results..... | 105 |
| 4.2.1 | C-alpha Comparisons..... | 105 |
| 4.2.2 | Secondary Structural Element Comparison..... | 110 |
| 4.2.3 | Geometry and Sequence Comparisons | 115 |
| 4.2.4 | 3D Surface Comparisons | 119 |
| 4.2.5 | 3D Surface and Pocket Analysis..... | 125 |
| 4.3 | Discussion | 138 |
| 4.3.1 | Structural Similarity Searches Support a P-CR Tetramer Model of the CSC Particle.. | 138 |
| 4.3.2 | The P-CR Could Form Dimers Within the CSC | 143 |

| | | |
|---|--|-----|
| 4.3.3 | The P-CR Trimer Model of the CSC Particle..... | 144 |
| 4.3.4 | The P-CR Could Interact with CESA-Interacting Proteins | 146 |
| 4.3.5 | Interpreting the OsCESA8 P-CR and Catalytic Core Model Best Fit to the CatD SAXS Density Relative to the FxVTxK Motif Loop..... | 148 |
| 4.3.6 | Designing P-CR Mutants to Test P-CR Interactions Based on These Models..... | 150 |
| CHAPTER 5. ASSESSING THE IMPORTANCE OF SPECIFIC RESIDUES IN THE CELLULOSE SYNTHASE PLANT-CONSERVED REGION USING COMPLEMENTATION OF THE <i>ARABIDOPSIS THALIANA</i> TEMPERATURE-SENSITIVE <i>ATCESA1 RSW1-1</i> MUTANT BACKGROUND | | 155 |
| 5.1 | Introduction..... | 155 |
| 5.2 | Materials and Methods..... | 158 |
| 5.2.1 | Site-Directed Mutagenesis and <i>Arabidopsis thaliana</i> transformation..... | 158 |
| 5.2.2 | Growth Conditions..... | 159 |
| 5.2.3 | Gene Expression Analysis | 161 |
| 5.2.4 | <i>rsw1-1</i> Phenotyping and Complementation..... | 161 |
| 5.3 | Results..... | 163 |
| 5.3.1 | Sequence Conservation of the P-CR Across the Plant Kingdom | 163 |
| 5.3.2 | Probing Potential Interactions of the P-CR with the Catalytic Core | 168 |
| 5.3.3 | Homology Modeling of the <i>Arabidopsis thaliana</i> AtCESA1 Isoform P-CR using the OsCESA8 P-CR Structure. | 169 |
| 5.3.4 | Transgenic Expression of <i>Atcesal</i> P-CR Mutants..... | 173 |
| 5.3.5 | Complementing the Phenotype of the <i>Arabidopsis thaliana rsw1-1</i> Mutant Background..... | 172 |
| 5.4 | Discussion..... | 196 |
| 5.4.1 | Mutations that Fully or Partially Complement | 198 |
| 5.4.2 | Mutations that Poorly Complement..... | 200 |
| 5.4.3 | Future Work..... | 203 |
| 5.5 | Acknowledgments..... | 206 |
| APPENDIX..... | | 207 |
| REFERENCES | | 215 |

LIST OF TABLES

| | |
|---|-----|
| Table 1.1 Table of allelic mutants in PCW <i>CESAs</i> | 32 |
| Table 1.2 Table of allelic mutants in SCW <i>CESAs</i> | 37 |
| Table 1.3 Table of phosphorylation sites in CESA..... | 42 |
| Table 2.1 Data collection and refinement statistics | 65 |
| Table 4.1 Top five Dali results ranked by Z-score | 107 |
| Table 4.2 Top five FATCAT results ranked by p-value..... | 109 |
| Table 4.3 Top five deconSTRUCT results ranked by align score | 112 |
| Table 4.4 Top five PDBeFold results ranked by PDBeFold Q-score..... | 115 |
| Table 4.5 Top five BioXGEM 3D-BLAST results ranked by E-value..... | 117 |
| Table 4.6 Top five iSARST results ranked by Str. diversity | 119 |
| Table 4.7 Top five 3D-SURFER 2.0 results ranked by Euclidean distance..... | 122 |
| Table 4.8 Molecular contacts of the P-CR in its crystal structure PDB ID: 5jnp characterized using PDBePISA. | 123 |
| Table 4.9 Interface surface characterization of the molecular contacts described in Table 4.8 characterized by PDBePISA | 123 |
| Table 4.10 Top five PDBePISA interface similarity search results ranked by Q-score | 125 |
| Table 4.11 Top three VisGrid cavities and protrusions based on the 3D-SURFER 2.0 results . | 127 |
| Table 4.12 Top ten Fpocket pockets ranked by Fpocket score..... | 133 |
| Table 5.1 Residues selected for mutation in the P-CR | 157 |
| Table 5.2 Primers used to generate <i>Atcesal</i> mutations and measure expression by RT-qPCR . | 159 |
| Table 5.3 Seed stocks for <i>Atcesal</i> and P-CR mutants in the <i>Arabidopsis thaliana</i> <i>rsw1-1</i> background..... | 161 |

LIST OF FIGURES

| | |
|--|-----|
| Figure 1.1 Bar diagram of the domains and signature sequences of cellulose synthases in bacteria (BCSA) and plants (CESA) | 22 |
| Figure 1.2 Sequence comparison between <i>Arabidopsis thaliana</i> AtCESA1 and rice (<i>Oryza sativa</i>) OsCESA8 using Clustal Omega HHM sequence alignment | 23 |
| Figure 1.3 <i>Rhodobacter sphaeroides</i> bacterial cellulose synthase subunit A (BCSA) crystal structure (PDB ID: 4hg6)..... | 25 |
| Figure 1.4 Freeze-fracture transmission electron microscopy (FF-TEM) images showing six-fold symmetry of the CSC formed by CSC particles | 50 |
| Figure 2.1 Solution X-ray scattering experiments and determination of the P-CR 3D surface contour structure. | 63 |
| Figure 2.2 X-Ray crystal structure of the OsCESA8 P-CR (Ala399-Gln475) | 67 |
| Figure 2.3 Helical wheel diagram of the coiled-coil interactions between HX1 and HX2 | 68 |
| Figure 2.4 Best fit models of the P-CR in the OsCESA8 CatD..... | 72 |
| Figure 2.5 Supplemental Figure S1. P-CR secondary structure and apparent molecular weight analysis..... | 79 |
| Figure 2.6 Supplemental Figure S2. Reconstructing the P-CR SAXS molecular envelope..... | 79 |
| Figure 2.7 Supplemental Figure S3. Sequence coverage for mass spectrometry data of OsCESA8 P-CR crystals | 80 |
| Figure 2.8 Supplemental Figure S4. Electrostatic interactions between coiled-coil α -helices (HX1 and HX2)..... | 81 |
| Figure 2.9 Supplemental Figure S5. The P-CR SAXS molecular envelope fit into the P-CR crystal lattice..... | 82 |
| Figure 2.10 Supplemental Figure S6. Alternative fit of the P-CR into the OsCESA8 CatD molecular envelope | 83 |
| Figure 2.11 Supplemental Figure S7. The noncrystallographic dimer of the P-CR structure | 83 |
| Figure 2.12 Supplemental Figure S8. Constructing a P-CR trimer from SAXS and crystallographic data..... | 84 |
| Figure 3.1 General models of tetrameric, dimeric and trimeric P-CRs forming CSCs with the dimeric ZnF and CSR | 87 |
| Figure 3.2 Hypothetical models of the CSC using P-CR tetramers to form the CSC particle | 89 |
| Figure 3.3 Hypothetical models of the CSC using P-CR dimers to form the CSC particle | 94 |
| Figure 3.4 Hypothetical models of the CSC using P-CR trimers to form the CSC particle..... | 100 |
| Figure 4.1 Top Dali result, an ion transport 2 domain protein of a voltage-gated sodium channel VGIC (PDB ID: 3vou-A), aligned to the P-CR by Dali | 107 |

| | |
|---|-----|
| Figure 4.2 Top FATCAT result HIV glycoprotein gp41 (PDB ID: 3k9a-A) aligned to the P-CR by FATCAT | 110 |
| Figure 4.3 Top deconSTRUCT result interleukin-1 receptor-associated kinase 4 (PDB ID: 2nru-A) superimposed to the P-CR by deconSTRUCT | 113 |
| Figure 4.4 Top PDBeFold result YocF sensor kinase (PDB ID: 3ehf-C) sub-structure superimposed to the P-CR by PDBeFold..... | 115 |
| Figure 4.5 Top BioXGEM 3D-BLAST result, a potassium channel KcsA-Fab complex in low concentration of Tl ⁺ (PDB ID: 1r3k-C), aligned to the P-CR by BioXGEM 3D-BLAST | 117 |
| Figure 4.6 Top iSARST result, a bifunctional P-450/ NADPH-P450 reductase (PDB ID: 4hgh-A), sub-structure superimposed on the P-CR by iSARST | 119 |
| Figure 4.7 Top 3D-SURFER 2.0 result, a VGIC potassium channel protein (PDB ID: 4rai-A) aligned to the P-CR by combinatorial extension | 122 |
| Figure 4.8 Top PDBePISA result, synapse-associated protein 102 (PDB ID: 1um7), sub-structure aligned to the P-CR by combinatorial extension | 125 |
| Figure 4.9 Largest cavities of the P-CR defined by the VisGrid algorithm using the 3D-SURFER 2.0 results | 129 |
| Figure 4.10 Largest protrusions of the P-CR defined by the VisGrid algorithm using the 3D-SURFER 2.0 algorithm results | 131 |
| Figure 4.11 Ten best ranked pockets of the P-CR found by the Fpocket algorithm..... | 134 |
| Figure 4.12 Best ranked pockets of the P-CR and VGICs found by the Fpocket algorithm shown with the best ranked cavities and protrusions of the P-CR using the VisGrid algorithm | 136 |
| Figure 4.13 VGIC-like P-CR tetramer models based on the ion transport 2 domain protein of a voltage-gated sodium channel (PDB ID: 3vou) crystallographic tetramer..... | 140 |
| Figure 4.14 Atomic models of a P-CR tetramer oriented perpendicular to the plasma membrane based on the catalytic core orientation..... | 142 |
| Figure 4.15 Atomic model of a P-CR dimer oriented perpendicular to the plasma membrane based on the catalytic core orientation..... | 143 |
| Figure 4.16 P-CR crystallographic trimer as stick and surface models of both sides of the P-CR long axis after 180° rotations, shown with significant pockets and cavities in the P-CR relative to this trimer | 145 |
| Figure 4.17 Atomic model of a P-CR trimer oriented perpendicular to the plasma membrane based on the catalytic core orientation..... | 146 |
| Figure 4.18 Models of the P-CR tetramer, both in the same plane as the catalytic core or below the catalytic core relative to the plasma membrane, dimer or trimer illustrating the occlusion or solvent exposure of Cav 1, Cav 2 or Cav 3 in various models..... | 148 |
| Figure 4.19 Overlay of the BCSA gating loop with the P-CR and catalytic core IF2 interfacial helix | 149 |
| Figure 5.1 Clustal Omega HHM multiple sequence alignments of CESA P-CRs, divided into classes and represented as sequence logos using the WebLogo algorithm | 164 |

| | |
|---|-----|
| Figure 5.2 Sequence conservation of the combined Clustal Omega HHM multiple sequence alignments of the P-CR modeled on the OsCESA8 P-CR structure..... | 167 |
| Figure 5.3 Potential contact of the OsCESA8 P-CR with the OsCESA8 catalytic core homology model based on their best fit to OsCESA8 CatD SAXS density | 169 |
| Figure 5.4 Homologous sequences between the AtCESA1 and OsCESA8 P-CRs allow high confidence homology modeling of the AtCESA1 P-CR to the limit of the OsCESA8 P-CR X-ray crystal structure..... | 171 |
| Figure 5.5 Best homology model of the AtCESA1 P-CR domain using RaptorX showing the Atcesal mutants constructed for the <i>Arabidopsis thaliana</i> <i>rsw1-1</i> complementation experiments | 172 |
| Figure 5.6 <i>Atcesal</i> transgene expression in the <i>rsw1-1</i> mutant background relative to <i>CBP20</i> | 174 |
| Figure 5.7 Complementation experiments of Col-0, <i>rsw1-1/AtCESA1</i> 2.2, <i>rsw1-1</i> and Asp438Ala/Tyr439Ala 7.6 seedlings grown for four days at the <i>rsw1-1</i> permissive temperature (22°C) then three days at the restrictive temperature (31°C)..... | 175 |
| Figure 5.8 Differences in primary root growth varied based on the mutant and its expression level after four days at the permissive temperature (22°C) followed by three days at the restrictive temperature (31°C) | 176 |
| Figure 5.9 Root morphology of control and transgenic mutants grown for four days at the permissive temperature (22°C) followed by three days at restrictive temperature (31°C)..... | 177 |
| Figure 5.10 Primary root growth of the Arg457Ala <i>Atcesal</i> transgene in <i>rsw1-1</i> after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures..... | 178 |
| Figure 5.11 Arg457Ala <i>Atcesal</i> transgene expression in the <i>rsw1-1</i> mutant background relative to <i>CBP20</i> | 179 |
| Figure 5.12 Primary root growth of the Val470Ala <i>Atcesal</i> transgene in <i>rsw1-1</i> after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures..... | 180 |
| Figure 5.13 Val470Ala <i>Atcesal</i> transgene expression in the <i>rsw1-1</i> mutant background relative to <i>CBP20</i> | 181 |
| Figure 5.14 Primary root growth of the Arg452Ala <i>Atcesal</i> transgene in <i>rsw1-1</i> after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures..... | 182 |
| Figure 5.15 Arg452Ala <i>Atcesal</i> transgene expression in the <i>rsw1-1</i> mutant background relative to <i>CBP20</i> | 183 |
| Figure 5.16 Primary root growth of the Glu430Ala <i>Atcesal</i> transgene in <i>rsw1-1</i> after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures..... | 184 |
| Figure 5.17 Glu430Ala <i>Atcesal</i> transgene expression in the <i>rsw1-1</i> mutant background relative to <i>CBP20</i> | 185 |

| | |
|---|-----|
| Figure 5.18 Primary root growth of the Phe431Ala <i>Atcesal</i> transgene in <i>rswl-1</i> after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures..... | 186 |
| Figure 5.19 Phe431Ala <i>Atcesal</i> transgene expression in the <i>rswl-1</i> mutant background relative to <i>CBP20</i> | 187 |
| Figure 5.20 Primary root growth of the Leu812Arg <i>Atcesal</i> transgene in <i>rswl-1</i> after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures..... | 188 |
| Figure 5.21 Leu812Arg <i>Atcesal</i> transgene expression in the <i>rswl-1</i> mutant background relative to <i>CBP20</i> | 189 |
| Figure 5.22 Growth rate comparison between controls and the poorly complementing Asp438Ala/Tyr439Ala 7.6 mutant, expressing near native levels, before and after the temperature was increased to the <i>rswl-1</i> restrictive temperature (31°C)..... | 190 |
| Figure 5.23 Primary root growth of the Asp438Ala/Tyr439Ala <i>Atcesal</i> transgene in <i>rswl-1</i> after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures..... | 191 |
| Figure 5.24 Asp438Ala/Tyr439Ala <i>Atcesal</i> transgene expression in the <i>rswl-1</i> mutant background relative to <i>CBP20</i> | 192 |
| Figure 5.25 Primary root growth of the Arg453Lys (equivalent to <i>fra6</i> in <i>Atcesa8</i>) <i>Atcesal</i> transgene in <i>rswl-1</i> after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures..... | 193 |
| Figure 5.26 Arg453Lys <i>Atcesal</i> transgene expression in the <i>rswl-1</i> mutant background relative to <i>CBP20</i> | 194 |
| Figure 5.27 Primary root growth of the Pro417Ala <i>Atcesal</i> transgene in <i>rswl-1</i> after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures..... | 195 |
| Figure 5.28 Pro417Ala <i>Atcesal</i> transgene expression in the <i>rswl-1</i> mutant background relative to <i>CBP20</i> | 196 |
| Figure 5.29 Root morphology and Primary root length of P-CR mutants grown for five days at the permissive temperature (22°C) followed by one day at the permissive (22°C) or restrictive (31°C) temperatures..... | 197 |
| Figure 5.30 <i>Atcesal</i> mutants that poorly complemented <i>Arabidopsis thaliana</i> <i>rswl-1</i> mutant, shown on the best homology model of the AtCESA1 P-CR domain using RaptorX..... | 200 |

LIST OF ABBREVIATIONS

any1: *anisotropy1* mutant

AP2M or $\mu 2$: ADAPTOR PROTEIN2 COMPLEX MEDIUM ADAPTIN subunit

BCSA: bacterial cellulose synthase protein A

bc: *brittle culm* mutant

BIN2: BRASSINOSTEROID INSENSITIVE2 kinase

CatD: catalytic domain of a CESA protein

Cav 1/2/3: cavities detected in the P-CR structure by 3D SURFER 2.0, numbered by rank

C α : protein C-alpha carbon

CBP20: CAP-BINDING PROTEIN 20

CC1/2: COMPANIONS OF CELLULOSE SYNTHASE1 and 2

CD: circular dichroism

cDNA: complementary DNA

CESA: plant cellulose synthase that synthesizes (1 \rightarrow 4)- β -D-glucans

cev1: *constitutive expression of VSP1* mutant

CME: clathrin-mediated endocytosis

CSC: cellulose synthase complex

CSI: CELLULOSE SYNTHASE INTERACTIVE PROTEIN

CSL: CELLULOSE SYNTHASE-LIKE

CSR: class-specific region

CSS: complexation significance score

elil: *ectopic lignification1* mutant

ern1: *enhanced resistance to necrotrophs1* mutant

ER: endoplasmic reticulum

exi: *exigua* mutant

FF-TEM: freeze-fracture transmission electron microscopy

fra: *fragile fiber* mutant

fxr1: *flupoxam resistance1* mutant

fxr2: *flupoxam resistance2* mutant

GFP: green fluorescent protein

Δ^iG : Gibbs free energy of solvation
 GT: glycosyltransferase
 h: hour
 HX1: N-terminal α -helix of the P-CR coiled-coil in the P-CR X-ray crystal structure
 HX2: C-terminal α -helix of the P-CR coiled-coil in the P-CR X-ray crystal structure
 Indels: insertions and deletions
irx: *irregular xylem* mutant
ixr1: *isoxaben resistant1* mutant
je5: *Atcesa3^{je5}* mutant allele
 kD: kilodalton molecular weight
 KIH: knob-in-hole
 KOR1/GH9A1/*rsw2*: endo-1,4- β -D-glucanase protein of the *KORRIGAN* gene
 LB: Luria-Bertani media
lew: *leaf wilting* mutant
 LOF: loss of function
 LP: loop region connecting HX1 and HX2 in the P-CR X-ray crystal structure
 MASC: microtubule-associated cellulose synthase compartments, synonymous with SmaCC
 MBP: maltose-binding protein
 mdeg: milli-degree
 MD: molecular dynamics simulations
mrel: *multiple response to expansion1* mutant
 MS: Murashige and Skoog
mur: *murus* mutant
nws2: *no wilt symptoms2* mutant
 PAGE: polyacrylamide gel electrophoresis
 PCW: primary cell wall
 PDB: protein data bank
 % Seq ID: percent sequence identity
prc1: *procuste1* mutant
 Prot 1/2/3: protrusions detected in the P-CR structure by 3D SURFER 2.0, numbered by rank
 PTM: post-translational modifications

P-CR: plant-conserved region
R_g: radius of gyration
RLK: receptor-like kinase
RMSD: root mean square deviation
rsw: *radial swelling* mutant
RT-qPCR: real-time quantitative reverse-transcription PCR
SAD: single-wavelength anomalous diffraction
SAXS: small-angle X-ray scattering
SCW: secondary cell wall
SDS: sodium dodecyl sulfate
SEC: size exclusion chromatography
SeMet: seleno-methionine
SmaCC: small CESA compartments, synonymous with MASC
SSE: secondary structural element
ssNMR: solid-state nuclear magnetic resonance
TEV: tobacco etch virus (protease)
than: *thanatos* mutant
THE1: THESEUS1
TMH: transmembrane α -helix
UDP-Glc: uridine diphosphate-glucose
VGIC: voltage-gated ion channel
WT: wild type
YFP: citrine yellow fluorescent protein
ZnF: plant specific N-terminal RING-type zinc finger domain

ABSTRACT

The processive plant cellulose synthase (CESA) synthesizes (1→4)-β-D-glucans. CESAs assemble into a six-fold symmetrical cellulose synthase complex (CSC), with an unknown symmetry and number of CESA isomers. The CSC synthesizes a cellulose microfibril as the fundamental scaffolding unit of the plant cell wall. CESAs are approximately 110 kDa glycosyltransferases with an N-terminal RING-type zinc finger domain (ZnF), seven transmembrane α-helices (TMHs) and a cytoplasmic catalytic domain (CatD). In the CatD, the uridine diphosphate glucose (UDP-Glc) substrate is synthesized into (1→4)-β-D-glucans. The ZnF is likely to facilitate dimers in the CSC. Recombinant class-specific region (CSR), a plant specific insertion to the C-terminal end of the CatD is also known to form dimers *in vitro*. The CSR sequence is the primary source of distinction between CESA isoforms and class structure. Also within the CESA CatD is a 125-amino acid insertion known as the plant-conserved region (P-CR), whose molecular structure was unknown. The function of the P-CR is still unclear, especially in the context of complete CESA and CSC structures. Thus, one major knowledge gap is understanding how multimeric CSCs synthesize multiple chains of (1→4)-β-D-glucans that coalesce to form microfibrils. The specific number of CESAs in a CSC and how interactions of individual CESA isoforms contribute to the CSC are not known. Elucidating the structure-function relationships of the P-CR domain, and with the consideration of the ability of CSR and ZnF domains to dimerize, it is possible to more completely model the structure of the CSC.

Recombinantly expressed rice (*Oryza sativa*) secondary cell wall OsCESA8 P-CR domain purifies as a monomer and shows distinct α-helical secondary structure by circular dichroism analysis. A molecular envelope of the P-CR was derived by small angle X-ray scattering (SAXS). The P-CR was crystallized and structure solved to 2.4 Å resolution revealing an anti-parallel coiled-coiled domain. Connecting the coiled-coil α-helices is an ordered loop that bends back towards the coiled-coils. The P-CR crystal structure fits the molecular envelope derived by SAXS, which in turn fits into the CatD molecular envelope. The best fit places the P-CR between the membrane and substrate entry portal. In depth analysis of structural similarity to other proteins, and 3D-surface structure of the P-CR, leads to hypotheses that it could function in protein-protein interactions as a dimer, trimer or tetramer in the CSC, that it could form protein-protein interactions with CESA-interacting proteins, and/or modulate substrate entry through its N- and/or C-terminus.

From modeling, hypothetically important residues within the P-CR or related to the P-CR through potential protein contacts were mutated in *Arabidopsis thaliana* *AtCESA1* constructs. These constructs were expressed in the temperature-sensitive *radial swelling* (*rsw*) *rsw1-1* mutant of *AtCESA1* to test for complementation of growth phenotypes at restrictive temperatures. Preliminary experiments indicate that some mutated CESA1 sequences fail to complement the *rsw1-1* phenotype, suggesting that specific functions of individual amino can be tested using this system.

CHAPTER 1. LITERATURE REVIEW

1.1 Introduction

Cellulose, a linear chain of (1→4)-β-D-glucan, is the most ubiquitous organic polymer on earth. Plant cellulose synthases (CESAs) evolved from prokaryotic ancestral genes (Nobles et al., 2001). Bacterial synthases make single chains of (1→4)-β-D-glucan that are guided through the periplasm by accessory proteins and are integrated at the peripheral membrane into crystalline forms secreted into the medium as biofilms or pellicles (Kondo et al., 2002). Bacterial cellulose synthases were first discovered in *Gluconacetobacter xylinus* (syn *Acetobacter xylinum*) (Saxena et al., 1990; Wong et al., 1990). Sequence motifs of importance were identified as the four catalytic signature motifs or U-motifs, conserved D, DxD, D, and Q/RxxRW motif residues (Saxena et al., 1995). The characterization of the crystal structure of the bacterial cellulose synthase complex from *Rhodobacter sphaeroides* was a major breakthrough in understanding the catalytic mechanism for glucose transfer, chain elongation, and product channeling, and assigned a molecular function to signature motifs (Morgan et al., 2013). The catalytic core structure of bacterial cellulose synthase protein A (BCSA), a variant of the classical nucleotide-binding fold (Rossmann et al., 1974), is conserved among several similar types of nucleotide-sugar transferases including plant CESAs (Olek et al., 2014). In plant CESAs, the N-terminus is larger than in BCSA, containing a plant-specific N-terminal RING-type zinc finger domain (ZnF), and has two plant-specific domains inserted in the catalytic domain (CatD). The function of the two plant-specific domains, the plant-conserved region (P-CR) and the class-specific region (CSR), are not known (Delmer, 1999; Carpita, 2011). CESAs also have distinct transmembrane α-helical (TMH) topology and a truncated C-terminus lacking a PilZ domain, compared to BCSA.

Plant cellulose synthesis differs from bacterial synthesis in the formation of a *para*-crystalline cellulose microfibril of (1→4)-β-D-glucan chains in the cell wall. Plants utilize two different sets of three CESA isoforms to synthesize cellulose microfibrils in either primary or secondary cell walls (Carpita et al., 2015). The primary cell wall (PCW) surrounds every plant cell, in dicotyledonous species, and is rich in cellulose, hemicellulose and pectin. Following PCW synthesis, secondary cell wall (SCW) synthesis produces increased cellulose deposition in a few

specific cell types. SCWs also incorporate xylans, a polymer of xylose, and lignin. The cellulose microfibril defines the structural architecture of the cell wall throughout growth.

The *para*-crystalline cellulose microfibril structure is likely reflected in the complex interactions observed between CESAs. These interactions are thought to be facilitated largely by the plant-specific sequences in CESA, not found in BCSA. Electron microscopy studies have imaged the distinct “rosette” or cellulose synthase complex (CSC), with characteristic six-fold symmetrical CSC particles (Herth, 1985; Delmer, 1999). The CSC was initially considered to be a “hexamer of hexamers” or a 36-isomer complex, which fit with earlier estimates of CSC size (Herth, 1985; Delmer, 1999), microfibril size and computational modeling of cellulose (Matthews et al., 2012). However, more recent estimates have shifted towards a core cellulose microfibril size of 18 or 24 (1→4)-β-D-glucan chains, although less ordered surface chains might also exist (Fernandes et al., 2011; Newman et al., 2013; Thomas et al., 2013; Oehme et al., 2015; T. Wang et al., 2016; Kubicki et al., 2018). It is inferred that the CSC would have either 18 or 24 CESAs in a CSC (Sethaphong et al., 2013; Nixon et al., 2016; Sethaphong et al., 2016; Turner et al., 2018; Haigler et al., 2019). The 24 CESA model was proposed early in CSC modeling (Bowling et al., 2008), which is consistent with the even-numbered CESA oligomer units (Taylor et al., 2003; Desprez et al., 2007; J. Wang et al., 2008; Atanassov et al., 2009; Timmers et al., 2009; Olek et al., 2014) and the CESA dimers for which the structure has been solved using small angle X-ray scattering (SAXS) (Kurek et al., 2002; Olek et al., 2014). The 18 CESA models of the CSC are more recent, relying on estimates of *ab initio* models of CESA TMHs fit to freeze-fracture transmission electron microscopy (FF-TEM) images of CSCs (Nixon et al., 2016), SAXS on recombinant AtCESA1 CatD reconstituted from inclusion bodies (Vandavasi et al., 2016) and consistency with a 1:1:1 ratio of CESA isoforms (Gonneau et al., 2014; Hill et al., 2014).

CESAs also directly interact with a number of CESA-interacting proteins that assist cellulose biosynthesis, regulation or trafficking. CSCs are trafficked to and from the plasma membrane via microtubule-associated cellulose synthase compartments (MASC) (Crowell et al., 2009)/small CESA compartments (SmaCC) (Gutierrez et al., 2009). CSC-microtubule interactions are mediated through CELLULOSE INTERACTIVE PROTEINS 1 and 3 (CSI1/POM-POM2 and CSI3), which are associated with cortical microtubules that define CSC trajectories and anisotropic cell growth (Y. Gu et al., 2010; Bringmann et al., 2012; S. Li et al., 2012; Lei et al., 2013, 2015). The COMPANIONS OF CELLULOSE SYNTHASE1 and 2 (CC1/2) proteins, alongside CSI, are

additional mediators between the CSC and microtubules in this system. CC1/2 appear to function primarily in CSC recycling during stress-induced cellulose deficiency, but their role in steady state recycling is unclear (Endler et al., 2015, 2016). Other CESA-interacting proteins likely function exclusively while CESA is active in the plasma membrane. For example, the endo-1,4- β -D-glucanase encoded by the *KORRIGAN* gene (KOR1/GH9A1/*rsw2*), likely cleaves individual (1 \rightarrow 4)- β -D-glucan chains to relieve tension on the microfibril or the entire microfibril in order to release the CSC from the membrane (Lane et al., 2001; Sato et al., 2001; Lei et al., 2014; Mansoori et al., 2014; Vain et al., 2014). The glycosphosphatidylinositol-anchored and N-glycosylated protein encoded by the *COBRA* genes also interacts with CESA and the microfibril near the plasma membrane by an unknown mechanism that controls or modifies cellulose crystalline architecture (Schindelman et al., 2001; Roudier et al., 2005; L. Liu et al., 2013; Sorek et al., 2014). During exocytosis from the Golgi apparatus, a number of proteins in the exocyst complex, responsible for forming MASCs/SmaCCs, directly interacts with CESA proteins and CSI1 (X. Zhu et al., 2018). Together with the exocyst complex, CSI1 and CESAs all interact with PATROL1, which is thought to prime the insertion of MASCs/SmaCCs near specific cortical microtubules. Regulation of CSC exocytosis is also facilitated through direct interactions between CESAs and the SHOU4 protein that blocks the fusion of MASCs/SmaCCs with the plasma membrane (Polko et al., 2018). Two distinct clathrin-mediated endocytosis (CME) pathways contain proteins that interact with CESAs. These were discovered first for the ADAPTOR PROTEIN2 COMPLEX MEDIUM ADAPTIN subunit (AP2M) or μ 2 pathway (Bashline et al., 2013) and then for the TPLATE complex pathway through interactions with the TWD-40 protein (Bashline et al., 2015), TPLATE and TPLATE complex muniscin-like proteins (Sánchez-Rodríguez et al., 2018). CESA-interacting proteins could indirectly couple CESAs to other protein complexes and might also affect CSC size estimates, so careful accounting for their presence and activity is necessary to elucidating the structure of the CSC.

In order to understand the principles that govern cellulose microfibril assembly, it is necessary to define the structure of CESAs, their CSCs and the molecular mechanisms used by CESA plant-specific sequences to form them into a complex. In this work, I approach these problems through structural biology of *in vitro* recombinant CESA domains, modeling CESA and CSCs, structure-structure comparisons of the P-CR, and 3D modeling of the P-CR. Amino acids identified as potentially critical components of the structure in these studies were experimentally

tested using transgenic complementation of the *radial swelling (rsw) rsw1-1 Atcesa1* mutant to determine the importance of these amino acids *in planta*. Determining the structure of the CESA proteins and the CSCs they form is critical to understanding the fundamental role they play in synthesizing the most abundant biomolecule on the planet.

1.2 CESA Primary and Secondary Structure

CESAs are ~110 kDa integral membrane proteins with seven TMHs, an N-terminal ZnF, a short C-terminal region and a large cytoplasmic CatD positioned between the second and third TMHs in contrast to the positioning of the bacterial catalytic domain between the fourth and fifth TMHs of BCSA (Figure 1.1) (Morgan et al., 2013; Olek et al., 2014; Slabaugh et al., 2014a, 2014b, 2016; Nixon et al., 2016; Rushton et al., 2017). Among CESA isoforms, sequences are well conserved as illustrated by a comparison between the AtCESA1 and OsCESA8 isoforms, which have a shared percent sequence identity (% Seq ID) of 68.2% and percent sequence similarity of 86.6% using the Clustal Omega sequence alignment algorithm (Sievers et al., 2011) (Figure 1.2). The catalytic core of the CatD contains the four critical catalytic motifs or U-motifs, containing four Asp (D) residues and a Q/RxxRW motif, that form the uridine diphosphate glucose (UDP-Glc) nucleotide substrate binding pocket (Saxena et al., 1995). These four broadly conserved catalytic sequences were used to identify genes encoding putative cellulose synthases in cotton (Pear et al., 1996). The initial D motif binds the UDP-Glc substrate at the uracil 4-dione and the Dx D motif binds to the diphosphate group. The next D in the TED motif is the catalytic base oriented in a position to access the end of the glucan chain and the Q/RxxRW motif forms a major part of the binding site for the terminal disaccharide (Morgan et al., 2013).

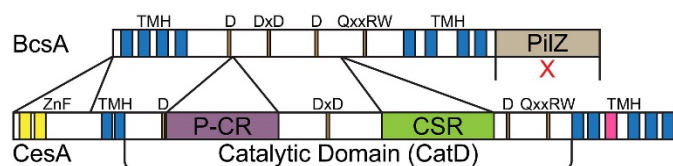


Figure 1.1 Bar diagram of the domains and signature sequences of cellulose synthases in bacteria (BCSA) and plants (CESA). Like BCSA, CESA possesses four catalytic signature motifs (orange) containing conserved D, Dx, D, and Q/RxxRW motif residues. Plants differ from bacterial synthases by the absence of the PilZ activator domain (beige) and the addition of three unique domains, a RING-type zinc finger domain (ZnF, yellow) in the extended N-terminus, and the Plant-Conserved Region (P-CR, purple) and Class-Specific Region (CSR, green) domains inserted into the catalytic core domain (CatD). Recent work has indicated that CESA might also differ from BCSA in its number of transmembrane helices (TMH, blue) with BcsA possessing 8 TMH and CESA likely possessing 7 TMH, with the putative TMH 5 (pink) hypothesized to occupy an interfacial orientation along the membrane (Slabaugh et al., 2016).

| | | |
|-------------------|---|------|
| ATCESA1 (829376) | MEASAGLVAGSYRRNELVRIRHESDGGTKPLKNMNGQICQICGDDVGLAETGDVFVACNECAFPVCRPCYERKDGTCQCPQCKTRFRHRHSGSPRVVGD | 100 |
| OSCESA8 (4342699) | MDGDA-----D-AVKSGRHGSQACQICGDDVGTAEAGDVFAACDVCGFVCRPCYERKDGTCQCPQCKTKYKRHKGSPAIRGE | 80 |
| ATCESA1 (829376) | EDDDVDVDDIENEFNYAQGANKARHQRHGEFSSSSR-----HESQPIPL--LTHGHTVSGEIRTPDTQSVRTTSGPLGSDRNAIS--SPYID | 184 |
| OSCESA8 (4342699) | EGETDADDVSDYNYPASGSADQKQKIADRMRSWRMAGGGDVGRPKYDSGEIGLTKYDSGEIPRGYIPSVNSQI--SGEIPGASPDHMMSPNTGIG | 178 |
| ATCESA1 (829376) | PRQPVPVRIVDPSK--DLNSYGLGNVDKWERVEGWKLKQEKNMQLMTGKYHE-----GKGGEIEGTSGNGEELQMAADTRLPMRSRVVPISSRLTPYRVV | 277 |
| OSCESA8 (4342699) | KRAPFPYVNHSPNPSREFSGSIGNVAWKERVGDGWLKQDKGAIPMTNGTSIAPSEGRGVGDIADSTDYNMEDALLNDETRQPLSRKVPSPSSRINPYRMV | 278 |
| ATCESA1 (829376) | <u>TMH-1</u> IILRLIILCFFLQYRTHPVKNAYPLWLTSVICEIWFALSWILDQFPKWYPINRETYLDRLALRYDRDGEPSQLVPVDVVFSTVDPLKEPPLVTANTVLS | 377 |
| OSCESA8 (4342699) | <u>TMH-2</u> IVLRLVLSIFLHYRITNPVRNAYPLWLLSVICEIWFALSWILDQFPKWYPINRETYLDRLALRYDRDGEPSQLAAVDIFVSTVDPMEPPLVTANTVLS | 378 |
| ATCESA1 (829376) | ILSVDPYVDKVCYVSDDGSAAMLTFALETSFAFARKWVPFCKFNIEPRAPEFYFAQKIDYLDKDIQPSFVKERRAMKREYEFKVRINALVAKAKIP | 477 |
| OSCESA8 (4342699) | ILAVDYPVDKVCYVSDDGSAAMLTFALETSFAFARKWVPFCKFNIEPRAPEWYFSQKIDYLDKDVHPSFVKDRRAMKREYEFKVRINGLVAKAKVP | 478 |
| ATCESA1 (829376) | EEGWTMQDGTTPWPGNTRDHPGMIQVFLHSGGLDTEGNELPRLIYVSREKRPGFQHHKKAGAMNALIRVSAVLTNGLYLLNVDCHYFNNSKAIKEAMC | 577 |
| OSCESA8 (4342699) | EEGWIMQDGTTPWPGNTRDHPGMIQVFLHSGGLDTEGNELPRLIYVSREKRPGFQHHKKAGAMNALIRVSAVLTNGLYMLNLDCHYINNSKALREAMC | 578 |
| ATCESA1 (829376) | FMMDPAGKGCYVQFPQRFDGIDLHRYANRNIVFFDINMKGLDGIQGPVYVGTGCCFNQALYGYDPVLTEEDLEPNIIKSCCGSRKKKSKSKKYN | 677 |
| OSCESA8 (4342699) | FLMDPNLGRSVYVQFPQRFDGIDRNDRYANRNIVFFDINRLGLDGIQGPVYVGTGCVFNRALYGYEPIKQKKK--GSFSLSLCGGRKKASKSKKSS | 676 |
| ATCESA1 (829376) | EKRRGINRSDSNAPLFNMEIDEGFE--GYDDERSILMSQRSVEKRFQSPVFIAATFMEQGGIPPTNPATLLKEAIHVISCYEDKTEWKGKEIGWIY | 775 |
| OSCESA8 (4342699) | DKKSNKHVDASVPVFNLEDIEEGVEGAGFDEKSLMSQMSLEKRFQSAFVASTLMEYGGVPQSATPESLLKEAIHVISCYEDKTEWGTIGWIY | 776 |
| ATCESA1 (829376) | SVTEDILTGFKMARGWISYICNPAPAFKGSAPINLSDRNLQVLRWALGSIEILLSRHCPIWYGYHGRLLERLIERIAYINTIVYPTISPIIAYCILPAF | 875 |
| OSCESA8 (4342699) | SVTEIDILTGFKMARGWRSYICMPKRPAPAFKGSAPINLSDRNLQVLRWALGSIEILLSRHCPIWYGYGGRKLFERFAYINTIYPLTSLIPLIYCVLPAL | 876 |
| ATCESA1 (829376) | CLITDRFIIEISNYASIWFILFISIAVTGILELRWGSVSIEDWWRNEQFVWIGGSAHLFAVFGLLKVLGAGIDTNTFTSKATDEGDGFAELYIFKW | 975 |
| OSCESA8 (4342699) | CLLTGKFIIEISNFASIFISLFIATGILEMRWGSVGIEDWWRNEQFVWIGGSAHLFAVFGLLKVLGAGIDTNTFTSKADEGDGFAELYMFKW | 976 |
| ATCESA1 (829376) | TALLIPPTTVLLVNLIGIVAGVSYAVNSGYQSWGPLEGKLFALWIAHLYPFLKGLLGRQNRTPITIVVWSVLLASIFSLWVRINPFDANPNANNFN | 1055 |
| OSCESA8 (4342699) | TTLLIPPTTILINLVGVVAGISYAINSGYQSWGPLEGKLFFAFWIVHLYPFLKGLMGRQNRTPITIVVWVAILLASIFSLWVRIDPFTTRVTGPDQT | 1056 |
| ATCESA1 (829376) | GKGGVF 1081 | |
| OSCESA8 (4342699) | CGINC- 1081 | |

Figure 1.2 Sequence comparison between *Arabidopsis thaliana* AtCESA1 and rice (*Oryza sativa*) OsCESA8 (NCBI GeneID in parenthesis) using Clustal Omega HHM sequence alignment (Sievers et al., 2011). Sequence conservation and similarity is shown below the sequence, defined by Clustal Omega as identical residues with a score of 1 in the Gonnet PAM 250 matrix (*), strongly similar properties with a score of >0.5 in the Gonnet PAM 250 matrix (:), weakly similar with a score of ≤ 0.5 and > 0 in the Gonnet PAM 250 matrix (.) and dissimilar with a score of 0 in the Gonnet PAM 250 matrix (blank). The N-terminal RING-type zinc finger domain (ZnF) and Zn binding Cys residues (-) are highlighted in yellow above the sequence. Predicted transmembrane α-helices (TMH) are underlined and labeled numerically. The plant-conserved region (P-CR, purple) and class-specific region (CSR, green) are highlighted above the sequence. Important catalytic motifs (orange) are highlighted above the sequence with critical U-motif (U1/2/3/4) residues indicated (+). The BCSA homologous finger helix (bold) and interfacial α-helices (IF1/2/3, blue) are highlighted above the sequence, including the former fifth TMH of CESA (IF3) that precedes the FxVTxK motif (pink) highlighted above the sequence. Known *Atcesa1* radially swelling1 (*rswl-1*) (Arioli et al., 1998) and *Atcesa8* fragile fiber6 (*fra6*) (Zhong et al., 2003) mutations are indicated (red) above the sequence. Mutations (red text) created in Chapter 5 are shown above the sequence.

The CESA catalytic core has sequence homology to *Rhodobacter sphaeroides* BCSA for which there are several crystal structures in the protein data bank (PDB ID: 4hg6, 4p00, 5ejz, 5eiY, 5ej1) (Figure 1.3) (Morgan et al., 2013, 2016; Olek et al., 2014; Slabaugh et al., 2014a; Nixon et al., 2016). BCSA enzymatically polymerizes the growing glucan chain using a nucleophilic S_N2-like substitution reaction (Lairson et al., 2008). This is done by deprotonating the C4 hydroxyl of the acceptor glucan likely through the Asp of the third U-motif (TED motif) in BCSA and CESA. The activated-nucleotide UDP-Glc substrate transfers the donor glucan C1 hydroxyl and the activated C4 of the acceptor. The BCSA structure also revealed several new motifs in the catalytic core that are conserved in CESAs. The HXKAGN motif near the catalytic Asp of the third U-motif might implicate the His/Lys/Lys of CESAs in important catalytic functions. The QTPH and FFCGS motifs are both involved in coordinating the terminal glucan and the growing chain alongside the Q/RxxRW motif, where the His residue of QTPH motif and Cys/Gly/Ser of the FFCGS motif interact with glucans further along the chain. The YNEP motif is involved in binding the UDP-Glc substrate, where the Tyr and Glu residues interact with the UDP. Unlike other motifs, however, the Tyr of the YNEP motif does not have a clear alignment to similar residues in CESA. The processive elongation of (1→4)-β-D-glucans is accomplished through the cyclical interactions between the finger helix of BCSA that moves from a membrane-facing position, interacting with the terminal glucose, to a cytosol-facing position, interacting with the newly polymerized glucan. Binding to the new glucan is hypothesized to be a consequence of the gating loop retracting to release UDP (Yang et al., 2015; Knott et al., 2016; Morgan et al., 2016). With an open gating loop, it is possible for a new UDP-Glc to bind, closing the gating loop. The binding of a new UDP-Glc and closing of the gating loop could act in moving the finger helix to the membrane-facing position. This could place the new glucan in the post-turnover position of the terminal glucose, breaking and reforming interactions of the Q/RxxRW, QTPH and FFCGS motifs with the evolving glucan chain, and resetting the cycle (Yang et al., 2015; Knott et al., 2016; Morgan et al., 2016).

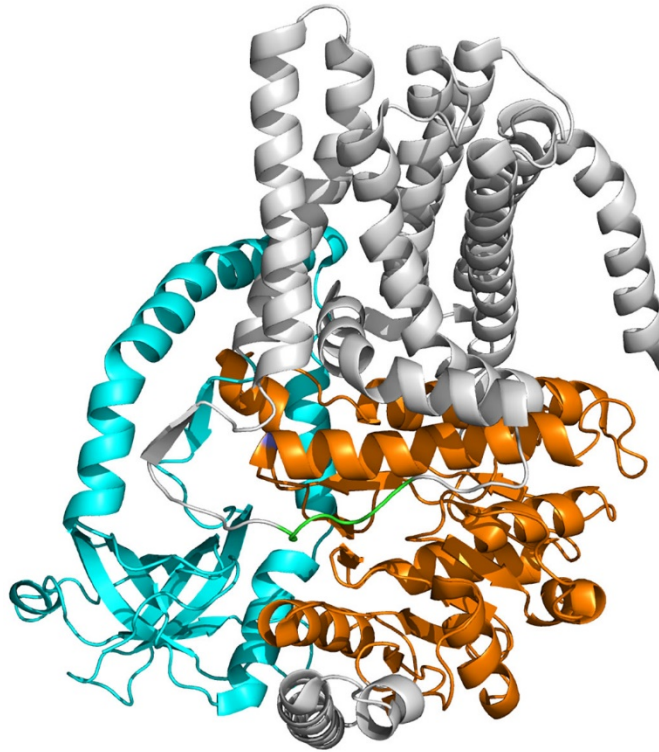


Figure 1.3 *Rhodobacter sphaeroides* bacterial cellulose synthase subunit A (BCSA) crystal structure (PDB ID: 4hg6) (Morgan et al., 2013). The homologous catalytic core shared by CESA and BCSA (orange), the non-homologous PilZ domain (cyan) not found in CESA and FxVTxK motif (green) are shown.

Computational and genetic analysis of what was previously predicted as the fifth TMH of CESAs showed that it likely folds into an interfacial helix instead of a TMH, where it could act in a fashion similar to the IF3 helix in BCSA (Slabaugh et al., 2014b, 2016; Pysh, 2015). Making the fifth TMH interfacial places a loop containing a highly conserved FxVTxK motif on the cytoplasmic side of the plasma membrane where it could act as gating loop similar to the gating loop and FxVTxK motif observed in BCSA. Analysis of the function of this motif took advantage of the *Arabidopsis thaliana* temperature-sensitive *rsw1-1* mutant of *AtCESA1* (Baskin et al., 1992; Arioli et al., 1998; Peng et al., 2000; Sugimoto et al., 2001; Williamson et al., 2001; J. Wang et al., 2008). The *rsw1-1* mutation creates an Ala549Val substitution in the catalytic core of CESAs, near the Asp of the second U-motif. This mutant CESA synthesizes cellulose microfibrils at or below the permissive temperature (23°C), however, at the restrictive temperature (31°C), cellulose synthesis stops. Lack of cellulose deposition corresponds with loss of CSCs from the plasma membrane, cessation of root elongation, and swelling of the root tip. Genetic complementation

with *AtCESA1* rescues normal growth of the root and cellulose synthesis at the restrictive temperature. In the case of the interfacial helix IF3 of CESA, *AtCESA1* sequences mutated in amino acids in the FxVTxK motif failed to complement *rswl-1* phenotypes at the restrictive temperature. For example, the Phe954Leu mutant did not complement the root length phenotype of *rswl-1* and had fewer functional CSCs in the plasma membrane and higher endomembrane accumulation (Slabaugh et al., 2014b), as measured by a citrine yellow fluorescent protein (YFP)-tagged *AtCESA6* (Paredes et al., 2006). This fluorescent fusion protein is expressed in the *procuste1* (*prc1*) *Atcesa6* null mutant (Desnos et al., 1996; Fagard et al., 2000) to monitor PCW CSC formation and behavior using confocal microscopy (Lei et al., 2012). The Val956Asp and Lys959Glu mutants of the FxVTxK motif did not rescue root length but did rescue the root swelling phenotype (Slabaugh et al., 2016). These mutants also showed lower cellulose microfibril anisotropy measured by cellulose microfibril alignment in confocal microscopy using the fluorescent cellulose-binding Pontamine Fast Scarlet dye (Anderson et al., 2010; Miart et al., 2014; Slabaugh et al., 2016). The Thr952Val mutant of this motif only partially rescued root length, but fully complemented the root swelling phenotype. In moss (*Physcomitrella patens*) PpCESA5, conserved residues in this motif showed partial complementation of the abnormal gametophore phenotype of a *Ppcesa5* knockout mutant (Goss et al., 2012), except for the Thr of the motif. Neither the Val956Asp mutant that did not rescue nor the Thr952Val mutant that did rescue showed disturbed cell wall architecture as measured by atomic force microscopy compared to the more disordered *rswl-1* cell wall (Slabaugh et al., 2016).

Gating loop interactions have also been hypothesized to be important in catalytic mechanisms in the closest CESA homolog CELLULOSE SYNTHASE-LIKE (CSL) glycosyltransferase CSLF6, a (1,3;1,4)- β -glucan synthase (Dimitroff et al., 2016). In this study, two CSLF6 residues facing the IF3 gating loop in a BCSA-based homology model were mutated. The Gly638Asp mutant, located a few residues upstream of the TED motif in the loop preceding the putative CESA finger helix, and the Tyr680Phe mutant, located just before the Q/RxxRW motif in the IF2 helix, both altered the degree of polymerization ratios of (1,3)- vs (1,4)- β -glucan linkages or the DP3:DP4 ratio, which indicates a modification of the catalytic turnover in these mutants.

1.3 Plant-Specific Domains of CESA

CESA genes are a subgroup of the CESA/CSL superfamily that encode processive glycosyltransferases. Multiple chains of (1→4)-β-D-glucan polysaccharides produced by *CESA* isomers hydrogen-bond to form the cellulose microfibril unique to plants. *CESA* genes belong to a subfamily of glycosyltransferases (GT), which is divided into two main superfamilies of nucleotide-sugar dependent enzymes, GT-A and GT-B, based upon sequence and catalytic fold, (Lairson et al., 2008). These proteins allow plants to generate the complex network of diverse polysaccharides within the plant cell wall (Carpita, 2011). The GT-A superfamily is sub-divided into families GT-1 to GT-4, with the GT-2 family containing the membrane-associated glycosyltransferases including CESAs, CSLs and other glycosyltransferases (Lairson et al., 2008).

Plant CESAs are distinguished from their bacterial homologs by three unique domains, one forming the extended N-terminus and two located in the catalytic core. The bacterial C-terminal PilZ domain that is responsible for synthesis activation via cyclic di-GMP is absent in plant CESAs (Figure 1.1 and 1.2) (Morgan et al., 2014). The novel plant-specific domains might replace the regulatory function of the PilZ or serve in assembly of CESA isoforms into oligomeric complexes, a requirement for plant cellulose synthases to produce microfibrils. For example, the N-terminal addition of the cytosolic ZnF was shown to function in redox-dependent homo- or heterodimerization between recombinant, truncated CESA N-terminal domains and might recruit CESAs into complexes (Kurek et al., 2002; Jacob-Wilk et al., 2006).

The sequences of the two domains within the plant CESA CatD, the P-CR and the CSR, are not similar to any sequences for which there are crystal structure and must be modeled separately against lower sequence identity template structures to create composite models (Olek et al., 2014). The CSR ranges between 100-150 amino acids across species, and was first thought to be hypervariable (Pear et al., 1996; Delmer, 1999), but instead forms a multi-clade sub-group structure that is conserved in CESA isoforms across plant species (Vergara et al., 2001; Penning et al., 2009; Carroll et al., 2011; Kumar et al., 2017). The sequence variation of the CSRs of these isoforms constitutes the primary distinction among the CESAs, defining the classes that are expressed in these groups. The CSR has been shown to facilitate a homodimeric interaction between CESAs, however, it is not known whether heterodimeric interactions can occur (Olek et al., 2014). The CSR has also been proposed as an intrinsically disordered domain based on molecular dynamics (MD) simulations (Scavuzzo-Duggan et al., 2018), but it might adopt a more

stable structure during interactions (Uversky, 2019), such as the homodimeric interaction observed *in vitro* (Olek et al., 2014). This dimer could also form one of the foundational interactions between CESAs in the CSC. *Ab initio* models of entire cotton (*Gossypium hirsutum*) or *Arabidopsis thaliana* CESAs 1, 3, 4, 6, 7 and 8 have been generated to approximate their structures based on a BCSA template (Figure 1.3) (Sethaphong et al., 2013, 2016; Nixon et al., 2016). However, substantial variation in the models of the P-CR and CSR between isoforms existed, particularly in the CSR. These models were also used for a homotrimer model of a CSC particle, but unlike the model proposed by Vandavasi et al. (2016), contacts were formed by the CSR instead of the P-CR. This illustrates the need for better structural models of the P-CR and CSR domains in the context of the full length CESA, and perhaps even models that include a plasma membrane complex with other CESA-interacting proteins.

In plants, early evidence for the CatD and short C-terminal region having class specificity came from swapping these domains between AtCESA1 and AtCESA3 isoforms, which conferred the ability to complement the null mutant of the donating isoform (J. Wang et al., 2006). Domain swaps of CSRs between *AtCESA1* and *AtCESA3* all complemented the *At rsw1-1* mutant background (Sethaphong et al., 2016). In domain swapping experiments of transgenic SCW AtCESAs (4, 7 and 8), expressed in their respective null mutant backgrounds, the CSR alone does not appear to confer all of the features that differentiate CESA isoforms, despite its class-based sequence diversification (Kumar et al., 2017). In these experiments, the CSR encompasses all of the second variable region but is slightly trimmed at the C-terminus. The results of swaps of the second variable region between isoforms differed across isoforms, showing *AtCESA4* and *AtCESA8* can complement *AtCESA7*. A more accurate definition of the CSR domain sequence boundaries could be useful in interpreting these results, which indicates that CESA isoform class specificity is a complex feature of CESA, which could span several regions in each CESA isoform. For example, the CESA N-terminus, in particular the first variable region in the N-terminal domain, appears to be linked to class specificity (Kumar et al., 2017). Swapping the CSR domains between isoforms in clade A and clade B CESAs of a moss (*Physcomitrella patens*), which lacks SCWs, shows complementation in null backgrounds and indicates their functional equivalence (Scavuzzo-Duggan et al., 2018). However, as mosses diverged early, it is possible that these results are not fully representative of the more complex flowering plant CESAs. These results imply that

class specificity is a complex characteristic of the entire CESA and CSC that likely requires unique contributions from the CSR.

The P-CR is an ~125 amino acid region with high sequence identity across all CESA isoforms (Vergara et al., 2001). The X-ray crystal structure of the P-CR shows that the first 77 amino acids form an anti-parallel coiled-coil with an unusually high proportion of aromatic amino acids (Rushton et al., 2017). Only the *Atcesa8* is mutated in the P-CR domain, resulting in the SCW derived *fragile fiber (fra)* phenotype, *fra6* (Zhong et al., 2003). This recessive mutation substitutes a Lys for a conserved Arg (Arg362Lys), resulting in loss of cellulose production in SCWs of homozygotes. As Lys and Arg are similar in size and charge, the *fra6* phenotype might indicate that the branched guanidino group of the Arg is important. It is unknown whether *fra6* gives rise to a properly folded CESA. The lack of a dominant negative phenotype in heterozygotes indicates that the *fra6* mutant CESA does not integrate or interfere with the CSC. Also there is no phenotype associated with over-expressing a transgenic *fra6 Atcesa8* isoform in wild-type (WT) *Arabidopsis thaliana*. Complementation experiments in SCW AtCESAs (4, 7 and 8) did not complement well for the conserved region 1 region containing the P-CR, except for a partial complementation when *AtCESA7* is used to complement the *Atcesa8* mutant (Kumar et al., 2017). The authors concluded that conserved region 1 likely interacts with conserved region 2 to form a functional catalytic core, which might not accurately reflect the P-CR as an independent domain insertion. When both the CSR and P-CR are swapped between clade A and clade B PpCESAs of moss a similar lack of complementation is observed (Scavuzzo-Duggan et al., 2018). A construct that swaps the P-CR specifically, without other catalytic core motifs, might show different results.

Multiple functions of the P-CR have been proposed related to interactions within a CESA, within the CSC, or in interactions with CESA-interacting proteins (Olek et al., 2014; Slabaugh et al., 2014a; Nixon et al., 2016; Sethaphong et al., 2016; Vandavasi et al., 2016; Rushton et al., 2017). One hypothesis for the function of the P-CR based on the best fit of the P-CR structure to the CatD SAXS structure is that it could interact with the catalytic core. Sequence connectivity puts the P-CR near the substrate entry and exit portal in the best fit model to the CatD (Olek et al., 2014; Rushton et al., 2017). Computational models of CESAs derived from a combination of homology modeling and MD have also utilized the P-CR in theoretical CSC contacts in models generated by SymmDock (Schneidman-Duhovny et al., 2005a, 2005b), focusing primarily on trimeric and tetrameric models (Nixon et al., 2016). Some support comes from a SAXS surface

model of an AtCESA1 CatD fit with similar computational models as above, which suggests the P-CR could mediate a trimeric interaction to form the fundamental CSC particle (Vandavasi et al., 2016). However, the biochemical reconstitution of this sample from inclusion bodies and the 3D alignment of this structure with the OsCESA8 CatD models indicates it might not serve as a biological contact in this manner (Olek et al., 2014; Rushton et al., 2017). One final consideration is that the P-CR could be important to the known interactions between CESAs and CESA-interacting protein(s), for which there is no known binding site (Rushton et al., 2017).

1.4 The *CESA* Gene Family

In the eudicot *Arabidopsis thaliana* and the grass rice (*Oryza sativa*), there are ten partially redundant *CESA* genes and several dozen *CSL* genes in six groups (Richmond et al., 2000; Hazen et al., 2003; Penning et al., 2009). The multigene families of plant *CESAs* (Richmond et al., 2000) are expressed differentially during plant growth; at least three isoforms are expressed during PCW formation (AtCESA1, AtCESA3 and AtCESA6 in *Arabidopsis thaliana*) and a different set of three isoforms is expressed during SCW synthesis (*AtCESA4*, *AtCESA7* and *AtCESA8* in *Arabidopsis thaliana*) (Tanaka et al., 2003; Appenzeller et al., 2004; Brown et al., 2005). Characterization of mutants for each of these isoforms reveals similar, non-additive phenotypes within their set, indicating they are non-redundant and perform specific roles within the same CSC (Taylor et al., 2003, 2008).

1.4.1 Primary Cell Wall (PCW) *cesa* Mutants

The critical nature of the PCW in plant growth and development leads to embryo or gametophyte lethality in many of the severe or null mutants of genes involved (Table 1.1) (Sechet et al., 2018). Null or severe mutants of *Atcesal* are embryo-lethal when homozygous, such as the *rswl-1*, *rswl-2*, *rswl-20* and *rswl-45* mutants that all have mutations in the CatD (Arioli et al., 1998; Beeckman et al., 2002; Gillmor et al., 2002; Persson et al., 2007). The *rswl-10* transposon insertion line in *Atcesal* is a leaky allele that leads to a loss of function (LOF) (Fagard et al., 2000; Desprez et al., 2007; Hematy et al., 2007), which is also less severe than the temperature-sensitive *rswl-1* CatD mutation (Arioli et al., 1998). Both are commonly used in *Atcesal* mutant complementation experiments due to their ability to progress through development. Several *Atcesal* mutations have also been shown to confer resistance to the C17 herbicide (Z. Hu et al., 2016). Other cellulose

synthase inhibitors like triazofenamide and flupoxam, part of the triazole carboxamides herbicide family, also have resistance mutations in AtCESA1, including *flupoxam resistance2 (fxr2)* mutants that are identical or close to *C17-tolerant* substitutions (Shim et al., 2018). Resistance to the cellulose synthesis inhibitor quinoxiphen has been identified in the *aegeus*, (Harris et al., 2012), *lycos* (Sethaphong et al., 2013) and *P2P5* mutants (Tateno et al., 2016). Both *aegeus* and *lycos* are less severe than other *Atcesa1* mutants that display a dwarfed phenotype, altered cellulose content and reduced microfibril crystallinity. The *aegeus* mutant also showed increased CSC velocity in the plasma membrane. The *anisotropy1 (any1)* mutant of AtCESA1 shows anisotropic growth, a dwarf phenotype, reduced cellulose crystallinity, and decreased CSC velocity in the plasma membrane but no reduction in cellulose content (Fujita et al., 2013).

Table 1.1 Table of allelic mutants in PCW *CESAs*

| Gene | Allelic mutants | Mutation | Structural Phenotype | Reference |
|----------------|---------------------------------|----------------------|---------------------------------------|---|
| <i>Atcesal</i> | <i>rsw1-1</i> | ---- | Ala549Val in the CatD | (Arioli et al., 1998) |
| <i>Atcesal</i> | <i>rsw1-2</i> | ---- | Gly631Ser in the CatD | (Gillmor et al., 2002) |
| <i>Atcesal</i> | <i>rsw1-20</i> | ---- | Asp780Asn in the CatD | (Beeckman et al., 2002) |
| <i>Atcesal</i> | <i>rsw1-45</i> | ---- | Glu779Lys in the CatD | (Beeckman et al., 2002) |
| <i>Atcesal</i> | <i>rsw1-10</i> | Transposon insertion | LOF | (Fagard et al., 2000) |
| <i>Atcesal</i> | <i>C17-tolerant 3D</i> | ---- | Val297Met between TMH1 and TMH2 | (Z. Hu et al., 2016) |
| <i>Atcesal</i> | <i>C17-tolerant 18A1 fxr2-4</i> | ---- | Ser307Leu in TMH2 | (Z. Hu et al., 2016) (Shim et al., 2018) |
| <i>Atcesal</i> | <i>C17-tolerant 3G</i> | ---- | Leu872Phe in TMH3 | (Z. Hu et al., 2016) |
| <i>Atcesal</i> | <i>C17-tolerant 3F</i> | ---- | Ser892Asn in TMH4 | (Z. Hu et al., 2016) |
| <i>Atcesal</i> | <i>C17-tolerant 14V</i> | ---- | Lys945Arg at the end of the IF3 helix | (Z. Hu et al., 2016) |
| <i>Atcesal</i> | <i>C17-tolerant 20C1 fxr2-2</i> | ---- | Pro1010Leu between TMH5 and TMH6 | (Z. Hu et al., 2016) (Shim et al., 2018) |
| <i>Atcesal</i> | <i>C17-tolerant 9R fxr2-1</i> | ---- | Gly1013Arg in TMH6 | (Z. Hu et al., 2016) (Shim et al., 2018) |
| <i>Atcesal</i> | <i>C17-tolerant 14U</i> | ---- | Gly1013Glu in TMH6 | (Z. Hu et al., 2016) |
| <i>Atcesal</i> | <i>C17-tolerant 3E</i> | ---- | Ala1018Val in TMH6 | (Z. Hu et al., 2016) |
| <i>Atcesal</i> | <i>C17-tolerant 1B</i> | ---- | Ala1023Thr in TMH6 | (Z. Hu et al., 2016) |
| <i>Atcesal</i> | <i>fxr2-3</i> | ---- | Gly1009Asp between TMH5 and TMH6 | (Shim et al., 2018) |
| <i>Atcesal</i> | <i>aegeus</i> | ---- | Ala903Val in TMH4 | (Harris et al., 2012) |
| <i>Atcesal</i> | <i>lycos</i> | ---- | Gly620Glu in the CatD | (Sethaphong et al., 2013) |
| <i>Atcesal</i> | <i>P2P5</i> | ---- | Arg292Cys between TMH1 and TMH2 | (Tateno et al., 2016) |
| <i>Atcesal</i> | <i>any1</i> | ---- | Asp604Asn in the CatD | (Fujita et al., 2013) |
| <i>Atcesa3</i> | <i>rsw5</i> | ---- | Pro1056Ser in the C-terminal domain | (J. Wang et al., 2006) |
| <i>Atcesa3</i> | <i>elil-1</i> | ---- | Ser301Phe in TMH2 | (Cano-Delgado et al., 2003) |
| <i>Atcesa3</i> | <i>elil-2</i> | ---- | Ala522Val in the CatD | (Cano-Delgado et al., 2003) |
| <i>Atcesa3</i> | <i>elil-3</i> | Transposon insertion | LOF | (Cano-Delgado et al., 2000) |

Table 1.1 continued

| Gene | Allelic mutants | Mutation | Structural Phenotype | Reference |
|----------------|--|----------------------|---|---|
| <i>Atcesa3</i> | <i>cevl</i> | ---- | Gly617Glu in the CatD | (Ellis et al., 2001) |
| <i>Atcesa3</i> | <i>mrel</i> (<i>Atces3^{je5}</i> or <i>je5</i>) | ---- | Gly916Glu in the IF3 helix | (Benfey et al., 1993) |
| <i>Atcesa3</i> | <i>than</i> (aligns with <i>Atcesa7 fra5</i>) | ---- | Pro578Ser between the second U-motif and CSR in the CatD (aligns to the QTPH motif of BCSA) | (Daras et al., 2009) |
| <i>Atcesa3</i> | <i>repp3</i> (aligns with <i>Atcesa7 fra5</i>) | ---- | Pro578Leu between the second U-motif and CSR in the CatD (aligns to the QTPH motif of BCSA) | (Feraru et al., 2011) |
| <i>Atcesa3</i> | <i>C17-tolerant 2C fxr1-3</i> (aligns with <i>Atcesa6 ixr2-2</i>) | ---- | Ser983Phe at the end of TMH5 | (Z. Hu et al., 2016) (Shim et al., 2018) |
| <i>Atcesa3</i> | <i>C17-tolerant 8P fxr1-2</i> | ---- | Ser1037Phe in TMH7 | (Z. Hu et al., 2016) (Shim et al., 2018) |
| <i>Atcesa3</i> | <i>fxr1-1</i> | ---- | Ser1040Lys in TMH7 | (Shim et al., 2018) |
| <i>Atcesa3</i> | <i>ixr1-1</i> | ---- | Gly998Asp in TMH6 | (Heim et al., 1989; Scheible et al., 2001) |
| <i>Atcesa3</i> | <i>ixr1-2</i> | ---- | Thr942Ile between the IF3 helix and TMH5 | (Heim et al., 1989; Scheible et al., 2001) |
| <i>Atcesa3</i> | <i>ixr1-3</i> | ---- | Gly998Ser in TMH6 | (Shim et al., 2018) |
| <i>Atcesa3</i> | <i>ixr1-4</i> | ---- | Arg806Lys in the Q/RxxRW motif in the CatD | (Shim et al., 2018) |
| <i>Atcesa3</i> | <i>ixr1-5</i> | ---- | Arg276His in TMH1 | (Shim et al., 2018) |
| <i>Atcesa3</i> | <i>ixr1-6</i> | ---- | Ser377Phe next to the first U-motif in the CatD | (Sethaphong et al., 2013) |
| <i>Atcesa6</i> | <i>prc1-1</i> <i>prc1-4</i> | Transposon insertion | LOF | (Desnos et al., 1996) |
| <i>Atcesa6</i> | <i>prc1-9</i> <i>prc1-19</i> | Transposon insertion | LOF | (Fagard et al., 2000) |
| <i>Atcesa6</i> | <i>ixr2-1</i> | ---- | Arg1064Trp in TMH7 | (Desprez et al., 2002) |
| <i>Atcesa6</i> | <i>ixr2-2</i> (aligns with <i>Atcesa3 C17-tolerant 2C</i> and <i>fxr1-3</i>) | ---- | Ser1002Phe at the end of TMH5 | (Shim et al., 2018) |

The temperature-sensitive radial-swelling root phenotype can also appear when *AtCESA3* is mutated, as observed *rsw5* mutant (Table 1.1) (J. Wang et al., 2006). In fact, the temperature-sensitive radial-swelling root phenotype, selected for in the original screen for anisotropic growth in cytoskeletal mutants (Baskin et al., 1992), can be found in several other mutated genes acting upstream of or associated with PCW synthesis. Examples include the KOR1 endo-1,4- β -D-glucanase (*rsw2*) (Lane et al., 2001; Sato et al., 2001), an α -glucosidase II involved in N-linked glycan processing in the endoplasmic reticulum (ER) and Golgi apparatus secretory pathway (*rsw3*) (Burn et al., 2002), a meiosis-related separase (*rsw4*) (Wiedemeier et al., 2002; Wu et al., 2010), a microtubule organization-related protein (*rsw6*) (Bannigan et al., 2006), a microtubule-associated kinesin-5 motor protein (*rsw7*) (Wiedemeier et al., 2002), *rsw8* (Wiedemeier, 1998), dynamin-related protein 1A involved in CME (*rsw9*) (Collings et al., 2008), a ribose 5-phosphate isomerase required for *de novo* synthesis of the CESA UDP-Glc substrate and other pyrimidine nucleotides (*rsw10*) (Howles et al., 2006), a mRNA splicing DEAH-box helicase (*rsw12*) (Howles et al., 2016) and an α -glucosidase I also involved in N-linked glycan processing in the ER and Golgi apparatus secretory pathway (*rsw13*) (Gillmor et al., 2002). While the shared temperature-sensitive radial-swelling phenotype and associated cellulose deficiency in some *Atcesa1* and *Atcesa3* mutants can also occur when these other proteins are mutated, as observed in the *rsw2*, *rsw3*, *rsw9*, *rsw10*, *rsw11*, *rsw12* and *rsw13* mutants, some like the *rsw4* and *rsw7* mutants do not show this cellulose deficiency. These other proteins can also show separate phenotypes related to their functions in other pathways, at different stages of development. Thus, both developmental and disturbed feedback mechanisms can be intimately tied to cellulose synthesis.

Alongside the *rsw5* mutation, mutant alleles of *AtCESA3* have various phenotypes including *ectopic lignification1* (*eli1*), *constitutive expression of VSP1* (*cev1*), *multiple response to expansion1* or *Atcesa3^{je5}* (*mre1/je5*), *thanatos* (*than*), *regulators of PIN polarity3* (*repp3*), C17 herbicide resistance, *flupoxam resistance1* (*fxr1*) and *isoxaben resistant1* (*ixr1*) mutants (Table 1.1). The recessive *eli1* mutants, show reduced cellulose synthesis and atypical lignin deposition in tissues that are generally not lignified (Cano-Delgado et al., 2000, 2003). This ectopic lignin deposition is also associated with isoxaben treatments (Cano-Delgado et al., 2003). In another stress response-related mutation, the recessive *cev1* mutant displays constitutive expression of genes encoding the JA-responsive vegetative storage proteins1/2 and thionin2.1, as well as, the JA-/ethylene-responsive plant defensin1.2 and basic endochitinase B genes (Ellis et al., 2001,

2002). This results in *cev1* plants that are dwarf and have shorter hypocotyls, shorter roots, reduced cellulose content and abnormal anthocyanin accumulation compared to WT, which phenocopies isoxaben and 2,6-dichlorobenzonitrile cellulose synthesis inhibitor treatments. The *mre1* mutant, formerly known as *Atcesa3^{je5}* or *je5*, of AtCESA3 is also dwarfed, reduced in cellulose content and has aberrant cell responses, but was less severe than the *eli1* and *cev1* mutants of the allelic series (Benfey et al., 1993; Pysh et al., 2012, 2015). The *mre1* mutant, in particular, has become a preferred background for transgenic complementation of *Atcesa3* mutants and fluorescence experiments utilizing a green fluorescent protein (GFP)-tagged *AtCESA3* transgene (Desprez et al., 2007). The *than* mutant shows semi-dominant and dominant-negative phenotypes that are homozygous lethal, but cause dwarfism, radial swelling, ectopic lignification and reduced cellulose synthesis in heterozygotes (Daras et al., 2009). The similar *repp3* mutant substitution shows a link between cellulose synthesis, gravitropic responses and auxin signaling through PIN-FORMED protein auxin transporters (Feraru et al., 2011). In this study, the *repp3* mutant, or chemical interference with the cell wall, restored PIN-FORMED protein localization and gravitropic response. C17 herbicide resistance has been observed in *Atcesa3* mutants, as well as in the *Atcesa1* C17 resistance mutants (Z. Hu et al., 2016). Flupoxam herbicide resistance has also been observed in both *Atcesa1* and *Atcesa3* mutants, with the *Atcesa3 fxr1* mutants, which are again that are identical or close to *C17-tolerant* substitutions (Shim et al., 2018). Semi-dominant inheritance of isoxaben and thiazolidinone herbicide resistance, associated with *Atcesa3* and *Atcesa6* mutant isoforms, was originally found in the *Atcesa3 ixr1* mutants, which did not alter plant growth or cellulose synthesis (Heim et al., 1989; Scheible et al., 2001; Sethaphong et al., 2013; Shim et al., 2018). Surprisingly, *ixr1* substitutions occur across the entire sequence including the N-terminal TMH region, the CatD and C-terminal TMH6, with several occurring in or near conserved catalytic motifs.

The primary cell wall *Atcesa6* mutations are less severe, but the AtCESA6 isoform is partially redundant with AtCESA2, AtCESA5 and AtCESA9, such that a single-knockout mutant in any of these is not as severe as mutations in *Atcesa1* or *Atcesa3* PCW CESAs (Desprez et al., 2007; Persson et al., 2007). Triple mutants knocking out *Atcesa2*, *Atcesa5* and *Atcesa6* result in a seedling-lethal phenotype (Desprez et al., 2007) and triple mutants knocking out *Atcesa2*, *Atcesa6* and *Atcesa9* show a gamete-lethal phenotype (Persson et al., 2007). Still, *Atcesa6* null mutants, known as *prc1* (Table 1.1), have decreased cell elongation and reduced cellulose in roots and dark-

grown hypocotyls, similar to *rsw1* phenotypes and isoxaben/2,6-dichlorobenzonitrile treatments associated with all cell types (Desnos et al., 1996; Fagard et al., 2000; MacKinnon et al., 2006). As a non-lethal and light-sensitive phenotype, the *prc1* null mutants are commonly used for PCW complementation experiments. In particular, the *Atcesa6 prc1-1* mutant has been used as the background for the expression of an *AtCESA6* transgene tagged with an N-terminal YFP (Paredes et al., 2006), providing a system to image CSC formation, motility, colocalization and trafficking, and sometimes as a reporter for other PCW CESAs and CESA-interacting proteins (Lei et al., 2012). An N-terminal GFP-tagged *AtCESA6* (Desprez et al., 2007) and N-terminal enhanced YFP-tagged *AtCESA6* (Park et al., 2019) have also been developed in the *prc1-1* background for similar analyses. Using this N-terminal enhanced YFP-tagged *AtCESA6* it has been shown that mutating the catalytic Asp in the first U-motif of *Atcesa6* to Asn retains *Atcesa6* in the ER and precludes normal activity in the CSC (Park et al., 2019). This is in contrast to a Gln to Glu mutation in the Q/RxxRW motif that traffics and synthesizes cellulose normally. Like *Atcesa3* mutants, *Atcesa6* mutants can confer isoxaben resistance through the *isoxaben resistant2 (ixr2)* mutants (Desprez et al., 2002; Shim et al., 2018).

1.4.2 Secondary Cell Wall (SCW) *cesa* Mutants

Mutants in SCW CESAs, such as *Atcesa4*, *Atcesa7* and *Atcesa8* isoforms, are also less severe than most PCW *cesa* mutants (Table 1.2) (Carpita et al., 2015). This is due to their expression and activity occurring during fiber and xylem cell differentiation. Mutants in these genes were originally isolated in a screen for *irregular xylem (irx)*, sometimes referred to as a collapsed xylem, phenotype (Turner et al., 1997).

Table 1.2 Table of allelic mutants in SCW *CESAs*

| Gene | Allelic mutants | Mutation | Structural Phenotype | Reference |
|----------------|--|----------------------|--|--|
| <i>Atcesa4</i> | <i>irx5-1</i> | Transposon insertion | LOF | (Taylor et al., 2003) |
| <i>Atcesa4</i> | <i>irx5-2</i> | ---- | Trp995stop in TMH6 | (Taylor et al., 2003) |
| <i>Atcesa4</i> | <i>irx5-3</i> | ---- | Gln263stop at the end of TMH2 | (Taylor et al., 2003) |
| <i>Atcesa4</i> | <i>irx5-4</i> | Transposon insertion | LOF | (Alonso et al., 2003; Brown et al., 2005) |
| <i>Atcesa4</i> | <i>irx5-5 nws2</i> | Transposon insertion | LOF | (Hernández-Blanco et al., 2007) |
| <i>Atcesa4</i> | <i>irx5-6</i> | Transposon insertion | LOF | (Kumar et al., 2018) |
| <i>Atcesa4</i> | <i>exi2</i> | ---- | Trp939stop in TMH5 | (Rubio-Díaz et al., 2012) |
| <i>Oscesa7</i> | NC0259 | Transposon insertion | LOF | (Tanaka et al., 2003) |
| <i>Oscesa7</i> | ND8759 | Transposon insertion | LOF | (Tanaka et al., 2003) |
| <i>Oscesa7</i> | <i>S1-24</i> | ---- | Cys40Tyr in the ZnF | (D. Wang et al., 2016) |
| <i>Atcesa7</i> | <i>irx3-1</i> | ---- | Trp859stop at the end of TMH4 | (Taylor et al., 1999) |
| <i>Atcesa7</i> | <i>irx3-4</i> | Transposon insertion | LOF | (Brown et al., 2005) |
| <i>Atcesa7</i> | <i>irx3-6</i> | Transposon insertion | LOF | (Kumar et al., 2015b) |
| <i>Atcesa7</i> | <i>irx3-7</i> | Transposon insertion | LOF | (Kumar et al., 2015b) |
| <i>Atcesa7</i> | <i>exi5</i> | ---- | Trp954stop at the start of TMH6 | (Rubio-Díaz et al., 2012) |
| <i>Atcesa7</i> | <i>mur10-1</i> | ---- | Trp444stop in the P-CR | (Bosca et al., 2006) |
| <i>Atcesa7</i> | <i>mur10-2</i> | ---- | His734Tyr in the CatD | (Bosca et al., 2006) |
| <i>Atcesa7</i> | <i>fra5</i> (aligns with <i>Atcesa3</i> than and <i>repp3</i>) | ---- | Pro557Thr between the second U-motif and CSR in the CatD | (Zhong et al., 2003) |
| <i>Oscesa9</i> | <i>bc6</i> | ---- | Arg588Gly between the second U-motif and CSR in the CatD | (Kotake et al., 2011) |
| <i>Oscesa9</i> | <i>bc88</i> | ---- | Pro421Leu in the CatD | (Rao et al., 2013) |
| <i>Atcesa8</i> | <i>fra6</i> | ---- | Arg362Lys in the P-CR | (Zhong et al., 2003) |
| <i>Atcesa8</i> | <i>irx1-1</i> | ---- | Asp683Asn in the third U-motif in the CatD | (Turner et al., 1997; Taylor et al., 2000) |

Table 1.2 continued

| Gene | Allelic mutants | Mutation | Structural Phenotype | Reference |
|----------------|---|----------------------|--|---|
| <i>Atcesa8</i> | <i>irx1-1</i> | ---- | Asp683Asn in the third U-motif in the CatD | (Turner et al., 1997; Taylor et al., 2000) |
| <i>Atcesa8</i> | <i>irx1-2</i> | ---- | Ser679Lys between the CSR and third U-motif in the CatD | (Taylor et al., 2000) |
| <i>Atcesa8</i> | <i>irx1-4</i> | Transposon insertion | LOF | (Kumar et al., 2018) |
| <i>Atcesa8</i> | <i>irx1-5</i> | Transposon insertion | LOF | (Brown et al., 2005) |
| <i>Atcesa8</i> | <i>irx1-7</i> | Transposon insertion | LOF | (Kumar et al., 2017) |
| <i>Atcesa8</i> | <i>exil-1</i> | mRNA splice variant | LOF | (Rubio-Díaz et al., 2012) |
| <i>Atcesa8</i> | <i>exil-2</i> | ---- | Gly508Glu between the second U-motif and CSR in the CatD | (Rubio-Díaz et al., 2012) |
| <i>Atcesa8</i> | <i>lew2-1</i> <i>ern1</i> <i>irx1-6</i> | ---- | Trp217stop in TMH2 | (Z. Chen et al., 2005; Hernández-Blanco et al., 2007) |
| <i>Atcesa8</i> | <i>lew2-2</i> | ---- | Leu792Phe between TMH3 and TMH4 | (Z. Chen et al., 2005) |
| <i>Oscesa4</i> | <i>bc7</i> | ---- | exon 10 and intron 10 deletions | (Yan et al., 2007) |
| <i>Oscesa4</i> | <i>bc11</i> | ---- | Gly858Arg in the IF3 helix | (B. Zhang et al., 2009) |

Mutant alleles of *AtCESA4* identified as *irx5* result in reduced cellulose content and thinner cell walls in xylem and interfascicular region cells depositing SCWs (Table 1.2) (Alonso et al., 2003; Taylor et al., 2003; Brown et al., 2005; Hernández-Blanco et al., 2007; Kumar et al., 2018). The *irx5-5* also shows enhanced disease resistance, a phenotype associated with abscisic acid (ABA)-responsive, defense-related genes that are distinct from some of the PCW mutants that also induce defense responses (Hernández-Blanco et al., 2007). The *exigua* (*exi*) mutants are SCW CESA mutants that result in small, dark-green leaves with a normal shape, defects in cell expansion and reduced cellulose content, which include *Atcesa4 exi2* (Rubio-Díaz et al., 2012). Mutations in *Oscesa7* (equivalent to the *AtCESA4* class) alter the SCW in rice resulting in the *brittle culm* (*bc*) mutant phenotype, characterized by reduced cellulose content, dwarfism, small leaves, withered

leaf tips and thin culms in allelic mutants (Tanaka et al., 2003). The *Oscesa7 SI-24* mutant shows a *bc*-like mutant phenotypes that is less severe (D. Wang et al., 2016), perhaps interfering with the proposed interactions of ZnFs between CESAs in the CSC. Mutations attempting to remove catalytic activity but retain CSC interactions have been made for *Atcesa4* in the critical Asp residues of the first (D motif; Asp334Asn), second (DxD motif; Asp501Asn) and third (TED motif; Asp748Asn) U-motifs. Each of these mutants shows differences in their ability to complement *Atcesa4* null mutants (Kumar et al., 2018). This was repeated in *Atcesa7* and *Atcesa8* SCW isoforms, which all indicate the first U-motif is less critical for activity as the first U-motif Asp mutants show better complementation than the second or third U-motif Asp mutants that poorly complement, do not complement or produce even more severe phenotypes.

In the *Atcesa7* isoform, *irx3* mutants also display the characteristic collapsed xylem phenotype (Turner et al., 1997) when AtCESA7 is knocked out (Table 1.2) (Taylor et al., 1999; Brown et al., 2005; Kumar et al., 2015b). The *exi* mutant phenotype also appears in the *Atcesa7 exi5* mutant (Rubio-Díaz et al., 2012). The *murus* (*mur*) mutants were isolated in a screen for altered quantity and distribution of cell wall monosaccharides (Reiter et al., 1997). Of these, *mur10* is mutated in AtCESA7 with phenotypes of slow growth, dwarfism, dark-green leaves, low seed-set, decreased fucose and xylose, and increased arabinose (Bosca et al., 2006). The rice *bc* mutant phenotype can be observed in *Oscesa9* (equivalent to the *AtCESA7* class) as well, through the *bc6* and *bc88* mutants (Kotake et al., 2011; Rao et al., 2013).

Mutations that disrupt interfascicular fiber cell development and SCW architecture have been characterized in the *fragile fiber* (*fra*) mutants, representing mutations in AtCESA7, AtCESA8 and several cytoskeleton-associated proteins, but not the AtCESA4 isoform (Carpita & McCann, 2015). These mutants have stems that are unable to stand upright and have lower mechanical strength, observed in a decreased force needed to break stems in a three-point bend test, and aberrant cellulose deposition in SCWs (Zhong et al., 1997). Other mutants with fragile fiber phenotypes include a kinesin-like motor protein (*fra1*) (Zhong et al., 2002), a katanin-like microtubule severing protein (*fra2*) (Bichet et al., 2001; Burk et al., 2001; Zhong et al., 2002), a signaling-related type II inositol polyphosphate 5-phosphatase (*fra3*) (Zhong et al., 2004) and a putative protein with GTP binding motifs also related to root hair defects (*fra4*) (Schiefelbein et al., 1990; H. Wang et al., 1997; Y. Hu et al., 2003). A phosphoinositide phosphatase, also known as suppressor of actin domain phosphatase (*fra7*) (Zhong, Burk, et al., 2005), and a xylan

glucuronyltransferase responsible for GlcA addition onto xylans (*fra8*) (Zhong, Pena, et al., 2005) have also been characterized. In *Atcesa7*, the *fra5* mutant results in a semi-dominant reduction of fiber cell wall thickness and cellulose content (Zhong et al., 2003). The *fra5* mutation aligns with the *than* mutant of the PCW *Atcesa3* isoform (Daras et al., 2009). This might cause misfolding, interference with functional CSC formation and dominant-negative phenotype observed in *than*. The *fra6* mutant, mutated in the AtCESA8 isoform, results in a recessive inheritance pattern of reduced fiber cell wall thickness and cellulose content (Zhong et al., 2003).

Some *Atcesa8* mutants exhibit the *irx* phenotype, collapsed xylem, reduced cellulose content and thinner SCWs (Table 1.2) (Turner et al., 1997; Taylor et al., 2000, 2003). Mutations in AtCESA8 occur in the *irx1-1* and *irx1-2* mutants. Transposon-insertional null mutants of *Atcesa8* include *irx1-4*, *irx1-5* and *irx1-7* (Brown et al., 2005; Carroll et al., 2012; Kumar et al., 2017, 2018). Unlike mutations in AtCESA6, where mutating the first U-motif resulted in ER retention (Park et al., 2019), mutating the third U-motif of AtCESA8 in the *irx1-1* mutant leads to cellulose deficiency but CESAs still colocalize with microtubules in the plasma membrane indicating normal trafficking (Gardiner et al., 2003). Similar *Atcesa8* mutants in the first, second and third U-motifs were created with the assumption that they do not interfere with other non-catalytic functions or progression through the CESA lifecycle (Kumar et al., 2018). However, it is still unclear whether a lack of complementation in these mutants could be manifested through improper transport as in the first U-motif mutant of an *Atcesa6* isoform (Park et al., 2019). Still, these mutated CESA sequences were expressed in *Atcesa8* null mutant backgrounds and showed complementation, indicating that the first U-motif is less critical than the second or third U-motif for catalysis. Similar results were obtained with corresponding complementation experiments for *Atcesa4* or *Atcesa7* (Kumar et al., 2018). *Atcesa8 exi1* allelic mutants also result in small, dark-green leaves with a normal shape, defects in cell expansion and reduced cellulose content (Rubio-Díaz et al., 2012). Unlike other *exi* mutants in *Atcesa4* or *Atcesa7* that are complete knockouts, *exi1* mutants showed the weakest *exi* phenotypes and might not abolish catalytic activity. *Atcesa8 leaf wilting (lew)* mutants can induce drought and osmotic stress tolerance, likely through defense-related and ABA-responsive gene expression, but produce a strong collapsed xylem phenotype (Z. Chen et al., 2005). The *lew2-1* mutation has also been identified in other defense response screens, identifying it as *enhanced resistance to necrotrophs1 (ern1)* and later *irx1-6* (Hernández-Blanco et al., 2007). The rice *bc* mutant is mutated in *Oscesa4*, an AtCESA8 equivalent isoform, with

alleles *bc7* and *bc11*. The *bc7* mutation has a ~10% reduction in cellulose content, however, it does not show the dwarfed phenotype associated with impaired growth and development of other *bc* mutants (Yan et al., 2007). The *bc11* allele has reduced cellulose, disrupted SCW architecture and dwarfism (B. Zhang et al., 2009).

1.4.3 Post-Translational Modifications of CESAs and CESA-Interacting Proteins

Several CESA isoforms and CESA-interacting proteins have had post-translational modifications (PTM) identified, some of which can be associated with roles in modifying CESA activity, plasma membrane mobility, stability and trafficking (S. Li et al., 2014; McFarlane et al., 2014; Kumar et al., 2015a, 2016; Jones et al., 2016; Speicher et al., 2018; Turner et al., 2018; Polko et al., 2019). Phosphoproteomic surveys conducted on *Arabidopsis thaliana* (Nuhse et al., 2004), rice (*Oryza sativa*) (Nakagami et al., 2010) and maize (*Zea mays*) (Facette et al., 2013) have identified the majority of CESA phosphorylation sites (Table 1.3), summarized primarily in the PhosPhAt 4.0 database (Zulawski et al., 2013). Other plant phosphoproteomic surveys can be found in other databases (Schulze et al., 2015). CESA N-terminal domains, prior to the first predicted TMH, contains the most phosphorylation sites across both PCW and SCW CESAs but some CatD phosphorylation sites exist. N-terminal domain phosphorylation sites have been linked to microtubule interactions and anisotropic cell expansion by phosphonull (Ala) and phosphomimic (Glu) mutants expressed in an *Atcesa1 rsw1-1/Atcesa6 prc1-1/AtCESA6:YFP* transgene background to monitor CSC colocalization (S. Chen et al., 2010), through an indirect interaction with a CSI protein (Y. Gu et al., 2010; Bringmann et al., 2012; S. Li et al., 2012; Lei et al., 2013, 2015). Similar experiments on *Atcesa3* phosphonull and phosphomimic mutants support this microtubule association with phosphorylation across isoforms (S. Chen et al., 2016). AtCESA1 CSR phosphorylation has been shown to increase CSC activity in phosphomimic mutants, while phosphonull mutants decrease activity (S. Chen et al., 2010). Phosphorylation appears to regulate AtCESA5 activity through microtubule interactions in dark-grown hypocotyls under the control of the red/far-red PHYTOCHROME B photoreceptor (Bischoff et al., 2011), which is also involved with the dark-dependent phenotype of the *Atcesa6 prc1-1* mutant (Fagard et al., 2000). This light/dark transition might also manifest through light-induced increases in phosphorylation of AtCESA1 and AtCESA3 that are correlated with cellulose synthesis (Boex-Fontvieille et al., 2014). In the SCW AtCESA7 isoform, N-terminal domain phosphorylation increases protein

degradation by the 26S proteasome, perhaps by destabilizing its structure (Taylor, 2007). Whether this is independent of potential AtCESA7 microtubule interactions is unclear. Finally, large-scale proteomics experiments have identified several Ser/Thr phosphorylations in CESA-interacting proteins KOR1, CSI1 and CC1/2 (Speicher et al., 2018).

Table 1.3 Table of phosphorylation sites in CESA

| Protein | Structural Phenotype | Reference |
|---------|--|---|
| AtCESA1 | Ser24 before the ZnF in the N-terminal domain | (Nuhse et al., 2004; Speicher et al., 2018) |
| AtCESA1 | Ser134, Ser152, Thr157, Ser162, Thr165, Thr166, Ser167, Ser179, Ser180 and Thr243 between the ZnF and TMH1 | (Nuhse et al., 2004; Speicher et al., 2018) |
| AtCESA1 | Ser686 and Ser688 in the CSR | (Nuhse et al., 2004; Speicher et al., 2018) |
| AtCESA3 | Ser3 and Thr7 before the ZnF in the N-terminal domain | (Nuhse et al., 2004; Speicher et al., 2018) |
| AtCESA3 | Ser130, Ser139, Thr143, Ser144, Ser151, Tyr169, Ser170, Ser171, Ser176, Ser211, Thr212, Ser216, Ser226, Thr227 and Ser249 between the ZnF and TMH1 | (Nuhse et al., 2004; Speicher et al., 2018) |
| AtCESA3 | Ser671 in the CSR | (Nuhse et al., 2004; Speicher et al., 2018) |
| AtCESA6 | Ser11 before the ZnF in the N-terminal domain | (Nuhse et al., 2004; Speicher et al., 2018) |
| AtCESA2 | Ser11 before the ZnF in the N-terminal domain | (Nuhse et al., 2004; Speicher et al., 2018) |
| AtCESA5 | Ser11 before the ZnF in the N-terminal domain | (Nuhse et al., 2004; Speicher et al., 2018) |
| AtCESA5 | Ser122, Thr124, Ser126, Ser229 and Ser230 between the ZnF and TMH1 | (Nuhse et al., 2004; Speicher et al., 2018) |
| AtCESA4 | Ser135, Ser136 and Ser139 between the ZnF and TMH1 | (Nuhse et al., 2004; Speicher et al., 2018) |
| AtCESA4 | Ser374 between the first and second U-motifs in the CatD | (Nuhse et al., 2004; Speicher et al., 2018) |
| AtCESA7 | Ser180, Ser181 and Ser185 between the ZnF and TMH1 | (Nuhse et al., 2004; Speicher et al., 2018) |

Kinases involved in cell wall biosynthesis have been correlated with CSC activity, however, evidence for their interaction with the CSC and precise activity against these CESA phosphorylation sites, CESA-interacting proteins or upstream elements is limited (McFarlane et al., 2014; Jones et al., 2016; Speicher et al., 2018; Polko et al., 2019). Direct evidence for kinase

activity on CESA1 sites comes from genetic and *in vitro* experiments on the BRASSINOSTEROID INSENSITIVE2 (BIN2), a glycogen synthase kinase3-like kinase (Sánchez-Rodríguez et al., 2017). When constitutively active in the *bin2-1* mutant, BIN2 leads to decreased cellulose content and CSC motility, which correlates with decreased brassinosteroid synthesis. *In vitro*, BIN2 was shown to phosphorylate AtCESA1 Ser162, which primed Thr157 phosphorylation by BIN2. The nearby Thr165, Thr166 and Ser167 phosphorylation sites in AtCESA1 were not phosphorylated by BIN2, nor were sites in AtCESA3, AtCESA5 or KOR1, whose phosphorylation sites are known to be involved in cellulose biosynthesis and protein regulation. This indicates that some kinases have selective activity at specific phosphorylation sites and CESA isoforms, which in this case is responsive to brassinosteroids. It is unclear how phosphorylation modulates activity of CESAs, as well as, activities of CESA-interacting proteins. More complete CESA and CESA-interacting protein structures with different phosphorylation states are needed to elucidate the precise molecular mechanisms. Several Ser/Thr receptor-like kinases (RLKs) have also been associated with many cell wall related phenotypes that could be directly or indirectly involved in cellulose biosynthesis. Mutations of THESEUS1 (THE1) RLK, aligned to the *Catharanthus roseus* RLK-like family suppresses the *Atcesa6 prc1-1* null mutant ectopic lignification phenotype but not the cellulose deficiency phenotype (Hematy et al., 2007; Merz et al., 2017). PCW mutants stacked with *the1* mutants also support THE1 acting as a part of sensory mechanism for cell wall integrity, often linked to aberrant cellulose biosynthesis, which leads to ectopic lignification in many CESA and CESA-interacting protein mutants. The leucine-rich repeat RLK MALE DISCOVERER 1-INTERACTING RECEPTOR LIKE KINASE2, also known as LEUCINE-RICH REPEAT KINASE INDUCED BY SALT STRESS, controls root growth direction (twisting) and salt tolerance in a AtCESA6- and THE1-dependent manner, implying that it acts upstream of THE1 in stress response signaling pathways (Van der Does et al., 2017). The FEI1 and FEI2 leucine-rich repeat RLK, *fei* for the Chinese word for fat, have also been implicated in sensing cell wall integrity and signaling through the ethylene biosynthesis pathway (Xu et al., 2008), which is a known phenotype of *Atcesa3* mutants such as *cevl* (Ellis et al., 2002). Double mutants of *fei1/fei2* show reduced cellulose content and radial root-tip swelling under high sucrose growth conditions (Xu et al., 2008). Other proteins like the fasciclin-like arabinogalactan-protein SALT OVERLY SENSITIVE5 also appear to be involved in the interactions between cellulose synthesis and hormone signaling (Basu et al., 2016). Other RLKs,

such as FERONIA (Cheung et al., 2011; Shih et al., 2014; Feng et al., 2018), FER homologs ANXUR1 and ANXUR2 (Boisson-Dernier et al., 2013), HERCULES RECEPTOR KINASE1 (Guo et al., 2009) and Ser/Thr wall-associated kinases (Wagner et al., 2001) are also considered candidates for activity against CESA and CESA-interacting protein phosphorylation sites (McFarlane et al., 2014; Jones et al., 2016; Speicher et al., 2018; Polko et al., 2019).

Other PTMs such as N-glycosylation (S. Li et al., 2014; Kumar et al., 2015a), ubiquitination (Callis, 2014; McFarlane et al., 2014; Kumar et al., 2015a; Speicher et al., 2018) and SUMOylation (McFarlane et al., 2014) could regulate activity/conformational or degradation of CESA, the CSC or CESA-interacting proteins, however, only S-acylation (palmitoylation) has been demonstrated to occur (Hemsley et al., 2013; Kumar et al., 2016). S-acylation occurs on both PCW and SCW isoforms on a cluster of conserved Cys residues early in the CSR and two conserved Cys residues in the C-terminal region after TMH7 but not the eight Cys residues in the ZnF or other Cys residues across CESA isoforms (Kumar et al., 2016). While the C-terminal region Cys residues are individually important for function, as Cys to Ser mutants cannot complement the cellulose deficiency of *irx* mutants, the CSR Cys residues appeared redundant no phenotype in single mutants and gradually decreasing complementation of an *irx*-like phenotype in the quadruple mutant. Both C-terminal and CSR Cys residues appear fully acylated and do not interfere with CSC interactions as both AtCESA4 and AtCESA8 can be co-immunoprecipitated with acylated AtCESA7. Unacylated AtCESA7 CSR or C-terminal region integrated into a CSC, mimicked through Cys to Ser substitutions, however, produces MASCs/SmaCCs that do not properly colocalize with microtubules or insert into the plasma membrane. This is despite apparently normal CESA and MASCs/SmaCC behavior in the Golgi apparatus. It is unclear how this acylation alters the structure of CESAs or precise interactions within the CSC or with CESA-interacting proteins, such as CSR dimerization (Olek et al., 2014) or the intrinsically disordered domain structure predicted by MD experiments (Scavuzzo-Duggan et al., 2018). The importance of the short C-terminal sequence following the final TMH was known through the *rsw5 Atcesa3* knockout and domain swaps of the C-termini of AtCESA1 and AtCESA3 (J. Wang et al., 2006). N-glycosylation was initially hypothesized for CESAs because α -glucosidase I (*rsw13*) (Gillmor et al., 2002) and α -glucosidase II (*rsw3*) (Burn et al., 2002) mutants, involved in N-linked glycan processing in the ER and Golgi apparatus secretory pathway, show cell wall defects and impaired cellulose biosynthesis. In addition, a cytokinesis-defective mutant in a mannose-1-phosphate

guanylyltransferase (Nickle et al., 1998; Lukowitz et al., 2001), an α -1,2-mannosyltransferase (*lew3*) (M. Zhang et al., 2009) and two GT 75 family glycosyltransferases STELLO1 and 2 (Y. Zhang et al., 2016), had similar phenotypes. However, Peptide N-glycosidase F treatment assays of PCW CESAs indicated CESAs are not N-glycosylated (Gillmor et al., 2002). However, N-glycosylation of KOR1 was shown to be important for KOR1 cellulase activity (Liebminger et al., 2013). Little direct evidence of CESA ubiquitination has been found, despite the link between ubiquitins and plasma membrane-embedded protein internalization or protein degradation that is important for CESA turnover (McFarlane et al., 2014). AtCESA6 has been identified in ubiquitination screens but below the confidence threshold, and therefore requires further investigation (Manzano et al., 2008). The CESA homolog CSLD, a mannan synthase with high homology to CESAs in the N-terminal domain, has been shown to be ubiquitinated (F. Gu et al., 2016), which might indicate similarity in the phosphorylation and ubiquitination sites between these homologs (Speicher et al., 2018). The ubiquitin E3 ligase activity of most RING-type zinc finger domains and sequence similarity of the ZnF with known ubiquitin E3 ligases, within the tolerance of homology-based modeling, has led some to hypothesize the CESA ZnF might also possess this activity but this has not been identified (Callis, 2014; Kumar et al., 2015a). No evidence for SUMOylation has been found (McFarlane et al., 2014).

1.5 Synthesis, Trafficking and Life Cycle of CESAs and the CSC

Progress in the understanding of CESA protein synthesis, trafficking, degradation and recycling has been greatly assisted through the development of fluorescently labeled CESAs and the identification of numerous CESA mutants, CESA-interacting protein mutants and mutations in upstream pathways (Bashline et al., 2014; McFarlane et al., 2014; S. Li et al., 2014; Carpita et al., 2015; Kumar et al., 2015a; Speicher et al., 2018; Polko et al., 2019). Proper CESA protein synthesis and folding is a critical first step in cellulose biosynthesis and several CESA mutants might be compromised at this step, such as the *than* mutant (Daras et al., 2009). The AtCESA6 Asp395Asn mutation of the catalytic Asp of the first U-motif is retained in the ER (Park et al., 2019), while other catalytic U-motif mutants behave differently (Gardiner et al., 2003). This implies transport mechanisms from the ER to the Golgi apparatus can be interrupted for some CESA mutants. Individual CESAs are transported to the Golgi apparatus by either the coat protein complex II vesicle system or direct channeling through ER-Golgi apparatus membrane

connections (Neumann et al., 2003; Park et al., 2019). Within the Golgi apparatus the first CSCs begin to appear (Haigler et al., 1986), which is believed to occur with the assistance of the Golgi-localized STELLO1 and 2 proteins (Y. Zhang et al., 2016). CESAs are also S-acylated in the Golgi apparatus as part of the controls that regulate MASC/SmaCC progression and interactions with microtubules (Kumar et al., 2016). The involvement of other CESA-interacting proteins is not clear, some of which are known to be post-translationally modified in the Golgi apparatus, such as KOR1 (Liebminger et al., 2013). KOR1 is trafficked in small vesicles that resemble MASCs/SmaCCs (Robert et al., 2005), however, proteomic experiments on plant Golgi apparatus and Golgi apparatus-derived vesicles, separated by flotation centrifugation and free-flow electrophoresis, show CESAs and KOR1 are enriched in different fractions, indicating they might be trafficked by different mechanisms and only interact at the plasma membrane (Okekeogbu et al., 2019). CSCs interact with the exocyst complex to form the characteristic MASCs/SmaCC vesicles (X. Zhu et al., 2018). From the Golgi apparatus, MASCs/SmaCCs are trafficked along cortical actin using the MYOSIN XI motor protein for vesicle motility (Sampathkumar et al., 2013; W. Zhang et al., 2019). CSC insertion and trajectory is defined by specific sites on cortical microtubules, facilitated by CSC interactions with CSI1 (Y. Gu et al., 2010; Bringmann et al., 2012; S. Li et al., 2012; Lei et al., 2015), where they pause for insertion (Crowell et al., 2009; Gutierrez et al., 2009). Both CSI1 and exocyst proteins interact with the PATROL1 protein, for which the precise molecular activity is unknown, but is thought to prime the MASCs/SmaCC vesicles for plasma membrane insertion (X. Zhu et al., 2018). The *shou4* mutants show increased CSC exocytosis and hyper-accumulation of CSCs in the plasma membrane, and are suppressors of *Atcesa6 prc1-1*, *sos5* and *feil/fei2* mutants, but not other CESA-interacting protein mutant phenotypes like *cobra* (Polko et al., 2018). Integral membrane SHOU4 proteins, which lack significant homology to any known proteins, appear to block CSC insertion through direct interactions between CESAs and the SHOU4 N-terminal cytosolic domain.

After the CSC is inserted into the plasma membrane, it maintains interactions with CSI1 and cortical microtubules that define the alignment and trajectory of CSCs, cellulose deposition and the axis of anisotropic cell expansion (Crowell et al., 2009; Gutierrez et al., 2009; Y. Gu et al., 2010; Bringmann et al., 2012; S. Li et al., 2012; Lei et al., 2015). Cellulose synthase-microtubule uncoupling1 and 2 (CMU1/2) proteins interact with microtubules and are necessary for proper cortical microtubule spacing and maintenance of CESA-microtubule interactions, and might act in

stabilizing these interactions through the mechanical force of active cellulose synthesis and CSC propulsion (Z. Liu et al., 2016). Alongside CSI1, the CESA-interacting proteins CC1/2 interact with microtubules. During stresses that impair cellulose synthesis, such as high salt stress, these proteins function to promote the reassembly of microtubules through their N-terminal domains and recruit MASC/SmaCC vesicles through their C-terminal interactions with CESAs (Endler et al., 2015, 2016). At the plasma membrane, KOR1 (Lane et al., 2001; Sato et al., 2001; Lei et al., 2014; Mansoori et al., 2014; Vain et al., 2014) and COBRA (Schindelman et al., 2001; Roudier et al., 2005; L. Liu et al., 2013; Sorek et al., 2014) are actively modifying the cellulose microfibril through 1,4- β -D-glucan chain cleavages and maintenance of the crystalline architecture, both of which affect the motility of the CSC complex. CESA/CESA-interacting protein kinases are also active in the plasma membrane, such as the BIN2 kinase, known to phosphorylate AtCESA1 and alter CSC motility (Sánchez-Rodríguez et al., 2017). Other kinases are integral membrane proteins and/or known to be involved in cell wall integrity signaling, implicating them in plasma membrane associations, which also often lead to changes in the trafficking and endocytosis of CSCs (McFarlane et al., 2014; Jones et al., 2016; Speicher et al., 2018; Polko et al., 2019). Lipid kinases that generate specific phosphorylated derivatives of the phosphatidylinositol lipid in the plasma membrane component of signal transduction mechanisms and have also been shown to alter CESA trafficking through the phosphatidylinositol 3- and 4- kinases (Fujimoto et al., 2014). Phosphatidylinositol 3-kinase inhibition results in failure to traffic AtCESA3, and perhaps the CSC, from the Golgi apparatus into MASCS/SmaCCs, while Phosphatidylinositol 4-kinase inhibition results in abnormal CSC internalization from the plasma membrane, and might be involved early in the CME pathway responsible for CSC internalization. Finally, within the plasma membrane of plants, there are several phytosterols that alter cellulose biosynthesis and could be involved direct interactions with the CSC and/or CESA-interacting proteins through lipid microdomains (lipid rafts) (Schrack et al., 2012).

CME-based CSC internalization is critical to plant cell growth and development as it helps regulate the quantity and location of cellulose synthesis in balance with CSC exocytosis and other forms of CSC regulation (McFarlane et al., 2014; S. Li et al., 2014; Carpita et al., 2015; Kumar et al., 2015a; Speicher et al., 2018; Polko et al., 2019). Several proteins involved in CME have been shown to make direct interactions with CESAs during this process and are thought to recognize the CSC cargo and recruit the rest of the CME subunits. The CME protein AP2M interacts with

the CatD of AtCESA3 and AtCESA6 and the *ap2m-1* mutant results in increased CSC concentration at the plasma membrane (Bashline et al., 2013). Other CME proteins that are a part of the large adaptor TPLATE complex interact with CESAs but have a distinct and additive effect on improper CSC endocytosis compared to the *ap2m-1* mutant, such as the TWD-40 protein (Bashline et al., 2015). Subsequent TPLATE complex subunits that interact with CESAs were identified as TPLATE and TPLATE complex muniscin-like proteins, which are thought to act early in CME cargo recognition (Sánchez-Rodríguez et al., 2018). Following endocytosis, CSCs are often degraded in lytic vacuoles, which can be coordinated with the transition from PCW CESA isoforms to SCW (Watanabe et al., 2015, 2018). Polyubiquitinated proteins are trafficked to lytic vacuoles and this could be the fate of ubiquitinated CESAs (Callis, 2014; McFarlane et al., 2014; Kumar et al., 2015a; Speicher et al., 2018). Alternatively, degradation can be promoted through phosphorylation thought to destabilize CESA with degradation by the 26S proteasome (Taylor, 2007). Thus, CESA degradation might involve multiple pathways and signals, likely related to upstream pathways triggering the regulation of cellulose biosynthesis and the stability of the CESA domain. Early estimates of CSC lifespan in the plasma membrane using FFTEM following protein synthesis inhibition on green algae (*Oocystis*) CSCs (Robinson et al., 1980) and moss (*Funaria hygrometrica*) (Rudolph et al., 1989) cells estimated the CSC half-life to be less than 30 minutes. CESA-specific detection by western blot of cotton (*Gossypium hirsutum*) GhCesA1 also suggested a half-life of less than 30 minutes (Jacob-Wilk et al., 2006). Other estimates of the lifespan of AtCESA1, AtCESA3 and AtCESA6 PCW isoforms detected by isoform-specific antibodies show 80% of CESAs are retained after 48 hours following inhibition of protein synthesis in etiolated seedlings, but both light and temperature significantly reduced this lifespan (Hill et al., 2018a). Dynamic estimates of CSC plasma membrane residence time, based on density and insertion rates of YFP tagged AtCESA6, however, is only 7-8 minutes (Sampathkumar et al., 2013). Together these results imply that CESA degradation might be highly regulated and differ between species, cell type, cell life cycle or through variety of external stimuli. Furthermore, CESAs could spend more time in MASCs/SmaCCs following exocytosis or endocytosis, rather than being degraded. This is consistent with the idea that CSC recycling is a common and important aspect of CSC trafficking, whose machinery could hold functional CSCs in a pool inside of the cell (Bashline et al., 2014; McFarlane et al., 2014; S. Li et al., 2014; Carpita et al., 2015; Kumar et al., 2015a; Speicher et al., 2018; Polko et al., 2019). This recycling could be facilitated by the microtubule-

associated and CESA-interacting CC1/2 (Endler et al., 2015, 2016) and CSI1 (Lei et al., 2015) proteins, which have been linked to recycling under stress but not necessarily under steady-state conditions.

1.6 CESA Integration into a CSC

CESA oligomerization has been observed using FF-TEM showing CESA proteins forming six “particles” that align in a larger circular, six-fold symmetrical structure or CSC termed a “rosette” (Giddings et al., 1980; Mueller et al., 1980; Herth, 1985; Delmer, 1999) (Figure 1.4). *In vitro* recombinant CESA proteins form a variety of homooligomers whose atomic structures or functions are still unclear *in vivo* (Olek et al., 2014; Vandavasi et al., 2016; Rushton et al., 2017). Cellulose microfibrils can be synthesized using a single recombinant SCW hybrid aspen (*Populus tremula x tremuloides*) PttCESA8 isoform (Purushotham et al., 2016) or a clade A PpCESA5 of moss (Cho et al., 2015, 2017) isolated in proteoliposomes, indicating that only one isoform is needed to form a functional CSC. However, it remains to be seen if this is true for all isoform classes. Genetic evidence for homooligomerization *in vivo* exists from intragenic complementation of different *Atcesa3* mutants (Pysh, 2015).

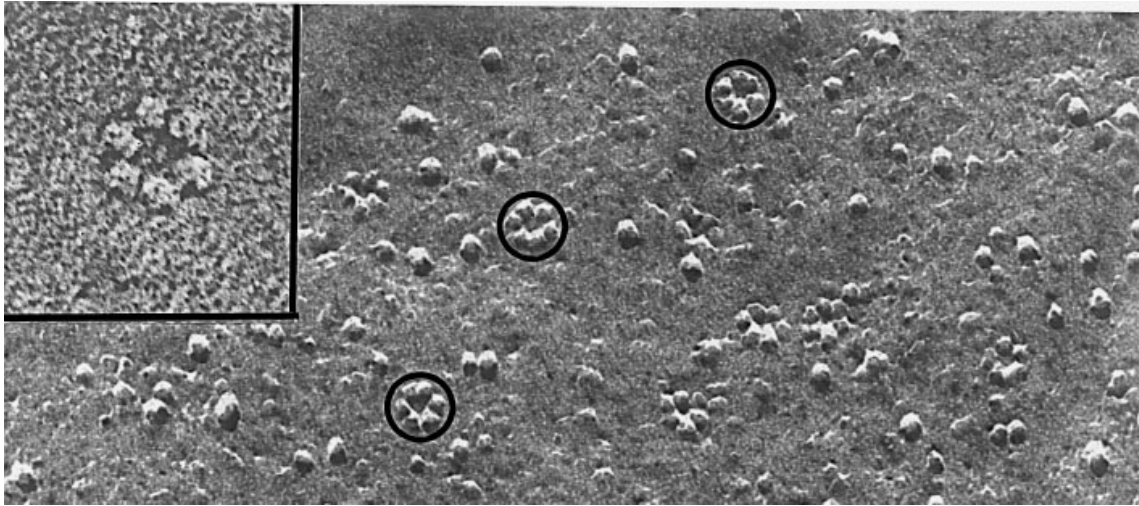


Figure 1.4 Freeze-fracture transmission electron microscopy (FF-TEM) images showing six-fold symmetry of the CSC formed by CSC particles. Adapted from C. Haigler in Delmer et al., 1999. Published legend: “Freeze fracture replicas of rosettes associated with cellulose microfibril biogenesis. The rosettes after the fracture event exist in the leaflet of the plasma membrane bilayer that is nearest the cytoplasm (the PF face). In the main micrograph, several rosettes are shown (three surrounded by circles) in the plasma membrane of a differentiating tracheary element of *Zinnia elegans*; differentiating tracheary elements deposit abundant cellulose into patterned secondary wall thickenings. The inset shows one rosette at higher magnification and after high resolution rotary shadowing at ultracold temperature with a minimum amount of platinum/carbon.” (Main micrograph, 222,000 x; inset, 504,545 x; both micrographs courtesy of Mark J Grimson and Candace H Haigler, Department of Biological Sciences, Texas Tech University, Lubbock, Texas.)

Assuming cellulose microfibril formation requires several CESA monomers to oligomerize from at least three specific CESA isoforms, early models of CSC formation hypothesized a 1:1:1 ratio of CESA isoforms (Taylor et al., 2003). However, expression of CESA isoforms, supposedly in the same CSC, varies greatly in different cells, tissues and across species, such as in *Arabidopsis thaliana* (Richmond et al., 2001; Brown et al., 2005), barley (*Hordeum vulgare*) (Burton et al., 2004), rice (*Oryza sativa*) (Tanaka et al., 2003), maize (*Zea mays*) (Holland et al., 2000; Dhugga, 2001; Appenzeller et al., 2004), cotton (*Gossypium hirsutum*) (A. Li et al., 2013) and aspen (*Populus tremula*) (X. Zhang et al., 2018). This variable expression, combined with models that assume a necessity to form differential contacts in CSC particles led to the hypothesis of a 1:2:3 ratio of the three CESA isoforms in a CSC particle forming a 36 CESA CSC (Ding et al., 2006). Quantitative immunoblotting of SCW CESAs from *Arabidopsis* stem protein extracts (Hill et al.,

2014) or coimmunoprecipitations of PCW CESAs from seedling protein extracts (Gonneau et al., 2014) using isoform-specific antibodies indicated a 1:1:1 ratio. Quantitative proteomics performed with label-free mass spectrometry also supports the 1:1:1 ratio of SCW CESA isoforms in *Arabidopsis thaliana* (AtCESA4, AtCESA7 and AtCESA8) and Norway spruce (*Picea abies*), but also showed that aspen (*Populus tremula*) had a 2:1:3 ratio in developing xylem and a 3:1:8 ratio in aspen tension wood (X. Zhang et al., 2018). Thus, there is still a great deal of uncertainty about the stoichiometry of CESAs in a CSC and it is unclear whether CSCs tolerate variation in the number of CESA isoforms, to what degree and the mechanisms that determine this variation (Yu, 2018). However, none of these measurements represents the ratio of CESA isoforms found in the plasma membrane-embedded CSCs but rather the whole cell extract, which includes CESAs from the Golgi apparatus (Haigler et al., 1986) and other CESA trafficking vesicles, such as MASCs/SmaCCs. These microtubule-associated MASCs/SmaCCs also interact with actin to alter cytosolic distribution during delivery to the plasma membrane and exocytosis of CSCs for degradation and turnover (W. Zhang et al., 2019). Thus, if not all CESA proteins are incorporated into functional CSCs in the plasma membrane then these ratios could be misleading.

The result of swapping protein domains between the three SCW isoforms in *Arabidopsis thaliana* (AtCESA4/7/8) or creating targeted catalytic base mutations of these isoforms, using complementation of their respective null backgrounds, indicates class structure is a complex feature of CESAs (Kumar et al., 2017, 2018). AtCESA7 is implicated to be a more critical isoform, as when AtCESA7 donates domains to AtCESA4 or AtCESA8, the cellulose deficiency is better restored by genetic complementation, which is greater for AtCESA4 followed by AtCESA8 (Kumar et al., 2017). This is consistent with similar experiments for larger domains in the swaps (Hill et al., 2018b). The authors concluded that this is possibly due to the position of AtCESA7 within a CSC particle and/or within the CSC unit itself. However, in mutants with specific catalytic residues knocked out, results show that Atcesa4 mutants complement better than Atcesa7 mutants and Atcesa8 mutants complement the worst, indicating that AtCESA4 is the least active protein in synthesis, AtCESA8 is the most active and AtCESA7 is intermediate (Kumar et al., 2018). In these experiments, incomplete knockdown of catalytic activity is possible but is difficult to fully explain. This complexity might reflect a complex organization or structural packing that requires *in vivo* heterooligomer function, compared to *in vitro* homooligomers (Cho et al., 2015, 2017; Purushotham et al., 2016). Alternatively, this complexity could be manifested in the symmetry of

the *in vivo* CSC through many possible interactions. Pseudo-symmetrical and asymmetrical interactions are reflected through important binding surfaces that can be hypothesized by modeling the interactions of isoforms within the CSC. Differential clashes between isoforms in CSC packing could also contribute to the selection of specific isoforms. In this way the critical isoform, like AtCESA7 (Kumar et al., 2018), might be regarded as the isoform capable of forming all of the contacts necessary to assemble the natural CSC, while the other less critical isoforms cannot. Alternatively, the combination of isoforms might be needed to orient domains in a way that allows them to interact with plant plasma membrane and cell wall associated components that *in vitro* systems do not require or possess. One final consideration is that the cellulose microfibril bundling which has been observed in cell walls could be a result of direct or indirect interactions between CSCs (W. Chen et al., 2015; Y. Zhang et al., 2015; T. Zhang et al., 2016; D. Ye et al., 2018). If this is critical to natural cellulose and cell wall synthesis then perhaps isoform class specificity plays a role in the interaction between CSCs, perhaps even forming transient CSC super-complexes.

1.7 Conclusion

Cellulose microfibril synthesis in plants is complex, requiring three isoforms and undefined protein-protein interactions between CESAs and CESA-interacting proteins to form the *in vivo* CSC. With the lack of an atomic model for the CSC, CESA or CatD structure, homology models between CESAs and BCSA have provided insight into the core structure and likely similarity of the catalytic mechanisms. Significant differences exist in the sequences and domain structures of CESA compared to BCSA, perhaps related to the distinct formation of CSCs and cellulose microfibrils. Domain structure differs through the loss of the PilZ activator domain, the gain of an extended N-terminus that includes a ZnF, insertion of the P-CR and CSR into the CatD catalytic core and altered TMH topology. The ZnF forms homo- and heterodimers, while the CSR has been implicated in forming homodimers. Both are redox-dependent and thought to be important in CSC formation and function. Little is known about the function of the P-CR. Neither the CSR nor P-CR structures are known independently or within the CatD, and neither have significant homology to domains with known structure or function. This has limited structural modeling to *ab initio* techniques combined with alignments to CESA homology models of the CatD catalytic core.

To address these knowledge gaps, the first hypothesis that I tested was whether *de novo* structures of the CESA domains could be solved using structural biology techniques, such as X-ray crystallography and SAXS. Recombinant OsCESA8 P-CR monomers were independently folded in solution and crystallized. The P-CR X-ray crystal structure, its fit to SAXS models and atomic models of the CatD catalytic core, allowed me to generate new hypotheses about the function(s) of the P-CR. From the best fit of the P-CR to the CatD SAXS density it is possible that the P-CR interacts with the CatD catalytic core and potentially interacts with substrate entry/turnover. Also observed were homodimer and trimeric crystal contacts, which could reflect current hypotheses of P-CR interactions in the CSC, including interactions with CESA-interacting protein(s), or could represent crystallization artefacts (Chapter 2).

Combining hypotheses for the function of the P-CR with current hypotheses on the function of CESA specific domains and interactions in the CSC from the literature, generates several new and previously implied models. These are based on the organization of the CSC as CESA tetramers or dimers forming a 24-chain microfibril, or CESA trimers forming an 18-chain microfibril with six-fold symmetry, and known interactions, but unknown binding sites, with CESA-interacting proteins (Chapter 3).

The lack of sequence homology between the P-CR and all known structures in the PDB indicates that the P-CR sequence is not well represented. Hypothetically, if similar structures, as opposed to sequences, exist in the PDB it might be possible to predict important surfaces in the P-CR and potential function(s) of the P-CR. From protein-protein structure comparisons of the P-CR structure to other structures in the PDB only the voltage-gated ion channel (VGIC) tetrameric pore-forming domain structures showed consistent alignment to the P-CR. This result, combined with surface structure analysis, has generated new hypotheses for how the P-CR hypothetically interacts as a dimer, trimer, tetramer or monomer docked with the catalytic core near the substrate entry portal, as it related to the CSC and CESA-interacting proteins (Chapter 4).

From these analyses, potentially important features in the OsCESA8 P-CR structure were identified in the *Arabidopsis thaliana* PCW AtCESA1 isoform for *in vivo* functional analysis. Through point mutations, nine *Atcesal* mutant constructs were designed to disturb CESA structural features and potentially cellulose synthesis, with their importance being judged on their ability to complement the temperature-sensitive, cellulose-deficient phenotype in the *rsw1-1* mutant background. Results from these experiments were then used to assess the importance in

AtCESA1, identifying those that have an apparent functional role and those that were unexpectedly unimportant (Chapter 5).

CHAPTER 2. RICE CELLULOSE SYNTHASEA8 PLANT- CONSERVED REGION IS A COILED-COIL AT THE CATALYTIC CORE ENTRANCE

This chapter has been published and is excerpted here mostly in the original form.

Rushton, P. S., Olek, A. T., Makowski, L., Badger, J., Steussy, C. N., Carpita, N. C., & Stauffacher, C. V. (2017). Rice Cellulose SynthaseA8 Plant-Conserved Region Is a Coiled-Coil at the Catalytic Core Entrance. *Plant Physiol*, 173(1), 482-494.

Author Contributions: N.C.C. and C.V.S. designed the experiments. P.S.R. cloned the P-CR, purified protein for crystallization and CD experiments, and solved the P-CR X-ray crystal structure. A.T.O. cloned and purified P-CR protein for SAXS experiments. L.M. conducted SAXS experiments and data analysis. J.B. also analyzed SAXS data and fit P-CR and current CESA/BCSA models to SAXS molecular envelopes. C.N.S. assisted in solving the P-CR X-ray crystal structure. P.S.R., N.C.C., C.V.S, L.M. and J.B. collaborated in writing the manuscript.

2.1 Introduction

The P-CR has not been definitively shown to have a distinct function in CSC assembly or in either (1→4)-β-D-glucan chain or microfibril synthesis. The catalytic core domain homologous to bacterial sequences is predicted to adopt the same nucleotide-binding fold as seen in the BCSA structure (Olek et al., 2014). However, the P-CR and CSR domains lack significant similarity to any of the known structures in the PDB (Berman et al., 2000) and, consequently, also lack SCOP (Murzin et al., 1995), CATH (Sillitoe et al., 2015) or InterPro (Mitchell et al., 2015) database classification. A general location for the P-CR and CSR domains has been proposed from solution structures of recombinant CatDs of OsCESA8 obtained by small-angle X-ray scattering (SAXS) (Olek et al., 2014). The SAXS-derived molecular envelope showed a two-domain structure, with the extended domain participating in dimerization. The best fit of the CESA8 catalytic core domain homology model to this envelope placed the catalytic core in the central portion of the envelope, with the CSR filling the extended domain and the P-CR capping the catalytic core domain distal to the CSR. Ab initio modeling of a cotton (*Gossypium hirsutum*) CESA1 catalytic domain (Sethaphong et al., 2013; Slabaugh et al., 2014b) produced a structurally conserved catalytic core but offer models for the P-CR and CSR domains inconsistent with the SAXS-derived solution structure (Olek et al., 2014).

Here, we report the structural characterization of an OsCESA8 recombinant P-CR domain using SAXS and X-ray crystallography. The crystal structure is an uncommon variant of a coiled-coil domain formed by anti-parallel α -helices and a large connecting loop, with aromatic residues stacked between the helices in the hydrophobic core. The P-CR crystal structure fits within a cylindrical, elongated molecular envelope for this domain derived by SAXS. In turn, this molecular envelope for the P-CR fits into the previously determined molecular envelope of the entire catalytic domain of the OsCESA8 in a position that is consistent with the volume assigned to the P-CR in previous work (Olek et al., 2014). The combination of the crystallographic structure of the P-CR, the homology model of the CESA8 catalytic core and two independent SAXS molecular envelopes for the CESA8 P-CR and CatD have produced a more detailed topological model for this portion of the CESA8 structure.

2.2 Materials and Methods

2.2.1 Cloning and Heterologous Protein Expression

For SAXS analysis a pBluescript SKII vector (Stratagene) containing the rice (*Oryza sativa*) Cellulose Synthase8 cDNA, *OsCESA8* (LOC_Os07g10770, formerly called OsCESA13), was used as a template to clone the P-CR, A₃₉₉ to V₅₂₃. The 375 bp P-CR DNA insert was amplified by PCR using a forward primer (5'- GCGCCATGACTAGTGACTCGGCAATGCTGACTT-3') and reverse primer (5'- GTGGTGCTCGAGTTACTAGACCAGGCGGGGAA-3') ordered from Integrated DNA Technologies, Inc. The P-CR was ligated into a modified pETM-11 (EMBL) vector, containing an N-terminal 6x His-tag, using *SpeI* and *XhoI* restriction sites. The empty pETM-11 plasmid was obtained from EMBL (<http://www.embl-heidelberg.de>). The *OsCESA8* P-CR pETM-11 plasmid construct was transformed into competent *Escherichia coli* [Rosetta2 (DE3); Novagen] cells using heat shock. Positive transformants were selected with 50 $\mu\text{g mL}^{-1}$ kanamycin and 25 $\mu\text{g mL}^{-1}$ chloramphenicol and then used to inoculate 5 mL Luria-Bertani media (LB) for overnight growth. The 5 mL overnight culture was used to inoculate a 500 mL LB culture that was grown to 0.6 to 0.8 OD₆₀₀ then cooled to 16°C before recombinant protein induction with 0.2 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) for 13 h.

For circular dichroism and crystallization experiments a higher yield construct was desired, so a vector that includes a cleavable maltose binding protein (MBP) and His tag was used to

produce a recombinant OsCESA8 P-CR domain. The rice (*Oryza sativa*) Cellulose Synthase8 full-length cDNA sequence template was synthesized with *Escherichia coli* optimized codons and cloned into the pUC-57 vector using the *SmaI* restriction site (Bio Basic Inc.). The *OsCESA8* P-CR, Ala399 to Val523, was amplified by PCR using a forward primer (5'-TACTTCCAATCCAATGCCGCTATGCTGACCTTCG-3') and reverse primer (5'-TTATCCACTTCCAATGCTATCATTAGACCAGACGAGGC-3') ordered from Integrated DNA Technologies, Inc. This was cloned into a pEV-L9 vector (E. Harms, R. Kuhn, unpublished), containing an N-terminal 8x His-tag, maltose binding protein fusion and tobacco etch virus (TEV) protease site using 'ligation independent cloning' (Aslanidis et al., 1990). The *OsCESA8* P-CR pEV-L9 plasmid construct was transformed into competent *Escherichia coli* (DE3) cells using heat shock. Positive transformants were selected with 50 $\mu\text{g mL}^{-1}$ kanamycin then used to inoculate 10 mL of Luria-Bertani media (LB) for overnight growth. The 10 mL overnight culture was used to inoculate a 2 L LB culture that was grown to 0.4 to 0.6 OD₆₀₀ then cooled to 20°C before recombinant protein induction with 0.2 mM IPTG for 18 h.

For crystal phasing seleno-L-methionine-substituted (SeMet) (Ramakrishnan et al., 1997) P-CR protein was obtained using the same *OsCESA8* P-CR pEV-L9 transformed cells and expression conditions with the substitution for LB broth of a media containing M9 salts pH 7.4, 0.4% glucose, 2 mM MgSO₄, 25 $\mu\text{g mL}^{-1}$ FeSO₄, 1 $\mu\text{g mL}^{-1}$ vitamins (Riboflavin, Niacinamide, pyridoxine monohydrochloride, thiamine), 40 $\mu\text{g mL}^{-1}$ basic L-amino acids, except methionine, and 40 $\mu\text{g mL}^{-1}$ seleno-L-methionine.

2.2.2 Protein Purification

For SAXS experiments frozen cell pellets containing the induced *OsCESA8* P-CR pETM-11 construct were suspended in buffer solution containing 50 mM HEPES[HCl], pH 8.0, 500 mM NaCl, 20 mM imidazole, 0.2% Nonidet P-40 detergent and 0.5X cOmplete, Mini, EDTA-free protease inhibitor cocktail (Roche), and then lysed by sonication. Cell debris was removed by centrifugation and the soluble protein supernatant was applied to Ni²⁺ 6- Fast Flow resin microbeads (GE Healthcare) and incubated for 1 h, pelleted and washed with a 50 mM HEPES[HCl], pH 8.0, containing 1 M NaCl and 20 mM imidazole. Protein was eluted from the microbeads by incubating the beads in an elution buffer solution containing 50 mM HEPES[HCl], pH 8.0, containing 0.5 M NaCl and 300 mM imidazole. Eluted protein was injected onto a

Superdex 200, 10/300GL (GE Healthcare) size exclusion chromatography (SEC) column, equilibrated in a 50 mM HEPES[HCl], pH 8.0, containing 0.5 M NaCl, to further purify the sample, producing a monomeric, homogeneous sample for SAXS analysis.

For circular dichroism and crystallization experiments frozen cell pellets containing the induced *OsCESA8* P-CR pEV-L9 construct were suspended in base buffer (50 mM HEPES[HCl], pH 8.0, 0.2 M NaCl) containing 20 mM imidazole, and 1X Halt™ Protease Inhibitor Single-Use Cocktail, EDTA-free (100X) (Thermo Fisher Scientific Inc.), then lysed using a French pressure cell. Cell debris was removed by centrifugation, and the soluble protein supernatant was injected onto a 5 mL HiTrap Chelating HP column (GE Healthcare) equilibrated at 4°C in base buffer containing 20 mM imidazole (binding buffer). The column was washed with binding buffer and the bound protein was eluted with a linear gradient of 20 to 500 mM imidazole in base buffer. Fractions containing the *OsCESA8* P-CR pEV-L9 protein were combined and dialyzed overnight at 4°C in binding buffer to remove excess imidazole, while simultaneously digesting the *OsCESA8* P-CR pEV-L9 protein with TEV protease (1:50 mg TEV protease to mg purified protein). Dialyzed, TEV cleaved protein was re-applied to the HiTrap Chelating HP column (GE Healthcare), equilibrated in binding buffer and the flow through was collected to isolate the untagged P-CR. P-CR samples were further purified by injection onto a HiLoad 16/600 Superdex 200 pg (GE Healthcare) SEC column equilibrated at 4°C in base buffer.

2.2.3 Circular Dichroism Analysis of P-CR Secondary Structure

Pure, TEV digested P-CR was concentrated to 1.5 mg ml⁻¹ in base buffer using Ultracel®-10K Amicon Ultra Centrifugal Filters (Millipore). P-CR samples were loaded into a quartz spectrophotometer cell with a 0.1-mm path length (Starna Cells, Inc.) and spectra collected on a Chirascan™ CD Spectrophotometer (Applied Photophysics LTD) at 22°C. P-CR spectra were collected by scanning between 195 and 260 nm, and the milli-degree (mdeg) output was converted to mean residue ellipticity. The P-CR mean residue ellipticity spectra was used to evaluate the proportion of secondary structure with the K2D deconvolution algorithm (Andrade et al., 1993) on the DichroWeb server (Whitmore et al., 2004).

2.2.4 SAXS Data Collection and Processing

All data were collected at the BioCAT undulator beamline (18ID) at the Advanced Photon Source (APS; Argonne, IL) with a 3.5-m SAXS camera as described previously (Fischetti et al., 2004). Data were collected using a sample cell consisting of a thin-walled quartz capillary held at an ambient temperature of 4°C. To minimize radiation damage, protein samples were made to flow through the X-ray beam at rate that limited X-ray exposure of any one protein to ~100 msec. At these exposure levels, the effect of radiation damage on radio- sensitive test proteins is undetectable. Typically, a data-set consisted of a series of three to five 2-second exposures from protein collected at the peak observed in samples separated by SEC as described above. Estimation of buffer scattering was derived by exposure of the flow through of the column. The two-dimensional scattering patterns were circularly averaged using BioCAT software and the resulting one-dimensional intensity distribution was plotted as a function of spacing $1/d$. Standard deviations of the observed data were calculated, with error propagation formulae used to calculate their effect on the final estimate of scattering from protein. Scattering patterns, including the generation of $\log(I)$ vs q plots, Kratky plots and Guinier plots, were visualized using display programs from the ATSAS software suite (Konarev et al., 2006). The Guinier plot appears approximately linear over the low- q portion of data used in this study with $R_g \sim 22.7 \text{ \AA}$ when estimated using the AutoRg function from the ATSAS data analysis software.

2.2.5 Reconstruction of Molecular Envelopes

The molecular envelope of the P-CR domain was reconstructed using programs from the ATSAS software suite (Konarev et al., 2006). The particle distance distribution function was calculated with GNOM (Svergun, 1992) using a maximum inter-atomic distance of 70 Å and data in the range $0.02 < q < 0.38 \text{ \AA}^{-1}$. Dummy residue models were generated to fit this data using the program GASBOR (Svergun et al., 2001) with the number of dummy residues set equal to the true number of amino acids in the P-CR construct. Forty independent models were aligned for averaging using the SUPCOMB program (Kozin et al., 2001). The forty aligned models were then averaged to provide a three-dimensional reconstruction of the shape of the P-CR with a partial specific volume set to the expected value from the P-CR sequence. Sub-averages of these runs were calculated in order to check convergence and reproducibility of the final average (Supplemental Figure S2).

Results were visualized using the graphics programs MIFit (<https://github.com/mifit/mifit>) and Chimera (Pettersen et al., 2004).

2.2.6 Fitting of Crystal Structures to Solution Scattering Reconstructions

The partial crystal structure of the P-CR was fit to the pseudo-atom grid object representing the P-CR volume using the interactive molecular graphics program, MIFit (<https://github.com/mifit/mifit>). Fitting was guided by the significant difference in the shape at the two ends of the cylindrical reconstruction, where one end is significantly larger and better accommodates the loop that connects the helical regions of the model. To develop a data-driven topology model of CESA, the P-CR SAXS molecular envelope containing the P-CR atomic model was combined with the CatD SAXS molecular envelope and the homology model for the CESA8 conserved catalytic core using MIFit, as described in previous work (Olek et al., 2014). Fitting constraints included matching the shapes of the reconstruction volumes and sequence continuity between the atomic coordinates for the P-CR structure and the homology model of the conserved catalytic core (described in the main text). In this new topological model the catalytic core model was refit to incorporate the explicit and implicit information from the P-CR and was not frozen at the position depicted in Olek et al. (2014).

2.2.7 Crystallization of P-CR

Pure, TEV digested P-CR was concentrated in base buffer to a range of protein concentrations, from 5 to 20 mg mL⁻¹ for initial crystallization screening then 7 mg mL⁻¹ for optimized crystals using Ultracel®-10K Amicon Ultra Centrifugal Filters (Millipore). Crystallization conditions were initially determined with 96-well sitting drop experiments with Classics Suite and JCSG Core Suite crystal screens (Qiagen). Experiments were conducted using a 1:1 protein to reservoir ratio quickly combined, sealed and incubated at 4°C with a duplicate at 22°C. After three days crystals were observed only in drops incubated at 4°C containing either 0.06 M Na₂HPO₄, 1.34 M K₂HPO₄ or 0.96 M trisodium citrate/dihydrate pH 7.0 reservoir solutions. Crystallization conditions were optimized to increase crystal size and diffraction quality resulting in optimal crystal growth in a 24 Cryschem M plate (Hampton Research) sitting drop experiment where 1 µL of P-CR protein was combined with 1 µL of 1.0 M trisodium citrate, dihydrate, pH 8.6 then incubated for 1 week at 4°C. Cryoprotection was optimized for P-CR crystals resulting in crystals being slowly

introduced to buffer containing 1.0 M trisodium citrate, dihydrate, pH 8.6, with incremental increases in glycerol concentration up to 20% (w/v). Cryoprotected crystals were then flash frozen in liquid nitrogen and the diffraction quality was evaluated at the Purdue University Macromolecular Crystallography Facility on a Rigaku-RU200 rotating anode generator with an R-Axis IV++ detector. P-CR crystals with the highest diffraction quality were stored in liquid nitrogen for later data collection at the APS synchrotron source.

To ensure crystals contained intact P-CR protein, crystals were dissolved in 2X SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) buffer and run on 15% SDS-PAGE gel. The single band corresponding to the P-CR molecular weight (14.8 kDa) was excised for each sample and submitted to the Purdue Proteomics facility for mass spectrometry fingerprinting.

2.2.8 X-ray Data Collection, Processing and Three-Dimensional Atomic Modeling

Data were collected on the high diffraction quality, cryo-protected P-CR crystals at the APS (Argonne, IL) on beamline 23-ID-B. Data were collected in 0.5° oscillation steps over 360° with X-rays at 12.659 keV energy (0.98 Å) using a Mar 300 CCD detector with a 250 mm detector distance. Frames were indexed in the F23 space group, then integrated and scaled to a cutoff of $I/\sigma > 2$ using the HKL2000 software (Otwinowski et al., 1997). Single-wavelength anomalous diffraction (SAD) (Dauter et al., 2002) frames of the SeMet derivative P-CR crystals were collected at a peak energy of 12.659 KeV (0.9794 Å). Phenix software (Adams et al., 2010) was used to phase the SAD frames via the AutoSol module (Terwilliger et al., 2012), which placed four SeMet residues in the asymmetric unit and auto fit two P-CR monomer sequences. The two structures were iteratively refined to initial structures containing a chain with 70 of the 125 total residues, with the remaining C-terminal residues unbuilt.

The initial P-CR structure based on SAD phases was iteratively re-built with Coot (Emsley et al., 2004) and refined in Phenix (Adams et al., 2010) to improve geometry and extend the C-terminus. At this resolution the two P-CR monomers in the asymmetric unit appear identical so NCS restraints were applied. The results showed modest improvement over independent refinement of the two molecules, but did not sufficiently improve the models to allow extension of the C-terminus. Attempts were made to resolve the C-terminal electron density, lacking in the initial structure, by first processing and solving the structure in all lower symmetry space groups, none of which provided any improvement. Using the F23 space group data a polyalanine model

was extended by three residues into weaker density in the 2Fo-Fc map (2σ) but was unsuccessful as the subsequent omit maps indicated no significant density improvements. Molecular replacement guided by Rosetta software (Rohl et al., 2004) through the MR-Rosetta module (DiMaio et al., 2011; Terwilliger et al., 2012) in Phenix (Adams et al., 2010) was then tried, using a poly-alanine backbone model of the original 70 residues to reduce phase bias in the molecular replacement. Iterative re-builds and refinement of the MR-Rosetta model in Coot (Emsley et al., 2004) and Phenix (Adams et al., 2010) generated the best map but still lacked complete C-terminal electron density. The improved map allowed for a chain with 77 of the 125 total residues to be built, but the remaining C-terminal residues were not visible and are likely disordered in the crystal. At this point additional density could be assigned to waters and citrate molecules, and a phosphate and a glycerol molecule found coordinated on opposite sides of the three-fold crystallographic axis.

The P-CR crystal structure data and coordinates were deposited in the Protein Data Bank (PDB) (<http://www.rcsb.org/pdb/home/home.do>) under the PDB code: 5JNP.

2.3 Results

2.3.1 P-CR Solution Structure and Molecular Envelope

The *OsCESA8* P-CR domain (residues 399-523) fused with an N-terminal 8X-His tag/MBP protein was cloned and expressed in *E. coli*. Final purification on a size exclusion chromatography (SEC) column after TEV cleavage showed a narrow peak for a homogeneous protein with the predicted molecular weight of the P-CR monomer (14.8kD) (Supplemental Figure S1A). In addition, circular dichroism (CD) spectra (195-260 nm) of this P-CR sample indicated a well-folded protein with primarily α -helical secondary structure in solution (Supplemental Figure S1B).

For SAXS studies, an N-terminal 6X-His tagged P-CR domain was expressed and purified in a similar manner, again resulting in a single band with the monomeric molecular weight appropriate for the longer tag in this vector (18.2 kD). Analysis of solution scattering curves corresponding to P-CR monomers gave a radius of gyration (R_g) of 24.2 Å. The Kratky plot of the scattering data showed characteristic peaks and troughs expected for a compact protein structure with a single conformation in solution (Figure 2.1A,B). The molecular envelope of the P-CR was determined from reconstruction calculations using data in the resolution range between $0.02 < q <$

0.38 Å⁻¹. The final molecular volume (Figure 2.1D) was based on averaging dummy residue models obtained from 40 independent reconstruction runs (Supplemental Figure S2). The averaged P-CR molecular envelope showed an elongated, cylindrical structure with a maximum corner-to-corner distance of 70 Å (Figure 2.1C) corresponding to a length of the reconstructed shape of 60 Å (Figure 2.1D). The envelope differs significantly on each end of the long axis with one end displaying a wider, flattened surface, while the other shows a more compact structure featuring a distinct protrusion perpendicular to the long axis (Figure 2.1D). Combined data from the CD and SAXS experiments indicate that the P-CR possesses an independently folded, compact monomeric structure containing a significant amount of α -helical secondary structure.

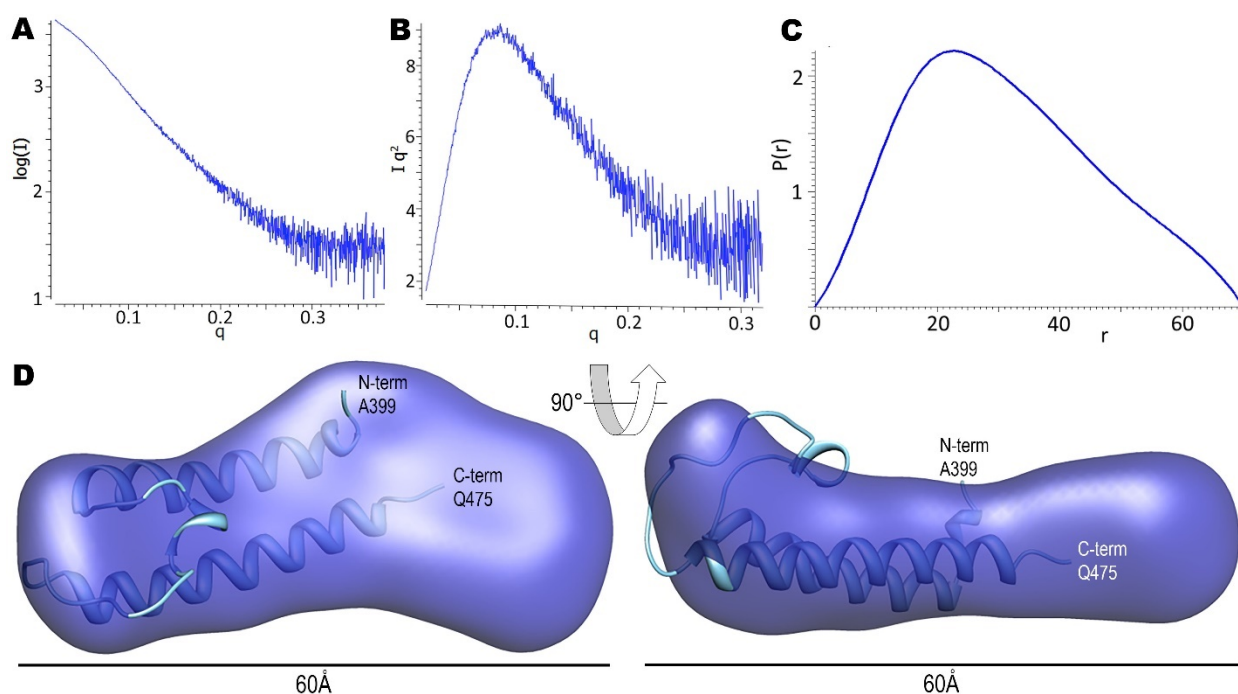


Figure 2.1 Solution X-ray scattering experiments and determination of the P-CR 3D surface contour structure. (A) The log (I) vs q plot of the raw data of the P-CR. (B) Kratky plot of the same P-CR data. (C) The p(r) pair distribution function for intensity data corresponding to P-CR as calculated by GNOM (Svergun, 1992). (D) Three-dimensional molecular envelopes of the P-CR (dark blue) reconstructed from the solution scattering data using the ATSAS software package (Konarev et al., 2006) with GASBOR (Svergun et al., 2001). Molecular surfaces generated from the grid objects that were obtained by averaging multiple reconstruction runs were rendered with Chimera (Pettersen et al., 2004). The P-CR crystal structure (light blue) is shown embedded in its best-fit orientation using MIFit (<https://github.com/mifit/mifit>).

2.3.2 P-CR Crystal Structure and Architecture

The recombinant P-CR protein, purified as described for the CD experiment, was crystallized in 1.0 M trisodium citrate/dihydrate, pH 8.6 at 4°C, producing tetragonal bipyramid shaped P-CR crystals with ~250x100x100 μm dimensions. Initial local data collection indicated that these crystals had an F23 space group with two P-CR molecules per asymmetric unit (Matthews coefficient 2.41, 49% solvent content) and unit cell parameters of $a=b=c=150.8 \text{ \AA}$ (Table 2.1). A survey of lower symmetry space groups confirmed F23 as the optimum choice. Data collected from the SeMet substituted P-CR crystals extended to 2.4 \AA and confirmed the crystal packing (Table 2.1). Of the possible eight seleno-methionine (SeMet) residues (four methionines per monomer), only four were observed in an anomalous scattering difference map. Using these four SeMet positions, the structure was solved using single wavelength anomalous dispersion (SAD) phasing (Dauter et al., 2002). The initial model revealed a two-fold symmetrical dimer in the asymmetric unit, where 70 of the 125 residues could be readily placed in the initial electron density map.

Table 2.1 Data collection and refinement statistics

| | OsCESA8 P-CR (399-523) SeMet (PDB: 5JNP) |
|---|--|
| Data collection | |
| Location | APS 23-ID-B |
| Energy (Å) | 0.9794 |
| Space group | F23 |
| Cell dimensions | |
| $a = b = c$ (Å) | 150.8 |
| $\alpha = \beta = \gamma$ (°) | 90 |
| Resolution (Å) | 45.47-2.40 |
| no. of reflections | 505517 |
| no. of unique reflections | 21575 |
| no. of <i>R</i> _{free} reflections | 1118 |
| CC _{1/2} | 96.3 (79.7) |
| <i>I</i> /σ | 101.4 (2.4) |
| Completeness (%) | 99.9 (100) |
| Redundancy | 45.2 (45.5) |
| Refinement | |
| total no. of atoms (avg <i>B</i>) | 1366 (86.03) |
| no. of protein atoms (avg <i>B</i>) | 1310 (85.06) |
| no. of waters (avg <i>B</i>) | 8 (85.52) |
| no. of ligand/ion atoms (avg <i>B</i>) | 48 (113.17) |
| <i>R</i> _{work} / <i>R</i> _{free} | 0.240/0.272 |
| rmsd for bonds (Å) | 0.009 |
| rmsd for angles (deg) | 1.24 |
| Molprobit overall score | 1.16 (100th percentile*) |
| Rotamer Outliers | 0% (0) |
| Ramachandran | |
| preferred | 100% (150) |
| generous | 0% (0) |
| disallowed | 0% (0) |
| * 100 th percentile is the best among structures of comparable resolution; 0 th percentile is the worst | |

Given that a significant number of residues in the C-terminal portion of the P-CR domain were missing in the electron density map, mass spectrometry experiments were done on thoroughly washed crystals to determine whether a cleaved P-CR molecule had crystallized. Coverage of the P-CR sequence by detected peptides was 95% and showed the full-length molecule was in the crystal (Supplemental Figure S3). SeMet-substituted P-CR crystals were also tested and showed

the appropriate shift for the selenium/sulfur substitution for all methionines detected.

Further refinement of the SAD-phased P-CR structure was done with Rosetta-guided phase improvement (DiMaio et al., 2011; Terwilliger et al., 2012) using a poly-alanine backbone model of the original 70 residues to reduce phase bias in the molecular replacement. This procedure generated an improved electron density map in which the structure could be extended to 77 residues with iterative cycles of building and refinement. Several ligands and water molecules were then placed and the model refined to give a final structure for residues 399-475 of the P-CR, with C-terminal residues 476-523 disordered in the crystal. The structure was evaluated using Molprobity (V. Chen et al., 2010) and deposited in the Protein Data Bank with the PDB code 5JNP (Table 2.1; Figure 2.2A).

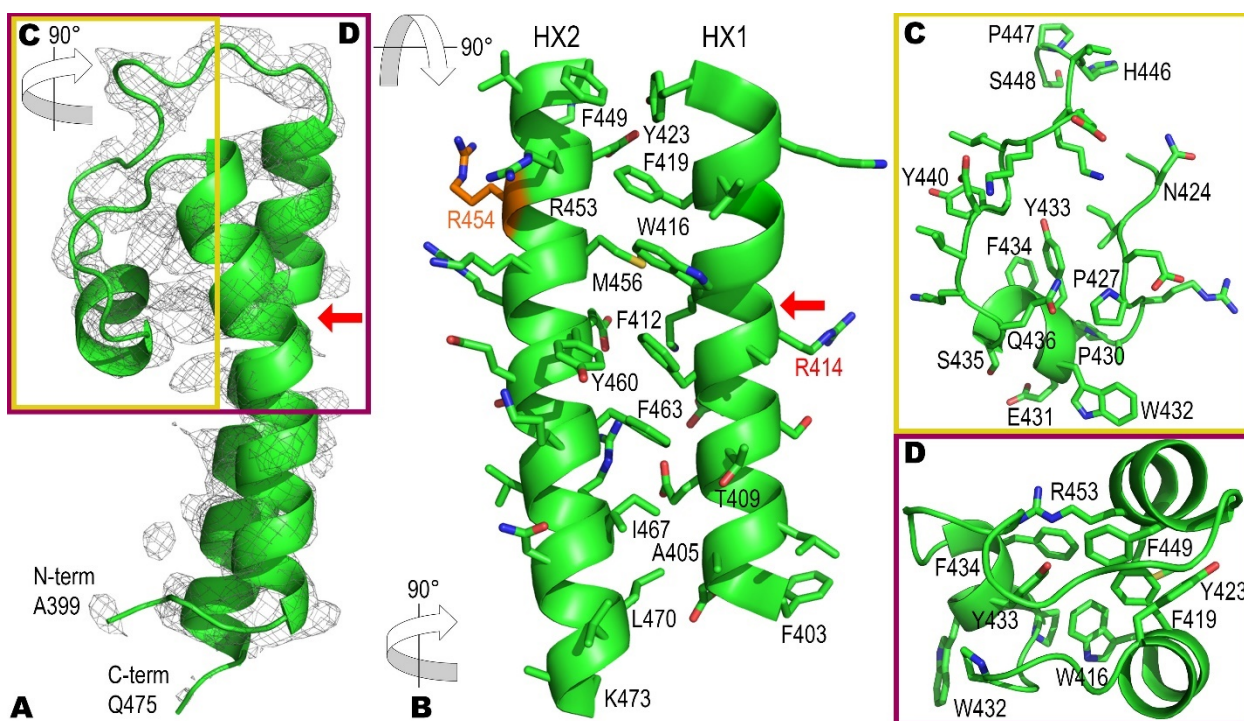


Figure 2.2 X-Ray crystal structure of the OsCESA8 P-CR (Ala399-Gln475). (A) Ribbon model of the P-CR crystal structure (green) encased in an electron density omit map (2σ) (gray mesh). Colored boxes and rotation arrows depict areas inset in C (yellow) and D (purple), respectively. The red arrow (A, B) points to the bend in HX1 between Arg414 and Pro418. (B) Details of the interactions between HX1 and HX2, showing R-groups and highlighting ‘knob-in-hole’ interactions. In this orientation, Arg414 and the stacked Arg454 and Arg458 are exterior facing functionalities; Arg454 (orange) is mutated to Lys in the cellulose-deficient *fra6* mutant found in the SCW *Atcesa8* isoform (Zhong et al., 2003). (C) Structure of the P-CR LP region, highlighted in the yellow box in A, rotated 90° relative to A to highlight loop structures. (D) Structure of the upper region of the P-CR hydrophobic core, highlighted in the purple box in A and rotated 90° relative to A to highlight the hydrophobic core. Aromatic residues from HX1, HX2 and the LP are tightly packed, forcing Trp432 to become solvent exposed.

As predicted by CD spectral analysis the N-terminal region of the P-CR structure is primarily α -helical with a few random coil loops. The structure is dominated by two long anti-parallel α -helices that are joined by a 25-residue connector loop that bends back towards the α -helices, forming several hydrophobic contacts (Figure 2.2A). The first helix spans ~ 31 Å between residues 403-423 (HX1) (Figure 2.2B), with a bend in HX1 occurring where backbone helix interactions do not occur between Arg414 and Pro418 (Figure 2.2A,B). The second large helix (HX2) runs anti-parallel to the HX1 spanning ~ 35 Å between residues 449-473 (Figure 2.2B). The hydrophobic core that stabilizes the structure of the P-CR coiled-coil is unusual for its seven

stacked aromatic residues, with only three aliphatic residues, a Thr and a Met occupying ‘knob-in-hole’ positions (Figure 2.2B and 2.3). These aromatic residues force the helices slightly farther apart, relative to standard aliphatic coiled-coils, to accommodate a larger hydrophobic core. The bend in HX1 between Arg₄₁₄ and Pro₄₁₈ also facilitates inter-helical contacts between these larger aromatic residues.

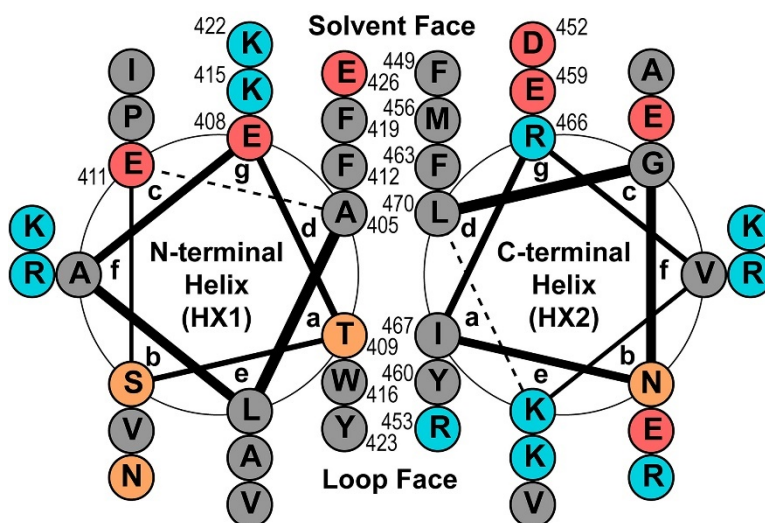


Figure 2.3 Helical wheel diagram of the coiled-coil interactions between HX1 and HX2. In an anti-parallel coiled-coil the ‘a’ and ‘d’ position of each helix intercalate to form the hydrophobic core. Compared to most coiled-coils the P-CR has an uncommonly high proportion of aromatic residues in the ‘a’ and ‘d’ positions, which are typically occupied by aliphatic residues. The ‘e’ and ‘g’ positions typically form salt bridges between helices, but in the P-CR only the solvent-exposed face appears to form these (Supplemental Figure S4). Figure illustrated with DrawCoil 1.0 (Grigoryan et al., 2008). Encircled residue abbreviations are filled with color based on R-group properties: hydrophobic (grey), polar neutral (orange), charged acidic (red) and charged basic (blue).

The 25-amino acid (424-448) loop region (LP), connecting the HX1 and HX2 helices is well ordered in the crystal by contacts it makes with the HX1 and HX2 (Figure 2.2A,C). A small α -helix within LP between residues 431-435 contains a triplet of conserved hydrophobic residues, Trp₄₃₂, Tyr₄₃₃, and Phe₄₃₄. The Tyr₄₃₃ and Phe₄₃₄ face inward toward the α -helices and the aromatic stack of the hydrophobic core between HX1 and HX2 (Figure 2.2D). The third conserved residue, Trp₄₃₂, is fully solvent exposed and shows only weak electron density (Figure 2.2C,D).

Although ‘knob-in-hole’ interactions form the basis of coiled-coil structures, electrostatic interactions can also determine the orientation of helices within the coiled-coil or restrict

interactions to specific binding partners (Grigoryan et al., 2008). The anti-parallel helices of the P-CR include three regions with potential electrostatic interactions along the length of the coiled-coil, although most of these interactions are ~ 4 Å and longer than a typical H-bonding distance (Supplemental Figure S4). All of these amino acids are located on the solvent face of the coiled-coil and opposite the side that interacts with LP.

The DALI (Holm et al., 2010), 3D-surfer (La et al., 2009) and BioXGEM 3D-Blast (Tung et al., 2007) algorithms all showed consistent alignment of the P-CR with other coiled-coil proteins. The closest matches were the pore-forming domains of several voltage-gated sodium channel structures (Catterall, 2014). These structures also have numerous aromatic residues stacked between the helices and a large connector loop that occupies approximately the same position as does the P-CR. However, because the biological role of helices lining sodium channels appears distinct from the possible roles of the helices in the P-CR, this similarity is not likely to extend beyond the structural level.

2.3.3 Comparison of the P-CR Crystallographic Structure with the SAXS Data

The partial X-ray crystallographic structure of the P-CR was fit into the SAXS molecular envelope of the protein in solution. The best-fit orientation of the coiled-coil of the P-CR crystal structure into the SAXS molecular envelope left a substantial density that would contain the 48 unobserved C-terminal residues and any flexible region of the N-terminal tags (Figure 2.1D). SAXS-based reconstruction methodology involves the imposition of relatively flat protein density that is reliable at low resolution (~ 20 Å). In contrast to atomic X-ray crystallographic data, the reconstructed molecular shapes are relatively insensitive to atomic movements on a scale that would render atoms invisible in the 2.4 Å resolution electron density map. For example, a SAXS-based envelope should resolve atomic groups fluctuating with displacements of ~ 5 Å, but atoms associated with structural variations of this magnitude would be invisible in a crystallographic map. As noted earlier, the Kratky plot (Figure 2.1B) shows a relatively large maxima at $q \sim 0.08$ Å⁻¹ and small rise at higher q , suggesting that the P-CR structure was well folded under the conditions used for the solution scattering experiments. Furthermore, calculations of scattering patterns with CRY SOL (Svergun et al., 1995) from the crystal structures of both the monomer and the pair of molecules in the crystallographic unit gave poor fits to the scattering data, and the resulting values of R_g were much smaller than those indicated by the direct analysis of SAXS data ($R_g = 24.2$ Å according to

data analysis with GNOM (Svergun, 1992) but $\chi = 29.1$, $R_g = 14.6 \text{ \AA}$ and $\chi = 32.3$, $R_g = 18.3 \text{ \AA}$ for the crystal monomer and dimer models respectively). Because the crystal structure contains only two thirds of the P-CR sequence, lack of agreement would be expected as the SAXS scattering data corresponds to the complete structure rather than the partial model. In addition, a disorder propensity calculation for the P-CR sequence using the GLOBPLOT2 server (Linding et al., 2003) (<http://globplot.embl.de>) predicted that the N-terminal sequence is more ordered than the C-terminal sequence and shows a sharp disorder maximum at the point in the sequence where the electron density in the crystallographic map disappeared. This predicted flexibility suggests that the C-terminal domain might occupy multiple conformational states within the confines of the crystal lattice, rendering it invisible in the atomic resolution electron density map. Therefore, the isolated P-CR molecule appears to contain two loosely linked structural domains with a C-terminal domain that is intrinsically less well ordered than the N-terminal domain.

2.3.4 Molecular Docking of the P-CR with the Catalytic Core

By combining the crystallographic structure of the P-CR sequence with the SAXS-derived molecular envelope of the entire P-CR volume, a homology-based chain trace of the catalytic core, and our previous reconstruction of the CESA8 CatD molecular envelope (Olek et al., 2014), we developed a detailed topological model of the CESA8 structure. The P-CR model was aligned with the SAXS-based P-CR reconstruction, representing the partial specific volume of the entire P-CR domain. The most plausible fit placed the P-CR loop region into the end of the volume with the distinct protrusion perpendicular to the main axis of the envelope, which then encompassed the helical coiled-coil (Figure 2.1D). Unoccupied envelope regions presumably include the C-terminal residues that are disordered in the crystal (Lys₄₇₆-Val₅₂₃). Using this best fit and the crystal structure of the P-CR as a guide, the molecular SAXS envelope was placed into the F23 crystal structure, resulting in good packing with no significant molecular overlaps (Supplemental Figure S5).

To place the P-CR in the context of the entire catalytic domain, the crystallographic P-CR structure within its SAXS molecular envelope was fused with the CESA8 catalytic core homology model (lacking the P-CR and CSR), and both objects were fit to the SAXS-derived molecular envelope of CESA8 CatD obtained in previous work. In this earlier interpretation, only three plausible orientations of the CESA8 catalytic core homology model were identified, all of which

occupied the central region of the CESA8 CatD molecular envelope as this is the only portion of the envelope wide enough to contain the catalytic core (Olek et al., 2014). These placements leave two disparate volumes on either side of the catalytic core that must be occupied by the P-CR and CSR domains. CESA8 sequence continuity required that the N-terminus of P-CR connect directly to the catalytic core two residues after the first catalytic residue (Asp₃₉₆) that binds the uracil in the UDP-Glc substrate, a relatively fixed point in the structure (Morgan et al., 2013). Constraining the N-terminus of the P-CR model to this end point eliminated most translational degrees of freedom in the placement of the P-CR relative to the catalytic core and determined that the P-CR domain occupied the density as shown in Figure 2.4A. Fitting of the P-CR molecular volume into the CESA8 CatD volume was achieved by matching the length of the P-CR SAXS reconstruction to the length of the larger CESA8 catalytic domain (Figure 2.4B). The unoccupied extended domain of the CatD molecular envelope was then consistent with the position on the catalytic core where the CSR domain would be inserted. In this new topological model, the catalytic core model was not frozen at the position determined in previous work, but was refit to incorporate the new constraints and structural information from the P-CR data.

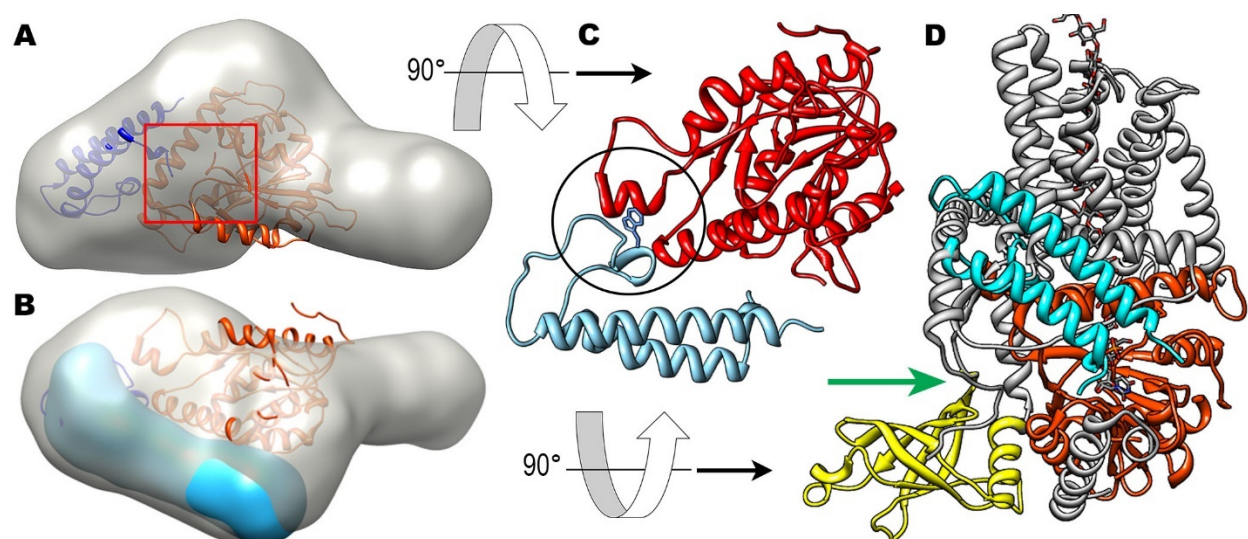


Figure 2.4 Best fit models of the P-CR in the OsCESA8 CatD. (A) The P-CR fit in the three-dimensional molecular envelope of the CatD (gray) (Olek et al., 2014), using MIFit (<https://github.com/mifit/mifit>). The CatD homology model (orange) and P-CR crystal structure (navy blue) are shown in their best fit, which places the N-terminus of the P-CR near the junction with the catalytic core (red square). (B) Rotated view shows the three-dimensional molecular envelope of the P-CR (blue) overlaid in the best-fit orientation. (C) A potential hydrophobic contact with Trp432 (stick) between the catalytic core and the P-CR LP region (black circle), illustrated here with P-CR (light blue) and the OsCESA8 CatD homology model (red). (D) The CatD homology model alignment with BCSA the P-CR coiled-coil (light blue) is overlaid onto the BCSA crystal structure. CESA homologous regions (orange) and non-homologous regions (grey) place the termini near the substrate entry pathway of BCSA (green arrow) (Morgan et al., 2013, 2014). The position of the BCSA PilZ domain (yellow) does not overlap that of the P-CR, which is found on the opposite side of the catalytic core domain.

The modeling of the P-CR domain with the CESA8 catalytic core placed some of the P-CR LP residues in a hydrophobic pocket in the catalytic core, eliminating the solvent exposure of Trp₄₃₂ (Figure 2.4C). As a result, the hydrophilic residues on the opposite face of the coiled-coil were placed in a solvent exposed position on the exterior of the CESA8 catalytic core. Several of the hydrophilic residues of the P-CR near the junctions of HX1 and HX2 with the LP region, the most distal point from the start of the P-CR domain insertion, were near the predicted membrane bilayer based on the position of the transmembrane α -helices in the homologous BCSA structure (Figure 2.4D). However, the best fit did not indicate any hydrophobic membrane or transmembrane associations.

In an independent trial, attempts were made to fit the P-CR structure into the small extended domain of the CESA8 CatD molecular envelope, which we previously predicted to contain the

CSR domain (Olek et al., 2014). Attempts to fit the P-CR structure into this extended domain resulted in severe clashes between the P-CR volume and the model of the catalytic core domain, demonstrating that this volume is too small to accommodate the P-CR (Supplemental Figure S6). Any movement of the catalytic core domain required to allow the P-CR to occupy the small domain, while eliminating the overlaps with the catalytic core, pushes the catalytic core and CSR domains substantially out of the molecular envelope. For this reason, this placement of the P-CR domain appears inconsistent with the SAXS data and confirms the placement in the position shown in Figure 2.4.

2.4 Discussion

2.4.1 The P-CR Domain is a Unique Structure in Plant Cellulose Synthases

Solving the crystal structure of the P-CR and docking its location within the OsCESA8 catalytic domain represents a major step in understanding the assembly of a cellulose synthase complex. The P-CR domain of CESA is a conserved feature of plant cellulose synthases not found in their bacterial homologs. Structural characterization of the P-CR shows this domain comprises an anti-parallel coiled-coil that incorporates an uncommonly high proportion of aromatic residues into a ‘knob-in-hole’ stabilizing structure. This aromatic stack expands the hydrophobic core between helices that forms several contacts with a large connecting loop, a rare feature for coiled-coils. The CESA8 catalytic core is a strongly conserved, novel variant of a classic nucleotide-binding fold found in several types of processive glycosyl transferases (Olek et al., 2014). Modeling based on sequence and structural data shows that the P-CR locks into this catalytic core, aligning the coiled-coil domain along the surface with one end at the entrance to the active site of the catalytic domain and the other near the predicted membrane-spanning domains (Figure 2.4).

2.4.2 Data-Driven Modeling of the OsCESA8 Catalytic Domain

Previous atomic models of the CESA catalytic domain have been based solely on *ab initio* computational modeling of the core, P-CR and CSR domains (e.g. (Sethaphong et al., 2013)). We have derived a three-dimensional model of the CESA8 catalytic domain by combining experimental structural information from SAXS-derived solution structures of the recombinant CESA8-CatD and P-CR, and the crystal structure of the P-CR. Explicit assumptions for this approach were (i) that the CESA8 and BCSA catalytic cores are structurally similar, and (ii) that

the P-CR domain adopts the same fold in isolation as in the full-length structure. Regarding assumption (i), the catalytic core domains of bacterial type II synthases homologous to plant CESAs are remarkably similar in structure based on RMSD and TM-scores (Olek et al., 2014). In addition, key sequence motifs that are critical for substrate binding and catalysis, as well as those that interact with the developing polysaccharide chain are common between BCSA and CESAs. Although the exact shape of the catalytic core of CESA might differ slightly from that of BCSA, the elements integral to synthesizing the polysaccharide chain are unlikely to vary substantially between these structures. While we cannot completely rule out the possibility that the P-CR structure is refolded in the context of the CESA structure, robustness of Kratky plots of the solution structure of the P-CR (Figure 2.1) and the fit of the crystal structure into this component are indicative of a well-folded P-CR structure. It seems unlikely that such an energetically favorable and stable structure would adopt a different conformation in the CESA8 catalytic domain.

Plausible arrangements for the P-CR and catalytic core components within the CESA CatD envelope are restricted by the dimensions of the SAXS-based envelope for the P-CR (Figure 2.1 and 2.4). Our P-CR reconstruction resulted in an elongated molecular shape of the same length as the long dimension of the larger CESA8 CatD structure obtained in our earlier SAXS-based reconstruction (Olek et al., 2014). We showed previously that the bulky catalytic core must occupy a specific orientation within the central region of the CESA envelope (Olek et al., 2014). This placement precludes positioning the P-CR within the smaller domain because the elongated molecular envelope would then penetrate into the space occupied by the catalytic core (Supplemental Figure S6). The only non-occluded space in which the P-CR envelope can be accommodated lies against one edge of the larger domain in the CESA envelope (Figure 2.4). Thus, the shapes and dimensions of the catalytic core and P-CR are sufficient to map out their respective spaces within the larger domain of the CESA CatD envelope. The unfilled space of the smaller domain is then of an appropriate size to contain the CSR. In addition, we demonstrated previously that dimerization is mediated by interaction surfaces on the smaller domain (Olek et al., 2014); we have also demonstrated that the P-CR shows no propensity to dimerize (Supplemental Figure S1A).

An atomic level model for the larger CESA domain is obtained via consideration of the fit of the P-CR crystal structure into the SAXS-derived P-CR molecular envelope. Placement of the P-CR partial crystal structure at the end of the SAXS-based P-CR envelope is dictated by the shape

of the envelope (Figure 2.1). A systematic evaluation determined the most plausible fits of the catalytic core into the CESA envelope with the requirement that the chain trace is almost completely contained within the CESA envelope (Olek et al., 2014). However, only one orientation of the catalytic core and of the PCR envelope/model within the CESA8 envelope allows the required sequence connectivity between these two components (Figure 2.4). Since the point of connection of the P-CR sequence to the CESA sequence lies immediately adjacent to the first key sequence motif (SDDG), we consider this attachment point fixed relative to the catalytic core. In fact, it would be difficult to justify a model that required a shift in the section of the polypeptide incorporating this motif because that change would eliminate a critical interaction with the UDP substrate.

Although the level of detail incorporated into this assembled model is limited by the relatively low resolution of the SAXS-based envelopes and potential differences between the BCSA and CESA catalytic cores, the model incorporates all available information in the most direct and plausible way. The end product is a topological model of the CESA8 catalytic domain that is sufficiently detailed to make testable hypotheses using site-directed mutagenesis to probe potential interaction surfaces.

2.4.3 Potential Functions of the P-CR Domain

Of the two plant-specific sequences in the catalytic domain, only the P-CR has not been previously assigned a possible functional role *in vivo*. Because the bacterial synthase lacks the P-CR and CSR these plant-specific sequences are unlikely to function directly in catalysis. The N-terminal helix of the P-CR coiled-coil connects with the catalytic core two residues after the first nucleotide-binding catalytic motif at Asp₃₉₆ near the substrate entry site, but the coiled-coil extends well away from this catalytic amino acid and is docked into hydrophobic contacts distant from the active site. However, docking of the coiled-coil domain of the P-CR at the entrance to the active site might suggest a role in control of substrate delivery in the absence of the bacterial PilZ domain (Morgan et al., 2013). In BCSA synthase, cyclic di-GMP activation of the PilZ domain interacts with a ‘gating loop’ to uncover the entrance to the active site (Morgan et al., 2014). An FxVTxK motif in the gating loop coordinates the UDP moiety of the nucleotide sugar in the bound state (Morgan et al., 2016). The FxVTxK motif is conserved in all plant CESAs, but traditional prediction of eight transmembrane helices (TMH) would place the motif in a large extracellular loop between

TMH5 and TMH6. In a more recent evaluation of predictions of transmembrane domain structure, TMH5 is proposed to occupy an interfacial position, placing the FxVTxK motif in a cytosolic location where it could function similarly as the bacterial enzyme (Slabaugh et al., 2016). Substitution of the Val and Lys residues of the motif results in severe disruption of cellulose synthesis (Slabaugh et al., 2016). Recent work employing limited proteolysis of tagged CESAs confirms that the FxVTxK motif is located in the cytosol and not extracellular (C. Haigler and coworkers, personal communication). Given this new placement of TMH5, an interaction with the P-CR might mediate substrate delivery.

The P-CR could also play a role in the assembly of the cellulose synthase complex by mediating contacts with catalytic domains of neighboring CESAs in the complex. Recombinant soluble OsCESA8 catalytic domains are thought to dimerize through their CSR domains (Olek et al., 2014). As described earlier, dimers of P-CR are observed in the crystal, but their relative orientation is problematic in the context of the full catalytic domain, which would force the P-CR loop region to clash significantly with the other monomer's catalytic core (Supplemental Figure S7). More recently, recombinant Arabidopsis CESA1 catalytic domains renatured in detergent from inclusion bodies were shown by SAXS modeling to form trimeric structures (Vandavasi et al., 2016). Attempts to fit *ab initio* models based on the cotton (*Gossypium hirsutum*) CESA1 catalytic domain (Sethaphong et al., 2013) using either P-CR or CSR symmetrical contacts exhibited displacements of the CESA model out of the molecular envelope (Vandavasi et al., 2016). However, the SAXS molecular envelope of native recombinant monomeric CESA8 CatDs determined by Olek et al. (2014) was assembled into a trimer with the appearance and dimensions that resemble the trimeric assembly depicted in Vandavasi et al. (2016) (Supplemental Figure S8A). In this interpretation the extended CSR domains fill the corners of the trimer, and the P-CR domains form the contacts between the catalytic domains, consistent with the Vandavasi model but contrary to CSR trimeric contacts proposed by Sethaphong et al. (2016). The P-CR contacts in this model conform to a three-fold contact in the crystal structure formed between pairs of Arg residues (Arg₄₅₄ and Arg₄₅₈) that are mediated by a phosphate ion (Supplemental Figure S8B). However, the trimeric model stabilized by P-CR domains is likely to be possible only when the membrane spanning domains are deleted from the recombinant protein. Expanding the P-CR trimer structure using the model of a single P-CR docked to the CESA8 CatD results in a severe clash with the currently proposed position of the membrane domains (Supplemental Figure S8C).

Thus, while a tight interaction of P-CR domains in isolated catalytic domains explains formation of homomeric trimers *in vitro*, the clashes that would result with the membrane spanning domains in the full-length protein brings into question a biological relevance *in vivo*.

Another possibility for the role of the P-CR is in binding to accessory proteins. The membrane-associated endo (1→4)-β-D-glucanase KORRIGAN (KOR1/GH9A1/*rsw2*) required for proper cellulose synthesis is one such candidate, but little information is known about the function or location of this protein in the CSC (Lei et al., 2014; Mansoori et al., 2014; Vain et al., 2014). Microtubule binding proteins that interact with CESA, cellulose synthase interactive proteins (CSII and CSI3) (S. Li et al., 2012; Lei et al., 2013) are additional candidates, but the location of the P-CR at the substrate entry portal to the active site would seem to make these interactions less likely. A location at the entrance to the active site might also suggest an association with proteins such as sucrose synthase, which has long been proposed to form metabolic channels to deliver UDP-Glc into the active site (Amor et al., 1995).

Arg₄₅₄ participates in an unusual stacked trimeric contact in the crystals, stabilized by a phosphate ion (Figure S8B). This residue is interesting as a non-synonymous *fra6* mutation in the *Atcesa8* isoform converts the Arg₄₅₄ to Lys, resulting in disruption of normal cellulose synthesis (Zhong et al., 2003). However, Arg₄₅₄, which participates in the crystallographic trimer *in vitro*, is oriented toward the predicted hydrophilic exterior of the HX1-HX2 coiled-coil (Figure 2.2B), where interaction with other synthase complex components is also possible. This finding underscores the potential importance of this P-CR residue in synthase function but does not yield specific information on the types of interactions in which it might participate.

2.4.4 Uncertainty Remains About the Number of CESAs per Cellulose Synthase Complex

The number of (1→4)-β-D-glucan chains in a cellulose microfibril, and so the number of CESA catalytic units in a synthase complex, remains undetermined. Classical measurements of microfibril diameters between 3 and 4 nm indicated 36 chains were assembled by the hexameric rosette complexes, indicating a hexamer of CESA hexamers. More recent solid-state nuclear magnetic resonance (ssNMR) measurements of cellulose crystalline cores and surface chains suggest that cellulose microfibril diameters range from 2.3 to 3.0 nm, corresponding to 18 to 24 glucan chains (Fernandes et al., 2011; Newman et al., 2013; Thomas et al., 2013), and potentially a smaller CESA oligomer. Genetic functional analyses of *CESA* mutants established that three

distinct isoforms of CESAs are required for optimal cellulose synthesis and are developmentally regulated to be specific for PCW and SCW synthesis (Taylor et al., 2004, 2008). Quantitative antibody-labeling studies indicated these isoforms are present in 1:1:1 ratios (Gonneau et al., 2014; Hill et al., 2014), prompting these authors to propose that a rosette comprises six specific heterotrimer interactions that accounts for the stoichiometry and for synthesis of an 18-chain microfibril. Consistent with this interpretation, image-enhancement techniques applied to the external portions of the transmembrane domains of CESA rosettes observed in freeze-fracture transmission electron microscopy (FF-TEM) give high proportions of triangular particle counts in these hexameric structures (Nixon et al., 2016). Vandavasi et al. (2016) argued that two homotrimers of each of the three isoforms also preserves the stoichiometry. However, neither model explains how microfibrils larger than 18 chains can be made. On the other hand, there is a wealth of data supporting formation of both homo- and heterodimers *in vivo* (Taylor et al., 2003; Desprez et al., 2007; Timmers et al., 2009) and in even-numbered monomeric units (Atanassov et al., 2009), which provides many options for preserving this isoform stoichiometry in homo- and heterodimer combinations. While the homotrimer formation of the isolated catalytic domains appears to involve coupling of the P-CR, specific models of this interaction are unlikely to represent a biologically relevant interaction given the structural clashes predicted when the catalytic domain is placed in context of the predicted location of the membrane spanning domains (Supplemental Figure S8C). The structure of the P-CR and its predicted location in the catalytic domain has given us new insight into the formation of the cellulose synthase complex, but we need an atomic resolution structure of an entire CESA alone or in an intact rosette, in order to resolve these questions at the core of understanding cellulose synthesis in plants.

2.5 Supplemental Materials

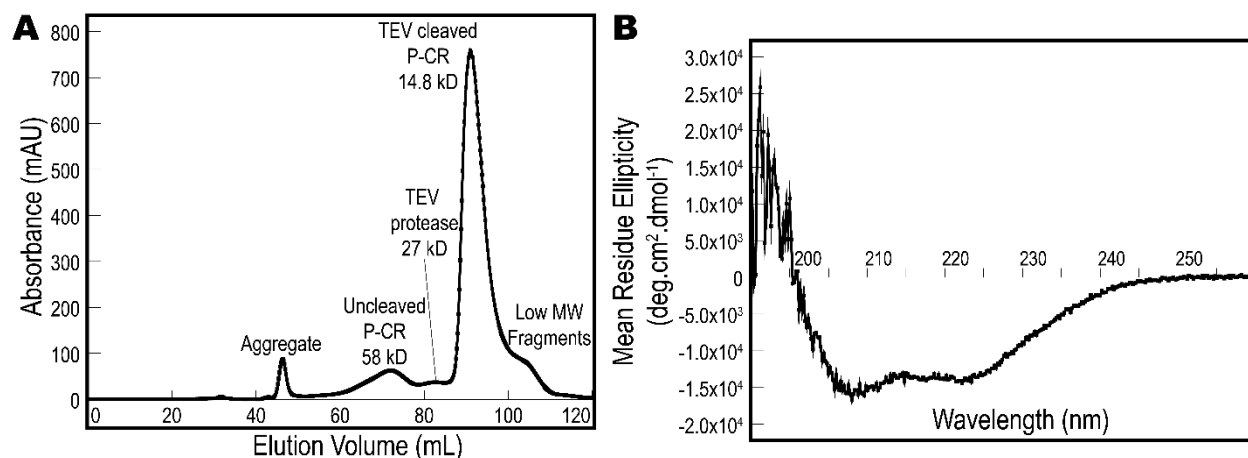


Figure 2.5 Supplemental Figure S1. P-CR secondary structure and apparent molecular weight analysis. (A) The SEC chromatogram of the TEV protease cleaved P-CR showed a single well-formed elution peak corresponding to the TEV cleaved P-CR, along with other small contaminate peaks. (B) The mean residue ellipticity of the P-CR between 195-260 nm is consistent with a primarily α -helical secondary structure, characterized by the strong negative peaks at 208 and 222 nm. Analysis of the spectra shows 59% α -helical, 8% β -strand and 33% random coil secondary structure using the K2D deconvolution algorithm (Andrade et al., 1993) on the DichroWeb server (Whitmore and Wallace, 2004).

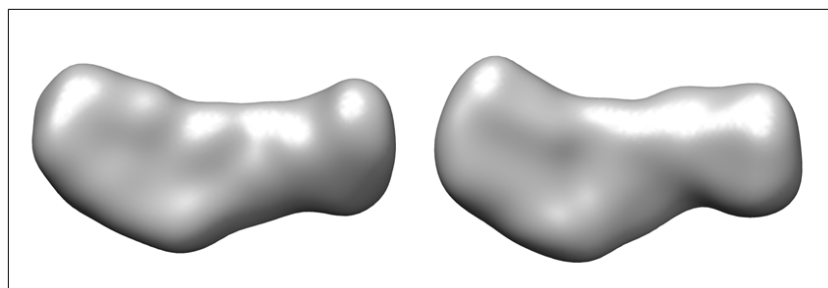


Figure 2.6 Supplemental Figure S2. Reconstructing the P-CR SAXS molecular envelope. The final SAXS reconstruction of the P-CR molecular envelope was based on the average of 40 dummy residue models obtained with GASBOR with a mean NSD=1.35. Two independent sets of 20 models were selected using even and odd serial numbers from the set of 40 models. The reconstructions from these two disparate subsets of models are similar (boxed reconstructions), indicating that the reconstruction based on all 40 bead models (right) is a reliable average.

A OsCesA8 P-CR crystals

3 out of 4 methionines detected – all oxidized

Protein sequence coverage: 95%

01 SNA^{AM}LT^{FD}A₁₁ LAETSEFARK₂₁ WVPFVK^{KY}NI₃₁ EPRAPEWYFS₄₁ QKIDYLKDKV₅₁
HPSFVKDRRA₆₁ MKREYEEFKV₇₁ RINGLVAKAQ₈₁ KVPEEGWIMQ₉₁ DGTPWPGNNT₁₀₁
RDHPGMIQVF₁₁₁ LGHSGGLDTE₁₂₁ GNELPRLV₁₂₈

B OsCesA8 P-CR selenomethionine crystals

3 out of 4 methionines detected – unoxidized, oxidized, and SeMet

Protein sequence coverage: 96%

01 SNA^{AM}LT^{FD}A₁₁ LAETSEFARK₂₁ WVPFVK^{KY}NI₃₁ EPRAPEWYFS₄₁ QKIDYLKDKV₅₁
HPSFVKDRRA₆₁ MKREYEEFKV₇₁ RINGLVAKAQ₈₁ KVPEEGWIMQ₉₁ DGTPWPGNNT₁₀₁
RDHPGMIQVF₁₁₁ LGHSGGLDTE₁₂₁ GNELPRLV₁₂₈

Figure 2.7 Supplemental Figure S3. Sequence coverage for mass spectrometry data of OsCESA8 P-CR crystals. OsCESA8 P-CR crystals washed in reservoir solution were run on SDS-PAGE and the single bands produced were submitted for mass spectrometry proteomic analysis. Matching peptide fragments are shown in red, undetected residues are shown in black. (A) OsCESA8 P-CR crystals showed excellent coverage over the entire OsCESA8 P-CR sequence. 3 of 4 methionines were oxidized (blue). (B) OsCESA8 P-CR selenomethionine crystals also showed excellent coverage over the entire OsCESA8 P-CR sequence but still could not detect the 4th methionine present in the RAMK peptide fragment. In this case the same 3 of 4 methionines as in A were shown to be a mixture of oxidized methionine, unoxidized methionine and selenomethionine substituted (green).

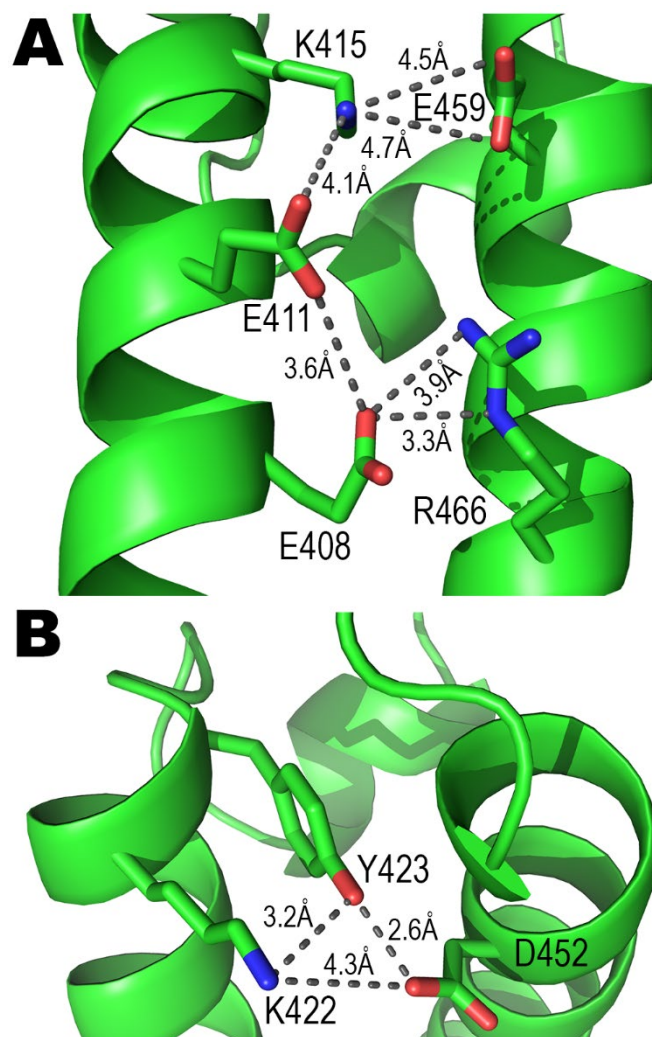


Figure 2.8 Supplemental Figure S4. Electrostatic interactions between coiled-coil α -helices (HX1 and HX2). Several weak hydrogen bonds between helices are found on the solvent face of the P-CR, defined as the opposite side as the interactions between loop region and the coiled-coil hydrophobic core. (A) Near the N- and C- termini of the coiled-coil Glu411 bridges Lys415 and Glu408 on HX1, which then form interactions between the helices with Glu459 and Arg466, respectively. (B) At the opposite end, closer to the loop region, the hydroxyls of Tyr423 and Asp452 bridge with Lys422 in a three center interaction.

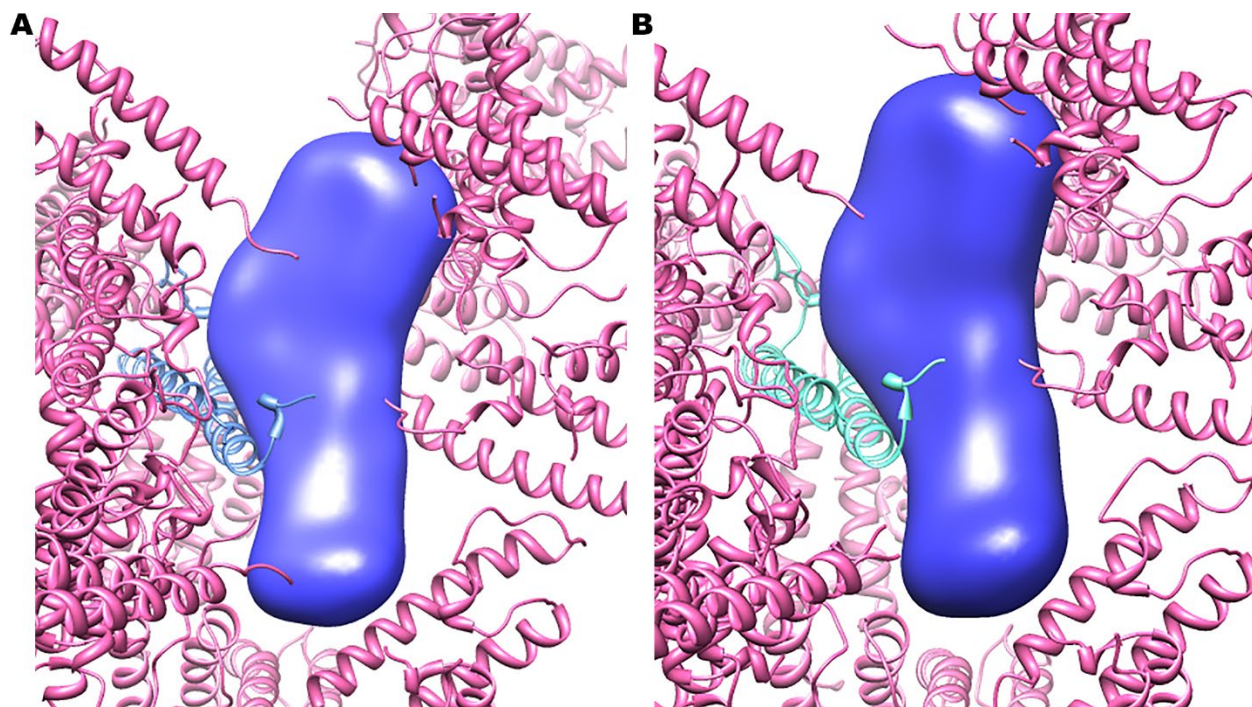


Figure 2.9 Supplemental Figure S5. The P-CR SAXS molecular envelope fit into the P-CR crystal lattice. The SAXS envelope (blue envelope) packs neatly into the crystal cell lattice when overlaid on either of the independent molecules in the crystal cell. (A) Fit over molecule A shown with molecule B (blue C α trace) and molecules related by crystallographic symmetry (pink). (B) Fit over molecule B obtained by applying the NCS symmetry, shown with molecule A (cyan) and molecules related by crystallographic symmetry (pink). Full symmetry expansion of the SAXS envelope in the crystal lattice showed that there are no significant overlaps between these molecular volumes within the resolution of the SAXS data.

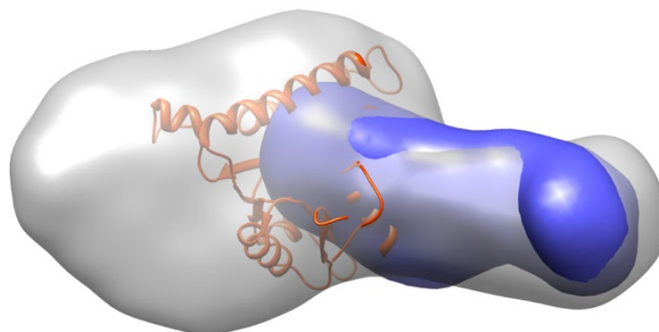


Figure 2.10 Supplemental Figure S6. Alternative fit of the P-CR into the OsCESA8 CatD molecular envelope. Docking trials in which the P-CR volume (blue surface) is placed so as to fill the smaller domain in the OsCESA8 CatD volume (grey surface), previously assigned to the CSR (Olek et al., 2014). This placement results in the P-CR protruding far into the central region occupied by the catalytic core (orange ribbon, showing a representative pose with the model based on the chain trace of BCSA).

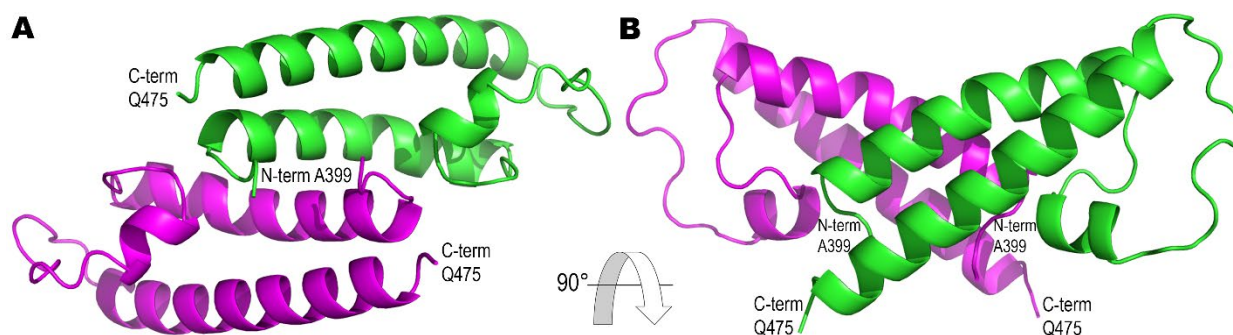


Figure 2.11 Supplemental Figure S7. The noncrystallographic dimer of the P-CR structure. The asymmetric unit of the F23 shows two identical P-CR monomers (green and purple, respectively) interacting across a two-fold symmetrical contact between coiled-coil helix HX1 for each monomer. The LP region and upper portions of the coiled-coil helices, distal to the N and C terminus of each monomer, extend beyond the N-terminal junction of the other P-CR monomer with the 1st signature sequence containing the 1st catalytic Asp residue. In the context of the full-length protein this would cause extensive clashes with the catalytic core, and the structural differences between the CESAs and BCSA catalytic cores required to accommodate this dimer appear unlikely given the high sequence conservation of the catalytic core.

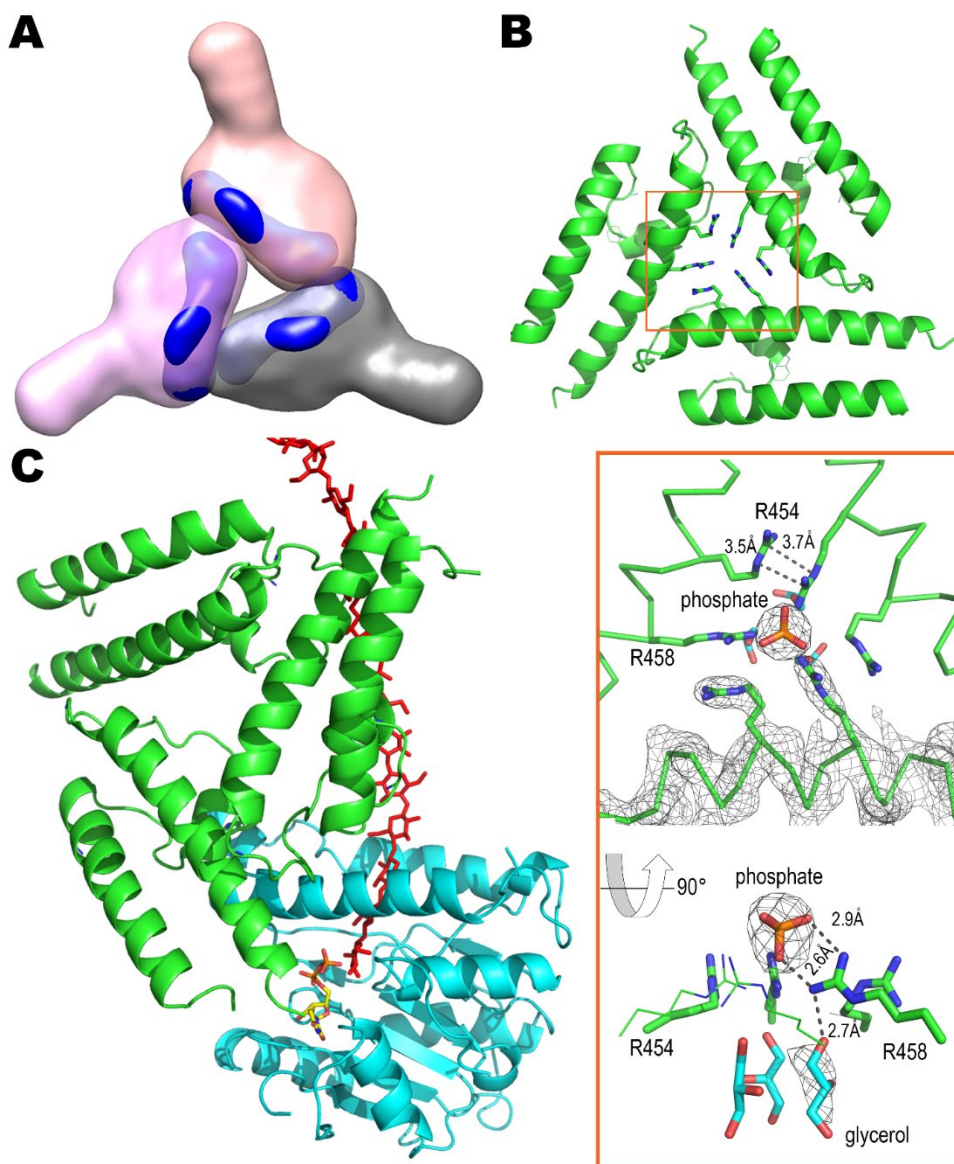


Figure 2.12 Supplemental Figure S8. Constructing a P-CR trimer from SAXS and crystallographic data. (A) The P-CR and subsequent best fit to the OsCESA8 CatD SAXS molecular envelopes (this work, Olek et al., 2014) can be aligned to a three-fold axis to approximate the overall SAXS envelope of the AtCESA1 catalytic domain trimer (Vandavasi et al., 2016). In this configuration the contacts are mediated by the P-CR. (B) Using the best fit alignment of the crystallographic structure of the P-CR to the SAXS generated model, the trimeric contacts overlay those around the three-fold in the crystal. Arg454 and Arg458 participate in an unusual stacked interaction, stabilized by a phosphate ion (insert). (C) Taking the best model of the P-CR and CatD based on our combined SAXS and crystallographic evidence and superimposing this on the BCSA catalytic core (Morgan et al., 2013) gave the model first shown in Figure 5D. Only the catalytic core (blue) and P-CR (green) of OsCESA8 are shown here. Adding the trimeric P-CR interaction to this model (three green P-CR domains) would require the other two P-CRs to protrude into the membrane, indicating this trimeric model is unlikely to be a biologically relevant structure.

CHAPTER 3. HYPOTHETICAL MODELS OF THE CELLULOSE SYNTHASE COMPLEX (CSC)

3.1 Introduction

Given the higher organization of plant CESAs into the CSC, along with various CESA-interacting proteins compared to BCSA, one or more of the CESA plant-specific domains, the ZnF, P-CR or CSR, might facilitate protein-protein contacts in the CSC. *In vitro* cellulose microfibril synthesis is also dependent on the presence of the ZnF (Purushotham et al., 2016). The ZnF has long been considered to form a fundamental dimer in the CSC and, *in vitro*, can form homo- or heterodimers between CESA isoforms (Kurek et al., 2002; Jacob-Wilk et al., 2006). *In planta* techniques such as bimolecular fluorescence complementation and yeast two-hybrid analysis of CESA isoforms have shown that homo- and heterodimerization is possible between most isoforms (Carroll et al., 2012; Miart et al., 2014; Xia et al., 2014). The CSR has been shown to form the dimer interface *in vitro* for recombinant CatD (Olek et al., 2014; Rushton et al., 2017). Questions still remain about its molecular structure, ability to form heterodimers, whether it performs additional functions in CESA or CSC organization and why the class-specific sequence organization evolved. The P-CR has been hypothesized to form dimeric, trimeric and tetrameric CSC contacts *in vivo* (Sethaphong et al., 2013, 2016; Nixon et al., 2016; Vandavasi et al., 2016; Rushton et al., 2017). Because the P-CR sequence is highly conserved, it seems unlikely to have the ability to make any differential interactions and thus, might perform the same function and the same interactions in all CESA isoforms. The P-CR and CSR could also have roles in CESA interactions with CESA-interacting proteins like KOR1 (Lane et al., 2001; Sato et al., 2001; Lei et al., 2014; Mansoori et al., 2014; Vain et al., 2014), the glycoposphatidylinositol-anchored and N-glycosylated “polysaccharide chaperone” protein of the *COBRA* genes (Schindelman et al., 2001; Roudier et al., 2005; L. Liu et al., 2013; Sorek et al., 2014), CSI1/3 (Y. Gu et al., 2010; Bringmann et al., 2012; S. Li et al., 2012; Lei et al., 2013, 2015), CC1/2 (Endler et al., 2015, 2016), PATROL1 (X. Zhu et al., 2018), SHOU4 (Polko et al., 2018), AP2M (Bashline et al., 2013) or some of the TPLATE complex proteins (Bashline et al., 2015; Sánchez-Rodríguez et al., 2018). These CESA-interacting protein contacts could occur specifically with one domain, or be spread across multiple domains of CESA or CESAs in the CSC and include potential indirect interactions. Therefore, it is important to consider

potential binding partners to account for the number and potential location(s) of all of the interactive surfaces within a CESA.

The integration of at least three different isoforms into the CSC might be an important factor in CSC formation and structure. As discussed in Chapter 1.6, domain swap and point mutation experiments indicate isoforms might vary in the degree to which they can complement the role of other isoforms (Kumar et al., 2017, 2018; Hill et al., 2018b). The conclusion was that one isoform is more critical than the others, due to hierarchy in the ability to substitute domains and isoforms in complementation experiments, in this case SCW AtCESA7. Both other isoforms (AtCESA4 and AtCESA8) appear to be more similar in their ability to substitute for one another. They concluded that this might reflect the position of these isoforms in the CSC. This critical isoform might assist in forming CSC symmetries or be capable of different functions compared to the other isoforms. Placing CESA isoforms within CSC models reveals that often one of the positions has interactions or symmetries distinct from the other two positions, which might account for the apparent functional differences between isoforms. Finally, while various models of CESA stoichiometry have been proposed in different species (Taylor et al., 2003; Ding et al., 2006; X. Zhang et al., 2018), for the following models a 1:1:1 ratio is assumed, consistent with results from *Arabidopsis thaliana* (Taylor et al., 2003; Gonneau et al., 2014; Hill et al., 2014; X. Zhang et al., 2018). It is possible that other ratios still reflect the critical isoform necessity but vary the ratio of other isoforms, which could even be a heterogeneous mixture of various isoform ratios within a cell. Given these constraints, a general model for tetrameric, dimeric and trimeric CSC can be generated (Figure 3.1), with which, numerous interactions and class-based iterations can be hypothesized.

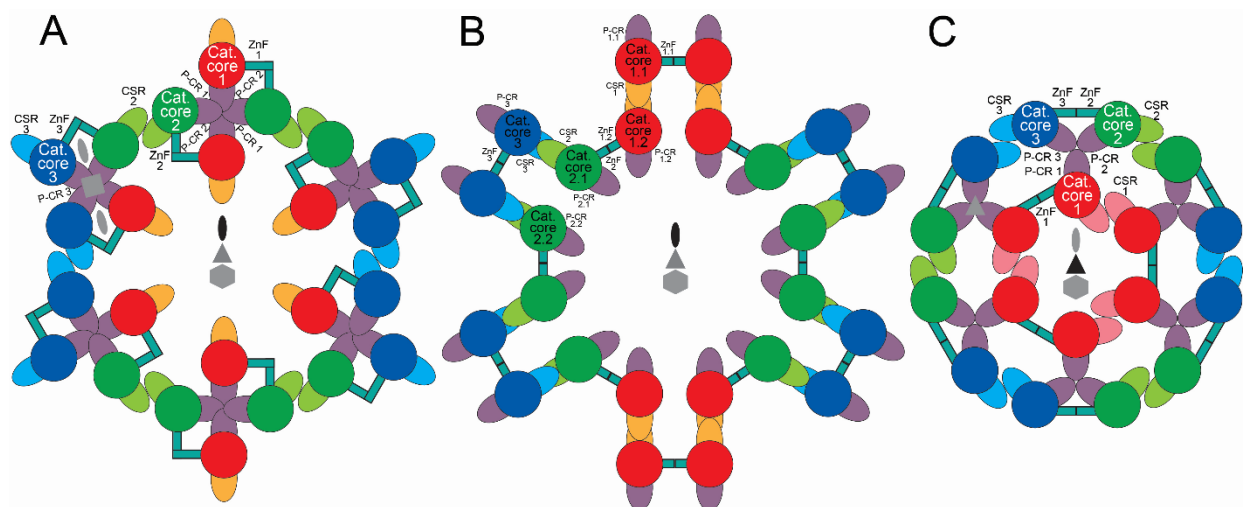


Figure 3.1 General models of (A) tetrameric, (B) dimeric and (C) trimeric P-CR (light purple ellipse) forming CSCs with the dimeric ZnF (narrow teal boxes) and CSR. The circular catalytic core of CESA is colored by its class for the critical isoform class labeled 1 (red) or less critical isoforms labeled 2 and 3 (green and blue). The CSR ellipse is colored light green or blue for the less critical isoforms and differentially for the critical isoform to reflect different models or differential functions within a model (yellow, orange, light pink). Rotational symmetries of the models are labeled near their respective symmetry axes shown as a two-fold (ellipse), three-fold (triangle), four-fold (square) or six-fold (hexagon). Models also vary in their amount of true symmetry (black symbol) and pseudo-symmetry (gray symbol), defined as symmetry that breaks when isoform identity or theoretical CSR orientation is considered.

3.2 Tetrameric P-CR Models of the CSC Particle

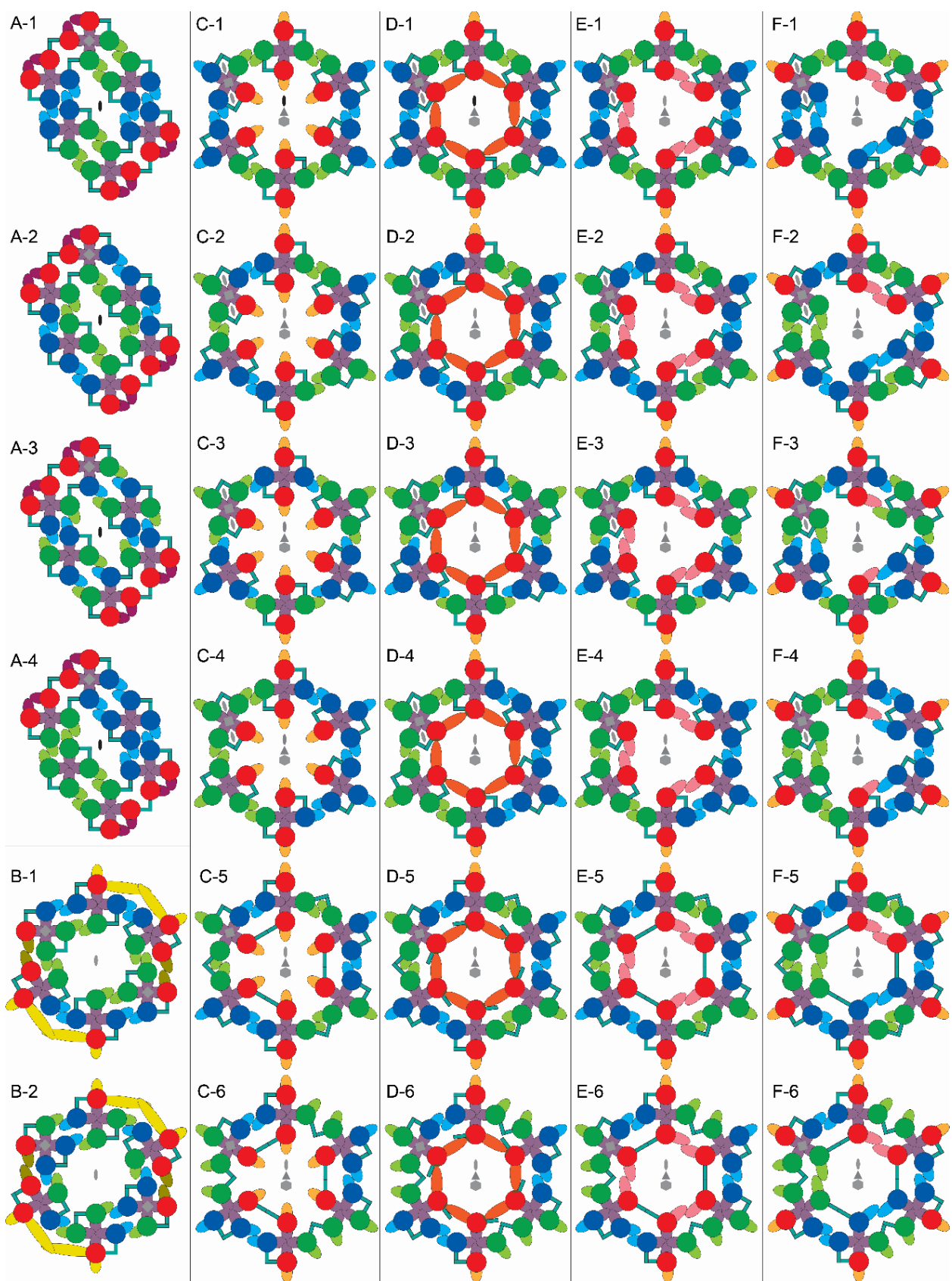
One hypothesis is that the CSC uses tetrameric P-CR interactions to form six-fold symmetrical CSC particles with 24 CESAs (Bowling et al., 2008). The 24-CESA model is derived from cellulose microfibril models that predict a 24-chain microfibril of hexagonal or rectangular shape (Ding et al., 2006; Fernandes et al., 2011; Newman et al., 2013; Thomas et al., 2013; Oehme et al., 2015; T. Wang et al., 2016). The 24-chain microfibril model is consistent with multiple sources of experimental evidence, however, the potential for twists and amorphous chains allows for multiple interpretations. First, isolated cellulose from Sitka spruce (*Picea sitchensis*) mature earlywood (mostly SCW) was measured in small-angle neutron scattering, wide-angle X-ray scattering, ^1H spin-diffusion in ^{13}C 2D ssNMR and surface chain deuteration Fourier-transform infrared spectroscopy experiments (Fernandes et al., 2011). These results indicated a 36-chain model did not the fit the data, an 18-chain model could fit some of the data, but a 24-chain model fit best.

Similar experiments on PCW cellulose microfibrils extracted from celery (*Apium graveolens*) collenchyma were also more consistent with a 24-chain model compared to an 18-chain model (Thomas et al., 2013). Wide-angle X-ray scattering and ssNMR experiments on mung bean (*Vigna radiata*) PCW cellulose microfibrils were compared to simulated diffractograms of different microfibril shapes and sizes, which showed a better fit to an 18-chain model but did not exclude 24-chain models (Newman et al., 2013). MD simulations of microfibrils compared to ssNMR data were also more consistent with the smaller diameter 18- and 24-chain models compared to the larger 36-chain models but could not discern between 18- or 24-chain models (Oehme et al., 2015). Finally, extensive analysis of 2D and 3D ssNMR spectra of PCW cellulose (Oehme et al., 2015; T. Wang et al., 2016) from *Arabidopsis*, *Brachypodium*, *Poa annua* was compared to other published values and the authors argued strongly for a microfibril of at least 24-chains (T. Wang et al., 2016). Even-numbered microfibrils models are also favored by the observation of even-numbered CESA oligomers (Taylor et al., 2003; Desprez et al., 2007; J. Wang et al., 2008; Atanassov et al., 2009; Timmers et al., 2009; Olek et al., 2014). An 18 CESA CSC with apparent six-fold symmetry cannot be built with tetramers. MD modeling and fitting to class-averaged CSC structure models, derived from FF-TEM images, has been done using a 24 CESA CSC model with the P-CR as the tetrameric contact aligned parallel to the plasma membrane (Nixon et al., 2016). Based on the fit to the low-resolution class-averaged structure, the tetrameric model showed a slightly poorer fit than the trimeric CSC particle model but a much better fit than the hexameric, pentameric and dimeric CSC particle models. Thus, an oligomer based on tetrameric P-CR contacts is a reasonable particle model, especially if the P-CR packing was tighter or the model is built using the P-CR long axis aligned perpendicular to the plasma membrane as opposed to parallel; neither orientation has been observed experimentally.

On the other hand, reconciling the 1:1:1 ratio of CESA isoforms in *Arabidopsis thaliana* (Gonneau et al., 2014; Hill et al., 2014; X. Zhang et al., 2018) within a tetrameric particle of six-fold symmetrical CSC is difficult, even assuming the P-CR is capable of making heterooligomeric tetramers between isoforms, as it always creates a 2:1:1:2:1:1 ratio of CESA isoforms within CSC particles unless isoforms are distributed highly asymmetrically. One way to solve this conundrum is to place a single isoform in a unique, often symmetry-breaking or pseudo-symmetrical position, with potentially unique properties relative to other isoforms. This is consistent with the idea that equivalent mutations in different CESA isoforms produce phenotypes with different levels of

severity, despite the CESA6-like isoforms showing genetic redundancy (Desprez et al., 2007; Persson et al., 2007; Kumar et al., 2017, 2018). However, the fact that a single CESA isoform is capable of forming a functional CSC *in vitro* (Cho et al., 2015, 2017; Purushotham et al., 2016) is not consistent with genetic evidence that three isoforms are required. Some isoforms might facilitate protein-protein interactions with other CESA-interacting proteins that are necessary for normal cellulose synthesis but not necessary for *in vitro* synthesis, perhaps through preferential positioning on the periphery of the CSC. There are many reasonable models, with a few illustrated in Figure 3.2, showing how the critical CESA isoform (red) might fill a critical position and facilitate interactions perhaps not available to other isoforms (green and blue).

Figure 3.2 Hypothetical models of the CSC using P-CR tetramers (light purple ellipse) to form the CSC particle. The circular catalytic core of a single CESA is colored by its class for the critical isoform class (red) or less critical isoforms (green and blue). The CSR (ellipse) is colored light green or blue for the less critical isoforms and differentially for the critical isoform to reflect different models or differential functions (light/dark yellow, light/dark orange and pink). Completely filled interactions as shown in A and B lead to either broken six-fold symmetry, using the critical isoform CSRs as interactions within a particle (scarlet in A) or long-distance CSR interactions (yellow in B). This long-distance CSR interaction could also be a CSR that does not interact within the CSC or is in an unbound state (yellow, alternate conformation in B). Introducing a two-fold symmetry in the handedness of the CSR interactions leads to the potential for a true two-fold symmetry in CSCs (C,D,E,F-1) but breaks symmetry of particles forming alternating outside/inside CSR interactions between particles (C,D,E,F-2,3,4,5). CSR dimers of the less critical isoforms can be homodimers (A-1,2,4; B-1; C,D,E,F-1,2,4,5), heterodimers (A,C,D,E,F-3; B-2) or have minimum CSR interactions (C,D,E,F-6) in all of the models. Considering minimum interactions implies these isoforms do not form interactions within the CSC or are unbound. ZnF (narrow teal boxes) dimeric interaction are shown within particles (B; C,D,E,F-1,2,3,4) or both within and between particles (A; C,D,E,F-5,6). For six-fold symmetrical particle models centrally placing the critical isoform could imply a six-fold symmetry of the CSR at the center (C, light orange), a unique asymmetric dimer of the CSR with the catalytic core and/or ZnF (D, dark orange), or a three-fold symmetrical interaction of CSR dimers forming a CSC inner ring (E, pink). Alternatively placing the CSR of the critical isoform on the six points of the CSC implicates the CSR in a critical role outside of core CSC oligomerization (F, light orange). The remaining two critical isoforms are shown making CSR homodimers within the CSC.



The tetrameric particle models shown in Figure 3.2 use the P-CR to generate four-fold pseudo-symmetrical particles with different isoforms and assumes P-CR contacts are non-differentiated between isoforms. The CSR was used to form some of the contacts between particles with the CSR of the critical isoform oriented in a unique position or symmetry in the CSC. The other isoforms could be either homodimeric (Figure 3.2A-1,2,4; B-1; C,D,E,F-1,2,4,5,6), heterodimeric (Figure 3.2A,C,D,E,F-3; B-2), or have the minimum number of interactions within a CSC (Figure 3.2C,D,E,F-6). In these models, the CSR contacts are mostly shown as dimers consistent with their *in vitro* dimerization (Olek et al., 2014). However, all of the variable regions likely facilitate or limit interactions and positions of the isoforms within the CSC. The ZnF is used to form both homo- and heterodimers and is assumed not to differentiate between isoforms, with both inter- and intra-particle interactions. This possibility is due to the cytosolic N-terminus being a flexible domain as implied in other work (PDB ID: 1weo, unpublished) (Kurek et al., 2002; Jacob-Wilk et al., 2006; Purushotham et al., 2016). The different models also vary in the amount of true symmetry (black symbol) and pseudo-symmetry (gray symbol) they possess. Here pseudo-symmetry is defined as symmetry that breaks when isoform identity is considered or a symmetry that breaks when the theoretical CSR orientation and angle are considered. Pseudo-symmetry would not be observed in low resolution structure, such as the CSCs observed from the extracellular surface by FF-TEM (Nixon et al., 2016). In order to create a model where all interactions are filled using either CSR homo- or heterodimers and ZnF dimers, it requires that the CSRs pack in a two-fold but not a six-fold symmetry (Figure 3.2A,B). The heart of the problem is that with 24 isoforms packed with dimers of the ZnF and CSR, and tetramers of the P-CR, an inner and outer ring of interactions are created with two different diameters. Thus, in order to have all of these domains make a symmetrical interaction, it requires either a collapse of the six-fold pseudo-symmetrical CSC if CSR contacts are equal in length throughout the model (Figure 3.2A) or a large variability in CSR or ZnF dimensions within a model (Figure 3.2B) (ZnF models not shown). This is illustrated by introducing two opposing 60° rotations of two particles that could pack to approximate six-fold pseudo-symmetry of the CSC. Here the critical isoform CSR is considered highly flexible and capable of extending to at least two different lengths, a long (yellow) and short (dark yellow) distance to complete the interactions of the outer ring. The CSRs placed at the long-distance interaction could also be unbound or interacting with other proteins outside the CSC (yellow, alternate conformation).

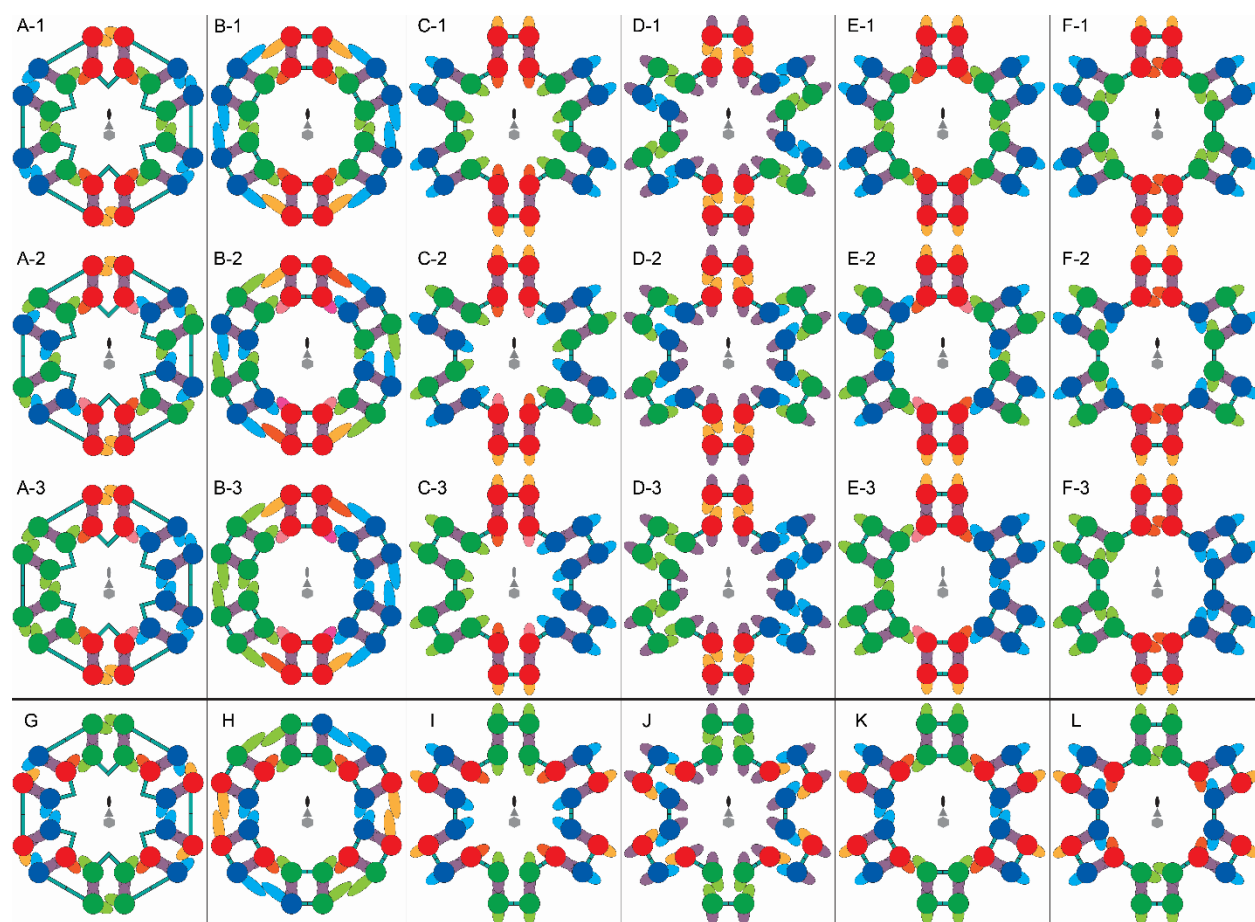
Iterating complete six-fold symmetry or 60° rotations between particles seems more likely and demands either the seemingly distant contact as in Figure 3.2B or the preferred model where the outermost CSRs do not form dimers within the CSC. The first of these models places the critical isoform interior to the CSC to either be consistent with the six-fold symmetry of the CSC, generated by CSR contacts on the periphery, or to form a hexameric interaction along with their ZnFs that assists the CSR interaction forming six-fold symmetry (Figure 3.2C). A second model uses the interior-positioned critical isoform to form an asymmetric, isoform-specific dimeric contact with the neighboring critical CESA isoform using the CSR of one isoform to contact a region not attributed to the P-CR or CSR in the other isoform (Figure 3.2D). Of note is that, in this set of models, the ZnF could be considered as part of the contacts for the internal dimers between critical isoforms. This requires an extra assumption that the ZnF is either interacting with the critical isoform CSR or at least not clashing with it in a stable unbound state. However, there is no evidence for this interaction. The best model uses the critical isoform CSR to form dimers similar to other isoforms, but in a novel contact that breaks the two-fold symmetry of the CSC (Figure 3.2E). This model is preferred because it maintains the ability of the critical isoform to make CSR homodimers. The only caveat is that the critical isoform CSR must be flexible enough to have at least two conformations relative to the other CSR in the dimer at the center of the CSC. In this way, the critical isoform is perhaps critical because it has this CSR flexibility compared to other isoforms. In an alternative model, the critical isoform can be viewed as critical because it fills all of the outermost six-fold symmetrical points of the tetramer, being unbound or interacting with proteins outside of its CSC, while also making dimeric contacts along with the other isoforms on the interior (Figure 3.2F). These are all symmetric models of the CSC built on P-CR tetramers, but it is possible that asymmetric interactions using different orientations, angles and lengths between variable regions and CSRs could assemble a highly pseudo-symmetrical six-fold viewed from the extracellular surface. However, this seems unlikely given the lack of large sequence insertions between isoforms.

3.3 Dimeric P-CR Models of the CSC Particle

A CSC model assuming 24 CESAs and (1→4)-β-D-glucan chains can also be achieved using P-CR, CSR and ZnF dimers to construct the CSC (Figure 3.3). Four CESA are assumed to pack tightly into a particle that interacts with other particles to assemble the CSC. Using the same

assumptions for the CSR and ZnF as used in the tetrameric P-CR model, it is possible to make new models that generate slightly different hypotheses regarding their symmetry, potential interactions and the position of the critical isoform. As no dimeric models for a 24 CESA CSC were generated to fit to the class-averaged FF-TEM CSC images by Nixon et al., (2016), it is not possible to predict accurately the packing needed within and between particles. In this case, building P-CR dimers along the long or short axis of the P-CR coiled-coil relative to the plasma membrane matters but unlike tetrameric models, because these models use CSR and ZnF dimers to complete the particle connectivity, the interaction constraints of these domains would also matter. These constraints make for more complicated models. The models shown in Figure 3.3 use P-CR homodimers, heterodimers or both, even though only a P-CR homodimer has been observed in the P-CR structure crystallographic asymmetric unit (Rushton et al., 2017). However, the junction of the P-CR with the catalytic core as determined by sequence is at a point near the substrate entry portal and catalytic sites of CESA. This puts other constraints on dimeric contacts mediated by the P-CR and would require significant adjustments from the orientation of the P-CR extracted from the SAXS data (Olek et al., 2014; Rushton et al., 2017). Assuming some flexibility exists that allows changes in the P-CR orientation relative to the catalytic core, a different packing or other P-CR dimeric interactions, the models of Figure 3.3 are possible.

Figure 3.3 Hypothetical models of the CSC using P-CR dimers (light purple ellipse) to form the CSC particle. The circular catalytic core of a single CESA is colored by its class for the critical isoform class (red) or less critical isoforms (green and blue). The CSR (ellipse) is colored light green or blue for the less critical isoforms and differentially for the critical isoform to reflect different positions and/or function within a model (light/dark orange and pink). Possible models for completely filled interactions (A,B,G,H), minimally connected (C,D,I,J) and intermediate connectivity (E,F,K,L) are shown. Two fundamental symmetries describe the critical isoforms position as interacting (A,B,C,D,E,F) or not interacting (G,H,I,J,K,L) with itself. CSR dimers of the less critical isoform can be homodimers (A,B,D,E,F-1,3; G,H,I,J,K,L) or heterodimers with each other (A,C,D,E,F-2) in all of the models with CSR dimers. For the first set of models less critical isoforms are arranged in an inner and outer ring segregation (A,B,C,E,F-1), a polarized position within particles (A,B,C,E,F-2) and a polarized position within the CSC (A,B,C,D,E,F-3). Assuming the CSR is the fundamental dimer within particles in a minimally connected model, rather than the P-CR, makes homodimers show a polarized position within CSC particles (D-1) and heterodimers show an inner and outer ring segregation in the CSC (D-2). In the second set of symmetries the critical isoform does not interact with itself only the less critical isoforms, while the less critical isoforms can also form homodimeric CSRs (G-L), heterodimeric CSRs with each other (not shown) or show polarization of isoforms within the CSC (not shown). ZnF (narrow teal boxes) homo- and heterodimeric interactions are shown within and between particles. Completely filled dimeric interactions of both the CSR and ZnF leads to long range interactions, assumed to be the ZnF (A,G) or potentially different variable length CSRs dimers (B,H). Assuming minimal dimeric interactions linking the isoforms, but including the functionally critical and known dimeric interaction of the ZnF, implies the CSR (C,I) or P-CR (D,J) could have another role outside of core CSC oligomerization or be unbound. When intermediate connectivity of the CSR and ZnF is considered, models could have inner isoforms make CSR dimers within the CSC, while peripheral isoforms could have another role outside of core CSC oligomerization or be unbound, with CSR dimers between (E,K) or within (F,L) particles.



Using P-CR, ZnF or CSR dimers to assemble particles within the CSCs six-fold symmetry creates two different dimers, those within the particle and those between particles. These dimers have different angles between the catalytic core, P-CR and CSR, shown here as $\sim 90^\circ$ angle to fulfill the particle symmetry and $\sim 120^\circ$ to fulfill six-fold pseudo-symmetry of the CSC (Figure 3.1). The models are divided into two major symmetries, models that assume the critical isoform forms interactions between the inner and outer ring of CSC particles (Figure 3.3A,B,C,D,E,F) or models that place the critical isoforms opposite one another within CSC particles, only interacting with the less critical isoforms through CSR or P-CR dimers (Figure 3.3G,H,I,J,K,L). The less critical isoforms can be homodimers (Figure 3.3A,B,D,E,F-1,3; G,H,I,J,K,L) or heterodimers with each other (Figure 3.3A,C,D,E,F-2) in all of the models with CSR dimers. This results in the less critical isoforms in positions that are separated into an inner and outer ring (Figure 3.3A,B,C,E,F-1), split evenly between inner and outer ring (Figure 3.3A,B,C,E,F-2) or polarized to sides of the CSC (Figure 3.3A,B,C,D,E,F-3). For the second set of fundamental symmetries, the less critical isoforms can be separated into an inner and outer ring (Figure 3.3G,H,I,J,K,L-1), separated to make primarily heterodimers (Figure 3.3G,H,I,J,K,L-2) or polarized across the CSC (Figure 3.3G,H,I,J,K,L-3). Several other configurations also exist (Appendix 1.1 Figure 1), and re-arrangements within all of these models, but re-arrangements ultimately create the same patterns and do not alter the unique position of the critical isoform.

Assuming complete connectivity of CSRs, ZnFs and P-CRs as in the tetrameric models, makes an outer ring of CESAs that would need a long, flexible connection (relative to other isoforms in the same CSC), which could be the ZnF or CSR (Figure 3.3A,B,G,H). The first model creates a situation where the critical isoform is unique relative to the other isoforms by making both inner (120°) and outer ring (90°) CSR dimers (Figure 3.3A-1). The other isoforms in the model make CSR homodimers in either ring respectively, but not both. In a second model, the CSR distinguishes the critical isoform which now only makes heterodimers that separate isoforms in the inner and outer rings (Figure 3.3G-1). The remaining non-critical isoform dimers are shown as homodimers with opposing inner or outer ring positions relative to their heterodimer with the critical isoform. Using CSR dimers, the other models can be distinguished by the less critical isoforms only forming heterodimers (Figure 3.3A-2) or homodimers in both inner and outer rings (Figure 3.3A-3). Adding a constraint that only P-CR homodimers are formed, which could be another aid in CSC assembly, limits the possible models (Figure 3.3A,B,C,E,F-2,-3).

While complete connectivity between all of the interaction domains of CESA is potentially possible through long range interactions, another possibility is that the CSC makes the minimum number of dimers needed to assemble the particles and connect them. Assuming a ZnF dimer is critical to CSC formation, based on *in vitro* studies (Kurek et al., 2002; Jacob-Wilk et al., 2006; Purushotham et al., 2016), models can be created where either the CSR (Figure 3.3C,I) or P-CR (Figure 3.3D,J) have another role outside of core CSC oligomerization or are unbound. The models that result from this assumption follow the same symmetry and organization as before with three models for each. However, it is not apparent how the critical isoform is distinguished when the CSR is no longer forming dimers within the CSC, using only the potentially non-differentiating homo- and heterodimers of the ZnF and P-CR (Figure 3.3C,I). If the P-CR domain does not form interactions within the CSC or is unbound then the critical isoform would be dependent on the capacity of the CSR to make homo- and/or heterodimers in various ways (Figure 3.3D,J). These models imply that the P-CR is capable of making one or more contacts with CESA-interacting proteins.

Perhaps the best models include both dimeric CSRs and CSRs that do not interact within the CSC, such that CSRs make the inner ring connection between particles while ZnF and P-CR make interactions within a particle (Figure 3.3E,K) or alternatively, CSRs make some contacts within a particle and the ZnFs make interactions between particles (Figure 3.3F,L). Isolating the CSRs to contacts between particles makes the most sense as the model creates symmetric dimeric P-CR and ZnF interactions within the particles compared to symmetrically mismatched ZnF and CSR dimers within the particles. On the other hand, having different interactions within a particle for the inner and outer ring could be important to maintaining overall CSC shape and distances between particles.

The dimeric models are subtle in what distinguishes an interaction within or between particles. These models, however, are interesting because of how the isoform identities and dimeric interactions break symmetry and establish unique positions for the critical isoform, which is less apparent in tetrameric models. In conclusion, the dimeric P-CR models of the 24 CESA CSC are more plausible than the tetrameric P-CR models, as there are currently limited structural models of a tetramer and no *in vivo* or *in vitro* experimental evidence for a P-CR tetramer specifically. A P-CR dimer has been observed *in vitro* (Rushton et al., 2017). The models where P-CRs do not interact to form CSC symmetry are also interesting as they suggest a novel role for the P-CR in

interactions with CESA-interacting proteins. Also given the high sequence conservation of the P-CR, all of the different isoform P-CRs could be capable of interacting with CESA-interacting proteins.

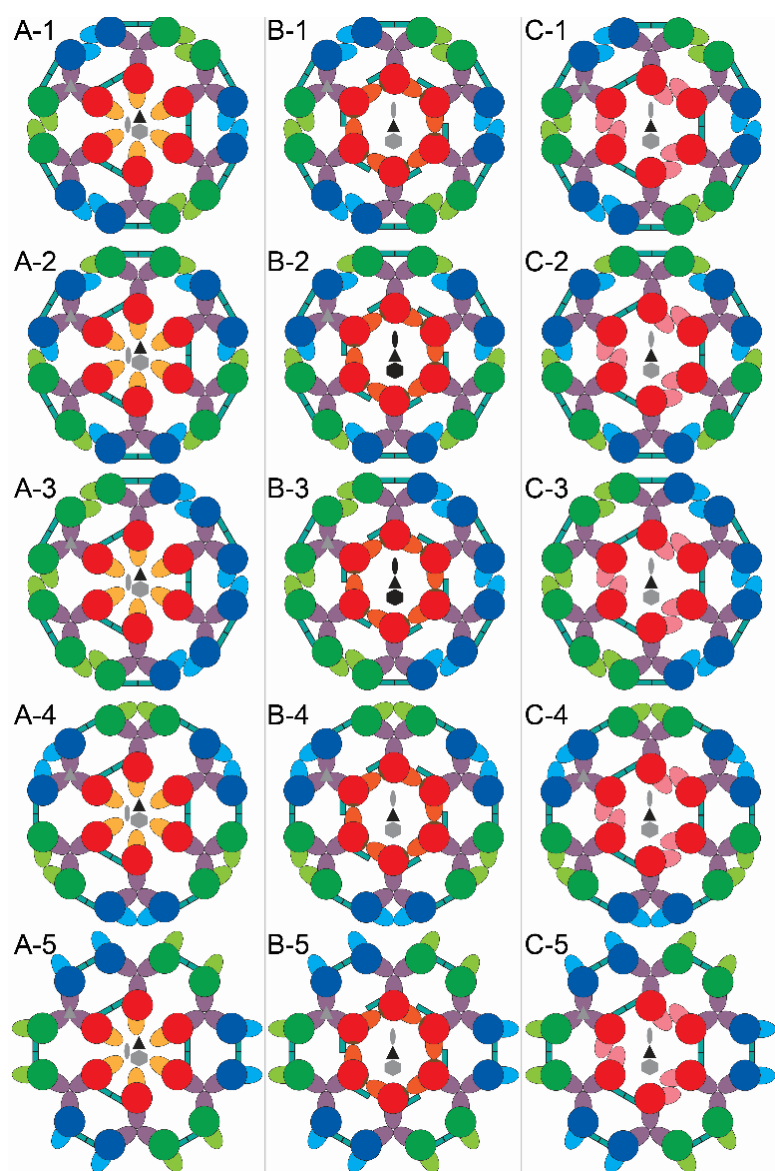
3.4 Trimeric P-CR Models of the CSC Particle

CESA homo- and/or heterotrimers within CSC particles have been proposed; the P-CR is often considered the primary candidate for this interface (Sethaphong et al., 2013, 2016; Nixon et al., 2016; Vandavasi et al., 2016; Rushton et al., 2017; Haigler et al., 2019), creating an 18 CESA CSC by then applying six-fold symmetry. A CSC built from trimeric CSC particles cannot produce 24-chain cellulose microfibrils with a hexameric array of CSC particles. An *Arabidopsis thaliana* CESA1 catalytic domain has shown apparent three-fold symmetry by SAXS after solubilization from inclusion bodies using detergent (Vandavasi et al., 2016). CESA1 *ab initio* models using either P-CR or CSR trimers were both fit to the SAXS density, but the P-CR trimer produced better alignment of the CatD overall, orienting all (1→4)-β-D-glucan chains in plane with each other. Other models use a CSR trimer to assemble an 18 CESA CSC, facing the P-CR outward to potentially interact with CESA-interacting proteins or other CSC particles (Sethaphong et al., 2016; Vandavasi et al., 2016). An experimentally observed P-CR trimer was formed in the P-CR crystal structure between neighboring asymmetric units, but without correlated data to *in vivo* function this could likely be a crystallization artifact (Rushton et al., 2017). Given the orientation of the P-CR relative to the catalytic core using the best fit of models to *in vitro* SAXS density for the entire CatD, the P-CR crystal structure trimer would create a non-sensical oligomer at an almost 90° perpendicular plane with the plasma membrane. However, with the flexibility of the last ~50 residues of the P-CR and the entire CSR, reorientation of these two domains relative to each other and the catalytic core is possible. If this flexibility were assumed to be in the P-CR relative to its position in the catalytic core homology model and its implied alignment with the TMHs, based on BCSA (Figure 1.3), it would be possible for a P-CR trimer to orient in plane with the plasma membrane. In this scenario, perhaps the P-CR is folded up against the catalytic core, as in the CatD SAXS model, when in solution and not forming trimeric contacts. This could occur when a CESA is in an inactive state or at low density, such as during protein trafficking or prior to CSC formation. Assuming a re-oriented P-CR trimer is the fundamental interaction of CSC particles creates a model similar to the *ab initio* models fit to class-averaged FF-TEM models of the CSC and CSC

particles (Nixon et al., 2016). Like the tetramer, the P-CR trimer in this model was built parallel to the plasma membrane using long axis of the P-CR coiled-coil, an alignment consistent with the P-CR trimer observed in the crystal structure. The trimer had the best fit of all the models to the class-averaged CSC particle model, averaged from the FF-TEM images, and packing of these models also fit well the six-fold symmetrical CSC. However, little detail was assigned to interactions between particles from the hypothetical molecular packing derived from SymmDock (Schneidman-Duhovny et al., 2005a, 2005b).

A P-CR trimer could form the fundamental contact within particles, with likely no distinction between homo- or heterotrimers. This then allows the ZnF and CSR to form dimeric contacts that complete the six-fold symmetrical CSC (Figure 3.4). However, this does not explain the necessity for three different isoforms or their 1:1:1 ratio. No heterooligomeric CESA has been observed *in vitro*, while homooligomeric CSCs are functional *in vitro* (Cho et al., 2015, 2017; Purushotham et al., 2016), indicating the three different isoforms only matter in the context of the entire CSC *in vivo*. The critical isoform might be needed to form a six-fold interface, or at least not clash with it, while maintaining the ability to function on the periphery as well (Figure 3.4A). *In vivo*, other isoforms might provide different functionalities not found *in vitro* and not necessary for *in vitro* CSC formation and microfibril synthesis, such as interactions with CESA-interacting proteins.

Figure 3.4 Hypothetical models of the CSC using P-CR trimers (light purple ellipse) to form the CSC particle. The circular catalytic core of a single CESA is colored by its class for the critical isoform class (red) or less critical isoforms (green and blue). The CSR (ellipse) is colored light green or blue for the less critical isoforms and differentially for the critical isoform to reflect different functions within a model (light/dark orange and pink). (A,B,C) Three different potential models for critical isoform interactions at the six-fold symmetrical center of the CSC are shown. (A) One potential model of critical isoform interactions as a hexameric interaction, or at least molecular packing that does not clash with six-fold symmetry as other isoforms might in this hypothesis, reinforced by ZnF (teal boxes) dimers between critical isoforms in a particle. (B) Another model could be the critical isoform makes asymmetric dimers with a neighboring critical isoform that might or might not be interacting with its ZnF, otherwise ZnFs could dimerize as in (A) or (C) assuming no overlap. (C) The third model is unique in CSR dimers between critical isoforms, alongside ZnF dimers, using the ZnF to connect peripheral isoforms within particles and the center isoform between particles. Peripheral CSR dimers between the less critical isoforms can homo- (A,B,C-1) or heterodimers between CSC particles (A,B,C-2). CSC organization could also polarize isoforms to opposite sides of the CSC (A,B,C-3). Peripheral ZnF interactions between CSC particles and CSR interactions within particles is also possible for homo- (A,B,C-4) and heterodimers (not shown). When minimal connectivity of the CSR and ZnF is considered, one model could be that inner isoforms form CSR dimers within the CSC, while peripheral less critical isoforms could have another role outside of core CSC oligomerization or be unbound (A,B,C-5).



In the P-CR trimeric model proposed, the critical isoform is the most occluded isoform, surrounded by other isoforms in the particle and buried in the six-fold interface of the entire CSC. The other two isoforms are on the periphery of the CSC, forming specific interactions for normal *in vivo* CSC activity. Both peripheral isoforms, or their redundant isoforms, could even have some specificity for certain CESA accessory proteins. Alternatively, the central isoform could be either or both of the less mutationally sensitive isoforms. The CSR and variable regions then distinguish the critical isoform through *in vivo* interactions of the CSC, with CESA-interacting proteins. The latter only re-organizes the identity of the isoforms at specific positions in these models but does not create any new symmetries. However, this only works if CSRs are assumed not to clash with the enforced six-fold symmetry of packing and peripheral CSC interactions (Figure 3.4A). Assuming the critical isoform is located at the center allows for additional models that assign functions to it within CSC organization (Figure 3.4B,C). One model assumes that the critical isoform is the only isoform capable of making asymmetric dimers between the CSR and a currently undefined catalytic core surface, which might even involve N-terminal ZnFs and/or variable regions (Figure 3.4B). Another model could be that the CSRs of the critical isoform are the only CSRs flexible enough to assemble a homodimer in the central position (Figure 3.4C) or the only CSR dimer used in CSC formation (Figure 3.4C-4). The peripheral less critical isoforms can then be assumed to make CSR homodimers (Figure 3.4A,B,C-1) or heterodimers (Figure 3.4A,B,C-2) between particles. Assuming CSR homodimers it is possible that the less critical isoforms are polarized in the CSC (Figure 3.4A,B,C-3). The ZnF is used to form peripheral interactions within particles and central interactions between particles (Figure 3.4A,B,C-1,2,3). Alternatively, the ZnF could only make interactions between particles, forming CSR homodimers (Figure 3.4A,B,C-4) or heterodimers (not shown) within particles. With the ZnF making interactions between particles it is possible that peripheral CESA isoforms could have another role outside of core CSC oligomerization or be unbound (Figure 3.4A,B,C-5). There is currently only experimental evidence for CSR dimers and assuming the isoforms do not have a unique CSRs that makes other interactions, the best trimeric model is one containing CSR dimers (Figure 3.4C).

3.5 Discussion

Limited experimental constraints and incomplete structural data have been used to generate an exhaustive set of symmetric models for the CSC based on P-CR dimeric, trimeric and tetrameric

interactions. The even distribution of isoforms within particles, ability to centrally position the critical isoform and potentially completely fill P-CR, CSR and ZnF interactions due to better packing of triangular trimeric particles around the six-fold center makes the P-CR trimer model more attractive than the dimeric or tetrameric models generated. Furthermore, CESA (Vandavasi et al., 2016) and P-CR (Rushton et al., 2017) homotrimers have been observed experimentally and the best fit of CSC particle models to class averaged CSCs from FF-TEM was a P-CR trimer consistent with the crystal structure (Nixon et al., 2016). Overall, this evidence points to the P-CR trimer as the best model for CSC organization, subsequently creating an 18-chain cellulose microfibril. P-CR homodimers have only ever been observed in the P-CR crystal structure, lacking the context of the entire CESA and potentially clashing with both catalytic cores in its best fit model to CatD SAXS data (Rushton et al., 2017). Based on this, the P-CR dimer is currently a better model than tetramers for a 24 CESA CSC as no P-CR tetramer has been experimentally observed. P-CR tetramers have only been hypothesized using the P-CR structural characterization in this study or *ab initio* modeling and tetrameric docking (Nixon et al., 2016). The role of the P-CR in interactions with CESA-interacting proteins remains to be determined.

Lack of clarity in cellulose microfibril dimensions, incomplete experimental structural models for the CSC and no clear or complete assignment of function to the P-CR, CSR and ZnF leaves all possibilities open. If interactions between particles, CSCs and CESA-interacting proteins are considered in the context of specific roles for isoforms, not necessarily critical for *in vitro* CSC formation, dimeric, trimeric and tetrameric models all have advantages and disadvantages. Still, there are many more possibilities and potential models, and considering the potential for asymmetric interactions and complex molecular packing will only make the true CSC structure all the more important. These models will now be evaluated at the molecular level for the P-CR using structural comparisons of the P-CR structures from Chapter 2 to other known structures (Chapter 4) and then probed with *in planta* mutagenesis (Chapter 5).

CHAPTER 4. STRUCTURAL COMPARISONS AND CHARACTERIZATIONS OF THE CELLULOSE SYNTHASEA8 PLANT-CONSERVED REGION REVEALS POTENTIALLY IMPORTANT RESIDUES FOR FUNCTION

4.1 Introduction

Discovery of the P-CR structure of the OsCESA8 was an important step in building a complete model of a plant cellulose synthase, however the function of this domain remains unknown. To generate new hypotheses about the function of the P-CR, its structure was compared to other known structures in the PDB using a variety of established algorithms. Four classes of protein structural comparison software were chosen to emphasize different structural features of the P-CR when represented in the context of any known function of a similar motif in another protein; protein C-alpha carbon comparisons (C α); secondary structural element comparison (SSE); 3D surface comparisons (surface); and geometry and sequence comparisons. Specific algorithms within each class were chosen to emphasize different approaches within these broad classes. Performance evaluations and benchmarking of these algorithms were also considered. C α comparisons were done using the Dali (Holm et al., 2016), and FATCAT (Y. Ye et al., 2004) algorithms. SSE comparisons were done using the deconSTRUCT (Z. Zhang et al., 2010) and PDBeFold (Krissinel et al., 2004) algorithms. Protein fragment sequence and geometry (Seq/Geo) comparisons were done using the BioXGEM 3D-BLAST (Tung et al., 2007) and iSARST (Lo et al., 2009) algorithms. Finally, surface comparisons were completed with the 3D-SURFER 2.0 (La et al., 2009) and PDBePISA (Krissinel et al., 2007) algorithms.

From these analyses, the best matches were several pore-forming domains of voltage-gated ion channel (VGIC) structures (Catterall, 2014). These VGICs showed significant statistical and visual similarity for the Dali (Holm et al., 2016), BioXGEM 3D-BLAST (Tung et al., 2007) and 3D-SURFER 2.0 (La et al., 2009) algorithms when compared to the P-CR (Rushton et al., 2017). Most other algorithms returned protein structures that were more traditional coiled-coil domain forms, featuring primarily small aliphatic residues in knob-in-hole (KIH) coiled-coil positions and lacking a connector loop between coiled-coil helices, which bends back and forms a contact with the coiled-coil in the P-CR. Little consistency was observed between the more traditional coiled-coils returned, and statistical/visual comparisons revealed them to be poor matches to the P-CR as

compared to the VGIC structures. The unusually high proportion of aromatic residues in the coiled-coil and presence of this connector loop limited the depth of structural coverage, making only a portion of the VGIC structures the closest structural comparison available. However, drastic differences in function between the P-CR and VGIC structures are likely because of the integral membrane nature of VGICs compared to the cytosolic function of the P-CR (Rushton et al., 2017).

The P-CR surface structure was characterized using the VisGrid (B. Li et al., 2008) and Fpocket (Schmidtke et al., 2010) algorithms. This highlighted that the top features of the P-CR and VGICs appear to be aligned between structures of a similar size, shape and molecular properties. From this analysis, novel models for P-CR interactions were hypothesized, and existing models were improved and more thoroughly characterized. In conclusion, the structural similarity searches and structural feature comparison provided inspiration for how the P-CR could make molecular contacts through structurally similar cavities, protrusions and pockets. I then designed mutants (highlighted in figures as orange stick models) that test the function of these features in cellulose synthesis and probe the function of the P-CR in the context of an entire CESA or CSC.

4.2 Results

4.2.1 C-alpha Comparisons

Among the most commonly used structural similarity software programs are the C α coordinate comparison algorithms. For this study, the Dali (Holm et al., 2016), and FATCAT (Y. Ye et al., 2004) algorithms were used to compare the P-CR structure (extracted from PDB ID: 5JNP) to other known structures in the PDB. Central to the C α class of structural comparison algorithms is the ability to align the query structure chain of C α coordinates to chains of C α coordinates extracted from a structural database such as the PDB or SCOP. The goal is to minimize the root mean square deviation (RMSD) in terms of the distance (Å) between the atomic centers of a query structure versus any given structure in the search database. C α algorithms belong to one of two schools based on their ability to alter the C α coordinates of a query structure within reasonable distances and tethering of these coordinates to the original structure in order to achieve a more optimal fit. For this study, both schools were evaluated. Dali utilizes a rigid body approach that does not allow for any modification of the backbone, while FATCAT considers the C α to be flexible within normal bond lengths and angles. Because only the C α coordinates are considered it is possible to

identify folds that are structurally similar but lack high sequence identity. This might provide insights into potential functions of a more general fold or important surfaces in a query structure with limited structural and functional annotation.

The P-CR (PDB ID: 5JNP) was submitted to the Dali protein structure comparison server (<http://ekhidna2.biocenter.helsinki.fi/dali/>) to compare rigid C α coordinates against the PDB database. The top alignments were retrieved from the PDB90 database containing a subset of structures that represents unique structures cutoff at the 90% sequence identity level. Because fewer redundant structures are retrieved, it is possible for more structural diversity to appear higher in the ranking list, which could otherwise be dominated by nearly identical structures that are heavily represented in the PDB. Dali ranks the structural similarity between the query and target structures using a Z-score. The Z-score is the number of standard deviations between the RMSD of aligned C α coordinates between the query and a structure in the database compared to the mean RMSD of random structural alignments in the database over the same length with the same or fewer numbers of gaps. Z-scores greater than 2 are considered statistically significant matches with the potential for structural similarity (Holm et al., 2008). The top five Dali results based on Z-score, are summarized in Table 4.1. All Z-scores are well above the significance cutoff and produce reasonable RMSD values that span an alignment over most of the P-CR structure. The shared % Seq ID is low in all five results and does not allow for confident side chain comparisons between structures. Interestingly, the top results were all different structures of VGICs, except for the fourth result, *FEMA* gene or Factor essential for Expression of Methicillin Resistance (PDB ID: 1lrz-A). Visual evaluation of the P-CR alignment to the top result, an ion transport 2 domain protein of a voltage-gated sodium channel VGIC (PDB ID: 3vou-A) (Figure 4.1), as well as other VGIC represented in Table 4.1, illustrates some of the key features shared between the structures. These shared features comprise coiled-coil domain structures of similar length, a similar number of aromatic residues at KIH positions, a relatively large and well-aligned connector loop between coiled-coil helices, and a small α -helical portion of the connector loop that bends back towards and interacts with the coiled-coil structure. This structure also has two aromatic residues on the small α -helix in its loop connecting coiled-coil helices. These aromatics align with a similar orientation to the solvent exposed Trp432 residue and sequential, buried aromatics in the P-CR small α -helix of the connector loop between coiled-coil helices. Several aromatic and charged amino acids are present near the Asp439/Tyr440 residues in the random coil next to the small α -

helix in this loop (Figure 4.1). These residues form the four-fold symmetrical contact in tetrameric VGICs, which is not observed in the P-CR structure. However, the functional importance of the small helix in VGICs, acting as ion gate and tetrameric interaction site, does hint that this structure could be an important contact(s) within CESA or the CSC.

Table 4.1 Top five Dali (Holm et al., 2016) results ranked by Z-score.

| Dali rank | 1 | 2 | 3 | 4 | 5 |
|--------------|--|---|---|---|--|
| PDB ID-Chain | 3vou-A | 6bwj-B | 6eo1-A | 1lrz-A | 4h33-A |
| Z-score | 6.8 | 6 | 6.2 | 6.1 | 6 |
| RMSD (Å) | 2.6 | 3.4 | 2.9 | 3.7 | 2.9 |
| # aligned | 72 | 73 | 72 | 61 | 68 |
| % Seq ID | 3 | 5 | 10 | 8 | 9 |
| Molecule | Ion transport 2 domain protein, voltage-gated sodium channel | Transient receptor potential cation channel subfamily V | Cyclic nucleotide-gated potassium channel MLL3241 | <i>FEMA</i> Factor essential for Expression of Methicillin Resistance | KvLm voltage-gated potassium channel LMO2059: C-terminal pore module |

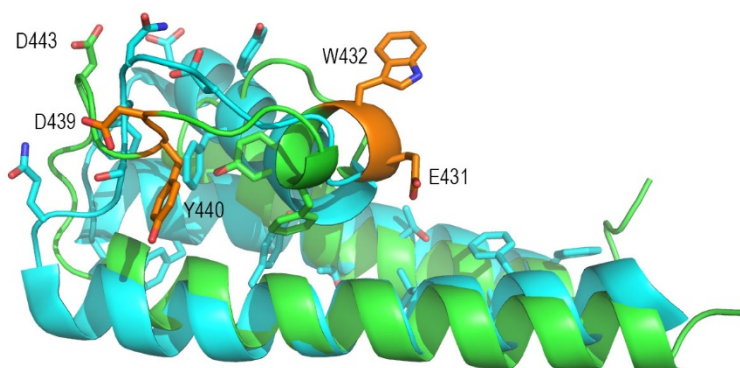


Figure 4.1 Top Dali (Holm et al., 2016) result, an ion transport 2 domain protein of a voltage-gated sodium channel VGIC (PDB ID: 3vou-A) (cyan), aligned to the P-CR (green) by Dali. An unusually high proportion of aromatic residues in KIH positions observed in both structures are shown as stick models for the ion transport 2 domain protein of a voltage-gated sodium channel VGIC. Aromatic and charged residues of the aligned connector loops and small α -helix also show similarities (stick models). Mutants made and validated in Chapter 5 that have alignment to these features are shown in orange.

Flexible structure alignment by chaining AFPs (Aligned Fragment Pairs) with Twists or FATCAT algorithm (Y. Ye et al., 2004) segments and aligns the query structure to a structure fragment database as opposed to fitting entire molecules. FATCAT attempts to minimize the RMSD between aligned molecules in a database using fragmented structures, making it possible to generate alignments to which an entire rigid body molecule might not conform. Smaller regions within proteins could be locally well-aligned but, because they contribute proportionally to the overall alignment and RMSD, the overall structural alignment sometimes forces them out of alignment due to rigid body connections. Thus, by aligning fragments it is possible to emphasize the regional alignments so long as sequential connections are preserved. By introducing “gaps and twists”, or rigid body movements of individual fragments relative to the entire starting molecule prior to reconnecting fragments, individual fragments can be aligned and reconnected to other fragments in an altered model of the original structure.

To prevent these model structures from over-refining, the number of gaps and twists contribute a significant penalty to the FATCAT score. The reported FATCAT score is the best score from all possible query fragments alignment against a structure within a database and includes contributions from the RMSDs of the individual fragments, length of alignments, and penalties for gaps and twists. To measure the significance of a structural alignment, p-values are calculated using the FATCAT score of the match compared to pre-computed scores from random alignments for the database chosen. P-values are typically considered significant when a 95% confidence is achieved, equivalent to a p-value of <0.05 . As a universal metric of structural alignment, FATCAT also reports the final RMSD of the theoretical structure aligned fragments (opt-RMSD) and the rigid body alignment of the molecules without fragmentation (chain-RMSD).

The top five FATCAT results, using the FATCAT server (<http://fatcat.burnham.org/fatcat/>) for the P-CR compared to the PDB90 database are summarized in Table 4.2 and ranked by lowest p-value. The FATCAT scores and p-values are significant for all five results, but shared % Seq ID is low for all results. Both opt and chain RMSD values were reasonable and both were slightly lower than the Dali RMSDs (Table 4.1) reported over a similar number of aligned Ca coordinates, although the introduction of gaps might explain this. None of the results required the introduction of a twist, however, visual inspection of the top FATCAT result, HIV glycoprotein gp41 (PDB ID: 3k9a-A), revealed that perhaps a twist predicted at Pro418 in the P-CR could have produced a better alignment (Figure 4.2). It appears FATCAT was only able to align small portions of larger

proteins to primarily the coiled-coil region of the P-CR. The top FATCAT result, although not completely composed of smaller and generally aliphatic residues at KIH positions like more traditional coiled-coils, does not contain as many aromatics in the KIH positions as the P-CR or the VGIC structures retrieved from Dali (Holm et al., 2016), 3D-SURFER 2.0 (La et al., 2009) and BioXGEM 3D-BLAST (Tung et al., 2007). This structure has only four consecutive Gly residues forming the connector loop between the coiled-coil helices as opposed to the 25-residue connector loop in the P-CR. Thus, without a significant connector loop, the top FATCAT alignment must contain a gap in this region and does not form a contact surface between the coiled-coil and connector loop as in the P-CR and VGIC structures. In conclusion, FATCAT supports the overall similarity of coiled-coil domains between the P-CR and more traditional coiled-coil structures through statistically significant alignments of these domains. However, a complete lack of alignment in the connector loop of the P-CR leads to a poor overall structural comparison between these protein domains with FATCAT as compared to other algorithms.

Table 4.2 Top five FATCAT (Y. Ye et al., 2004) results ranked by p-value.

| FATCAT rank | 1 | 2 | 3 | 4 | 5 |
|----------------|-----------------------------|---|-----------------|-------------------------------------|--------------------|
| PDB ID-Chain | 3k9a-A | 4abm-A | 2q1k-A | 4i0x-A | 1zva-A |
| FATCAT score | 133.93 | 129.13 | 114.16 | 131.12 | 128.47 |
| P-value | 8.81E-07 | 9.49E-07 | 1.29E-06 | 1.54E-06 | 2.97E-06 |
| opt-RMSD (Å) | 1.81 | 1.72 | 1.51 | 1.91 | 2.11 |
| chain-RMSD (Å) | 2.11 | 2.61 | 1.09 | 2.15 | 2.27 |
| # aligned | 74 | 76 | 73 | 74 | 74 |
| Twists | 0 | 0 | 0 | 0 | 0 |
| Gaps | 18 | 21 | 25 | 20 | 17 |
| % Seq ID | 1.4 | 5.3 | 2.7 | 6.8 | 2.7 |
| Molecule | HIV glycoprotein gp41 | Charged multivascular body protein 4B | AscE protein | ESAT-6- like protein MAB 3112 | E2 glycoprotein |

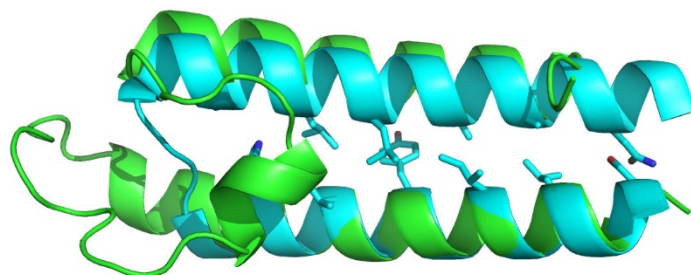


Figure 4.2 Top FATCAT (Ye et al., 2004) result HIV glycoprotein gp41 (PDB ID: 3k9a-A) (cyan) aligned to the P-CR (green) by FATCAT. Coiled-coil KIH residues of the HIV glycoprotein gp41 are shown as stick models.

4.2.2 Secondary Structural Element Comparison

When searching for structural similarity, one common idea is that similar folds, surfaces, pockets, and sometimes functions can be shared despite low sequence conservation (Krissinel, 2007). Inherent to this idea is that secondary structures and their orientation relative to one another are plastic in terms of whether a different residue will impact the overall fold or an interaction. Often a secondary structure element sequence can vary greatly in some regions provided the secondary structural element is not broken and key residues are allowed to form contacts, pockets and surfaces, sometimes clearly defined as conserved motifs. If we consider the arrangement of individual SSEs forming a pocket or surface that can be an intermolecular contact, then focusing an overall structural alignment on individual SSE assignments could be more accurate as opposed to using C α coordinates alone. By isolating and aligning SSE from the query it is also possible to emphasize the more stable α -helices and β -sheets as opposed to potentially poorly resolved and contextually flexible random coils that often connect SSE. However, maintaining the sequential arrangement and directionality of SSEs while limiting the introduction of gaps and rigid body movements between them is critical to preventing the selection of perfectly aligned SSEs at the cost of overall fold similarity.

The deconSTRUCT (Z. Zhang et al., 2010) and PDBeFold (Krissinel et al., 2004) algorithms were chosen as they employ similar calculations rooted in C α coordinate RMSD, focusing on SSE identification and alignment, but with some variation in overall approach. However, neither algorithm produced consistent results when compared to each other or any of the

other algorithms used. Visual inspection of the deconSTRUCT top result shows a structure with shorter helices aligned to the P-CR coiled-coil and a larger connector loop between helices. PDBeFold, on the other hand, produced top results featuring a traditional coiled-coil lacking a significant connector loop, but with strong alignment of the coiled-coil helices. The size of the P-CR and number of SSEs, might provide too little SSE context, with too many gaps and rigid body movements for an accurate overall alignment.

As described extensively in Z. Zhang et al. (2010), the deconSTRUCT algorithm is rooted in five steps: directional matching, sequential order checking, space layout checking, alignment of matched SSEs on the level of backbone atoms and alignment extension. This algorithm seeks to quickly reduce the number of possible matches by strictly eliminating structures with differing SSE directionality and sequential ordering. The P-CR structure was input into the deconSTRUCT algorithm server (http://epsf.bmad.bii.a-star.edu.sg/struct_server.html) and compared to a BlastClust (<ftp://ftp.ncbi.nih.gov/blast/documents/blastclust.html>) “100% id” database containing a representative 42,665 protein chains from the PDB database. The results from the deconSTRUCT alignments were ranked based on their deconSTRUCT align score (Z. Zhang et al., 2010), an algorithm-specific measure of how well SSEs align post-alignment extension and refinement. The top five results (Table 4.3) all showed similar deconSTRUCT alignment scores, alignment lengths and C α RMSDs. Alignment lengths spanned nearly the entire P-CR and showed similar C α RMSDs compared to other algorithms and deconSTRUCT alignment scores were reasonable based on published values, indicating high quality alignments.

Table 4.3 Top five deconSTRUCT (Z. Zhang et al., 2010) results ranked by align score.

| deconSTRUCT rank | 1 | 2 | 3 | 4 | 5 |
|-----------------------|--|--|------------------------------|----------------------|-------------------------------|
| PDB ID-Chain | 2nru-A | 1mlv-A | 3ody-X | 1m57-A | 2bl8-A |
| Align score | 48.26 | 46.33 | 45.88 | 45.42 | 44.85 |
| geometric Z-score | -2.62 | -1.99 | -2.88 | -1.5 | -8.41 |
| RMSD (Å) | 3.66 | 2.91 | 4.2 | 3.1 | 3.26 |
| # aligned | 66 | 58 | 66 | 58 | 58 |
| Avg length difference | 5.33 | 3.33 | 8 | 4 | 15 |
| Molecule | Interleukin-1 receptor-associated kinase 4 | Ribulose-1,5 biphosphate carboxylase/oxygenase large subunit N-methyltransferase | p38alpha y323q active mutant | Cytochrome C oxidase | Enterocine a immunity protein |

However, geometric Z-scores, which measure how likely the alignments are to occur randomly, based on a precomputed comparison of all the structures within the database, do not agree with the rankings. Z-scores in deconSTRUCT are represented as negative values, with lower scores indicating higher significance. Using a standard Z-score significance cutoff of -2 , identical to the Z-score significance cutoff of 2 that was used in the Dali comparisons (Holm et al., 2008), all four of the top results are near this cutoff, with only the fourth ranked structure, a cytochrome C oxidase (PDB ID: 1m57-A) falling below the cutoff. The fifth structure is interesting as the geometric Z-score is quite significant at a value of -8.41 , but inspection of this alignment reveals a long average length difference that make it a poor match.

Visual inspection of the top result, an interleukin-1 receptor-associated kinase 4 (PDB ID: 2nru-A), reveals some interesting similarities, but also why this structure is not highly significant (Figure 4.3). Fundamentally these structures show differences in the two long helices aligned to the P-CR coiled-coil helices. The P-CR forms a coiled-coil with its long helices, whereas the interleukin-1 receptor-associated kinase 4 (PDB ID: 2nru-A) has much fewer KIH-like interactions, over shorter α -helices and is more likely a portion of an α -helical bundle when taken into the context of the entire structure (not shown). However, the interleukin-1 receptor-associated kinase 4 (PDB ID: 2nru-A) has a loop that connects and bends back a short α -helix towards the two helices aligned to the P-CR coiled-coil. This short helix, however, does not form many

contacts with its coiled-coil helices and does not incorporate several aromatic residues as is seen in both the P-CR and VGICs. Furthermore, this loop is longer than the P-CR connector loop and forms mostly contacts internal to the complete structure. In general, though, the loop appears have a cavity with some visual similarity to the most significant cavity identified in the P-CR and VGICs using the Fpocket (Schmidtke et al., 2010) and VisGrid (B. Li et al., 2008) algorithms discussed in Chapter 4.2.5. In particular, the interleukin-1 receptor-associated kinase 4 (PDB ID: 2nru-A) loop aligned to the P-CR connector loop has a trio of charged residues (sequence: GluAspTyr) on a second short α -helix in this loop that aligns to the Asp439/Tyr440 residues mutated and validated in Chapter 5 (Figure 4.3 orange stick models). In conclusion, while the top deconSTRUCT result does not appear as consistent or statistically significant as the VGIC structures found in other algorithms, it might support the importance of specific structural elements and cavities.

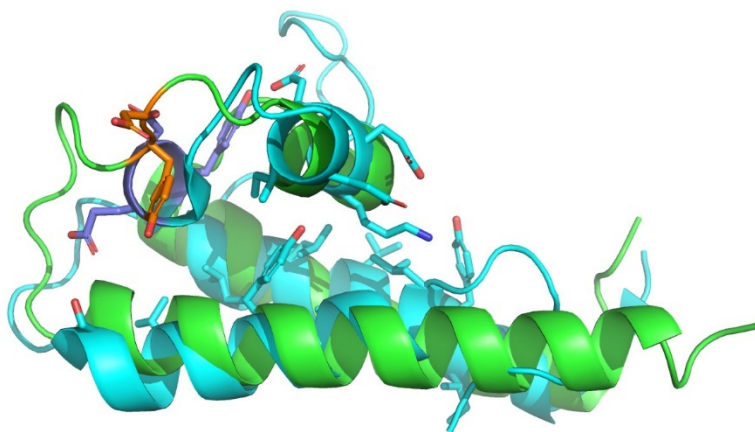


Figure 4.3 Top deconSTRUCT (Z. Zhang et al., 2010) result interleukin-1 receptor-associated kinase 4 (PDB ID: 2nru-A) (cyan) superimposed to the P-CR (green) by deconSTRUCT. Few KI and intermolecular contacts (shown as cyan stick models) are observed between the short helices aligned to the P-CR coiled-coil compared to VGIC structures. Three residues aligned to the connector loop in the P-CR, sequence: GluAspTyr (purple stick models) in 2nru-A, which align to the Asp439/Tyr440 P-CR mutants (orange) made and validated in Chapter 5.

The PDBeFold or secondary-structure matching algorithm (Krissinel et al., 2004) begins its alignment by first assigning and characterizing the SSE of the query, similar to deconSTRUCT (Z. Zhang et al., 2010). However, it differs in the way it processes database alignments. PDBeFold disregards connectivity of the query structure during initial alignment, length and directionality

refinements. Further refinements are based on C α RMSD comparisons between aligned SSE and connectivity is reconsidered, and penalties are assigned for losses of connectivity and size of gaps (Krissinel et al., 2004). The top five results for the P-CR using the PDBeFold server (<http://www.ebi.ac.uk/msd-srv/ssm/>) to search against the category termed “whole PDB archive” are summarized and ranked by Q-score in Table 4.4. A Q-score is the score given for the quality of the C α alignment maximized by the algorithm, including variables for length and C α RMSD. Q-scores showed a minor decrease down the top five results and were slightly below the typical values for structures that are known to be similar, which are generally >0.5. P-scores are a more complex measure of alignment quality that includes C α RMSD, number of aligned residues, gaps, number of matched SSE and SSE match score. Like the Q-scores, the P-scores were not indicative of well-matched structures for any of the results, and noticeably disagreed with the rankings of Q-scores, likely due to the inclusion of RMSD and number of aligned residues in the P-score calculations. Z-scores, which measures the likelihood of this result by chance from this database, all showed significance at values >2. Visual inspection of the results shows PDBeFold retrieved structures that primarily match to the coiled-coil portion of the P-CR. The best result, YocF sensor kinase (PDB ID: 3ehf-C) (Figure 4.4), shows that PDBeFold likely failed to account for the P-CR connector loop in its alignments, with only a short four residue loop connecting the helices. It produces a satisfactory alignment of the coiled-coil helices in terms of orientation and length, but the composition appears very different, being a traditional coiled-coil with primarily small, aliphatic residues forming KIH contacts. Because of the poor statistical results and failure to account for some key features of the P-CR, the PDBeFold results were not informative for analysis or mutant design.

Table 4.4 Top five PDBeFold (Krissinel et al., 2004) results ranked by PDBeFold Q-score.

| PDBeFold rank | 1 | 2 | 3 | 4 | 5 |
|---------------|--------------------|--------------|------------------------------------|--|--------------|
| PDB ID-Chain | 3ehf-C | 3fxe-B | 4nkg-B | 4ub8-Z | 2q1k-B |
| Q-score | 0.4986 | 0.4294 | 0.4214 | 0.4101 | 0.4097 |
| P-score | 2.544 | 2.799 | 2.045 | 2.024 | 2.62 |
| Z-score | 4.51 | 4.744 | 4.016 | 3.994 | 4.581 |
| RMSD (Å) | 1.597 | 1.634 | 2.044 | 2.328 | 1.767 |
| # aligned | 53 | 49 | 56 | 56 | 47 |
| % Seq ID | 3.774 | 8.163 | 3.571 | 5.357 | 2.128 |
| Molecule | YocF sensor kinase | IcmR protein | Serine/threonine-protein kinase N1 | Photosystem II reaction center protein Z | AscE protein |

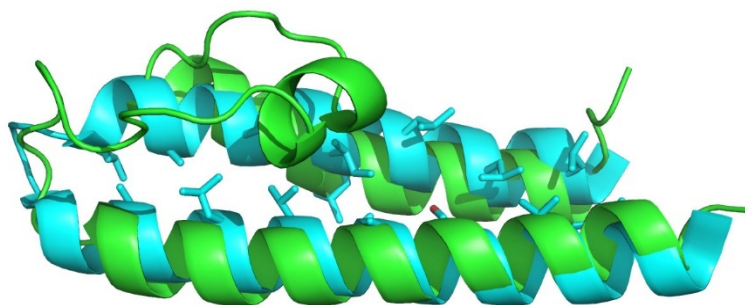


Figure 4.4 Top PDBeFold (Krissinel et al., 2004) result YocF sensor kinase (PDB ID: 3ehf-C) (cyan) sub-structure superimposed to the P-CR (green) by PDBeFold. Coiled-coil KIH residues of YocF sensor kinase are shown as stick models.

4.2.3 Geometry and Sequence Comparisons

Speed and accuracy are competing factors in protein structure alignment with a significant CPU and server cost. One approach is to reduce the 3D structural coordinates into a 1D sequence that captures the relevant structural information. Generally, this focuses on the angle between sequential $C\alpha$ coordinates and assigns a code to each residue or a collection of residues. This code is defined by training the algorithm against known structures, with sufficient structural input to accurately describe and capture the structural diversity in the PDB or SCOP databases. This performance can be benchmarked and used to justify the accuracy of an algorithm, and both BioXGEM 3D-BLAST (Tung et al., 2007) and iSARST (Lo et al., 2009) perform well compared to other algorithms. Once an algorithm-specific sequence code is defined for the query structure it

is then possible to implement fast and accurate BLAST-like comparisons using algorithm-specific substitution matrices. The BioXGEM 3D-BLAST algorithm retrieved another variant of a VGIC as its top result, emphasizing the similarities of the P-CR and these channel structure components. The iSARST algorithm retrieved results unlike any of the other programs with alignments to discontinuous sequences in the interior of other proteins. The complex folds surrounding the aligned fragments and highly discontinuous sequences retrieved suggests iSARST failed to make alignments that translate to the P-CR.

BioXGEM 3D-BLAST algorithm (Tung et al., 2007) uses the angles between C α coordinates of a five residue fragment to create its structural alphabet and substitution matrix. Using a set of 674 structural pairs, broken into 225,523 fragments, chosen for high structural similarity with low shared sequence identity, it was possible to develop a robust structural alphabet containing 23 states that define all of the structural components encountered in the 225,523 fragments. From this alphabet, a BLOSSUM-like structural alphabet substitution matrix was developed using methods similar to the BLOSSUM62 amino acid substitution matrix (Tung et al., 2007). A structural database can be transformed into the structural alphabet and queried using a BLAST-like algorithm, 3D-BLAST, to find the nearest pairs. The nearest pairs are defined in terms of an E-value significance factor with E-values reaching significance at $\leq 1 \times 10^{-10}$ and $\leq 1 \times 10^{-15}$ for results with <25% shared sequence identity values (Tung et al., 2007). The P-CR was entered into the BioXGEM 3D-BLAST server (<http://3d-blast.life.nctu.edu.tw/dbsas.php>) and compared against the entire PDB. The results are summarized in Table 4.5, showing that all the results retrieved were only a factor of 10 below the significant threshold of 1×10^{-10} , over nearly complete alignment lengths, but with modest % shared sequence identity values compared to typical values (Tung et al., 2007). The top result, a potassium channel KcsA-Fab complex in low concentration of Tl⁺ (PDB ID: 1r3k-C) (Figure 4.5), was a variant of a VGIC structure with similar characteristics to the VGICs found by C α comparisons of Dali (Holm et al., 2016) and 3D surface comparisons of 3D-SURFER 2.0 (La et al., 2009). The potassium channel KcsA-Fab complex in low concentration of Tl⁺ (PDB ID: 1r3k-C) sequences aligned to the small α -helix and connector loop regions are slightly different than those observed in the other VGICs but maintain similar functions in the structure. The hydrophobic core contact with the connector loop also contains fewer aromatics compared to the P-CR and other VGIC structures examined. The extended C-terminal helix of the VGIC is also shown here, illustrating some of the variation in length and

orientation that exists in the C-terminal coiled-coil helix, with better alignment to the P-CR in this helix than in other VGICs examined. This result supports the consistency of the VGIC matches found in this study and highlights similarly well-aligned regions that were targets in mutational design.

Table 4.5 Top five BioXGEM 3D-BLAST (Tung et al., 2007) results ranked by E-value.

| BioXGEM.3D-BLAST rank | 1 | 2 | 3 | 4 | 5 |
|-----------------------|--|--------------------------------|--------------------|----------------------|--|
| PDB ID-Chain | 1r3k-C | 1tk6-D | 3iq1-D | 2fjc-O | 1f30-I |
| E-value | 8.00E-12 | 1.00E-11 | 2.00E-11 | 2.00E-11 | 2.00E-11 |
| # aligned | 78 | 72 | 79 | 73 | 73 |
| % Gaps | 11.5 | 5.6 | 10.1 | 10.1 | 9.6 |
| % Seq ID | 35.9 | 36.1 | 30.4 | 35.6 | 38.4 |
| Molecule | Potassium channel KcsA-Fab complex in low concentration of Tl ⁺ | Iron-rich dpsA-homolog protein | DPS family protein | Antigen TpF1 protein | DNA protection during starvation protein |

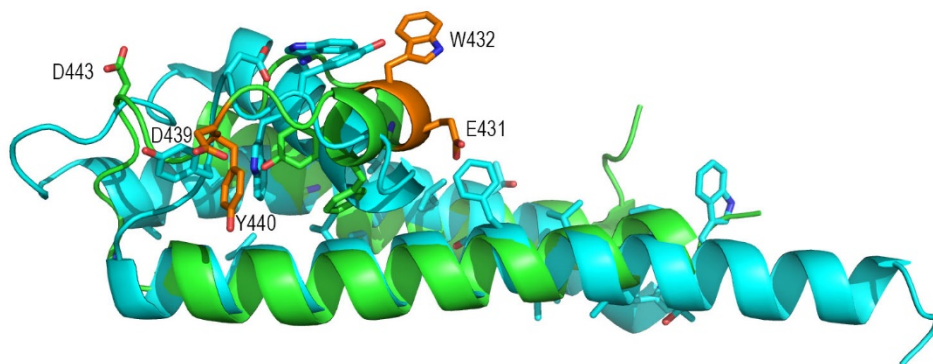


Figure 4.5 Top BioXGEM 3D-BLAST (Tung et al., 2007) result, a potassium channel KcsA-Fab complex in low concentration of Tl⁺ (PDB ID: 1r3k-C) (cyan), aligned to the P-CR (green) by BioXGEM 3D-BLAST. An unusually high proportion of aromatic residues in KIH positions observed in both structures are shown as stick models for the potassium channel KcsA-Fab complex in low concentration of Tl⁺ (PDB ID: 1r3k-C). Aromatic and charged residues of the aligned connector loops and small α -helix also show similarities highlighted as stick models. Mutants made and validated in Chapter 5 that have alignment to these features are shown in orange.

The SARST (Structural similarity search Aided by Ramachandran Sequential Transformation) algorithm (Lo et al., 2007) instead uses the sequential pattern of Ramachandran

angles to develop its structural alphabet. One thousand protein domains with single, continuous chains chosen randomly from the ASTRAL SCOP 1.67 40% identities subset (Fox et al., 2014) were used as a training set to define a set of 22 characters for the structural alphabet. This structural alphabet was then used to create a scoring matrix specific to the algorithm through a “regenerative approach” by which the matrix is improved through iterative refinement against increasingly larger and more diverse structural databases. The algorithm was further improved by integrating FAST (J. Zhu et al., 2005), TM-align (Y. Zhang et al., 2005) and SAMO (L. Chen et al., 2006) algorithms to post-refine the SARST results, known as the iSARST (Lo et al., 2009) web server.

The top five results from the iSARST web server (<http://sarst.life.nthu.edu.tw/iSARST/>) for the P-CR compared to the “100% identity non-redundant PDB structural database” using the TM-align post-refinement engine are shown in Table 4.6. The structures retrieved appear to be very similar in function, albeit with slightly different alignments to the P-CR and all with poor RMSD comparisons. All results showed poor structural diversity (Str. diversity) scores, a metric composed of alignment length and RMSD comparisons. Post-refinements using the FAST and SAMO algorithms produced similar results but with slightly worse statistics. Decreasing the complexity of the database only reduced the total number of results but did not change the top result. Visual inspection of the aligned portion of the top result, a bifunctional P-450/NADPH-P450 reductase (PDB ID: 4hgh-A) was interesting in that it does appear to follow the general fold of the P-CR with some secondary structure alignments (Figure 4.6). However, the alignment was highly discontinuous and contained several gaps and regions of poor alignment. The context of this alignment within the entire structure showed it was aligned to a largely interior, complex fold that does not have any similarity to the more solvent-exposed position of the P-CR in the CatD model (Figure 2.4) (Olek et al., 2014; Rushton et al., 2017).

Table 4.6 Top five iSARST (Lo et al., 2009) results ranked by Str. diversity.

| iSARST rank | 1 | 2 | 3 | 4 | 5 |
|----------------|--|--|---------------------------------|--|--|
| PDB ID-Chain | 4hgh-A | 3m4v-A | 4dufA | 2nnb-B | 1p0w-B |
| Size | 455 | 451 | 456 | 441 | 448 |
| # aligned | 64 | 66 | 70 | 67 | 64 |
| RMSD (Å) | 3.76 | 4.06 | 4.41 | 4.32 | 3.96 |
| Str. diversity | 31.86 | 32.48 | 32.76 | 32.83 | 32.89 |
| % Seq ID | 7.8 | 1.3 | 2.6 | 5.2 | 6.5 |
| Molecule | Bifunctional P-450/ NADPH-P450 reductase | Bifunctional P-450/ NADPH-P450 reductase | Cytochrome P450 BM3 variant 2G9 | Bifunctional P-450: NADPH-P450 reductase | Bifunctional P-450: NADPH-P450 reductase |

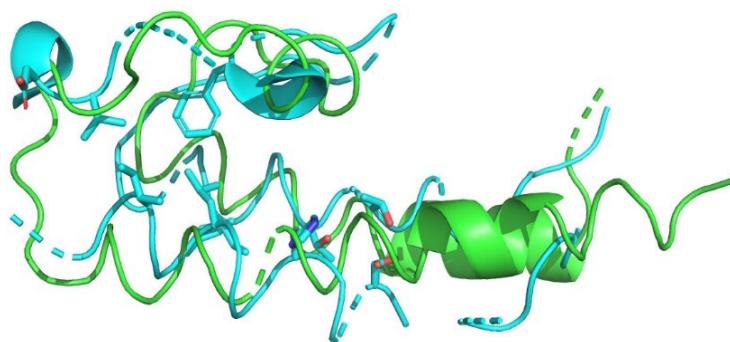


Figure 4.6 Top iSARST (Lo et al., 2009) result, a bifunctional P-450/ NADPH-P450 reductase (PDB ID: 4hgh-A) (cyan), sub-structure superimposed on the P-CR (green) by iSARST. Discontinuities caused by alignment to non-sequential fragments of the bifunctional P-450/ NADPH-P450 reductase (PDB ID: 4hgh-A) are shown as dotted lines. Coiled-coil KIH residues of 4hgh-A are shown as stick models.

4.2.4 3D Surface Comparisons

The surface structure of a protein is critical to its function and is the foundation of ligand-binding and protein-protein interactions. Thus, another approach to structural alignment is to reduce a structure to its molecular surface to compare it with other known protein surfaces, without considering the sequence, C α coordinates or SSE alignment. The basis for this is that similar structures are likely to have similar surfaces sometimes despite poor sequence conservation, poor C α alignment and/or differing SSE. Similarity can be assessed either by comparing the complete

surfaces of molecules or by comparing a theoretically important surface within a molecule to other local surfaces of proteins with known structures and functions. The P-CR was evaluated using both concepts, with complete surface comparisons done by the 3D-SURFER 2.0 algorithm (La et al., 2009) and individual surface comparisons of the crystallographic molecular interfaces done by the PDBePISA algorithm (Krissinel et al., 2007). This analysis showed that PDBePISA was not able to identify a statistically significant or visibly similar interface, likely due to the lack of full-length CESA structural and biological context in the crystallographic interfaces of the P-CR. On the other hand, 3D-SURFER 2.0 identified the same region of a VGIC as the top result as several other programs, further supporting the similarity between these domains.

The 3D-SURFER 2.0 algorithm (La et al., 2009) first voxelizes the query structure into a 3D surface mesh representing the molecular surface with MSROLL version 3.9.3 (Connolly, 1993) and then discretizes the mesh to create a cubic grid. 3D Zernike descriptors (Sael et al., 2008) are calculated from the cubic grid and used to compare the query structure to pre-computed 3D Zernike descriptors for a database. The comparison is made by calculating Euclidean distances derived between 3D Zernike descriptors. Smaller Euclidean distances between matches indicates less separation of the 3D Zernike descriptors and therefore a protein molecular surface with similar shape and size. This calculation is done independent of sequence identity or alignment length, such that molecular surfaces could contain more or less residues than the query, although a cutoff can be assigned for differences exceeding reasonable thresholds (La et al., 2009). The results from comparing the P-CR to the entire PDB database using the 3D-SURFER 2.0 server (<http://dragon.bio.purdue.edu/3d-surfer/>) are summarized in Table 4.7 ranked by Euclidean distance. Euclidean distances for all matches are below the significance cutoff (<10) and are similar to published examples of benchmarked structures (Sael et al., 2008; La et al., 2009). The top result, a VGIC potassium channel protein (PDB ID: 4rai-A) (Figure 4.7), showed better results with a difference of ~ 1 in Euclidean distance between it and the distances for the other four results. Although residue thresholds were not used in this analysis, the top result was closer in number of residues aligned to the P-CR, with 89 residues compared to the P-CR 75 residues. This could indicate at least some degree of similarity in structural packing. Like Dali (Holm et al., 2016) and BioXGEM 3D-BLAST (Tung et al., 2007) 3D-SURFER 2.0 returned a VGIC as its top result, although again not the identical PDB entry as either of two other top VGICs. Visual examination of the top result, trimmed and pair-wise aligned to the P-CR using the combinatorial extension

algorithm (Shindyalov et al., 1998) through the PyMOL (*The PyMOL Molecular Graphics System, Version 2.2*, 2019) CEalign command, revealed similar shared characteristics as those identified in the Dali (Holm et al., 2016) and BioXGEM 3D-BLAST (Tung et al., 2007) VGIC top results (Figure 4.7). Again, there are two aromatics on the aligned small α -helix in the connector loop that appear to align with conserved aromatics in the P-CR structure. The random coil regions that follows and aligns near Asp439/Tyr440, however, contains a poly-proline stretch that seems to take the place of the aromatic residues observed here in other VGICs. Unlike the BioXGEM 3D-BLAST (Tung et al., 2007) top result, a potassium channel KcsA-Fab complex in low concentration of Tl⁺ (PDB ID: 1r3k-C), the hydrophobic core contacting the connector loop contained more aromatic residues. This was in better agreement with the P-CR and the unusually high proportion observed in its KIH positions. Finally, the C-terminal coiled-coil helix in the VGIC potassium channel protein (PDB ID: 4rai-A) has a significant curvature that was not seen in the other VGICs, which across the VGIC structures indicates some structural heterogeneity is tolerated for this α -helix, but facilitate the same tetrameric interaction. This might indicate a lack of structural importance in the shape or surface around this helix in VGICs. A significant bend is also observed in the P-CR but in the N-terminal helix, at Pro418, aligned to the N-terminal helix of the VGICs and not the C-terminal helix where the 4rai-A bend is observed. One hypothesis could be that, like the VGICs, the P-CR surface and subsequent function might not depend on having a bent helix at Pro418, as some VGICs have a bend and others do not. This makes a Pro418 mutant that disturbs or removes this bend interesting. Thus, the top 3D-SURFER 2.0 result supports the consistency in the retrieval of VGICs with different algorithms and highlights some of the variation seen in the VGIC the connector loop and C-terminal coiled-coil helix.

Table 4.7 Top five 3D-SURFER 2.0 (La et al., 2009) results ranked by Euclidean distance.

| 3D-Surfer rank | 1 | 2 | 3 | 4 | 5 |
|--------------------|---------------------------|-------------------------------|--|---|------------------------------------|
| PDB ID-Chain | 4rai-A | 4djm-E | 2jsn-A | 2eyv-A | 4ro2-B |
| CATH ID | N/A | 2.60.40.360 | N/A | 3.30.505.10 | N/A |
| Euclidean distance | 2.128 | 2.905 | 3.045 | 3.087 | 3.099 |
| # Target Res. | 89 | 109 | 96 | 124 | 91 |
| Molecule | Potassium channel protein | <i>E. coli</i> chaperone DraB | Trafficking protein particle complex subunit 4 | v-crK sarcoma virus CT10 oncogene homolog isoform a | Proteasome subunit α type-3 |

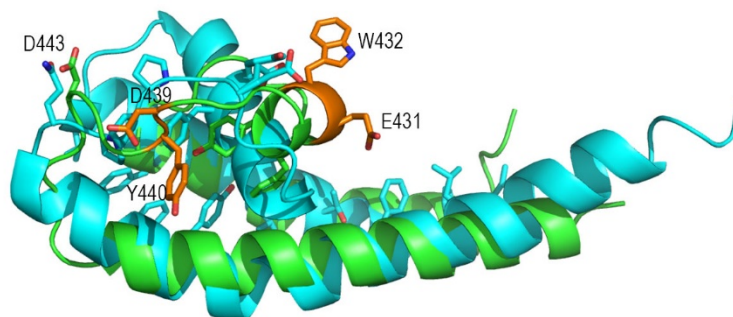


Figure 4.7 Top 3D-SURFER 2.0 (La et al., 2009) result, a VGIC potassium channel protein (PDB ID: 4rai-A) (cyan) aligned to the P-CR (green) by combinatorial extension (Shindyalov et al., 1998). An unusually high proportion of aromatic residues in KIH positions observed in both structures are shown as stick models for the VGIC potassium channel protein (PDB ID: 4rai-A). Aromatic and charged residues of the aligned connector loops and small α -helix also show similarities highlighted as stick models. Mutants made and validated in Chapter 5 that have alignment to these features are shown in orange.

PDBePISA (Krissinel et al., 2007) takes a different approach by using the intermolecular crystal contacts found between molecules in an atomic coordinate file to characterize and compare the structure to other known structures. This is based on the idea that contacts between separate molecules or protein chains within crystals are likely to utilize physiological surfaces and interfaces. The caveat, however, is that protein crystallization often uses high protein concentrations and precipitants to induce crystallization and a crystal will often have more contacts than the physiological assembly. Generally, the largest interfaces, with the lowest Gibbs free energy of solvation, are likely to be physiological contacts.

Several molecular contacts within the P-CR crystal are summarized in Table 4.8 and 4.9 (excluding small ligands from the list for clarity), which are derived from results using the PDBePISA server (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html). However, none of the contacts showed significant potential for a physiological interaction based on our current models (Rushton et al., 2017). Only the top three intra-molecular interfaces have an interface area sufficient for a relevant protein-protein contact ($>500 \text{ \AA}^2$), and the second and third largest interfaces were effectively the same (Table 4.8) (J. Chen et al., 2013). This is supported by the Gibbs free energy of solvation (Δ^iG) (kcal/mol), Δ^iG P-value, which indicate that only the top interface is large enough to be significant, as a Δ^iG P-value <0.5 indicates the interface is more hydrophobic and less likely to be a crystallization artefact than an average structure (Table 4.9).

Table 4.8 Molecular contacts of the P-CR in its crystal structure PDB ID: 5jnp characterized using PDBePISA (Krissinel et al., 2007).

| Int | Mol 1 | Surf. (\AA^2) | iN_{at} | iN_{res} | Mol 2 | Sym. | iN_{at} | iN_{res} | Int Area (\AA^2) |
|-----|-------|--------------------------|------------------|-------------------|-------|-----------------|------------------|-------------------|-----------------------------|
| 1 | B | 6201 | 94 | 26 | A | x,y,z | 96 | 26 | 951.2 |
| 2 | B | 6201 | 43 | 15 | B | -y+1/2,z,-x+1/2 | 46 | 13 | 445.1 |
| 3 | A | 6215 | 44 | 15 | A | -z+1,x,-y+1 | 47 | 13 | 444.3 |
| 4 | A | 6215 | 4 | 2 | B | z,x,y | 4 | 2 | 13.6 |
| 5 | B | 6201 | 1 | 1 | A | x,-y+1,-z+1 | 1 | 1 | 11.3 |
| 6 | B | 6201 | 1 | 1 | A | -x+1/2,-y+1/2,z | 1 | 1 | 10.7 |

Table 4.9 Interface surface characterization of the molecular contacts described in Table 4.8 characterized by PDBePISA (Krissinel et al., 2007).

| Int | Int Area (\AA^2) | Δ^iG (kcal/mol) | Δ^iG P-value | N_{HB} | N_{SB} | N_{DS} | CSS |
|-----|-----------------------------|------------------------|---------------------|-----------------|-----------------|-----------------|-------|
| 1 | 951.2 | -19.6 | 0.059 | 4 | 2 | 0 | 0.798 |
| 2 | 445.1 | -3.4 | 0.553 | 4 | 0 | 0 | 0.383 |
| 3 | 444.3 | -3.4 | 0.560 | 4 | 0 | 0 | 0.383 |
| 4 | 13.6 | 0.3 | 0.749 | 0 | 0 | 0 | 0.000 |
| 5 | 11.3 | 0.3 | 0.862 | 0 | 0 | 0 | 0.000 |
| 6 | 10.7 | 0.3 | 0.859 | 0 | 0 | 0 | 0.000 |

The top interface corresponds with the dimeric interface within the crystal asymmetric unit (Figure 2.12) and is likely to be stronger than the other two interactions alone, indicated by the complexation significance score (CSS). CSS measures the contribution of an interface to its assembly, in this case dimeric, in terms of the maximal fraction of the total Gibbs free energy of

solvation of the assembly. However, extending this dimer to a model with the P-CR connected to the catalytic core, as determined by the best fit of the P-CR to the CatD SAXS density, could form clashes between one P-CR and the catalytic core of a binding partner (Rushton et al., 2017). The second and third interface are effectively two sides of the trimeric interactions seen in the crystal, with one trimeric interaction for each of the two P-CRs in the asymmetric unit. Because the P-CRs are not identical within the asymmetric unit, although very close, they are identified as slightly different interfaces by PDBePISA. A trimeric interaction has a larger size and Δ^iG (kcal/mol) than a purely dimeric interaction as is represented here and summarized by the CSS fraction. Taking this into account for the trimeric interface produces a potentially strong interaction, which might not be reflected in the Δ^iG P-value of an individual interaction. Although the P-CR trimer based on this interaction also does not fit the CatD SAXS density and catalytic core homology model with an orientation appropriate for a CSC particle (Rushton et al., 2017), with some flexibility allowed this could be a functional interaction. In this scenario, the interface size would be within the normal range of typical protein-protein interaction (between 500-2000 Å²) (J. Chen et al., 2013), that could be combined with evidence for ionic coordinations and packing between monomers experimentally derived from the crystal.

Next, the top three P-CR interfaces were compared to pre-calculated interface characterizations of the entire PDB database and produced identical results in all cases due to the symmetry of these interactions. The structurally similar interfaces were first determined by comparing the shape, composition and Δ^iG (kcal/mol) of an interface to generate a list of structurally similar interfaces with >70% similarity based on the PDBePISA algorithm and its calibration parameters (Krissinel et al., 2007). This list was then compared to the P-CR using pair-wise structure comparisons by the PDBeFold or secondary-structure matching algorithm (Krissinel et al., 2004) to calculate Q-scores, based on alignment length and C α RMSD comparisons. The structures were ranked based on these Q-scores and the top five results along with the properties of their corresponding interface are summarized in Table 4.10. None of the Q-scores were indicative of good structural alignment, >0.5, but showed little variation over the top five results. Visual comparison of the top structure, synapse-associated protein 102 (PDB ID: 1u7m), to the P-CR (Figure 4.8), was done by pair-wise aligning to the P-CR with the PyMOL (*The PyMOL Molecular Graphics System, Version 2.2*, 2019) CEalign command (Shindyalov et al., 1998). Again, the visual comparison suggests little similarity as the top structure was a more traditional

coiled-coil with primarily small, aliphatic residue KIH interactions and showed no structural alignment to the coiled-coil connector loop in the P-CR.

Table 4.10 Top five PDBePISA (Krissinel et al., 2007) interface similarity search results ranked by Q-score.

| PDBePISA rank | 1 | 2 | 3 | 4 | 5 |
|----------------------------|--------------------------------|--------------|---------------------------|--|------------------------|
| PDB ID | 1u7m | 2q1k | 1pd3 | 2p58 | 3fxe |
| Target Interface | 1 | 1 | 1 | 3 | 1 |
| Q-score | 0.454 | 0.444 | 0.437 | 0.433 | 0.429 |
| Int Area (Å ²) | 1009.6 | 1073.2 | 1269 | 330.9 | 1162.4 |
| Δ^iG (kcal/mol) | -22.3 | -17.6 | -25 | -7.3 | -16.1 |
| % Seq ID | 10.2 | 6.4 | 3.9 | 8 | 8.2 |
| Molecule | Synapse-associated protein 102 | AscE protein | Nonstructural protein NS2 | Putative type III secretion protein YscE | Icmq and IcmR proteins |

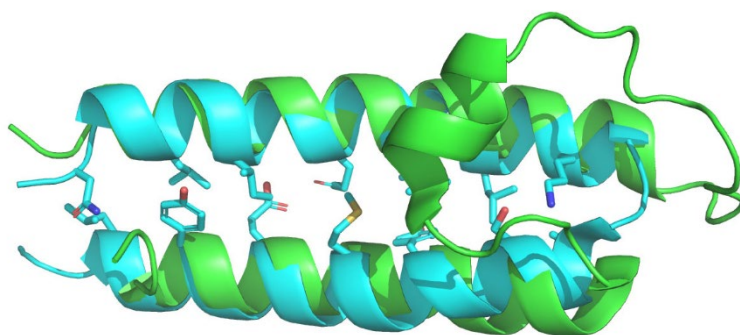


Figure 4.8 Top PDBePISA (Krissinel et al., 2007) result, synapse-associated protein 102 (PDB ID: 1u7m) (cyan), sub-structure aligned to the P-CR (green) by combinatorial extension (Shindyalov et al., 1998). Coiled-coil KIH residues of the synapse-associated protein 102 (PDB ID: 1u7m) are shown as stick models.

4.2.5 3D Surface and Pocket Analysis

Several surface cavities and protrusions in the P-CR were defined by 3D SURFER 2.0, while pockets were defined by Fpocket. The aromatic residues in the hydrophobic core of the P-CR, along with the long connector loop, illustrate a set of unique features compared to the more

traditional coiled-coil domains. These were identified in the Dali (Holm et al., 2016), BioXGEM 3D-BLAST (Tung et al., 2007) and 3D-SURFER 2.0 (La et al., 2009) structural comparison algorithms and were consistent in retrieving VGICs as the top result. Furthermore, these alignments highlighted the similarity in the surface cavities, protrusions or pockets between the P-CR and VGIC structures, specifically in the unusually high proportion of aromatic residues in the KIH positions of their coiled-coils and connector loop between coiled-coil helices.

3D-SURFER 2.0 specializes in surface analysis and integrates it with the VisGrid algorithm (B. Li et al., 2008) to characterize the largest cavities, protrusions and flat regions of the P-CR. In another approach, the P-CR was analyzed by Fpocket (Schmidtke et al., 2010), which uses alpha spheres to detect pockets, as opposed to a voxelizing the surface with a spherical probe. The three top VGIC results from previous structural similarity comparisons were also analyzed with Fpocket to compare their pockets with those found in the P-CR. Several similar cavities were found between the P-CR and VGICs. The largest and best rated pocket of the P-CR aligned to the top VGIC pocket, which is known to form the VGIC tetrameric interface (Catterall, 2014). Quantification of the surface area, shape and residue composition of this shared region also supports the possibility that this structure forms a protein-protein interaction in the P-CR. Several of the most prominent cavities, protrusions and pockets highlighted potentially important residues in the P-CR that were designed as mutations of the *Atcesa1* sequence (Chapter 5) to probe whether these formed surfaces critical to normal P-CR, CESA and CSC function *in vivo*.

The VisGrid algorithm (B. Li et al., 2008) identifies surface features by scanning the surface of a protein to identify the fraction of visible directions for a target position on the voxelized surface derived from the 3D-SURFER 2.0 algorithm (La et al., 2009). A protein cavity or hollow is defined as a cluster of target positions that have a small fraction of visible directions based on the criteria trained into the algorithm (B. Li et al., 2008). A protrusion is similarly defined as a cavity or hollow found in the negative image of the surface, while a flat region is a cluster of target positions with a roughly 0.5 fraction of visible directions. The surface area and volume of the surface associated with a cluster of target positions can then be calculated for the cavities, protrusions and flat regions. The top three cavities and protrusions ranked by the algorithm are listed in Table 4.11 with their surface areas, volumes and sequence of formative residues, summarized based on results using the 3D-SURFER 2.0 server (<http://dragon.bio.purdue.edu/3d-surfer/>). No significant flat region was detected.

Table 4.11 Top three VisGrid (B. Li et al., 2008) cavities and protrusions based on the 3D-SURFER 2.0 (La et al., 2009) results.

| VisGrid Rank | Surface Area (Å ²) | Volume (Å ³) | Residues |
|--------------|--------------------------------|--------------------------|--|
| Cavity 1 | 876 | 2085 | F412, K415, W416, F419, K422, Y423, F434, D452, R453, A455, K457, E459, Y460 |
| Cavity 2 | 791 | 1617 | E431, F434, S435, K437, Y440, R453, R454, K457, Y460, E461, K464 |
| Cavity 3 | 779 | 1610 | Y423, E426, Y433, Q436, I438, Y440, K442, K444, H446 |
| Protrusion 1 | 633 | 1085 | A399, L401, F403, L406, V471, A474, Q475 |
| Protrusion 2 | 1021 | 2033 | K421, N424, E426, R428, W432, K442, D443, V445 |
| Protrusion 3 | 363 | 390 | P447, K451, R454, R458 |

The largest cavities described in Table 4.11 are shown on the P-CR to illustrate residue compositions (Figure 4.9A,B,C) and molecular surfaces (Figure 4.9D,E,F). The largest cavity (Cav 1) is a deep, circular cavity between the coiled-coil α -helices opposite its contact with the connector loop, with Phe434 of a conserved aromatic trio in this loop forming the base of the cavity (Table 4.11; Figure 4.9A,D). Pro418, Arg453, Arg454 and Arg458 mutants were designed to explore the potential function of this cavity and surrounding residues. However, several other residues could be additional candidates for mutations to alter the surface charge distribution or occlude the cavity. The second largest cavity (Cav 2) was significantly smaller and not nearly as deep or circular (Table 4.11; Figure 4.9B,E). It is positioned at the interface between the coiled-coil α -helix HX2 and the C-terminal half of the connector loop. Unlike Cav 1, Cav 2 is likely more dynamic in shape and size due to being formed by random coil secondary structure and a coiled-coil helix, with fewer intermolecular constraints. However, there was sufficient intermolecular stability to measure electron density and observation in the X-ray crystal structure indicating that it is still quite stable and not disordered. This cavity is also centered on the P-CR crystal trimeric interaction (Figure 2.12). Several mutants can be hypothesized to alter this cavity, and potentially other nearby features simultaneously, however, it would likely be difficult to occlude, unlike Cav 1. There are several interesting features in this region for designing mutants, such as Glu431, Tyr440, Arg453 and Arg454 (aligned to *fra6* in *Atcesa8* (Zhong et al., 2003)) all of which are part of this cavity. Other nearby mutations could also affect this cavity, such as Trp432, Asp439 and Arg458. The third largest cavity (Cav 3) occurs almost entirely along the solvent-exposed face of the connector loop, spanning from both coiled-coil junctions to the small α -helix on the connector

loop (Table 4.11; Figure 4.9C,F). Although roughly the same size as Cav 2, it is comprised primarily of random coil and might be more dynamic in shape and size, in a way that might alter Cav 2 due to their shared structural elements. Again, changing the dimensions or surface charge of Cav3 is possible, but occluding it could be difficult. Due to this flexibility and the lack of similarity found in other structures, this cavity was less well explored in the mutants designed, but interesting phenotypes of Trp432, Asp439 and/or Tyr440 mutants could justify creating more mutants to explore this cavity in future experiments.

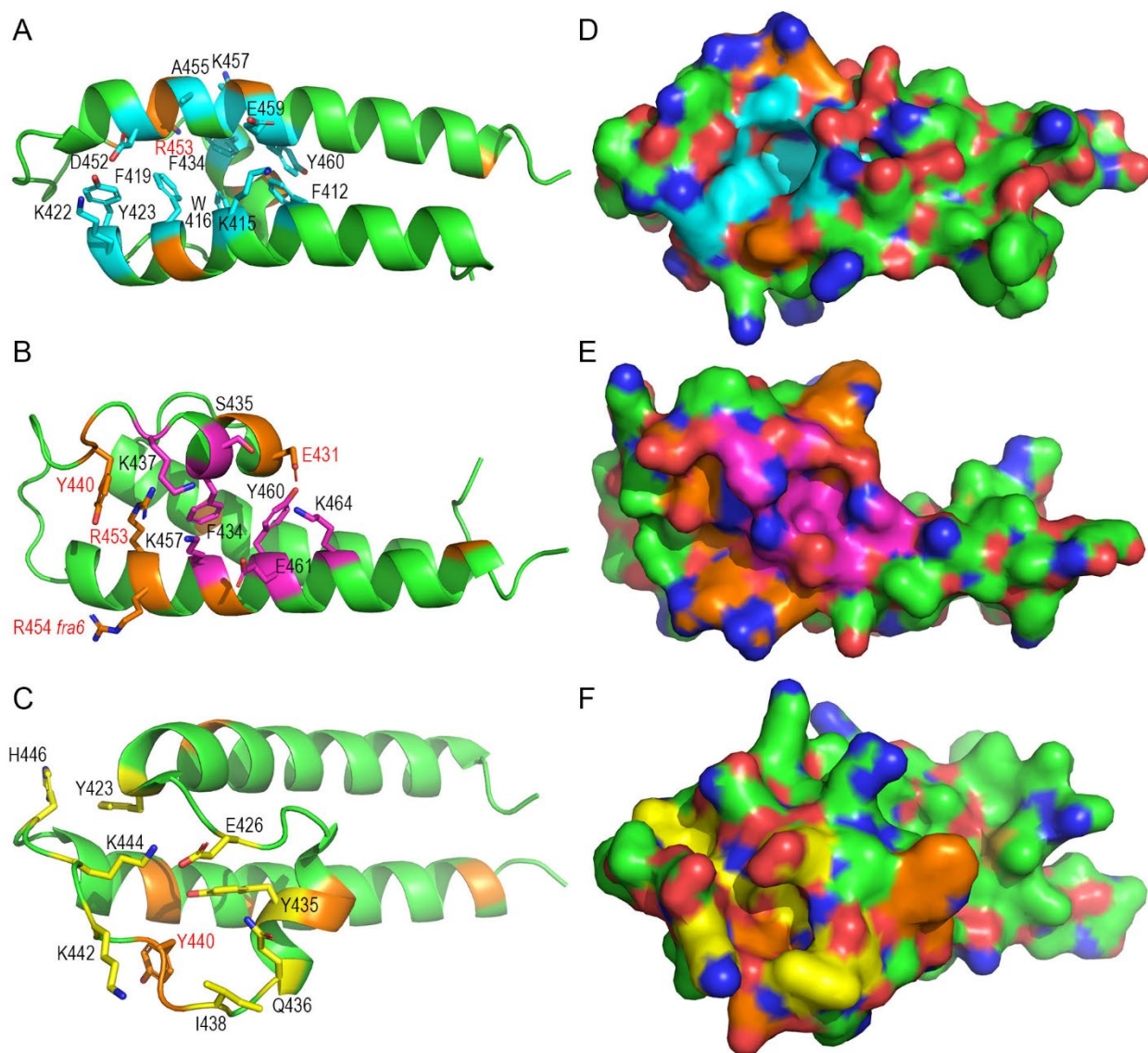


Figure 4.9 Largest cavities of the P-CR defined by the VisGrid (B. Li et al., 2008) algorithm using the 3D-SURFER 2.0 results. Residue compositions (A,B,C), illustrated as sticks, and molecular surfaces (D,E,F) are shown for the first (Cav 1) (A,D-cyan), second (Cav 2) (B,E-magenta) and third (Cav 3) (C,F-yellow) largest cavities. Mutants made and validated in Chapter 5 that have alignment to these features are shown in orange and labeled red, including the known cellulose-deficient mutant *fra6* found in the SCW *Atcesa8* isoform (Zhong et al., 2003)

Protrusions in this structure were generally smaller and often defined by one or a few critical residues. The largest protrusion (Prot 1) was odd in terms of its ranking given its size (Table 4.11; Figure 4.10A,D). Furthermore, due to its location, near the N-terminal junctions with CESA

and the disordered C-terminal portion of the P-CR in the crystal structure, this protrusion might be an artefact. Thus, only the designed Val471 mutant is likely to be informative for this protrusion. This region might be near the substrate entry portal based on alignment to homologous BCSA catalytic core models (Rushton et al., 2017). The second ranked, but largest protrusion (Prot 2) is interesting because it forms the edges of the third largest cavity, even sharing some of the same residues (Table 4.11; Figure 4.10B,E). Thus, Trp432 and Tyr440 mutants in this region are likely to probe both cavity and protrusion features, with the current Trp432 mutant acting as the best representative for this protrusion. Finally, the third ranked protrusion (Prot 3) has a very small surface area and volume, spanning only four residues, but it is interesting because it includes the Arg454 mutant (aligned to *fra6* in *Atcesa8* (Zhong et al., 2003)) (Table 4.11; Figure 4.10C,F). This protrusion also incorporates key residues of the P-CR crystallographic trimer using Arg454 and Arg458 (Figure 2.13), and partially forms Cav 2 discussed above (Figure 4.9B,E), so it is perhaps best considered a piece of Cav 2 in future analysis.

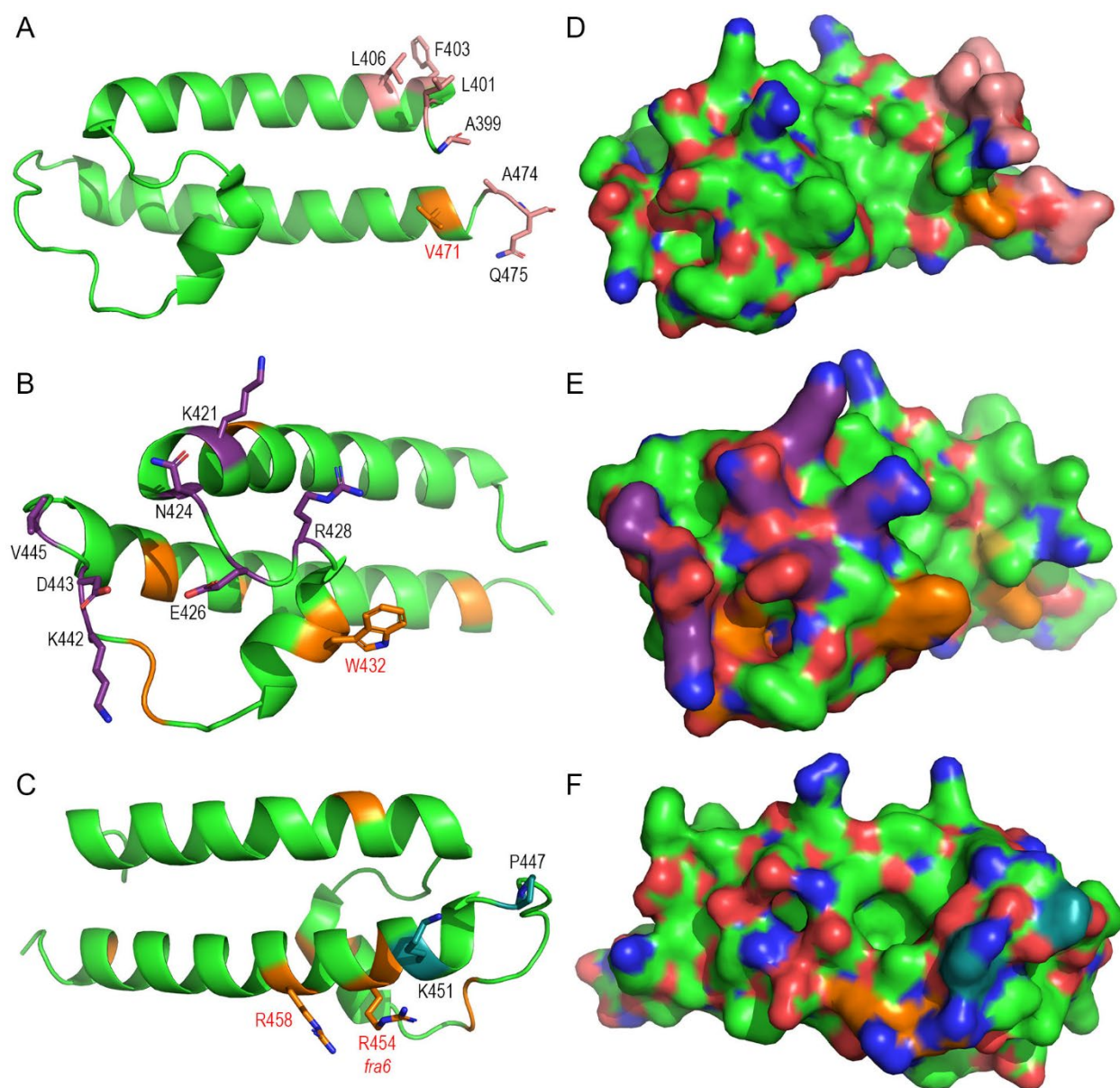


Figure 4.10 Largest protrusions of the P-CR defined by the VisGrid algorithm (B. Li et al., 2008) using the 3D-SURFER 2.0 algorithm (La et al., 2009) results. Residue compositions (A,B,C), illustrated as stick models, and molecular surfaces (D,E,F) are shown for the first (Prot 1) (A,D-pink), second (Prot 2) (B,E-purple) and third (Prot 3) (C,F-teal) ranked protrusions. Mutants made and validated in Chapter 5 that have alignment to these features are shown in orange and labeled red, including the known cellulose-deficient mutant *fra6* found in the SCW *Atcesa8* isoform (Zhong et al., 2003)

The Fpocket algorithm (Schmidtke et al., 2010) takes a different approach to cavity or pocket discovery and analysis by utilizing the inherent spaces that exist between residues to

generate alpha spheres that fill and describe these spaces. Unlike 3D-SURFER 2.0 (La et al., 2009) and VisGrid (B. Li et al., 2008), Fpocket does this without surface voxelation, which limits pocket discovery to regions that can fit the spherical probe typically set to a standard 1.4 Å radius. Thus, Fpocket is capable of finding more interior pockets and more accurately describes the spaces. Integral to this is the alpha sphere generation and refinement, where alpha spheres are defined as the space between a set of 4 residues forming a tetrahedron, with perfect packing of residues resulting in an alpha sphere with a radius close to the van der Waal radii. Thus, a long list of alpha spheres radii minus the van der Waal radii is generated to describe every space in the structure. The algorithm then filters out alpha spheres by size and polarity, clusters alpha spheres into similar groups, merges nearby pockets and removes small and hydrophobic pockets based on experimental optimizations of the algorithm (Schmidtke et al., 2010). This final list of pockets then has a list of descriptors generated for every residue contacting a pocket. These descriptors are based on the molecular properties currently identified and classified within the algorithm. The descriptors are used to calculate Fpocket scores and rank the pockets by comparing them to pre-calculated descriptor sets for residues in known binding pockets used in the training dataset.

Fpocket identified ten pockets in the P-CR based on the structure submitted to the Fpocket webserver at (<http://mobyli.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::fpocket>), summarized in Table 4.12 and illustrated in Figure 4.11. The alpha sphere clusters were extracted and independently measured with the Voss Volume Voxelator or 3V algorithm (Voss et al., 2010) to determine surface area, volume, sphericity and effective radius of the pockets. The 3V default probe was set at 3 Å and decreased until no changes in volume were observed. Surface areas and volumes of the pockets varied significantly between the top ten hits with the better ranking pockets tending to be larger. The Fpocket score did not perfectly rank them based on size, however, indicating pocket descriptors also played a major role in distinguishing pockets. The sphericity and effective radius did not differ much between ranks, perhaps indicating a lack of elongated pockets in the structure or a preference for more spherical alpha sphere clusters by the algorithm. Visual inspection of the pockets (Figure 4.11A,B,C,D) calls into question the validity of pockets ranked 2, 7 and 10 due to their proximity to the N-terminal junction with CESA and C-terminal junction with the disordered remainder of the P-CR, similar to the Prot 1 identified by VisGrid (B. Li et al., 2008) (Figure 4.10A,D). Only the Val471 designed mutant is near pocket 2, with no other

mutants designed near pockets 7 or 10. The remaining pockets are scattered throughout the structure primarily along the intermolecular contacts of the connector loop with coiled-coil helices.

Table 4.12 Top ten Fpocket (Schmidtke et al., 2010) pockets ranked by Fpocket score.

| Rank | Fpocket Score | Surface Area (Å ²) | Volume (Å ³) | Sphericity (Ψ) | Effective Radius (Å) | Residues |
|------|---------------|--------------------------------|--------------------------|----------------|----------------------|--|
| 1 | 29.3 | 330 | 348 | 0.72 | 3.16 | Y440, L441, F449, V450, R453, R454, K457, R458, E461 |
| 2 | 29.1 | 258 | 263 | 0.77 | 3.06 | M400, T402, A405, I467, L470, V471 |
| 3 | 23.9 | 247 | 239 | 0.75 | 2.9 | K422, Y423, N424, E426, D443, K444, V445, H446, F449 |
| 4 | 22.7 | 238 | 247 | 0.8 | 3.11 | I425, E426, Y433, Q436, I438, Y440, K442, K444, R453 |
| 5 | 21.0 | 301 | 306 | 0.73 | 3.05 | R414, K415, P418, F419, K422, Y423, D452, A455, M456, E459 |
| 6 | 19.5 | 211 | 210 | 0.81 | 2.98 | W416, V417, V420, K421, R428, A429 |
| 7 | 19.0 | 284 | 273 | 0.72 | 2.88 | F403, D404, A407, E408, V465, R466, G469, L470, K473 |
| 8 | 17.4 | 172 | 162 | 0.83 | 2.82 | P430, E431, F434, S435, K437, K457, Y460, E461, K464 |
| 9 | 13.3 | 240 | 247 | 0.79 | 3.09 | A455, R458, E459, E461, E462 |
| 10 | 11.8 | 245 | 238 | 0.76 | 2.91 | A399, M400, L401, A405, L406, T409, S410, A413 |

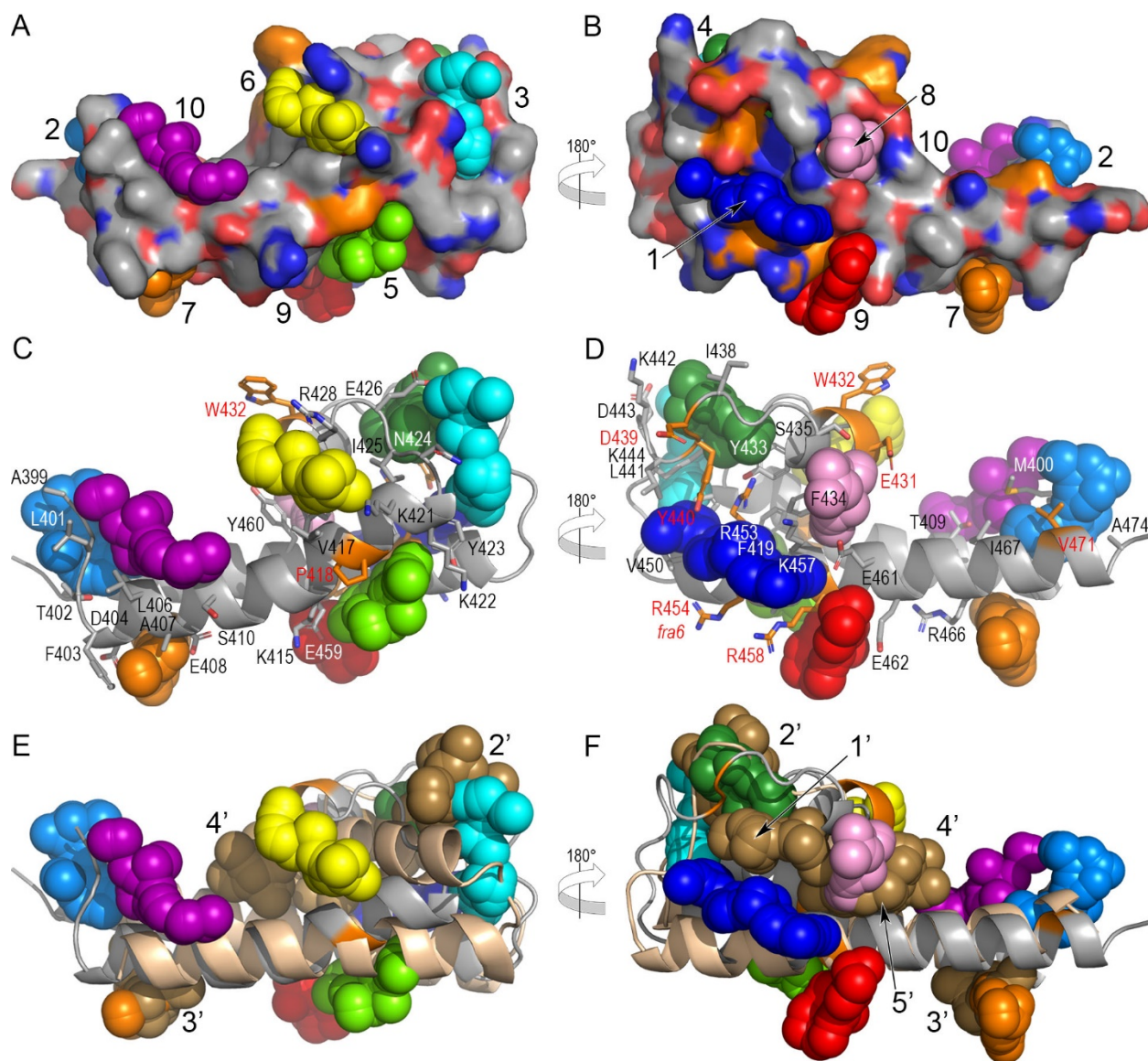


Figure 4.11 Ten best ranked pockets of the P-CR found by the Fpocket algorithm (Schmidtke et al., 2010). Pockets (spheres) are colored and labeled based on their rank and shown against the P-CR (grey; red and blue surface charges). Molecular surfaces (A,B) of the P-CR are shown surrounding the various pockets and residue compositions (C,D) that form the various pockets are labeled and illustrated as stick models. (E,F) Alignment of the P-CR pockets to the similar VGIC structure, ion transport 2 domain protein of a voltage-gated sodium channel (PDB ID: 3vou-A) (tan). The five pockets of the ion transport 2 domain protein of a voltage-gated sodium channel (PDB ID: 3vou-A) or 3vou pockets (brown) detected by Fpocket are labeled based on their rank (#') and indicate some structural similarity in their pockets when overlaid onto the P-CR structure (grey) and its pockets colored as in A,B,C,D. Mutants made and validated in Chapter 5 that have alignment to these features are shown in orange and labeled red, including the known cellulose-deficient mutant *fra6* found in the SCW *Atcesa8* isoform (Zhong et al., 2003)

To determine whether the most similar structure to the P-CR contained similar pockets, all three of the top VGIC structures identified were processed through Fpocket. This showed a consistent number and positioning of pockets between the identified VGIC structures. However, Fpocket detected only five pockets in the VGIC structures versus the ten found in the P-CR. One of these VGIC structures (PDB ID: 3vou) aligned to the P-CR (Figure 4.11E,F), illustrates that the top ranked VGIC pocket VGIC 1 (brown, 1) is close to P-CR pocket P-CR 1 (dark blue), VGIC 2 (brown, 2) is close to P-CR 3 (cyan), VGIC 4/5 (brown, 4 and 5) is close to P-CR 8 (pink), and VGIC 3 (brown, 3) is close to P-CR 7 (orange). The top pocket in the VGICs, VGIC 1, known to be critical for VGIC tetramerization (Catterall, 2014), is near the top P-CR pocket, P-CR 1. The P-CR 1 pocket also forms the P-CR crystal trimeric interaction (Figure 2.13), making this pocket one of the best candidates for P-CR protein-protein interactions.

Comparing the results of VisGrid to Fpocket for the P-CR (Figure 4.12) showed that they were generally in agreement and that mutants designed in this study could be defined in multiple contexts. Pockets are defined as empty spaces between residues (labeled P-CR 1-10 for the P-CR and 3vou 1-5 for the VGIC structure PDB ID: 3vou), while cavities and protrusions are the residues that describe the surface (labeled Cav 1-3 or Prot 1-3 for the P-CR). The largest discrepancy was that the top Fpocket pocket, P-CR 1 (Figure 4.12A), did not align to the top ranked VisGrid cavity Cav 1 (Figure 4.12B) but rather Cav 2 (Figure 4.12A). Fpocket pocket P-CR 5 (Figure 4.12B) instead aligned to Cav 1. Absolute ranking and size does not need to correlate to functionally important surfaces. Rather consistency in the ability to identify an important surface might lead to stronger predictions about the functional importance of a site. This reflects both the differences in cavity/protrusion and pocket detection, as well as how the algorithms differ in ranking them. Cav 1 has the largest surface area defined by VisGrid and is an oddly shaped circular, deep cavity, and fits the criteria well for VisGrid to identify it as a less visible surface. This cavity is so deep that its interior residues are also the surface on the opposite face, rotated 180° (Figure 4.12C), shown as Cav 1*. Although Cav 1 was ranked highest for being large, the comparable Fpocket pocket P-CR 5 in this same region could have been limited in size by the inability to be described fully by alpha spheres. This could be due to the filtering out of small, mostly hydrophobic interior pockets that could have been assigned to the deepest point of Cav1. Furthermore, Cav 1 seems to lack some of the nearby protrusions that Cav 2, Cav 3 and other Fpocket pockets have. Perhaps this was a mistake and P-CR 5 deserves a higher rank in light of the VisGrid ranking, but given the odd shape,

depth, more hydrophobic interior and lack of coordinating protrusions nearby, it makes sense that Fpocket would rank this lower using its descriptor-based ranking. Four mutants designed in this study are capable of exploring P-CR 5 and importance of Cav 1. In particular, Pro418 and Arg453 probe the interior and shape, while Arg454 (aligned to *fra6* in *Atcesa8* (Zhong et al., 2003)) and Arg458 probe nearby surface charges (Figure 4.12B). One of Fpocket's lowest ranked pockets, P-CR 9 is also near Cav 1 and primarily formed by Arg458 (Figure 4.12B). P-CR 9 is a less likely candidate for a binding site given the poor Fpocket score, based on its descriptors, relative to other P-CR pockets (Table 4.12). Cav 1 and P-CR 5 also do not have a corresponding VGIC pocket, so no correlation based on structural similarity and known function can be drawn.

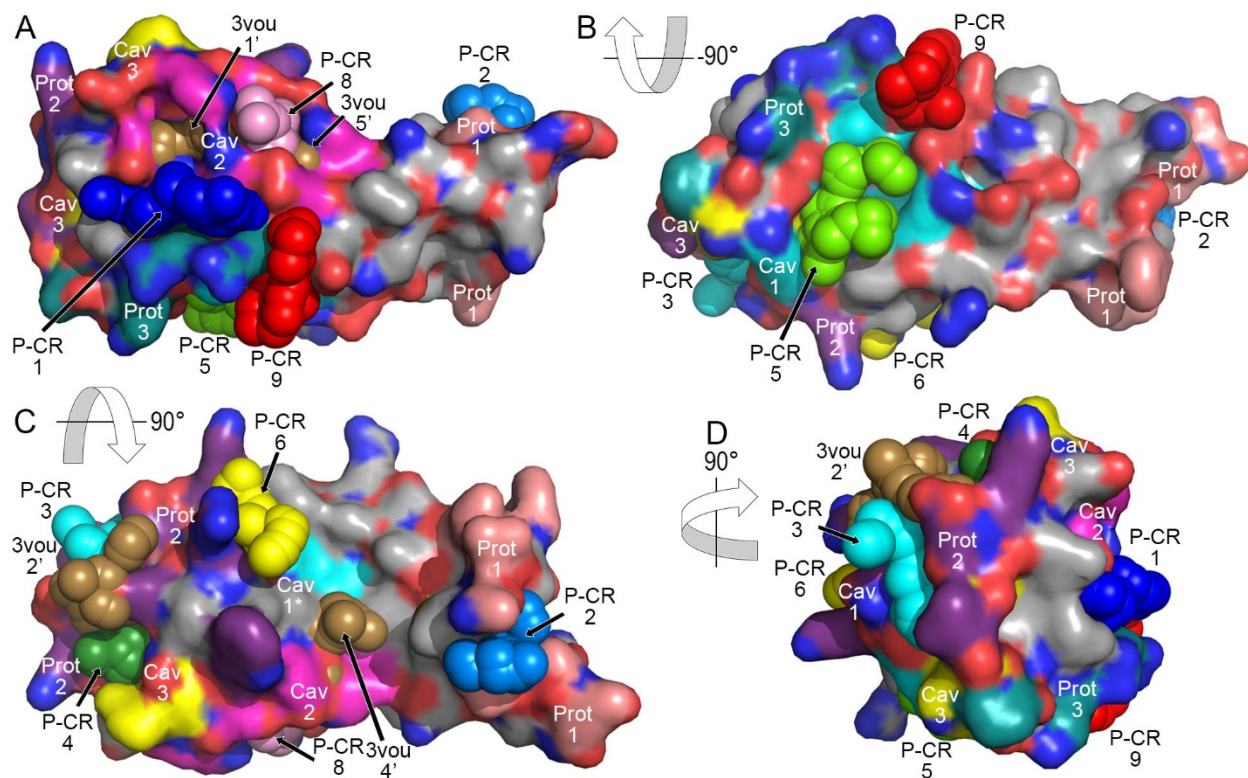


Figure 4.12 (A) Best ranked pockets of the P-CR and VGICs found by the Fpocket algorithm (Schmidtke et al., 2010) shown with the best ranked cavities and protrusions of the P-CR using the VisGrid algorithm (B. Li et al., 2008). The structure is rotated along the x-axis -90° (B) or 90° (C) and rotated along the y-axis 90° (D) relative to (A). P-CR Fpocket pockets are shown as spheres labeled as P-CR with a rank and colored as in Figure 4.11. VGIC pockets (brown) are shown for the ion transport 2 domain protein of a voltage-gated sodium channel (PDB ID: 3vou-A) and labeled 3vou with its rank. P-CR pockets 7 and 10, and 3vou pocket 3 were removed. VisGrid cavities are labeled Cav with their rank and colored as in Figure 4.9; protrusions are labeled Prot with their rank and colored as in Figure 4.10.

Given this analysis, P-CR 1 and Cav 2 (Figure 4.12A) become the most likely candidates for the location of a protein-protein interaction, which correlates with the results of the structural similarity searches. When comparing structures, the best P-CR alignments to VGICs show that the VGIC tetramerization interface, which includes the top Fpocket pocket in VGICs (3vou 1) (Figure 4.11F), also aligns to the most likely interface in the P-CR, P-CR 1 and Cav 2 (Figure 4.12A). Fpocket pocket P-CR 8 is also buried in Cav 2 and should perhaps be considered a continuation of this cavity. P-CR 8 also roughly aligns with VGIC pockets 3vou 4 and 5 (Figure 4.11F), which again should perhaps be considered a single pocket and then might have ranked higher. Cav 2 from VisGrid is surrounded by some residues from the Prot 2 and 3 protrusions and the Arg458 that defines P-CR 9. Being the primary contact of the P-CR trimeric interface also makes Cav 2 the best region to mutate in order to perturb a potential trimeric interaction (Figure 2.12).

The P-CR crystal trimeric interaction was explored with six mutant designs that, not surprisingly, also have relevance to other features found here and in previous work (Rushton et al., 2017). The Asp439/Tyr 440, Arg453, Arg458 and Arg454 (aligned to *fra6* in *Atcesa8* (Zhong et al., 2003)) mutants form the majority of P-CR 1 and are centered in Cav 2, while Glu431 and Trp432 are on the surface and form the protrusions that create P-CR 8. Again, P-CR 9 is defined by the Arg458 mutant and combined with the Arg454 mutant covers the defining Prot 3 residues. Together these regions are hypothesized to be important for P-CR function and are the most heavily explored regions by mutagenesis. The hypothetical importance of the structural features identified here is based on the structural similarity of the P-CR to VGICs and the necessity to form CSC protein-protein interfaces. In the absence of experimental evidence for how the P-CR fits into the structure of the entire CESA and CSC, these proposed protein-protein interfaces allow for hypothesis-driven experiments.

Cav 3 is located between the random coil secondary structure that forms the connector loop and is filled with Fpocket pockets P-CR 3 and 4 (Figure 4.12D). As discussed above, this is unlikely to be a stable cavity or pocket, meaning its dimension and/or orientation of charged residues might vary in solution. Cav3 also overlaps extensively with Prot 2 and is sandwiched between several other cavities and pockets. Cav3 encompasses the P-CR 4 and 5 pockets and also aligns to the VGIC pocket 3vou 3. The VGIC pocket 3vou 3 does not have an apparent structural role in VGICs (Catterall, 2014). Three of the designed mutants probe the stability and any potential function of Cav 3, but none are in the residues surrounding and forming P-CR 3 or 4 pockets. The

Tyr440 mutant sits at the base of Cav 3 and likely provides some stability with Asp439 also nearby. Trp432 represents one end of the elongated Cav 3, opposite the coiled-coil junctions. However, Trp432 is oriented away from the cavity, with Glu431 nearby and protruding slightly. Because of this, they are unlikely to have side chains that contribute to an interaction centered on Cav3. In conclusion, Cav 3 is one of less likely protein-protein interfaces in the P-CR compared to Cav 1 and Cav 2, but still has features that could be used in an interaction.

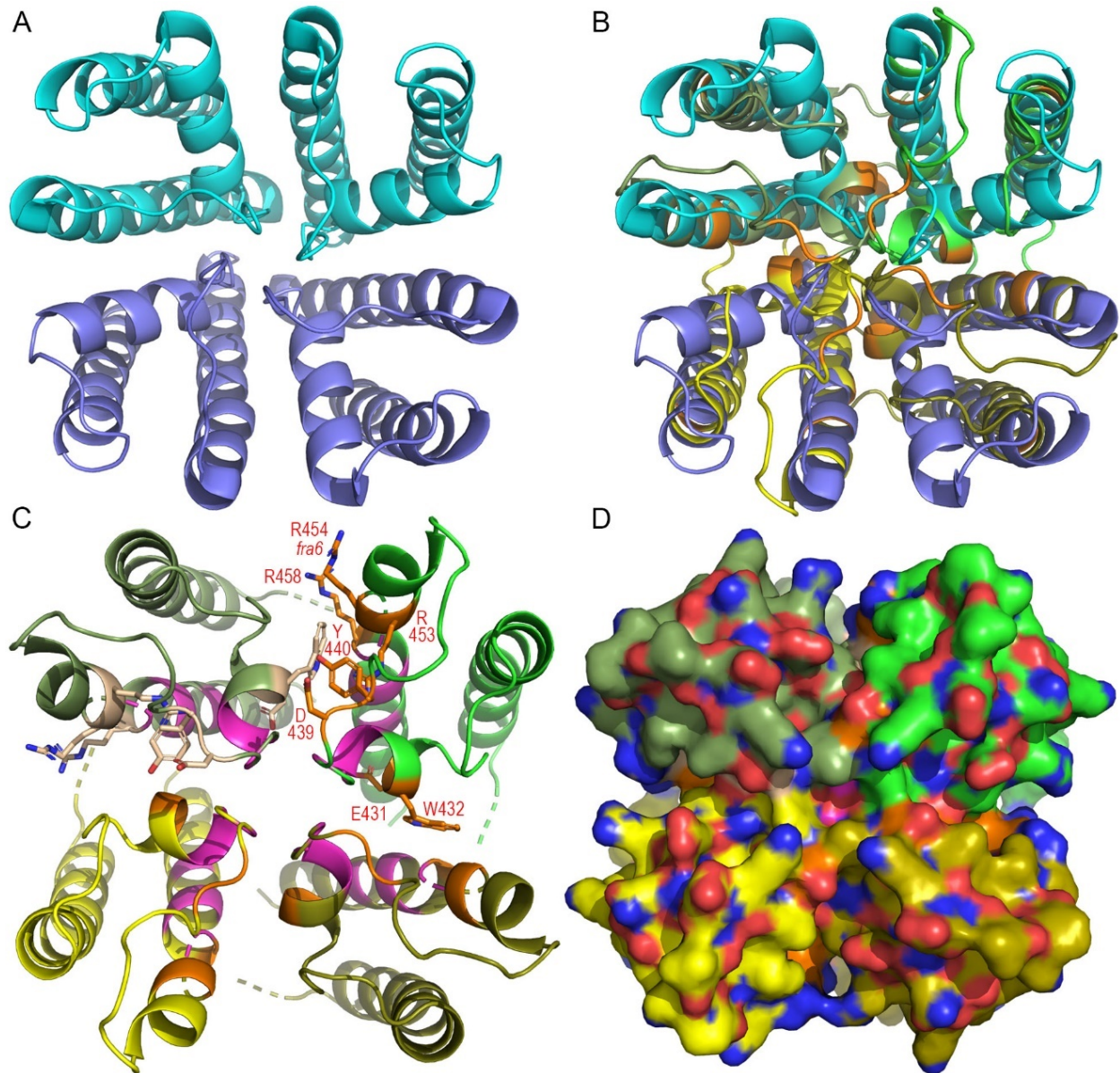
Finally, the best ranked protrusion from VisGrid, Prot 1, forms the majority of Fpocket pocket P-CR 2 (Figure 4.12C), but these features really lack the context of the CESA catalytic core connected to it at the P-CR N-terminus and the remainder of the disordered P-CR residues at the C-terminal limit of the crystal structure. There are also no significant VGIC pockets with an alignment to either Prot 1 or P-CR 2. However, this region could be important as it is positioned near the potentially homologous BCSA gating loop containing the FxVTxK motif. This is based on the best fit analysis of the P-CR structure and catalytic core homology model to the CatD SAXS density where the CESA FxVTxK motif loop is assumed to run between the P-CR and catalytic core. Therefore, a Val471 mutant was designed to represent this region. The Val471 lies on the side of a small, aliphatic groove that defines the side of the P-CR 2 pocket, opposite the larger N- and C-terminal protrusions. This region lacks significant features to hypothesize functionally important residues, being composed of mostly aliphatic residues, or might have features with functions not apparent without the context of an entire P-CR, CESA or CSC structure.

4.3 Discussion

4.3.1 Structural Similarity Searches Support a P-CR Tetramer Model of the CSC Particle

Having solved the P-CR structure, I looked for structurally similar protein domains to predict the function or functionally important residues of the P-CR. I used a variety of different algorithms to search for and quantify similarities between the P-CR and the other structures of the PDB. The most statistically relevant alignments consistently came from algorithms that retrieved voltage-gated ion channels (VGICs). VGICs will not likely provide a good direct correlation to potential P-CR functions (Rushton et al., 2017). However, they do present an interesting amount of similarity in their folds, with no other structures showing much similarity. Not surprisingly, the similar folds give rise to similar pockets, cavities and protrusions, which indicates VisGrid's P-

CR Cav 2 and Prot 2 regions, as well as, Fpocket pockets P-CR 1, P-CR 8 and P-CR 9 are similar to the VGIC pockets responsible for VGIC homotetramerization, shown as 3vou 1, 3vou 4 and 3vou 5 (Figure 4.12A and 4.13A). Even if the P-CR forms tetramers as in the tetrameric CSC particle models, it might not be analogous to the integral membrane VGIC tetramer unless it can adopt a similar configuration in the cytosol (Figure 3.2). Still, given the similarity of the structures, this is likely the best model for how a P-CR could form tetramers (Figure 4.13B).



After optimization of the P-CR tetramer by shifting each P-CR 2.5 Å away from the four-fold axis to remove main chain clashes and MD simulations to relax side chain orientations, as described in (Peterson et al., 2018), a reasonable P-CR tetramer model is created (Figure 4.13C,D). Based on this model the P-CR tetramer could be built from an asymmetric dimer where the connector loop, mostly around Prot 2, of one monomer interacts with Cav 2 of another monomer. This dimeric interaction is then iterated to make a tetramer. However, a tetramer has never been observed in CESA structures, making this only a hypothetical structure. A VGIC-like P-CR tetramer would use the P-CR long axis interaction that is perpendicular to the plasma membrane, which would fit the CSC particles better than the more extended model with the P-CR long axis aligned parallel to the plasma membrane used previously (Nixon et al., 2016). Considering this tetramer in the context of a CESA or CSC particle leads to two potential models for a CSC particle tetramer (Figure 4.14). The first model packs the P-CR tetramer against the catalytic core (Figure 4.14A,B), which would be similar to the P-CR fit to the CatD SAXS density (Rushton et al., 2017) (Figure 2.4). A second model is created by instead rotating the P-CR tetramer to a position below the catalytic core relative to the plasma membrane (Figure 4.14C,D). While not as tightly packed, this model maintains the orientation of the P-CR junctions with the catalytic core better than the first model. In conclusion, the difficulty associated with balancing the 1:1:1 isoform ratio over the six-fold symmetrical CSC makes tetrameric P-CR models less enticing, but theoretically they are possible with some flexibility in the orientations of specific isoforms and domains (Figure 3.2). Nonetheless, if a CSC particle is considered that is consistent with 24 CESA CSCs and 24 glucan chains in the core cellulose microfibril, then a VGIC-like tetramer is likely the best model for how P-CRs might form tetramers.

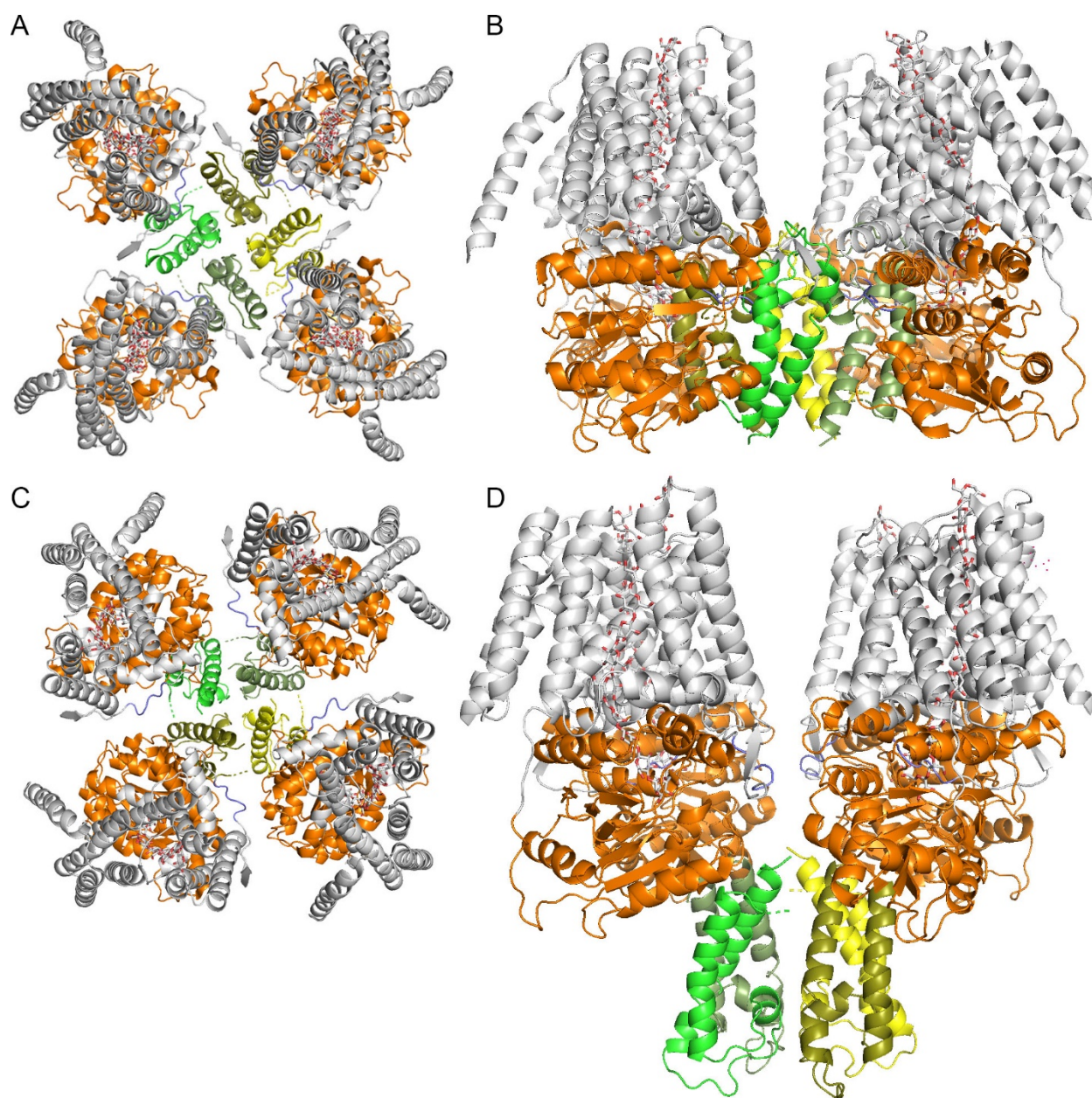


Figure 4.14 Atomic models of a P-CR tetramer (green and yellow) oriented perpendicular to the plasma membrane based on the catalytic core (orange) orientation. BCSA (PDB ID: 4hg6) TMH (grey), gating loop (FxVTxK motif: purple) and (1→4)-β-D-glucan chain (grey) are shown for reference. The P-CR tetramer could be on the same plane and packed against the catalytic cores (A,B) or below the catalytic cores (C,D), relative to TMHs and the plasma membrane based on the catalytic core (orange) orientation.

4.3.2 The P-CR Could Form Dimers Within the CSC

If the P-CR does not form tetramers, perhaps the pocket, cavity and protrusion similarities indicate that the P-CR might have an interface at Cav 2, but does not have enough similarity or context to predict oligomerization. A Cav 2 interaction could be consistent with the fundamental dimeric unit of the VGIC tetramer. This presents another model by which P-CR dimers could form a 24 CESA CSC (Figure 3.1 and 3.3) (Rushton et al., 2017). However, without experimental evidence for a dimer using Cav 2 it is less likely than a P-CR dimer based on the experimentally observed crystallographic dimer. A P-CR dimer would require an adjustment to the CatD model (Figure 2.4) in order to bring the interaction in plane with the plasma membrane, by either rotating the P-CR away from its position in the CatD model when in a dimeric interaction or by changing its position in the CatD model but maintaining its fit to the CatD SAXS density (Olek et al., 2014; Rushton et al., 2017). Orienting the P-CR crystal dimer perpendicular to the plasma membrane, based on the catalytic core orientation, produces a fundamentally reasonable model (Figure 4.15). However, the substantial overlap of P-CRs in this interaction packs the catalytic cores tightly and leads to several potential clashes.

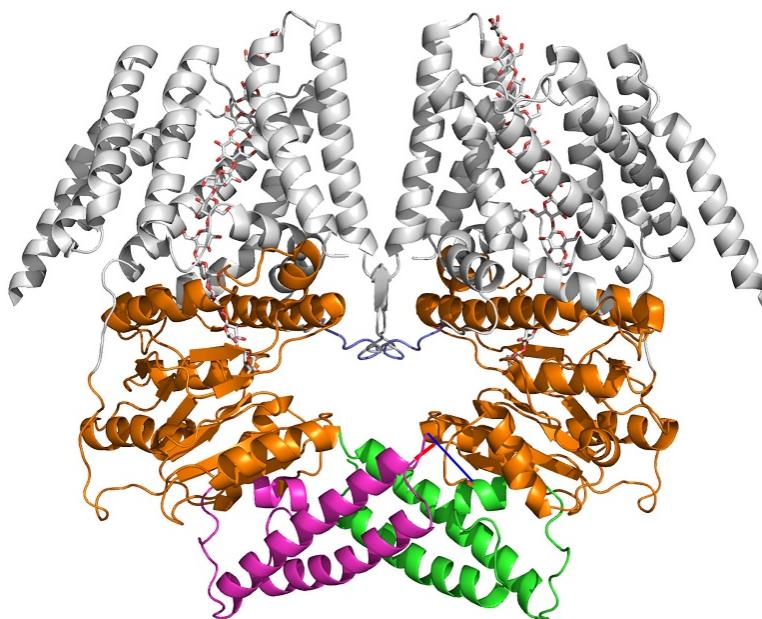


Figure 4.15 Atomic model of a P-CR dimer (green and violet) oriented perpendicular to the plasma membrane based on the catalytic core (orange) orientation. BCSA (PDB ID: 4hg6) TMH (grey), gating loop (FxVTxK motif: purple) and (1→4)-β-D-glucan chain (grey) are shown for reference. N-terminal (red line) and C-terminal (blue line) ends of the P-CR structure are near their catalytic core junctions.

4.3.3 The P-CR Trimer Model of the CSC Particle

A P-CR based trimer is one of the models of fundamental interactions within a CSC particle, consistent with an 18 CESA CSC and 18-chain cellulose microfibril (Nixon et al., 2016) (Figure 3.4). A P-CR trimer (Figure 2.12) (Rushton et al., 2017) and CESA trimer (Vandavasi et al., 2016) have also been experimentally observed. In the P-CR crystal structure, the trimer is formed between neighboring asymmetric units. Cav 2 faces the trimer axis and residues within and surrounding it form the trimeric contacts (Figure 4.16). The trimeric interaction is large enough, and well within the Gibbs free energy of solvation needed for a protein-protein interaction based on the PDBePISA analysis, especially considering the multiplicative effects of including all three interactions (Table 4.8 and 4.9). There were no proteins with similar trimeric interactions found in the structural similarity search, but this could be due to lack of similar proteins in the PDB in general. A P-CR trimer would require an adjustment to the CatD model (Figure 2.4) in order to bring the interaction in plane with the plasma membrane (Figure 4.17), by either rotating the P-CR away from its position in the CatD model when in a trimeric interaction or by changing its position in the CatD model but maintaining its fit to the CatD SAXS density (Olek et al., 2014; Rushton et al., 2017). Still, it is likely the best model for P-CR function and provides an experimental model to test the effects of disrupting the interface for most of the mutants described here. However, due to the overlapping location between this trimeric interface and the hypothetical VGIC-like P-CR tetrameric interface, using primarily Cav 2 and its surrounding features, it might be difficult to distinguish whether a mutant is interfering with a trimeric or tetrameric interface. Thus, both models must be considered simultaneously with only subtle differences in these interactions potentially discriminating between the models.

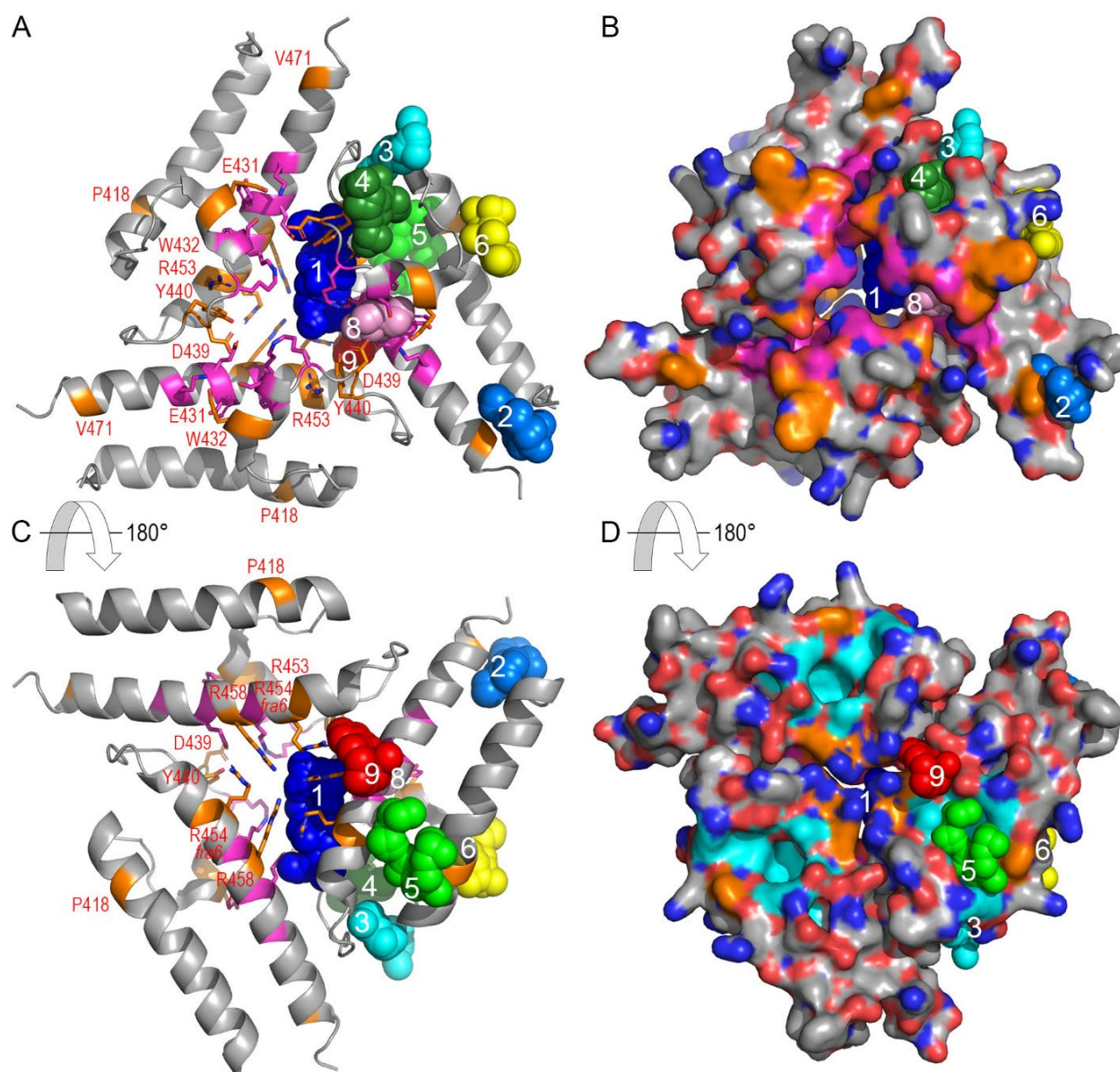


Figure 4.16 P-CR crystallographic trimer as stick (A,C) and surface (B,D) models of both sides of the P-CR long axis after 180° rotations (A,B vs. C,D), shown with significant pockets (spheres) and cavities in the P-CR relative to this trimer. Some of the P-CR mutants made and validated in Chapter 5 that are near or involved in the trimeric interaction are shown in orange and labeled red, including the known cellulose-deficient mutant *fra6* found in the SCW *Atcesa8* isoform (Zhong et al., 2003). Cav 2 (violet) is centered on the trimeric interaction with P-CR pockets 1 (blue), while P-CR pockets 4 (dark green), 8 (pink) and 9 (red) are nearby. Cav 1 (cyan) is filled with P-CR pocket 5 (green) and hypothetically could be an interaction site for a CESA-interacting protein without interfering with the trimeric interaction. Other pockets and P-CR mutants made and validated in Chapter 5 are shown for reference.

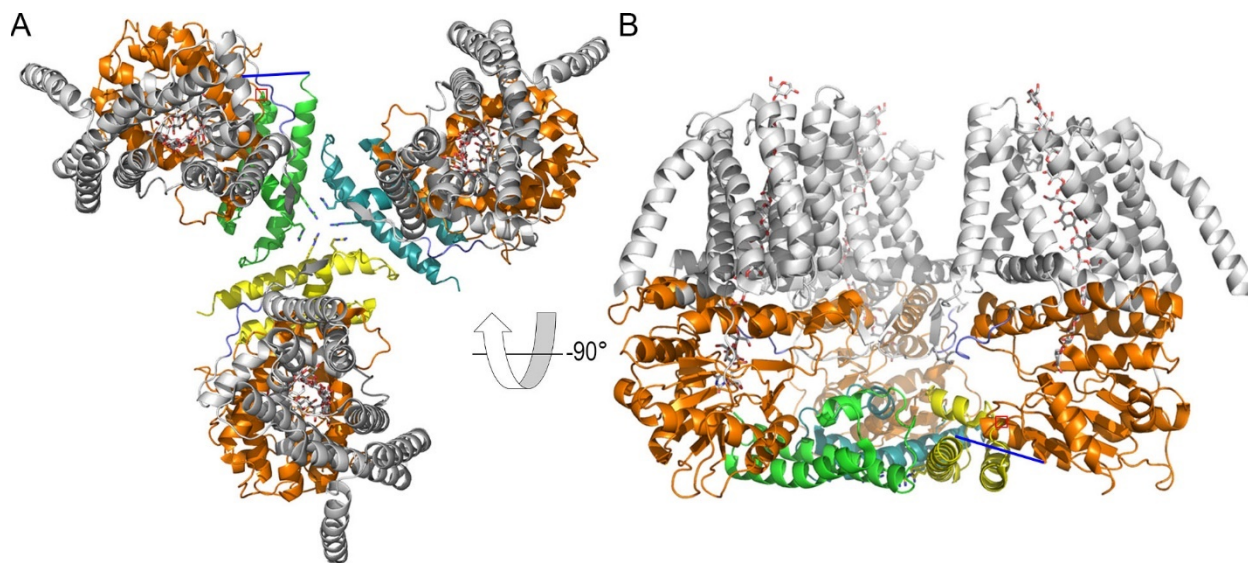


Figure 4.17 Atomic model of a P-CR trimer (green, teal and yellow) oriented perpendicular to the plasma membrane based on the catalytic core (orange) orientation. BCSA (PDB ID: 4hg6) TMH (grey), gating loop (FxVTxK motif: purple) and (1→4)-β-D-glucan chain (grey) are shown for reference. N-terminal (red box) and C-terminal (blue line) ends of the P-CR structure are near their catalytic core junctions.

4.3.4 The P-CR Could Interact with CESA-Interacting Proteins

Another possibility is that Cav 1, Cav 2 and/or Cav 3 are used to form asymmetric interactions with CESA-interacting proteins. Cav 2 appears to be the most likely interaction site for trimeric, tetrameric or CESA-interacting proteins, aligning to similar known protein-protein interaction sites in other proteins. Cav 1 is also a strong possibility for a heterogenous protein-protein interaction due to its high ranking in the 3D surface and pocket analysis. Given its location on the P-CR, ~90° away from the hypothetical trimeric (Figure 4.16D) or tetrameric P-CR interaction site at Cav 2 and ~180° away from the crystallographic dimer interaction, another interaction at this site would likely not occlude potential interactions at Cav 2 in a P-CR dimer or trimer (Figure 4.18). The P-CR tetramer faces Cav 1 away from the four-fold interaction, but the implications of the P-CR tetramer is that it could be on the same plane as the catalytic core or below the catalytic core relative to the plasma membrane. This creates drastically different models for how interactions could occur around Cav 1. In the first model, where the P-CR is tightly packed against the catalytic core on the same plane as the catalytic core, Cav 1 is largely occluded (Figure 4.18A). Conversely, in the second model where the P-CR tetramer is below the catalytic core relative to

the plasma membrane, Cav 1 is solvent-exposed and potentially available for other interactions. In the P-CR trimer models where the connector loop and Glu431/Trp432 face the membrane, Cav 1 would face the cytosol and potentially be available for interactions (Figure 4.18D). If the P-CR twisted 180°, facing the connector loop and Glu431/Trp432 facing the cytosol, Cav 1 would face the membrane. It is possible that interactions could occur with membrane-embedded CESA-interacting proteins such as KOR1 (Lane et al., 2001; Sato et al., 2001; Lei et al., 2014; Mansoori et al., 2014; Vain et al., 2014) or the glycoposphatidylinositol-anchored and N-glycosylated “polysaccharide chaperone” protein of the *COBRA* genes (Schindelman et al., 2001; Roudier et al., 2005; L. Liu et al., 2013; Sorek et al., 2014), while others, like CSI1 and CSI3, are cytosolic (Y. Gu et al., 2010; Bringmann et al., 2012; S. Li et al., 2012; Lei et al., 2013, 2015). Thus, assuming an interaction with a CESA-interacting protein at Cav1 does not interfere with the substrate portal, a P-CR dimeric, trimeric or tetrameric model are all possible. A CESA-interacting protein interaction at Cav1 is an attractive idea that allows the P-CR to perform multiple plant-specific roles compared to BCSA, but the role Cav 1 could play would depend heavily on the assumed orientation of the P-CR oligomer relative to the catalytic core and plasma membrane. Interactions with CESA-interacting proteins at Cav 3, unlike Cav 1, would be more likely to interfere with proposed dimeric, trimeric or tetrameric interactions of the P-CR without other structural adjustments and is the least likely based on its location, ranking and properties.

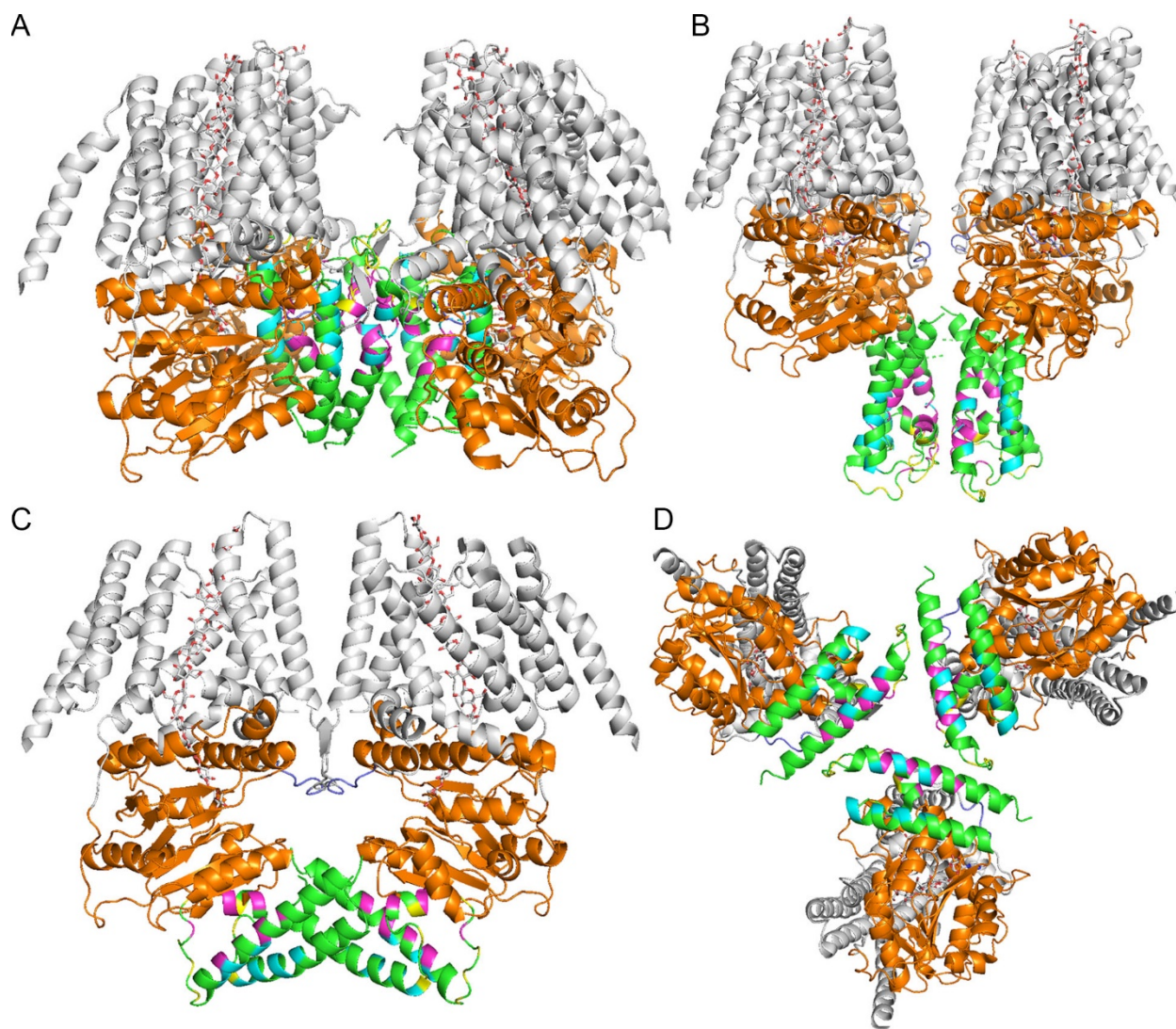


Figure 4.18 Models of the P-CR tetramer, both in the same plane as the catalytic core (orange) (A) or below the catalytic core (B) relative to the plasma membrane, dimer (C) or trimer (D) illustrating the occlusion or solvent exposure of Cav 1 (cyan), Cav 2 (violet) or Cav 3 (yellow) in various models. BCSA (PDB ID: 4hg6) TMH (grey), gating loop (FxVTxK motif: purple) and (1→4)- β -D-glucan chain (grey) are shown for reference.

4.3.5 Interpreting the OsCESA8 P-CR and Catalytic Core Model Best Fit to the CatD SAXS Density Relative to the FxVTxK Motif Loop

Inspecting the CatD best fit model overlaid with the BCSA structure reveals the BCSA gating loop is near the P-CR (Figure 4.19) (Rushton et al., 2017). The former fifth TMH of CESA, believed to form an interfacial helix (IF3) with a cytosolic loop region, contains the FxVTxK motif in its loop region that is also found in the BCSA gating loop (Figure 1.3, 2.4 and 4.19) (Slabaugh et al., 2014b,

2016). This loop is believed to occupy a similar location in CESA as the BCSA gating loop. It is tempting to hypothesize its function in substrate gating as in BCSA, however, CESA lacks the functionally well-described PilZ domain that tethers and interacts with this loop to open or close the substrate entry portal (Morgan et al., 2013, 2014). The precise location of this loop is also largely unknown in CESA as only *ab initio* models have been adjusted to a seven versus eight TMH model, with the former fifth TMH made interfacial and its loop cytosolic (Nixon et al., 2016). Even modeled against the weakly surface-bound BCSA gating loop, this CESA loop is limited in structural precision by its inherent flexibility in BCSA and predicted flexibility in CESA.

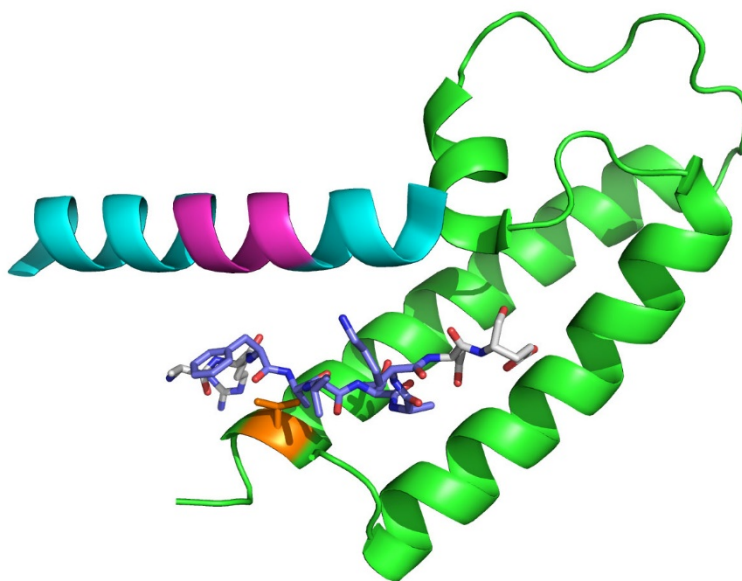


Figure 4.19 Overlay of the BCSA gating loop (stick model) with the P-CR (green) and catalytic core IF2 interfacial helix (cyan). This loop possesses the FxVTxK motif (navy) critical for function and also found in the loop following the IF3 helix (former fifth TMH) of CESA (Slabaugh et al., 2014b, 2016). The IF2 interfacial helix contains the Q/RxxRW catalytic motif (violet). The most prominent residue in the P-CR near this loop is Val471 (orange), which was mutated and validated in Chapter 5.

Even assuming the FxVTxK motif interaction is the same in CESA as in BCSA, where it interacts with the homologous IF2 helix of the catalytic core, it is unclear what other interactions it could be making to maintain this position. CESAs lack the PilZ domain of BCSA that, regardless of activation state, interacts with the BCSA gating loop. Thus, no structural model of a CESA exists that is known to stabilize this loop and it therefore might occupy a completely different location and have a different function in CESA. Although this loop was not included in the CatD

SAXS construct used to fit the homologous catalytic core and P-CR models, there is enough density to accommodate this in the SAXS envelope assuming it has the same relative location as in BCSA. This assumes the BCSA gating loop homologous sequences in CESA run between the P-CR and catalytic core in the CatD based on the best fit model (Figure 2.4) (Rushton et al., 2017). Because the P-CR junctions with the catalytic core are aligned close to the substrate entry portal, the P-CR N- and C-terminus will be the closest portion of the P-CR to this loop. The N-terminus of the P-CR is complete up to this junction, tethering the P-CR structure to a likely fixed position given the catalytic Asp residue only a few residues prior to the P-CR N-terminus. The C-terminus lacks a large portion predicted to be disordered, at least without any local structural context provided by the area around the P-CR junctions with the catalytic core. Still, assuming the potential for the P-CR to make interactions with the BCSA gating loop homologous sequences in CESA, the CatD SAXS density best fit model places Fpocket pocket P-CR 2, near the P-CR N- and C-terminus, at this location (Figure 4.11). This alignment shows that because the loop is elongated, it has a limited number of potential interactions, only packing against a few small aliphatic residues in the P-CR. Of these, the most prominent and closely packed residue, Val471, was chosen to potentially interfere with the hypothetical interaction of the P-CR with the BCSA gating loop homologous sequences in CESA, with an Ala mutant likely lacking the significant shape and hydrophobicity to replace the Val (Figure 4.19). It is possible the P-CR plays a role in coordinating the BCSA gating loop homologous sequences in CESA, perhaps in a way that at least stabilizes it like the PilZ domain and possibly even functions to alter substrate entry like in BCSA.

4.3.6 Designing P-CR Mutants to Test P-CR Interactions Based on These Models

Predicting the impact of specific mutations is often difficult because protein functions often involve several residues. Although it is possible to predict how a residue impacts local features, it is difficult to determine if a mutant phenotype is the result of the loss of a local or small portion of the P-CR or a more generalized function of the entire domain or CESA.

One approach is to consider the mutants for each cavity (Cav) in the P-CR and their impact on structure and hypothetical functions. Starting the analysis with Cav 1, the bend in the coiled-coil α -helix HX1 created by Pro418 appears to allow this cavity to form by bending the helices apart at this point. An alanine mutant is hypothesized to restrict this cavity by maintaining a normal α -helical configuration, potentially bringing the helices closer together. Cav 1 extends through to

both sides of the P-CR indicating there are no residues that would clash in a way that would prevent the cavity from collapsing if HX1 is straightened and produces a tighter packed coiled-coil. It is possible, however, that the hydrophobic core where the connector loop folds back and interacts with the coiled-coil near the Pro418 bend is so densely packed that it forces even a straightened helix to bend away. If disturbing this bend, regardless of whether the cavity is filled, alters the functionality of the P-CR it could be difficult to establish the reason until more structures can be solved. Modifying the Arg residues in and around Cav 1 could also produce a functional effect. An Arg453Ala mutant could increase the depth of Cav 1 or simply remove atoms necessary to a molecular contact in this cavity. Arg454 and/or Arg458 alanine mutants could alter the potential to coordinate contacts by changing the surface charge and shape of the Prot 3 protrusion near Cav 1. However, both Arg454 and Arg458 form important contacts that coordinate the trimeric three-fold axis using Cav 2 in the models proposed in Chapter 3 (Figure 4.16).

While a potential interaction could occur through Cav 1 between CESAs in the CSC, a potentially more attractive model is one where Cav 1 is responsible for CESA-interacting protein interactions. This could allow the P-CR to fulfil multiple functions through an unknown number of direct interactions with CESA-interacting proteins. In this model, Cav 2 or Cav 3 could make interactions within the CSC and Cav 1 could interact with CESA-interacting proteins. Thus, a Pro418Ala mutant that potentially collapses Cav 1 by removing the break and bend in the P-CR coiled-coil helix could eliminate this hypothetical CESA-interacting protein interaction site. An Arg453Ala mutant could also remove a large portion of the internal packing surface of Cav 1, potential for Van der Waals interactions and charge, or simply change the shape of the surface. Finally, Arg454 or Arg458 alanine mutants could affect important coordination sites and/or a molecular packing surface for an interaction. Reducing their size and removing their charge could weaken or disfavor an interaction. If all or some of these mutants altered a binding site for a CESA-interacting protein they might show phenotypes similar to mutants in the CESA-interacting proteins.

Although Cav 3 is a less likely interaction site due to its weaker significance, likely flexible nature and inability to co-exist with hypothetical trimeric or tetrameric interactions, it could be a CESA-interacting protein interaction site. Asp439 and Tyr440 mutants could alter the charge and packing surface of this long narrow cavity but are more significant to the structure of Cav 2 than Cav 3. Thus, Asp439 and Tyr440 alanine mutants are more likely to be interpreted as interrupting

a Cav 2 function. If, however, other Cav 2 mutants are not critical to CESA function, but Asp439 and Tyr440 are, then more mutants around Cav 3 could be justified in future experiments. A Trp432Ala mutant is interesting in the context of potential Cav 3 functions as it is a large protrusion at the end of the cavity, potentially critical for coordinating interactions. The very nature of this exposed hydrophobic residue, however, makes it interesting in multiple contexts including potential interactions with the catalytic core, as predicted by the P-CRs alignment to the catalytic core in the CatD SAXS density (Rushton et al., 2017). Glu431 might also be tied to Trp432 interactions but in the context of Cav 3 it is facing the opposite side. A Glu431Ala mutant could also interfere with the same coordination depending on whether the interaction is centered on Cav 3 or Trp432.

For Cav 2, discerning between potential trimers, based on crystallographic interactions, and VGIC-like tetrameric interactions, based on structural similarity to protein domains with known functions, is difficult. For both models, the mutants around Cav 2 are predicted to alter CSC particle oligomerization. The most prominent and odd feature of the trimer, not necessary for the tetramer, is the ionically coordinated three-fold symmetrical Arg458 interaction between three P-CRs (Figure 2.13 and 4.16C,D) (Rushton et al., 2017). The positive charges of these side chains necessitate at least one ion sitting on the three-fold axis at this tight junction that could be a phosphate, sulfate or phospholipid head group *in vivo*. In the crystal structure this was needed to prevent this interaction from becoming potentially repulsive. The well-defined density of the outward stretched Arg458 is supported by Arg454 that forms π - π stacking interactions with Arg458 that appear to stabilize it. Arg454 is a known cellulose-deficient mutant when mutated to Lys in the *Atcesa8* isoform *fra6* mutant (Zhong et al., 2003) Using the trimeric P-CR model could explain this mutation as destabilizing the outward-stretched Arg458. This could destabilize the trimeric interaction locally or destabilize the entire P-CR trimer, potentially leading to weaker or absent CSC particle trimers impacting the entire CSC. An Arg458Ala mutant alters the more prominent trimeric interaction here and would therefore be expected to show a similar phenotype to the *fra6* mutant in the *Atcesa8* isoform. Given the orientation of these Arg residues in the tetrameric model it is possible that they could form specific interactions with the neighboring P-CR and so also be disruptive to the CSC particle (Figure 4.13C).

In the P-CR trimer, another pair of residues appears to mediate contacts between monomers on the face where the coiled-coils interact with the connector loop. This contact interface (Figure

4.16A,B) spans a larger surface than the primarily Arg454 and Arg458 driven interactions on the other face of the trimer (Figure 4.16C,D). Here, Asp439 appears to form hydrogen bonds between neighboring P-CR molecules that surround the somewhat exposed hydrophobic core of the P-CR. The most exposed hydrophobic residue, Tyr440, forms both a hydrophilic and hydrophobic contact between P-CRs, indicating it could contribute to binding energies in multiple ways. However, the packing of neighboring molecules is less dense than expected and is mediated primarily by hydrophilic contacts, creating the P-CR 1 pocket that fills this space. Thus, the Asp439 and Tyr440 residues appear to be the most prominent and potentially critical residues at the center of the trimeric interaction on this side of the trimeric P-CR model. An Asp439 and Tyr440 double mutant, for example, could illicit the greatest effect, while only altering a small surface within the P-CR that is in a loop likely not to be structurally critical to the P-CR coiled-coil structure.

In the P-CR VGIC-like tetrameric model, Asp439 and Tyr440 are equally likely to be in a critical interaction surface as in the trimer, as they are centered on the contact between neighboring molecules (Figure 4.13C,D). Here hydrophobic packing appears to be most significant, with Glu431 and Trp432 of the neighboring connector loop now apparently inserting into Cav 2. The implication is that Trp432 and Tyr440 of opposing molecules form a strong hydrophobic interaction, with Glu431 and Asp439 of opposing molecules available to form ionic coordination with other nearby residues in the other molecule, respectively (Figure 4.13C). The tetrameric model presents a unique hypothesis about how Glu431 and Trp432 could interact with neighboring P-CRs. If Glu431 and/or Trp432 are critical for CSC particle formation, mutations in them could distinguish the trimeric and tetrameric models.

In conclusion, the mutants designed in Chapter 5, based on the P-CR structure and hypothetical models of its function, will be informative but alone might not be sufficient to discriminate between models. In particular, assigning specific roles to the Glu431 and Trp432 alanine mutants will be difficult given all of the potential contacts. Slightly more restricted to hypotheses that center on Cav 2, the Asp439 and Tyr440 alanine mutants can best be predicted to be critical for fundamental trimeric or tetrameric interactions within the CSC and could destabilize the CSC in either case. Similarly, in either model Arg453 forms back of Cav 2, a major occlusion to the rest of the hydrophobic core. Interfering with this could destabilize Cav 2 by opening it up further or by removing a critical ionic coordination in an interaction. Interpreting the Arg453Ala mutant alone might be similar to Asp439 and Tyr440 alanine mutants in terms of destabilizing the

CSC, but with subtle differences in the fact that Asp439 and Tyr440 are more exterior to Cav 2 while Arg453 is the most interior. Arg453 also forms a major portion of Cav 1, while Asp439 and Tyr440 are also surrounding Cav 3, which could have potentially different functional effects. Thus, correlating mutant phenotypes to known functions in CESA biology will be critical to determining the function(s) of residues and cavities.

Experiments designed to probe the significance of these mutants in CESA and CSC functions will narrow down the important residues in the P-CR structure and potentially functionally important regions, such as specific cavities. This can enable additional hypotheses to probe specific functions in CSC formation and/or CESA-interacting protein interactions. Through this it might be possible to create a more complete model of the CSC by describing its fundamental interactions, potentially distinguishing 18 vs 24 CESA CSC models and identifying the CESA-interacting protein interaction sites that are not yet well understood.

CHAPTER 5. ASSESSING THE IMPORTANCE OF SPECIFIC RESIDUES IN THE CELLULOSE SYNTHASE PLANT-CONSERVED REGION USING COMPLEMENTATION OF THE *ARABIDOPSIS THALIANA* TEMPERATURE-SENSITIVE *ATCESA1 RSW1-1* MUTANT BACKGROUND

5.1 Introduction

In Chapters 2 and 4, I evaluated the OsCESA8 P-CR crystal structure and its fit to the CatD SAXS density with the catalytic core homology model. In so doing, I identified several amino acid residues whose mutations might be predicted to disrupt the structure and function of the P-CR relative to the catalytic core in various CSC models. The OsCESA8 P-CR sequences were aligned with the AtCESA1 sequence to determine the cognate amino acids for mutation. To assess empirically the importance of these residues *in vivo*, I used the *radial swelling1-1* (*rswl-1*) mutant of *Atcesa1* in *Arabidopsis thaliana*, a temperature-sensitive mutant (Baskin et al., 1992; Arioli et al., 1998; Peng et al., 2000; Williamson et al., 2001; J. Wang et al., 2008) as a testbed for genetic complementation. In *rswl-1*, *Atcesa1* is functional at or below the permissive temperature (23°C) but ceases cellulose production at the restrictive temperature (31°C), correlating with loss of CSCs from the plasma membrane and severe swelling of the radicle tip (Arioli et al., 1998). Genetic complementation with *AtCESA1* restores WT cellulose synthesis and normal root-tip elongation at the restrictive temperature. Thus, genetic complementation of *rswl-1* with specifically modified *Atcesa1* constructs can simultaneously detect dominant-negative effects of the mutations at permissive temperatures, and the ability of the construct to complement at the restrictive temperature.

The OsCESA8 P-CR is an anti-parallel coiled-coil with an uncommonly high proportion of large, aromatic residues in knob-in-hole (KIH) positions and a 25-residue connector loop that bends back towards the α -helices, forming several hydrophobic contacts (Figure 2.2A) (Rushton et al., 2017). The P-CR starts with the first coiled-coil helix (HX1), spanning ~31 Å over residues 403-423 (Figure 2.2B). There is a bend in HX1 where the helix is broken by Pro418 (Figure 2.2A,B). The second coiled-coil helix (HX2) runs anti-parallel to HX1, spanning ~35 Å over residues 449-473 (Figure 2.2B). The hydrophobic core that stabilizes the structure of the P-CR coiled-coil is unusual for its seven stacked aromatic residues that force the HX1 and HX2 slightly

apart (Figure 2.2B and 2.3). The 25-amino acid loop region (LP) spanning residues 424-448, connecting the HX1 and HX2 helices, contains a small α -helix between residues 431-435 (Figure 2.2A,C). The best fit of the P-CR to the CatD indicates that the P-CR could interact with the catalytic core using the small α -helix on the LP (Figure 2.4 and 5.3) (Olek et al., 2014; Rushton et al., 2017). This interaction could have implications for the nearby substrate entry portal and FxVTxK motif containing loop on the interfacial former TMH 5, assuming this loop is similar to the BCSA gating loop (Slabaugh et al., 2016; Rushton et al., 2017). When compared to other structures in the PDB, the similarity to the voltage-gated ion channel (VGIC) pore-forming domain presents a new model for how CSC particles could interact and highlights potentially important surfaces on the P-CR. An *in silico* analysis of the P-CR surface suggested the importance of the cavities, protrusions and pockets found in and around hypothetical and crystallographic interactions.

From these analyses, and identification of conserved residues in the P-CR sequence, potentially critical residues were identified. The potentially important residues in the P-CR were selected for mutation (Table 5.1). Six single mutants and one double mutant were designed in the homologous residues of the AtCESA1 P-CR (Pro417Ala, Glu430Ala, Phe431Ala, Arg452Ala, Arg457Ala, Val470Ala and Asp438Ala/Tyr439Ala). An eighth mutant was designed in the BCSA homologous IF2 helix of the catalytic core near the potential interaction site with the P-CR (Leu812Arg), and a ninth mutant was created to correspond with the known *Atcesa8 fra6* mutant in the P-CR (Arg453Lys) (Zhong et al., 2003). The high sequence conservation of the P-CR, 86% sequence identity and 96% sequence similarity, and an equal number of amino acids (Altschul et al., 1997), produced a highly confident homology model of the AtCESA1 P-CR that indicates the hypothetically important residues selected should be consistent between the OsCESA8 and AtCESA1 sequences. However, insertions and deletions (indels) upstream of the P-CR in both sequences results in the AtCESA1 sequence being one residue offset in its alignment with the OsCESA8 sequence (Figure 1.2).

Table 5.1 Residues selected for mutation in the P-CR

| OsCESA8 <i>Atcesa1</i> | SSE | P-CR Surface | Structural Property in the P-CR | Hypothetical Importance | Results |
|-----------------------------------|---|------------------------------------|---|--|---|
| Pro418 Pro417Ala | HX1 | Cav 1 | Breaks α -helix hydrogen bonds at the bend in HX1 | [Interactions within the P-CR] | Does not complement <i>rswl-1</i> |
| Glu431 Glu430Ala | LP small α -helix | Cav 2 P-CR 8 | Primary surface feature of the LP small α -helix | [Interactions within the P-CR] [P-CR interactions with the catalytic core] [Interactions in the CSC particle] | Complements <i>rswl-1</i> , but with slight growth inhibition |
| Trp432 Phe431Ala | LP small α -helix | Prot 2 Cav 3 P-CR 8 | Primary surface feature of the LP small α -helix; solvent exposed in P-CR crystal | [P-CR interactions with the catalytic core] [Interactions in the CSC particle] | Complements <i>rswl-1</i> , but with slight growth inhibition |
| Asp439 Asp438Ala/ Tyr439Ala | LP coil near HX2 | Cav 2 Cav 3 P-CR 1 | Solvent exposed in P-CR crystal | [Interactions in the CSC particle] | Does not complement <i>rswl-1</i> |
| Tyr440 Tyr439Ala/ Asp438Ala | LP coil near HX2 | Cav 2 Cav 3 P-CR 1 P-CR 4 | Stacks against Arg453 at Cav 2 inside the P-CR hydrophobic core | [Interactions within the P-CR] [Interactions in the CSC particle] | Does not complement <i>rswl-1</i> |
| Arg453 Arg452Ala | HX2 | Cav 1 Cav 2 P-CR 1 P-CR 4 | Stacks against Tyr440 at Cav 2 inside the P-CR hydrophobic core | [Interactions within the P-CR] [Interactions in the CSC particle] | Complements <i>rswl-1</i> , but with slight growth inhibition |
| Arg454 Arg453Lys | HX2 | Cav 2 Prot 3 P-CR 1 | Aligned to <i>fra6</i> mutant of <i>Atcesa8</i> . Stacks with Arg457 in the P-CR crystal | [Interactions in the CSC particle] | Does not complement <i>rswl-1</i> |
| Arg458 Arg457Ala | HX2 | Prot 3 P-CR 1 P-CR 9 | Stacks with Arg454 (aligned to <i>fra6</i> mutant of <i>Atcesa8</i>) in the P-CR crystal | [Interactions in the CSC particle] | Complements <i>rswl-1</i> |
| Val471 Val470Ala | HX2 | Prot 1 P-CR 2 | Most significant protrusion near the modeled gating loop containing the FxVTxK motif (based on BCSA) | [P-CR interactions with the catalytic core] | Complements <i>rswl-1</i> |
| Leu813 Leu812Arg | IF2 α -helix in the catalytic core | N/A | Potentially interacting with the P-CR Trp432 in the combined model of the P-CR and CESA catalytic core with CatD SAXS density | [P-CR interactions with the catalytic core] | Complements <i>rswl-1</i> root swelling, but inhibits growth at restrictive and permissive temperatures |

*AtCESA1 sequences are one residue offset due to indels upstream of the P-CR and Leu813.

An *AtCESA1* promoter and complementary DNA (cDNA) sequence in an *Agrobacterium* vector (pFGC5941) was mutated to generate nine sequence variant vectors, including *fra6*, which were transformed into the *Arabidopsis thaliana* *rsw1-1* mutant background and a WT *AtCESA1* positive control (*rsw1-1/AtCESA1*). Several independent lines were isolated for all ten different plasmids by selection for transgenic herbicide resistance and then bred to homozygosity in the T3 generation. Transcript levels of the transgenes were measured in between five and seven independent lines of the Glu430Ala, Asp438Ala/Tyr439Ala, Arg452Ala, Leu812Arg homozygous mutants and a single *rsw1-1/AtCESA1* independent line using real-time quantitative reverse-transcription PCR (RT-qPCR). All lines showed variable *Atcesa1* transgene expression levels near the levels of a CAP-BINDING PROTEIN 20 (*CBP20*) reference gene control (Czechowski et al., 2005), as expected for transgenic plasmids containing a native *AtCESA1* promoter in seedlings active in PCW growth (Michael et al., 2008). Mutant lines representing a range of expression levels were tested for their ability to complement the *rsw1-1* temperature-sensitive root-growth and radially swollen root tip phenotype at 21°C and 31°C. Some *Atcesa1* mutants did not complement *rsw1-1* (Pro417Ala, Asp438Ala/Tyr439Ala and Arg453Lys/*fra6*), others completely restored the WT phenotype (Arg457Ala and Val470Ala) and a few complemented the radially swollen root tip phenotype of *rsw1-1* but had slight (Phe431Ala, Arg452Al and Glu430Ala) or severe (Leu812Arg) growth inhibition.

5.2 Materials and Methods

5.2.1 Site-Directed Mutagenesis and *Arabidopsis thaliana* transformation

An *Agrobacterium* pFGC5941 binary vector (TAIR: CD3-447) containing the *Arabidopsis thaliana* *AtCESA1* coding sequence driven by its native promoter, created in previous work (Slabaugh et al., 2014b), was used for these experiments. Primer design and site-directed mutagenesis was done using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs, Ipswich, Massachusetts) and NEBaseChanger algorithm (<http://nebasechanger.neb.com/>) for primer design (Table 5.2). Vectors were transformed into *Escherichia coli* (DH5α) cells using heat shock and positive bacterial transformants were selected with 50 µg mL⁻¹ kanamycin. Following sequence confirmation by sequencing and analytical restriction enzyme digest, vectors were transformed into *Agrobacterium tumefaciens* strain GV3101 by electroporation. Positively

transformed GV3101 were selected with 50 $\mu\text{g mL}^{-1}$ kanamycin, 35 $\mu\text{g mL}^{-1}$ gentamycin and 10 $\mu\text{g mL}^{-1}$ rifampicin. Seeds of the *rswl-1* mutant were obtained from TAIR (<https://www.arabidopsis.org/>) (TAIR: CS6554) and transformed using the floral dip method (Clough et al., 1998; X. Zhang et al., 2006).

Table 5.2 Primers used to generate *Atcesa1* mutations and measure expression by RT-qPCR.

| Name | F/R | Sequence (5' to 3') |
|-----------------------|-----|----------------------------------|
| Mutagenic Primers | | |
| AtCesA1_D438A/Y439A | F | GGCCCCTGAAgctTATTTTGCCC |
| | R | CTAGGTTCAATGTTGAATTTC |
| AtCesA1_E430A | F | AGGGCCCCTGcATTCTATTTTG |
| | R | AGGTTCAATGTTGAATTTCTTGC |
| AtCesA1_L812R | F | ACCAATCAATagaTCTGATCGTTTGAACCAAG |
| | R | GCAGATCCCTTGAACGCA |
| AtCesA1_R452A | F | TGTAAAGAGgacACGAGCTATGAAG |
| | R | AAAGACGGTTGGATCTTG |
| AtCesA1_P417A | F | GAAATGGGTAgCATTTTGCAAG |
| | R | TTTGCAAACCTCAGCGGTTTC |
| AtCesA1_F431A | F | GGCCCCTGAAgctTATTTTGCCC |
| | R | CTAGGTTCAATGTTGAATTTC |
| AtCesA1_R453K | F | TAAAGAGCGAaagGCTATGAAGAGAG |
| | R | CAAAAGACGGTTGGATC |
| AtCesA1_R457A | F | AGCTATGAAGgcaGAGTATGAAGAG |
| | R | CGTCGCTCTTTAACAAAAG |
| AtCesA1_V470A | F | AAATGCTCTTgcaGCCAAAGCAC |
| | R | ATCCTCACTTTAAACTCTTC |
| qPCR Primers | | |
| AtCesA1_pFGC5941_qPCR | F | CGTTTTATCTTAATTGGCTCTTCATTG |
| | R | TCCCCTCTTTCTACCTTCCC |
| AtCBP20_qPCR | F | ATTACGGTACTGGCTCATTGG |
| | R | TCGTCTCTTTGTCGTTTCCG |

5.2.2 Growth Conditions

Arabidopsis thaliana Columbia (Col-0), *rswl-1*, and transgenic seeds were surface-sterilized by mixing in 10% Clorox bleach (6.25% NaClO) and 0.05% Tween 20 solution with seeds for 10 min followed by 5X washes with a >10:1 volume of sterile water. The final wash was decanted and

seeds cold-treated in the dark at 4°C for at least two days before planting or plating. Seeds were sown on soil and grown under standard *Arabidopsis thaliana* germination and growth conditions in growth rooms and greenhouses maintained near 22°C. Healthy transgenic seedlings (~2 weeks post-sowing) were selected for the *bar* gene on the pFGC5941 vector by spraying with 20 mg L⁻¹ glufosinate-ammonium. For gene expression analysis and complementation experiments, individual seeds were sown in a grid pattern on agar plates containing autoclaved 0.5X Murashige and Skoog (MS) media or 2.15 mg L⁻¹, 2% sucrose, 0.5 g L⁻¹ MES, 8 g L⁻¹ agar at pH 5.7. For gene expression assays, plates were placed vertically in a growth room for seven days at 22°C under a 16/8 hour (h) light/dark photoperiod. For complementation experiments, plants were grown vertically for four days at 22°C and three days at 31°C with ~60% humidity, under a 16/8 h light/dark photoperiod at 150 μmol m⁻² s⁻¹ light intensity in a growth chamber. Independent lines carrying a transgene were self-crossed to homozygosity by the T3 generation (Table 5.3). For additional experiments, seeds were surface sterilized with 70% (v/v) ethanol in water, then 10% Clorox bleach (6.25% NaClO), and rinsed with several changes of sterile water. They were chilled at 4°C for three days. Seeds were grown on plates with 0.5X Murashige and Skoog (MS) media or 2.15 mg L⁻¹, 1% sucrose and 1% agar.

Table 5.3 Seed stocks for *Atcesal* and P-CR mutants in the *Arabidopsis thaliana* *rswl-1* background. Independent lines that produced homozygous individuals are numbered from random selections of the 1:2:1 segregating population. T3 generation seedlings were assigned a decimal number and seed stocks from homozygous individuals are listed corresponding with the individual plant used to isolate those seeds (separated by a /). Heterozygotes are not listed.

| Homozygous Independent Lines | | | | |
|-------------------------------------|--------------------|----------------------------|--------------------|--------------|
| Set 1 | | | | |
| <i>rswl-1/AtCESA1</i> | D438A/Y439A | E430A | L812R | R452A |
| 2.2 | 2.5/.6/.8 | 1.2/.3/.5/.6/.7 | 1.2/.5 | 1.7 |
| 7.4/.5/.8 | 3.1 | 2.1/.3/.6 | 2.1/.5 | 2.1/.5 |
| 8.4/.7/.8 | 4.5/.7/.8 | 3.7/.8 | 3.2/.3 | 3.7 |
| 9.1/.6/.8 | 5.4/.5 | 4.2 | 6.4/.7 | 4.1 |
| 10.1/.6/.8 | 6.8 | 5.1/.2/.3/.8 | 7.8 | 5.2/.3/.8 |
| | 7.4/.6 | 6.3 | 9.2/.6 | 7.8 |
| | 9.3/.4/.6/.7 | 7.5 | 10.1/.2/.4 | 8.4 |
| | 10.3/.4 | 8.1/.4/.6 | 12.1/.6 | 9.2/.3 |
| | 11.1/.2/.3/.6 | | | 12.6 |
| | 12.4/.6 | | | 13.3/.7 |
| | 13.3/.5/.8 | | | 15.1/.5/.7 |
| | 14.2/.6 | | | |
| 5 | 12 | 8 | 8 | 11 |
| Set 2 | | | | |
| P417A | F431A | R453K (<i>fra6</i>) | R457A | V470A |
| 1.4 | 1.5 | 1.1/.2/.6 | 1.3/.7 | 1.1/.3/.5/.6 |
| 3.2/.7 | 3.1/.3/.5/.7/.8 | 3.5 | 2.1/.2/.4/.6/.7/.8 | 2.2/.4/.6/.7 |
| 4.3/.6 | 4.1 | 5.1/.3/.4/.5/.7/.8 | 3.1 | 4.1 |
| 5.1/.6/.7 | 5.5/.6 | 7.1/.8 | 5.3/.5 | 5.3/.5/.8 |
| 6.4/.6 | 6.2/.4/.8 | 9.2/.8 | 6.2/.3/.4/.7 | 6.2/.5 |
| 9.1 | 7.2/.3/.6/.7 | 10.3/.8 | 8.1/.4 | 7.3/.5/.6/.8 |
| 10.1/.2/.5/.6/.8 | 8.7 | | 9.5/.6/.7 | 8.2/.3/.4/.8 |
| | 9.3/.5/.6/.7 | | | 10.7/.8 |
| 7 | 8 | 6 | 7 | 8 |

5.2.3 Gene Expression Analysis

Thirty seeds of Col-0, *rswl-1*, and a selection of independent lines representing a range of expression levels for each mutant, were sown on MS plates to create a single biological replicate. This was iterated to create a total of three biological replicates. Seven-day-old seedlings of each plate were harvested a few hours into the light phase of the light/dark cycle, flash-frozen using

liquid nitrogen and homogenized with steel beads. RNA was extracted using the TRIzol™ Reagent (Invitrogen) protocol and DNase-treated with the TURBO DNA-free™ kit (Invitrogen). cDNA was synthesized from RNA templates using the iScript cDNA synthesis kit (Bio-Rad). RT-qPCR was done using custom primers and SYBR™ Green PCR Master Mix (Applied Biosystems) with a StepOnePlus™ Real-Time PCR system (Applied Biosystems) (Table 5.2). For each of the three biological replicates, C_T values were measured and averaged for three technical replicates. Expression was averaged across the biological replicates as a fold change ratio compared to the *CBP20* (TAIR: AT5G44200) reference control gene (Czechowski et al., 2005), based on experimental and previously published data in the Arabidopsis eFP Browser (Winter et al., 2007; Michael et al., 2008). No significant amplification occurred using the RT-qPCR primers in the Col-0 or *rsw1-1* backgrounds.

5.2.4 *rsw1-1* Phenotyping and Complementation

Six seeds of Col-0, *rsw1-1*, *rsw1-1/AtCESA1 2.2* and one independent mutant line were sown on MS plates in a grid pattern for three of the Glu430Ala, Asp438Ala/Ty439Ala and Arg452Ala independent lines, and two of the Leu812Arg independent lines. This was repeated four times, rotating the position of each line to control for positional variance on the plate, creating a total of 24 samples for each line. In every experiment, plates were transferred to the growth chamber at the same time of day and light/dark cycle of the growth chamber. After three days of growth at 22°C, seedlings were measured on day three. Daily images of the plates at the same time of day were captured by scanning (EPSON PERFECTION V750 PRO) them through day seven. On day four, the growth chamber temperature was increased from 22°C to 31°C. Primary root length was measured from the scanned images by tracing with a segmented line set to scale using ImageJ software (<https://imagej.nih.gov/ij/>) (Schneider et al., 2012). Variance due to position on the grid was much smaller than the variance observed between plates. Thus, all samples (across four plates) were averaged and compared (n=13-24).

Additional experiments used three plates representing each mutation and on each plate twelve seeds were planted: four seeds of the pFGC5941 vector in Col-0, four seeds of a mutant and four seeds of *rsw1-1*. Plants were grown in a 16/8 h light/dark photoperiod at a temperature of 21°C for almost five days (112 h). At this time, ends of the roots were marked on the plates (time 0); one plate representing each mutation was left at 21°C, and two plates representing each

mutation were moved to 31°C (16/8 h light/dark photoperiod). After 8 h, one plate from 21°C and one plate from 31°C were scanned (EPSON PERFECTION V750 PRO). Micrographs of root tips were taken by OMAX camera connected to the Nikon SMZ1500 microscope with 2x magnification. Lengths of the roots were measured from the scans by computer program ImageJ2. The plate from 21°C was then returned to 21°C. After 24 h from time 0, the plate from 21°C and a second plate from 31°C were removed, scanned, photographed, and roots measured as before. Three lines of each mutation were screened, except for Leu812Arg for which two lines were screened. For Asp438Ala/Tyr439Ala, Glu430Ala, Arg452Ala, and Leu812Arg mutations, the same lines used in the initial experiments were re-evaluated under these conditions. For Pro417Ala, Phe431Ala, Arg453Lys, Arg457Ala, and Val470Ala mutations, three lines of each were chosen randomly.

5.3 Results

5.3.1 Sequence Conservation of the P-CR Across the Plant Kingdom

The P-CR is a highly conserved domain within CESA compared to the class-based sequence divergence of the CSR (Vergara et al., 2001; Penning et al., 2009; Carroll et al., 2011; Kumar et al., 2017). Although there are no insertions or deletions in the P-CR, several substitutions exist across the plant kingdom that might indicate residue properties less critical to a conserved function. Full-length CESA sequences were obtained from the KEGG orthology databases for CESAs (KO ID: K10999), totaling 38 sequences for each of the six classes, with a limit of four sequences from any one family to limit over-representation (Kanehisa et al., 2016) (Figure 5.1). Isoform numbering is independent of class definition and often varies between species. P-CRs were isolated and sorted into one of the six classes named for their corresponding *Arabidopsis thaliana* primary (AtCESA1, 3 and 6) and secondary cell wall CESAs (AtCESA4, 7, 8) isoform based on sequence alignments, using the Clustal Omega Hidden Markov Model multiple sequence alignment algorithm available through the Clustal Omega server (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) (Sievers et al., 2011) (Appendix 1.2 Supplemental Figures 1-6). Other more distant species were also included in this analysis. *Selaginella moellendorffii* (fern) and *Physcomitrella patens* (moss) are ancient plant lineages that synthesize cellulose microfibrils with CSCs, but do not possess the same sequence diversity or class-based architecture of their CESAs compared to seed plants (Roberts et al., 2007;

Harholt et al., 2012; Tran et al., 2016). A less complex structure exists in ferns and mosses indicating a precursor to the specialization that appears in seed plants (Norris et al., 2017). Furthermore, PpCESA5 is capable of *in vitro* cellulose synthesis using a single isoform (Cho et al., 2015, 2017), as is the single hybrid aspen (*Populus tremula x tremuloides*) PttCESA8 isoform (Purushotham et al., 2016). They have similar CESA sequence structures and all of the plant-specific regions, however, they form their own clade in phylogenetic analyses (Kumar et al., 2017). Thus, in this analysis, they were grouped separately (Appendix 1.2 Supplemental Figure 7).

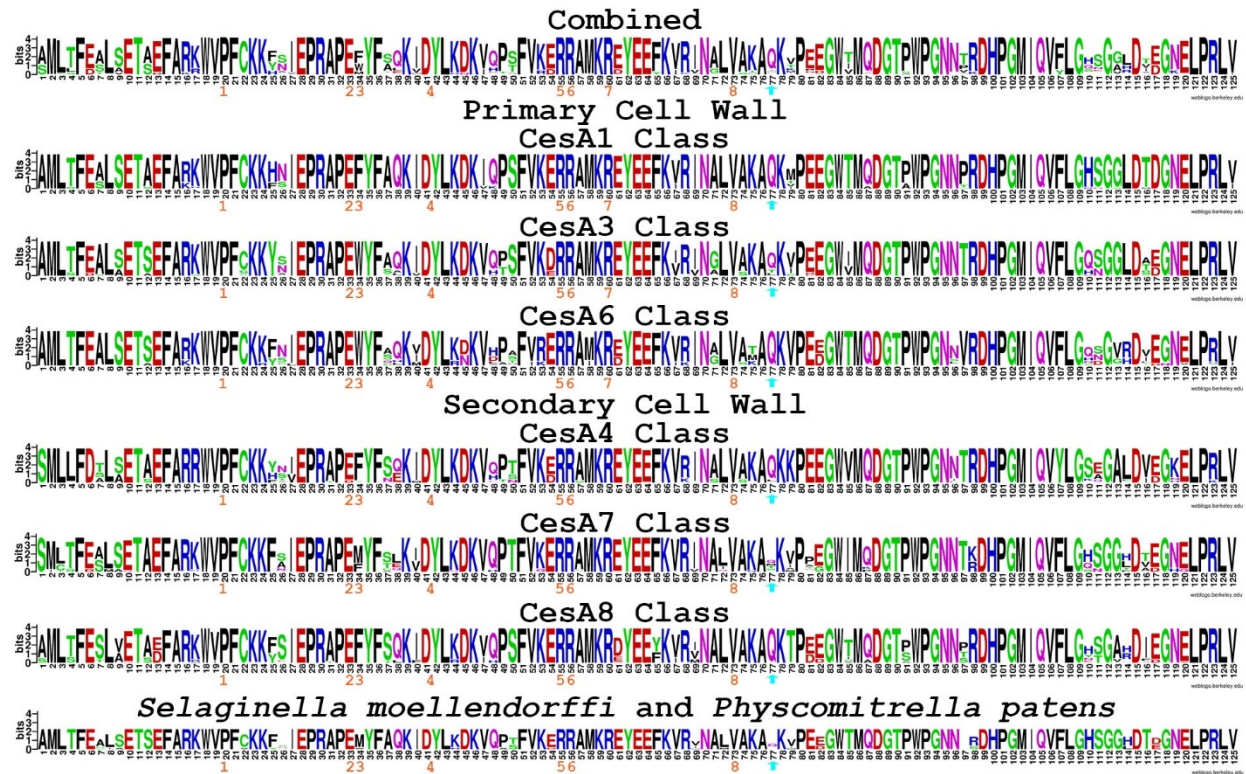


Figure 5.1 Clustal Omega HHM multiple sequence alignments of CESA P-CRs (Sievers et al., 2011) (Appendix Figures 1-7), divided into classes and represented as sequence logos using the WebLogo algorithm (Crooks et al., 2004). Letter height and potentially multiple letters illustrate variations in the sequence proportional to the number of sequences analyzed. Residues are colored based on their chemical properties defined as hydrophobic (black, A, V, L, I, P, W, F, M) polar (green, G, S, T, Y, C, Q, N), acidic (red, D, E) or basic (blue, K, R, H). P-CR residues targeted for mutation are numbered in orange. The limit of the P-CR X-ray crystal structure is indicated by a cyan arrow.

Sequence alignments were transformed into sequence logos using the WebLogo algorithm and server (<https://weblogo.berkeley.edu/logo.cgi>) (Crooks et al., 2004) (Figure 5.1). Letter height represents the proportional representation of a residue in an alignment compared to a random

residue, described as bits, with number of sequences and lack of gaps contributing to higher bit value. The P-CR showed a high sequence and property conservation with only a few variable regions. Different classes had slight variations compared to each other and some were more highly conserved than others, such as the CESA6 class. The combined sequence alignment of all classes, including moss and fern, illustrates that the sequence of the P-CR is highly conserved and likely to have a very similar fold and function across isoforms.

As expected, the residues that form the hydrophobic interactions between coiled-coil helices or hydrophobic core interactions with the connector loop in the P-CR structure were well conserved across P-CR sequences. The highest sequence conservation was observed around P-CR Cav 2, previously identified as a likely interaction surface and used to build a trimer from the P-CR X-ray crystal structure trimer contact, as well as in the hypothetical tetrameric interaction based on structural similarities with voltage-gated ion channel (VGIC) pore forming domain, discussed in Chapter 3. The P-CR is also highly conserved throughout the remaining disordered C-terminal residues. The ordered portion might be more conserved to maintain structure and structural contacts, while the less ordered sequence might act primarily as a linker between substructures or domains. The C-terminus of the P-CR at the limit of the OsCESA8 P-CR structure is predicted to be a flexible, disordered random coil using secondary structure prediction algorithms. Based on the sequence conservation, this portion might play an important role in CESA biology, perhaps related to the nearby substrate entry portal and UDP-coordination site at the junction of the P-CR with the catalytic core domain modeled in Olek et al. (2014). The ordered coiled-coil structure contains several non-conserved residues, perhaps due to a general structural role enforced by the surrounding residues and a location that does not interfere with other CESA or CSC interactions. Important structural and hypothetically important surface residues appear highly conserved. The C-terminus of the P-CR might have a much more defined structure, packing or function as part of a full length CESA.

Illustrating this sequence conservation within the P-CR structure highlights similarity between these domains across species (Figure 5.2). Identical residues and high property conservation is observed in the coiled-coil and hydrophobic core interactions (orange and yellow residues), which strongly supports a similar fundamental fold. The connector loop and the prominent features identified in the P-CR structure analysis are also highly conserved, in particular around Cav 2 (Figure 4.9B,E). The fact that many of the prominent, exposed residues of the

random coil structure of the connector loop are conserved is not necessarily expected, as loops are often less mutationally sensitive if they act as soluble linkers between motifs. This might indicate an important function for these random coil residues in the CSC. This analysis might also suggest that the side of the coiled-coil not interacting with the connector loop is not as critical to function as it shows much less sequence conservation (green and cyan residues). Only one of the mutants constructed in this study did not show complete or nearly complete conservation, found at position 34 of the P-CR and equivalent to Trp432 in the OsCESA8 P-CR (Figure 5.1). This residue has conserved properties as a large, hydrophobic residue and given its prominent, solvent-exposed location and potential as a hydrophobic interaction site on the P-CR surface, it was also chosen for mutation. The residues selected for mutation and the structural features they probe are highly conserved, but they do not appear to be fundamental to the P-CR coiled-coil or hydrophobic core structures and therefore might probe the hypothetical interactions of the P-CR.

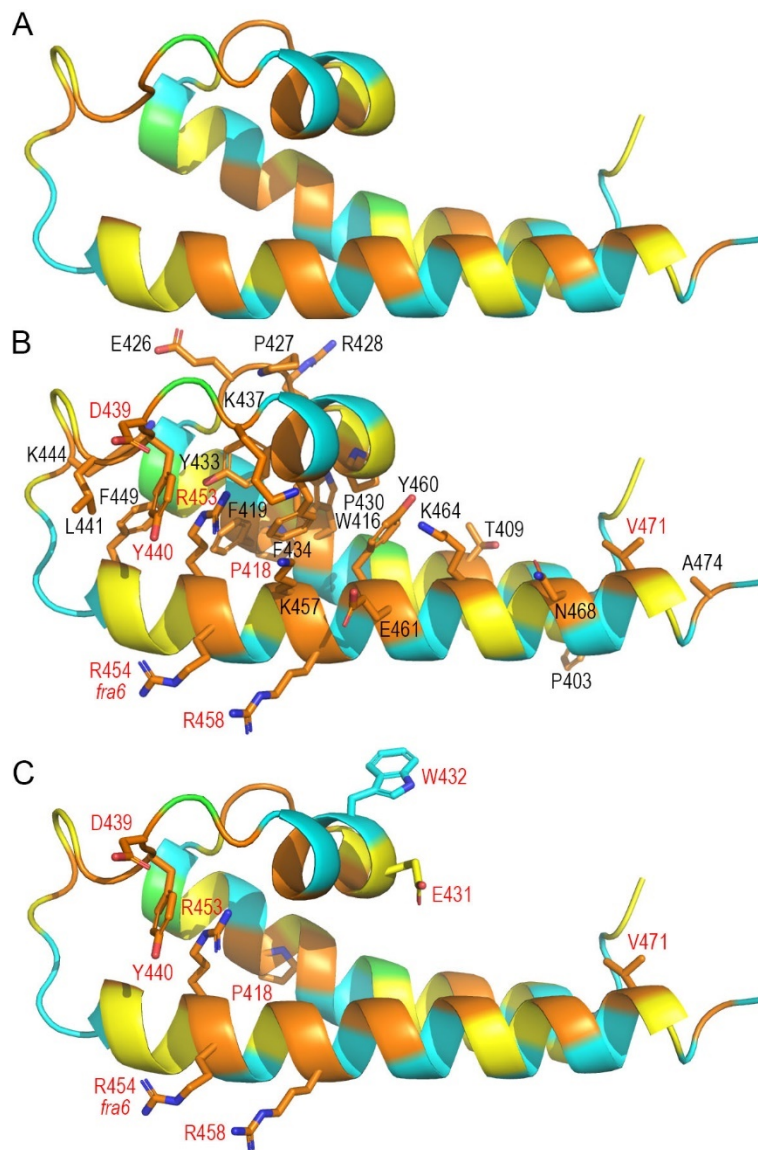


Figure 5.2 Sequence conservation of the combined Clustal Omega HHM multiple sequence alignments (Sievers et al., 2011) of the P-CR modeled on the OsCESA8 P-CR structure. (A) Residues are colored based on their sequence conservation and similarity defined by Clustal Omega as identical residues (orange), strongly similar properties (yellow, >0.5 in the Gonnet PAM 250 matrix), weakly similar (green, ≤ 0.5 and > 0 in the Gonnet PAM 250 matrix) and dissimilar (cyan, 0 in the Gonnet PAM 250 matrix). (B) Identical residues are labeled and have side chains shown as stick models, with mutations constructed in this study labeled in red. (C) Mutated residues shown as side chain stick models.

5.3.2 Probing Potential Interactions of the P-CR with the Catalytic Core

Based on the best fit model of the OsCESA8 P-CR X-ray crystal structure, SAXS density and catalytic core homology model to OsCESA8 CatD SAXS density, it was predicted that the P-CR is packed lengthwise against the catalytic core as described in Chapter 2. In this best fit model, the connector loop of the P-CR is packed between the P-CR coiled-coil and the catalytic core (Figure 2.5). Close inspection of this fit indicates that the only feature of the P-CR capable of interacting with the catalytic core is the small α -helix of the connector loop, with Glu431 and Trp432 at the apex and oriented toward the catalytic core near the end of the IF2 helix (Figure 5.3). In the best fit model, Trp432 sits up against the end of the IF2 helix where there is a highly conserved Leu in CESAs (Appendix 1.2 Supplemental Figures 1-7). Trp432, or any of the other large, hydrophobic residues conserved at this location in the P-CR, might form a strong hydrophobic interaction with Leu813 in OsCESA8. Nearby Leu817 and Arg816 also likely form a hydrophobic surface in this context that might contribute to the hypothetical molecular packing. This might explain the presence of a large, hydrophobic residue (Trp432 in OsCESA8) on the surface of the P-CR, which is unlikely to remain exposed in the context of a CESA or CSC given the cost to solubility and folding that such surface aromatics present. There is also the potential to form a hydrogen bond between Glu431 and the nearby Ser414 or backbone hydroxyls, which could act in specifically coordinating this interaction.

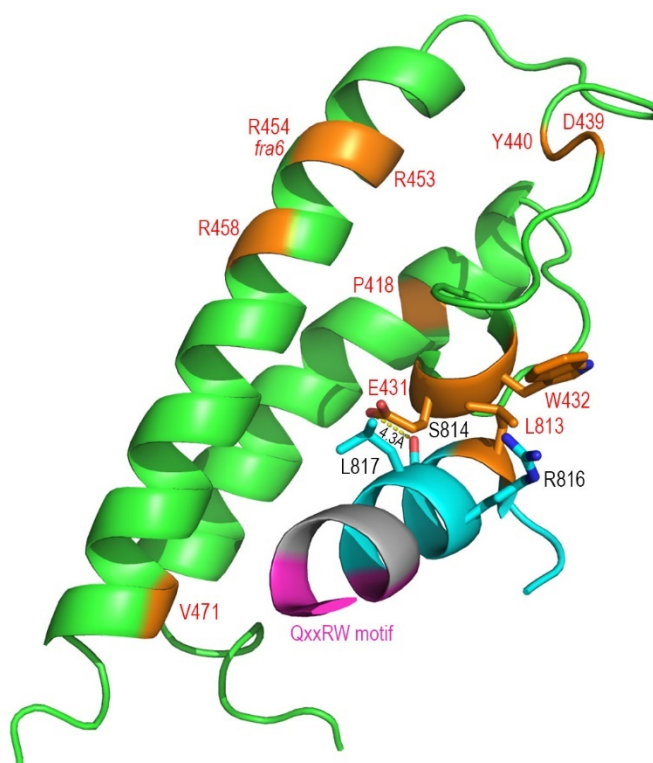


Figure 5.3 Potential contact (stick models) of the OsCESA8 P-CR (green) with the OsCESA8 catalytic core homology model (cyan) based on their best fit to OsCESA8 CatD SAXS density (Rushton et al., 2017). Trp432 and Glu431 are shown packed against the CESA IF2 helix and its Leu813, modeled on the homologous BCSEA catalytic core. P-CR mutants designed in this study are shown in orange and labeled in red. The critical catalytic motif Q/RxxRW of the homologous IF2 helix is shown in purple.

5.3.3 Homology Modeling of the *Arabidopsis thaliana* AtCESA1 Isoform P-CR using the OsCESA8 P-CR Structure.

To ensure that the AtCESA1 P-CR used in the *in vivo* complementation experiments does not contain any significant sequence differences compared to OsCESA8 for which we have a P-CR X-ray crystal structure, several AtCESA1 P-CR homology models were compared. The numbering of the AtCESA1 sequence is shifted one residue compared to OsCESA8 upstream of the CatD and P-CR leading to a shift in aligned residue numbering (Figure 1.2). The AtCESA1 P-CR sequence has 86% sequence identity (% Seq ID) and 96% sequence similarity to OsCESA8 across the P-CR sequence (Figure 5.1 and 5.4A). Homology modeling using threading-based algorithms was done using RaptorX (Kallberg et al., 2012), Phyre2 (Kelley et al., 2015), HHpred (Zimmermann et al., 2018) and MUSTER (Wu & Zhang, 2008). The AtCESA1 sequence was trimmed to the limit of

the OsCESA8 P-CR X-ray crystal structure and input into RaptorX (<http://raptorx.uchicago.edu/StructurePrediction/predict/>), Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2>), HHpred (<https://toolkit.tuebingen.mpg.de/>) and MUSTER (<https://zhanglab.ccmb.med.umich.edu/MUSTER/>) servers to create homology models based on known structures in the PDB. In all cases the AtCESA1 sequence aligned to the OsCESA8 P-CR sequence and X-ray crystal structure (PDB ID: 5jnp) as expected, and produced similar models with low C α RMSDs values (Figure 5.4B,C). Although fundamentally similar algorithms, each one has slightly different refinement optimizations that can result in different models. MUSTER stood out as the poorest model, because it altered the sequential register of the sequence structure alignment by adding an extra residue to the random coil region before the C-terminal coiled-coil helix, which is reflected in a larger C α RMSD (2.67 Å). This forces the remaining C-terminal residues to an n-1 position in the structural alignment in the P-CR, which is unrealistic in the highly conserved P-CR sequence. The three remaining models were much closer in C α RMSDs (RaptorX: 1.09 Å, Phyre2: 1.12 Å and HHpred: 1.17 Å). The only differences observed were primarily in the orientation of non-identical residues (Figure 5.4C). HHpred differed slightly in its orientation of dissimilar side chains of the residues enough that it is a poorer model for AtCESA1. RaptorX and Phyre2 produced almost identical models within the resolution of the OsCESA8 P-CR X-ray crystal structure. RaptorX produced the lowest RMSD and aligned a few dissimilar side chains slightly closer to the aligned residue in the OsCESA8 P-CR. For the purpose of modeling interactions based on the OsCESA8 P-CR, without any *a priori* reason to assume differences, the RaptorX model was chosen to represent AtCESA1 when designing mutants (Figure 5.5).

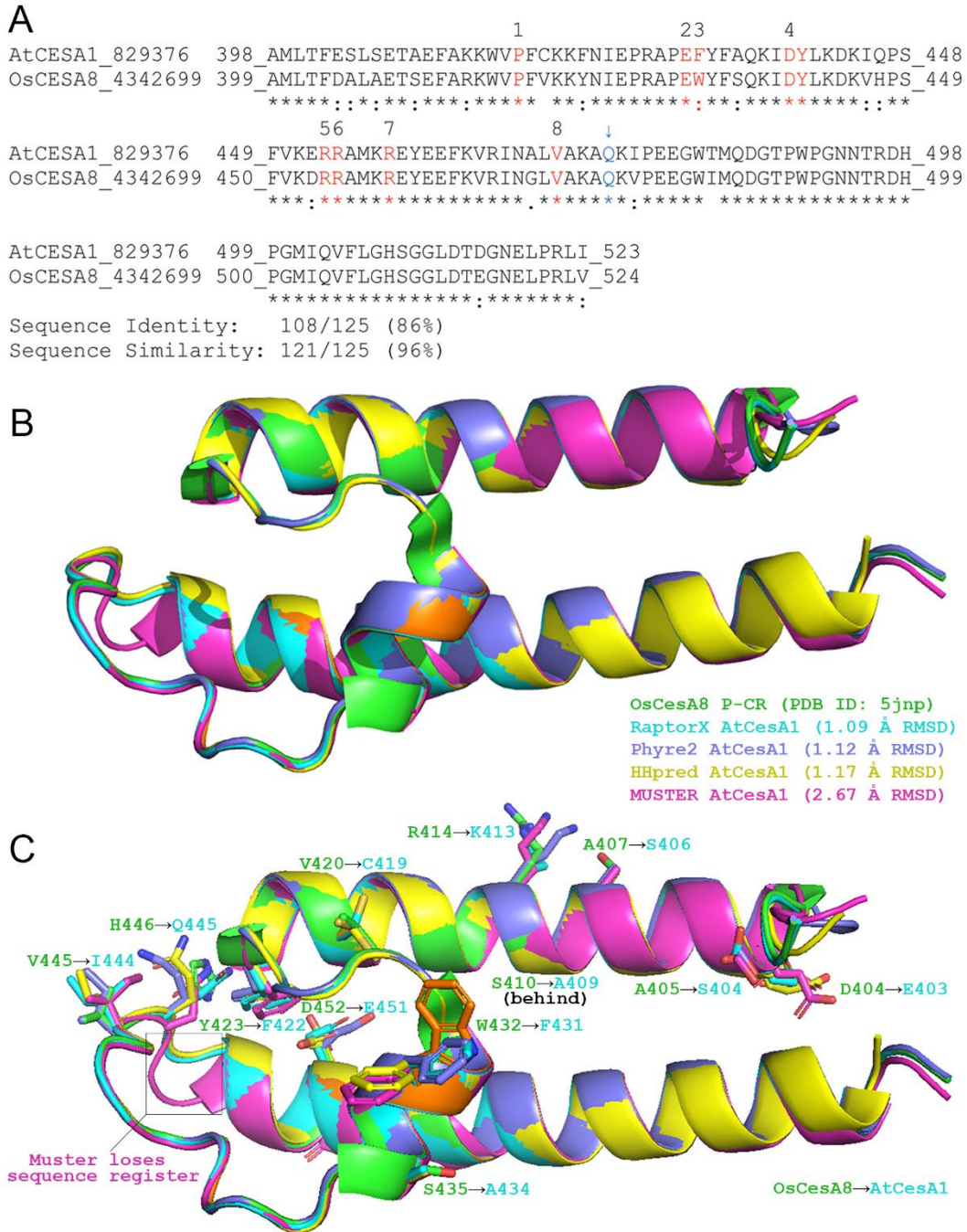


Figure 5.4 (A) Homologous sequences between the AtCESA1 and OsCESA8 P-CRs allow high confidence homology modeling of the AtCESA1 P-CR to the limit of the OsCESA8 P-CR X-ray crystal structure. (B) RaptorX (cyan) (Kallberg et al., 2012), Phyre2 (purple) (Kelley et al., 2015), HHPred (yellow) (Zimmermann et al., 2018) and MUSTER (violet) (Wu & Zhang, 2008) AtCESA1 models, based on threading against the OsCESA8 P-CR X-ray crystal structure (green) (PDB ID: 5jnp), are shown aligned to one another using the align command in PyMOL (*The PyMOL Molecular Graphics System, Version 2.2*). (C) Differences between the OsCESA8 and AtCESA1 sequence are shown as stick models and labeled with their OsCESA8 (green) and AtCESA1 (cyan) sequences.

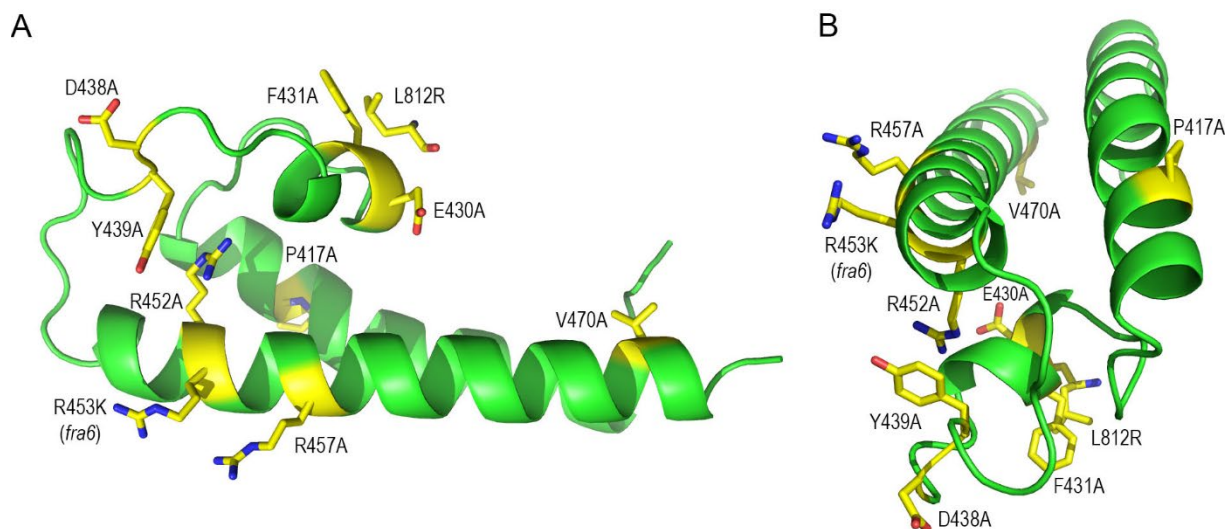


Figure 5.5 Best homology model of the AtCESA1 P-CR domain using RaptorX (Kallberg et al., 2012) showing the Atcesal mutants (yellow, stick models) constructed for the *Arabidopsis thaliana rsw1-1* complementation experiments. L812R is part of the BCSA homologous catalytic core, modeled relative to the P-CR from the best fit of the P-CR to the CatD (other catalytic core residues not shown).

In general, none of the sequence differences occurred close to the prominent features probed by mutagenesis and are likely to be equivalent as they appear to maintain similar properties and size, such as the balanced number of Ala to Ser conversions between structures. The Arg414 to Lys413, Val445 to Ile444 and two Asp to Glu substitutions were also observed but appear to fulfill the same structural role in AtCESA1. The His446 to Gln445 and Val420 to Cys419 substitutions could theoretically be more significant to function, although these residues have similar sizes and share properties. Perhaps the only difference of note is the Trp432 to Phe431 substitution, with some Met residues aligned in other sequences, although Met is still a relatively large hydrophobic residue. Despite these substitutions, the exposed nature in the P-CR X-ray crystal structure points to the importance of this residue. Altogether, there is no reason to believe any of the features or hypothetical function(s) of the P-CR or even the smaller features identified in the P-CR would vary between OsCESA8 and AtCESA1 (Figure 5.1). Thus, converting the mutants designed based on the OsCESA8 P-CR X-ray crystal structure and best fit model of the P-CR to OsCESA8 CatD SAXS density to an AtCESA1 P-CR is reasonable, with the RaptorX model best representing the AtCESA1 isoform (Figure 5.5). From this, a final list of mutants designed to test the importance of P-CR or P-CR related residues *in vivo* using AtCESA1 is: [1]

Pro417Ala [2] Glu430Ala [3] Phe431Ala [4] Asp438Ala/Tyr439Ala (double mutant) [5] Arg452Ala [6] Arg457Ala [7] Val470Ala [8] Leu812Arg [9] Arg453Lys (equivalent to *fra6* in *Atcesa8* (Zhong et al., 2003)).

5.3.4 Transgenic Expression of *Atcesa1* P-CR Mutants

Mutated P-CR sequences were generated by site-directed mutagenesis of an *Agrobacterium* pFGC5941 binary vector (TAIR: CD3-447) containing the *AtCESA1* gene driven by the *AtCESA1* promoter. For each of the ten constructs used in this study, between five and twelve homozygous independent lines were isolated in the T3 generation using herbicide selection and self-pollination (Table 5.3). Levels of transgene expression varied across each independent line but were comparable to the *CBP20* reference gene control (TAIR: AT5G44200) in seven-day-old seedlings (Figure 5.6). This is consistent with WT gene expression showing an approximately 1:1 ratio of *AtCESA1* (TAIR: AT5G32410) to *CBP20* (TAIR: AT5G44200) RNA, compared through the Arabidopsis eFP Browser of published expression data (Winter et al., 2007), using similar growth conditions and timescales as used in this experiment (Michael et al., 2008). The *rsw1-1/AtCESA1* 2.2 transgenic line, expressing an unmutated *AtCESA1* *Agrobacterium* vector, was used as a positive control. The Asp438Ala/Tyr440Ala and Arg452Ala mutants, as well as the Glu430Ala mutant to a lesser degree, provided a range of expression levels allowing for complementation experiments that test expression levels higher, proximate and lower relative to *CBP20*. The Leu812Arg mutant produced only a single independent line recovered with expression levels proximate to *CBP20*, as well as several independent lines with lower expression.

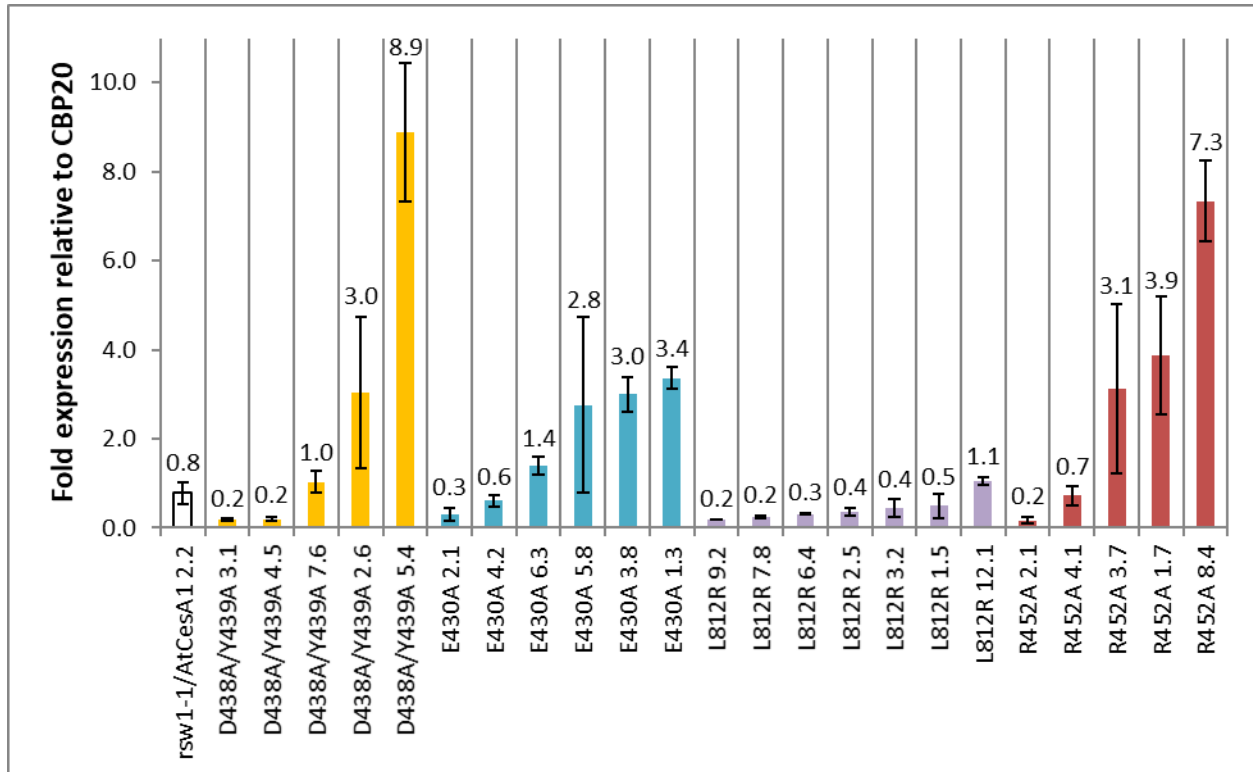


Figure 5.6 *Atcesal* transgene expression in the *rsw1-1* mutant background relative to *CBP20*. *rsw1-1/AtCESA1* 2.2 (white), Asp438Ala/Tyr439Ala (yellow), Glu430Ala (blue), Leu812Arg (purple) and Arg452Ala (red) independent lines are shown as mean values \pm standard deviation (error bars) above their bar graphs (n=3).

5.3.5 Complementing the Phenotype of the *Arabidopsis thaliana* *rsw1-1* Mutant Background

Using similar growth conditions as the expression analysis, control and mutant individuals were grown on MS media plates for four days at 22°C to initiate growth and root development (Figure 5.7). At four days, the temperature of the growth chamber was raised to 31°C, the restrictive temperature for cellulose synthase activity in the *rsw1-1* negative control. The *rsw1-1/AtCESA1* line 2.2 was used as a transgenic positive control together with Columbia (Col-0) for testing the *Atcesal* mutants. After three days at the restrictive temperature (31°C) (seven days total), the WT Col-0 and *rsw1-1/AtCESA1* 2.2 transgenic control had nearly reached the edge of their quadrant in the plate, demonstrating the greatest differences between mutant and control experiments before growing into another quadrant (Figure 5.7). Because one phenotype of *rsw1-1* is that root tip growth halts within 24 h, the level of complementation in the *AtCESA1* mutants was assessed

relative to the primary root length of *rswl-1*, Col-0 and *rswl-1/AtCESA1*. Complete recovery of WT phenotype was observed for the *rswl-1/AtCESA1* 2.2 positive control compared to the Col-0 control, in stark contrast to the short primary root lengths of the *rswl-1* negative control. Poor complementation was also observed in several mutants, shown here for the Asp438Ala/Tyr439Ala mutant. Repeating this analysis over additional replicates for representative independent lines with higher, proximate and lower expression levels relative to *CBP20* of the Asp438Ala/Tyr439Ala, Glu430Ala, Leu812Arg and Arg452Ala mutants resulted in statistically significant differences between primary root length phenotype of *rswl-1* of some of the independent mutant lines and the controls (Figure 5.8). Morphologically the radially swollen root tip phenotype of *rswl* was complemented for some of the independent mutant lines (Figure 5.9).

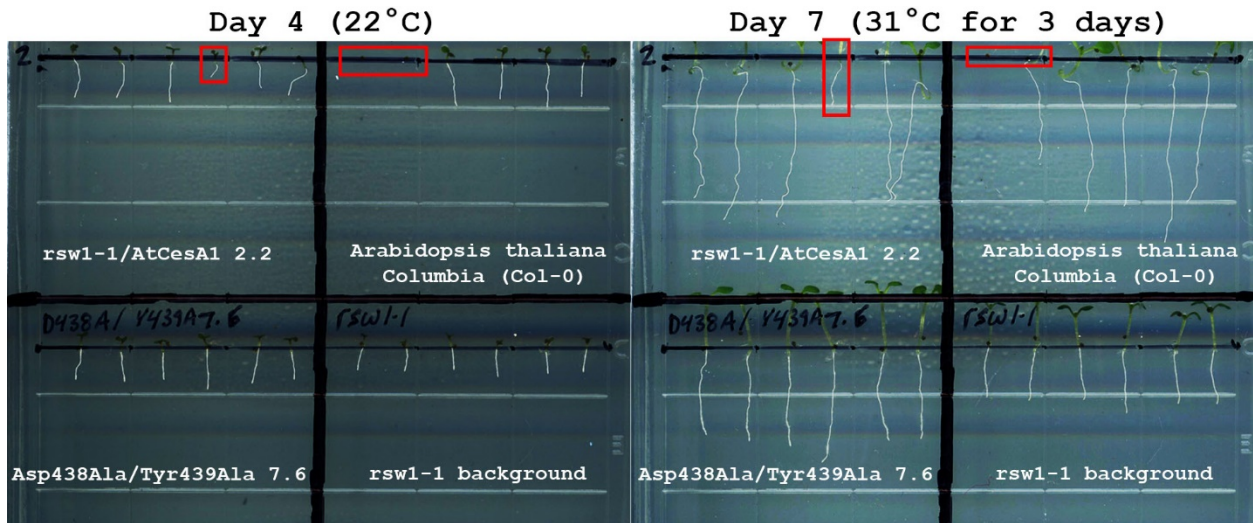


Figure 5.7 Complementation experiments of Col-0, *rswl-1/AtCESA1* 2.2, *rswl-1* and Asp438Ala/Tyr439Ala 7.6 seedlings grown for four days at the *rswl-1* permissive temperature (22°C) then three days at the restrictive temperature (31°C). Seedlings that failed to germinate or germinated late were removed from the analysis (red box). Seedlings grown at 22°C were similar at day 4, but at 31°C differences appeared between the Asp438Ala/Tyr439Ala 7.6 mutant, positive and negative controls, which were most dramatic by day seven. Scanned images like these were used to measure primary root length by setting them to scale with a ruler and tracing the primary root with a segmented line (white lines) in ImageJ (Schneider et al., 2012).

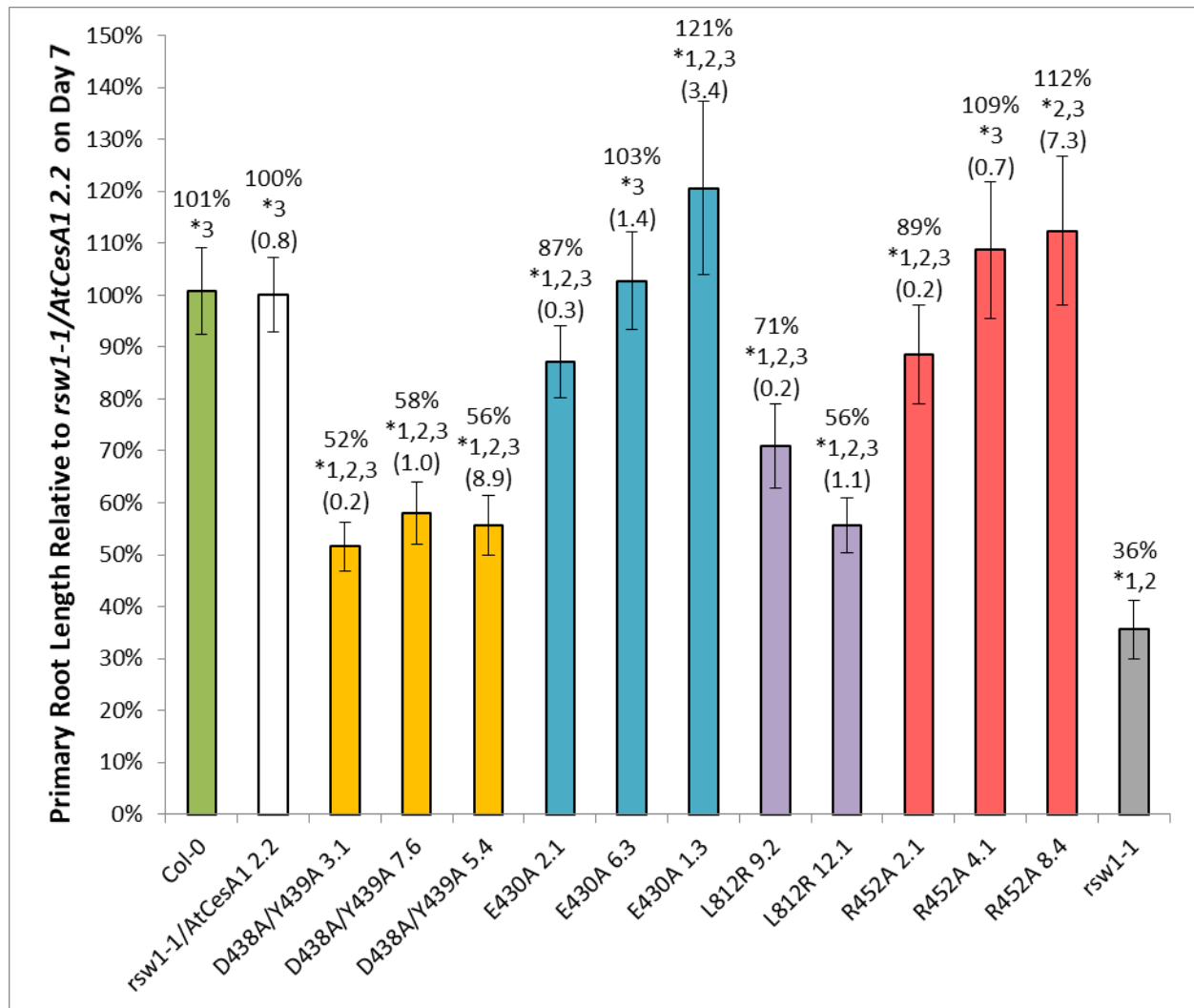


Figure 5.8 Differences in primary root growth varied based on the mutant and its expression level after four days at the permissive temperature (22°C) followed by three days at the restrictive temperature (31°C). Primary root length mean values \pm standard deviation (error bars) (n=13-24) were measured from scanned images and scaled to a percentage of the *rsw1-1/AtCESA1* 2.2 control mean value (labeled above each graph) to assess complementation compared to the *rsw1-1* negative control. Statistically significant differences (ANOVA $P < 0.05$) are labeled with an asterisk and number corresponding to differences between Col-0 (1), *rsw1-1/AtCESA1* 2.2 (2) and *rsw1-1* (3). Mean expression levels relative to *CBP20* are shown in parenthesis for each independent line tested.

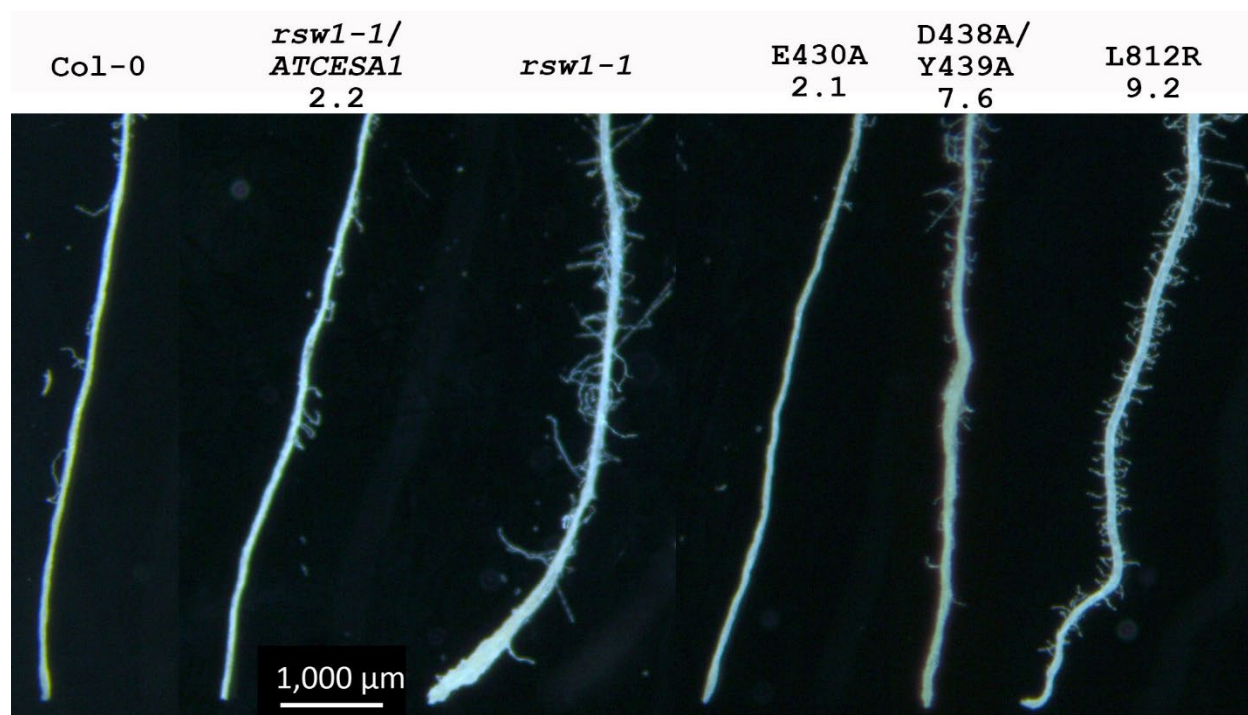


Figure 5.9 Root morphology of control and transgenic mutants grown for four days at the permissive temperature (22°C) followed by three days at restrictive temperature (31°C) (Arg452Ala not shown). Image was created as a composite of images.

P-CR mutants Arg457Ala and Val470Ala showed complete complementation of both the radially swollen root tip and growth inhibition phenotype over a range of expression levels tested at the restrictive (31°C) temperature in short-term growth experiments (Figure 5.10-5.13). Arg457Ala independent lines showed average expression levels of the transgene between ~2.1- and 5.8-fold greater than *CBP20* and all three lines complemented the *rsw1-1* phenotypes (Figure 5.10 and 5.11). Val470Ala independent lines showed average expression levels of the transgene between ~0.5- and 6.2-fold relative to *CBP20* and complementation of the *rsw1-1* radially swollen root tip phenotype in all three lines (Figure 5.12 and 5.13). Slight growth inhibition was observed in the independent line expressing the transgene at ~0.5-fold the level of *CPB20* expression (line 10-7) but not higher expressing lines, indicating insufficient transcript might lead to some amount of partial complementation.

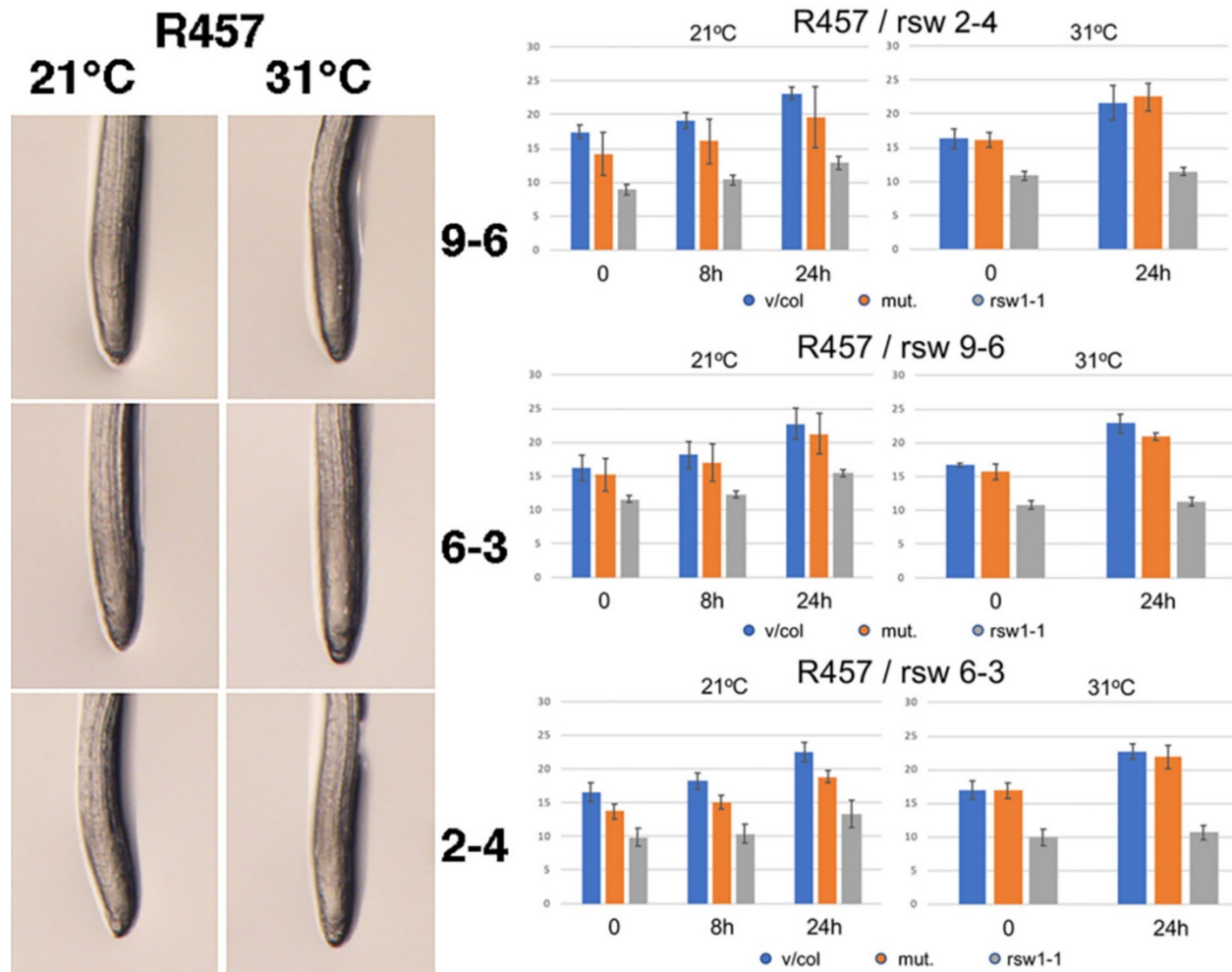


Figure 5.10 Primary root growth of the Arg457Ala *Atcesal* transgene in *rsw1-1* after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures. Three homozygous lines (9-6, 6-3 and 2-4) were tested. Primary root length mean values \pm standard deviation (error bars) (n=4) were measured from scanned images in mm. The pFGC5941 vector in Col-0, Arg457Ala *Atcesal* transgenes and *rsw1-1* control are shown as blue, orange and grey, respectively. (Experiments by Anna T. Olek)

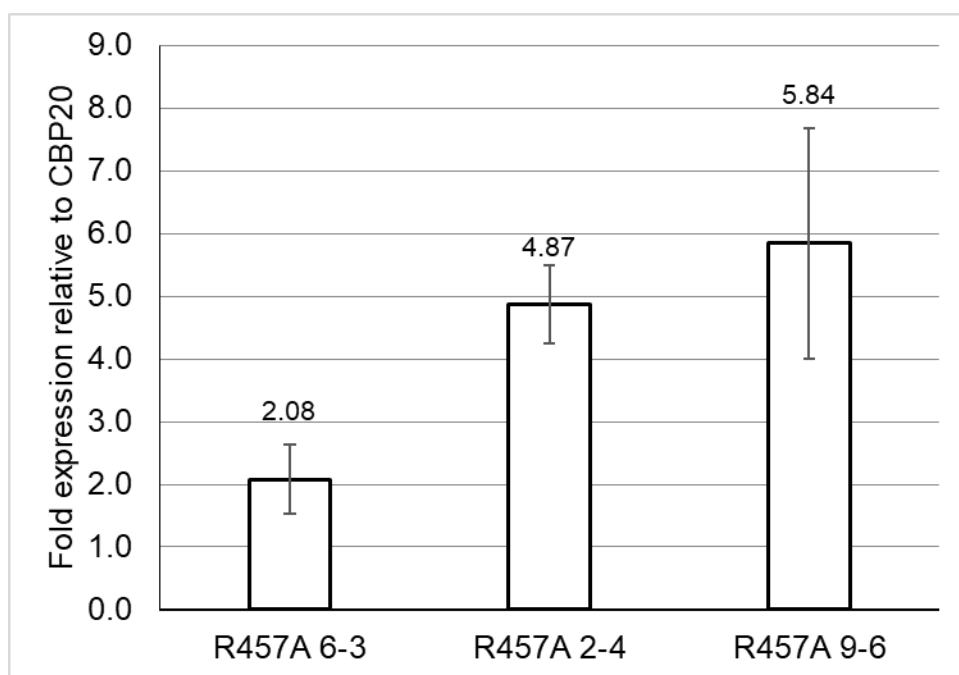


Figure 5.11 Arg457Ala *Atcesal* transgene expression in the *rsw1-1* mutant background relative to *CBP20*. Independent lines are shown as mean values \pm standard deviation (error bars) above their bar graphs (n=3).

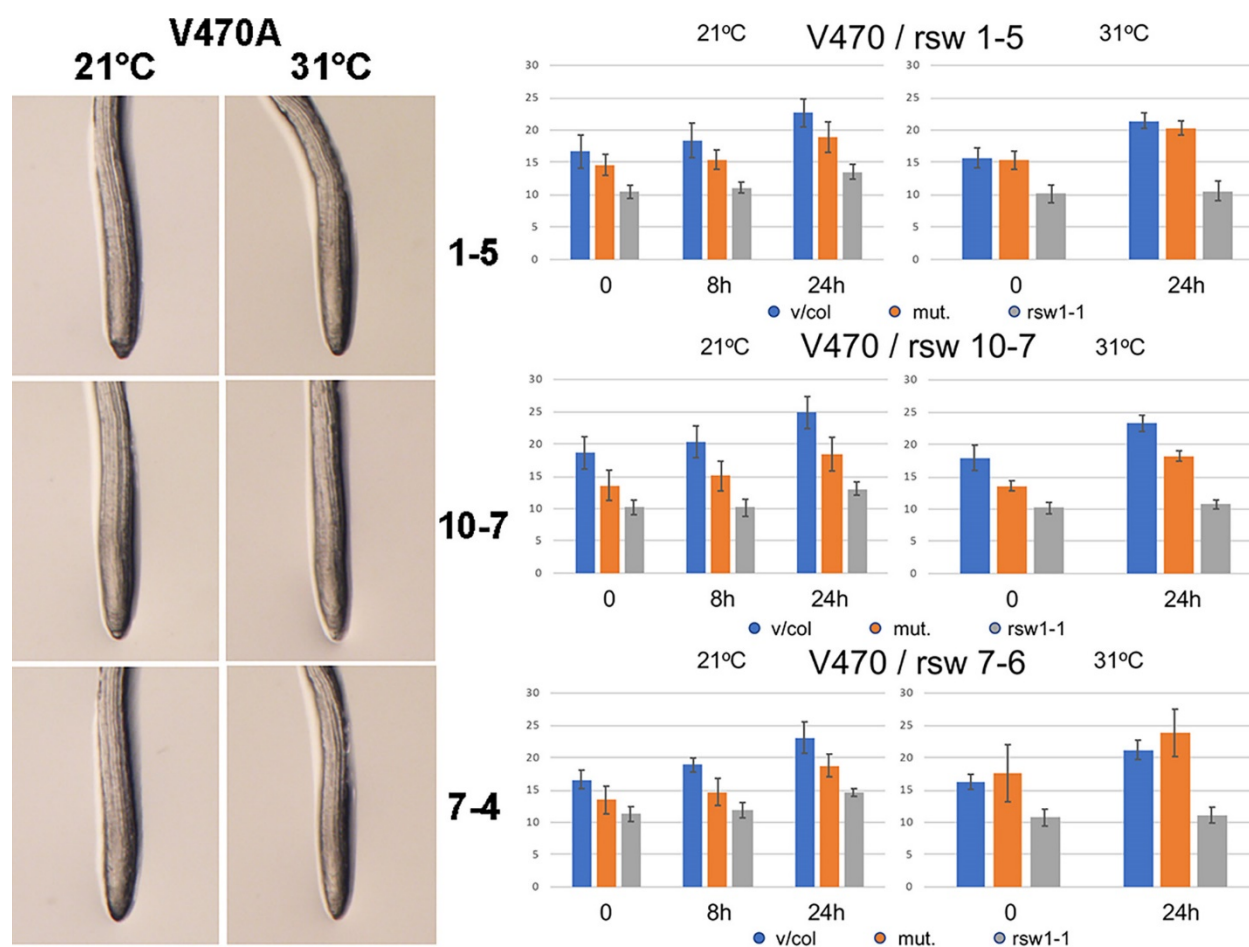


Figure 5.12 Primary root growth of the Val470Ala *Atcesal* transgene in *rsw1-1* after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures. Three homozygous lines (1-5, 10-7 and 7-4) were tested. Primary root length mean values \pm standard deviation (error bars) (n=4) were measured from scanned images in mm. The pFGC5941 vector in Col-0, Val470Ala *Atcesal* transgenes and *rsw1-1* control are shown as blue, orange and grey, respectively. (Experiments by Anna T. Olek)

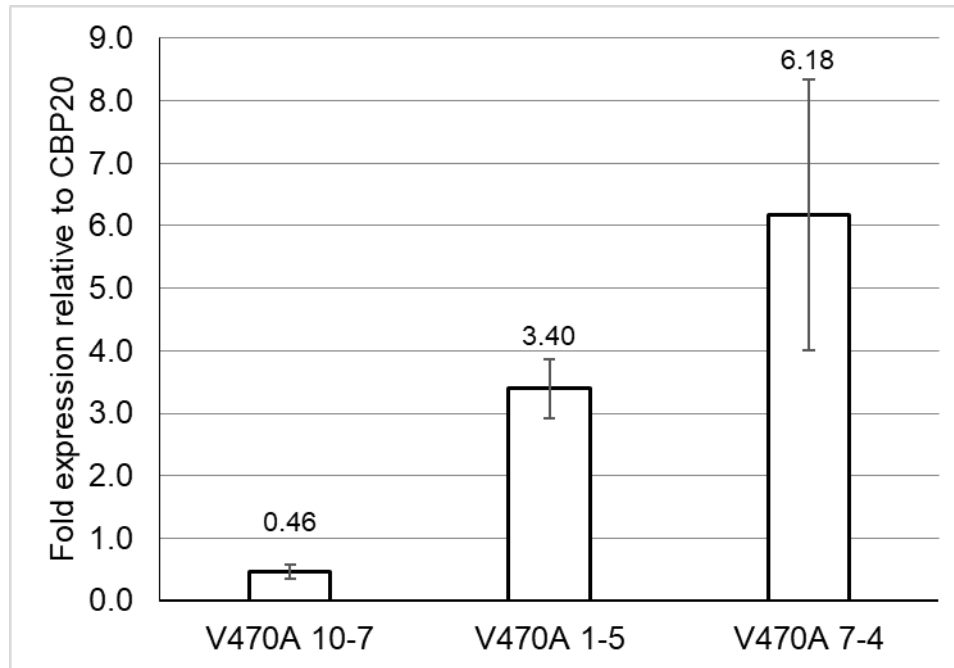


Figure 5.13 Val470Ala *Atcesal* transgene expression in the *rswI-1* mutant background relative to *CBP20*. Independent lines are shown as mean values \pm standard deviation (error bars) above their bar graphs (n=3).

Arg452Ala and Glu430Ala independent lines expressing these mutants at a range of expression levels complemented the *rswI-1* radially swollen root tip phenotype, but showed differences in the complementation of the primary root length phenotype of *rswI-1* (Figure 5.8). Mutants with lower expression levels resulted in significantly less complementation and slight growth inhibition relative to the *rswI-1/AtCESA1 2.2* positive control (ANOVA $P < 0.05$). Overexpressing the Glu430Ala mutant higher than *CBP20* resulted in growth that was significantly increased at the restrictive temperature (31°C). Morphologically, both mutants appeared similar to Col-0 and *rswI-1/AtCESA1 2.2* (Figure 5.9, Arg452Ala not shown), without the radially swollen root tip phenotype of *rswI-1*. Additional short-term growth experiments similarly concluded that Arg452Ala and Glu430Ala mutants complemented the radially swollen root tip phenotype of *rswI-1* at both the restrictive (31°C) and permissive (21°C) temperatures over a range of expression levels (Figure 5.14-5.17). Slight growth inhibition was observed at the restrictive (31°C) temperature for Arg452Ala lines (Figure 5.14 and 5.15) and Glu430Ala lines 2-1 and 6-3, but not line 1-3 (Figure 5.16 and 5.17). Again, Glu430Ala line 1-3 increased growth at the

restrictive temperature (31°C) compared to controls, but expressing the transgene ~3.4-fold greater than *CPB20* expression it is possible this due to overexpression.

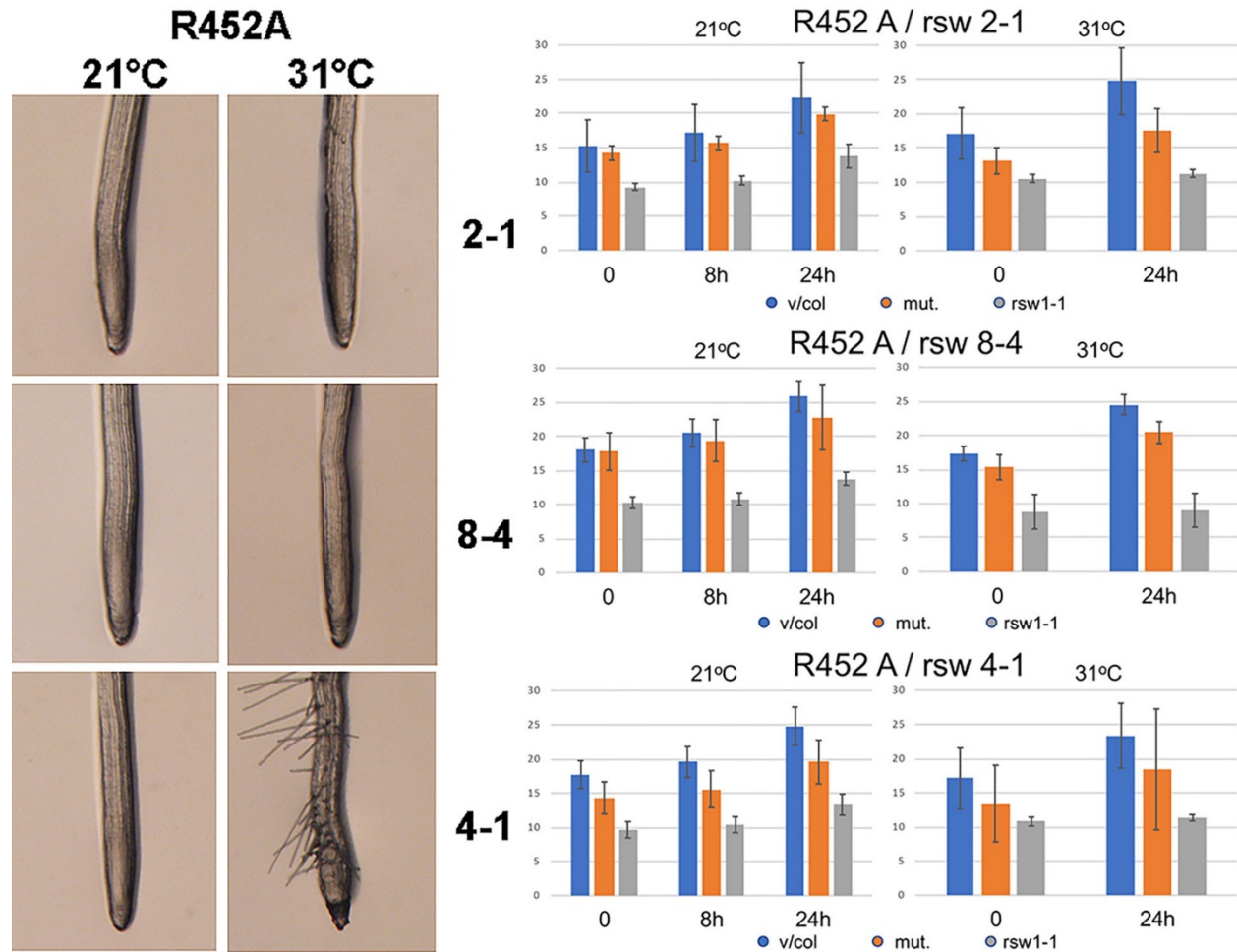


Figure 5.14 Primary root growth of the Arg452Ala *Atcesal* transgene in *rsw1-1* after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures. Three homozygous lines (2-1, 8-4 and 4-1) were tested. Primary root length mean values \pm standard deviation (error bars) (n=4) were measured from scanned images in mm. The pFGC5941 vector in Col-0, Arg452Ala *Atcesal* transgenes and *rsw1-1* control are shown as blue, orange and grey, respectively. (Experiments by Anna T. Olek)

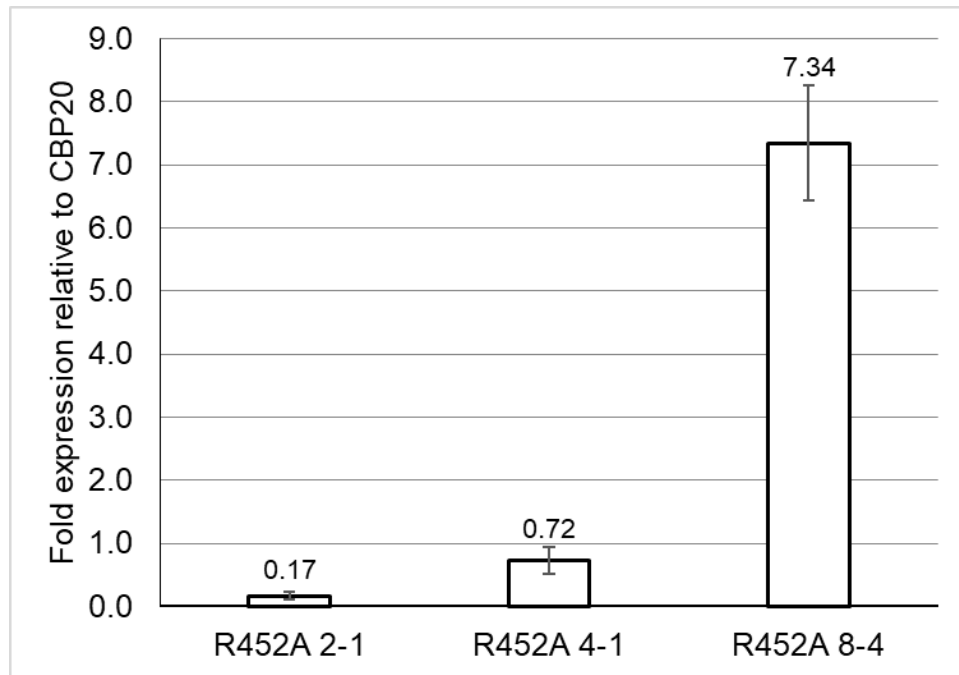


Figure 5.15 Arg452Ala *Atcesa1* transgene expression in the *rsw1-1* mutant background relative to *CBP20*. Independent lines are shown as mean values \pm standard deviation (error bars) above their bar graphs (n=3).

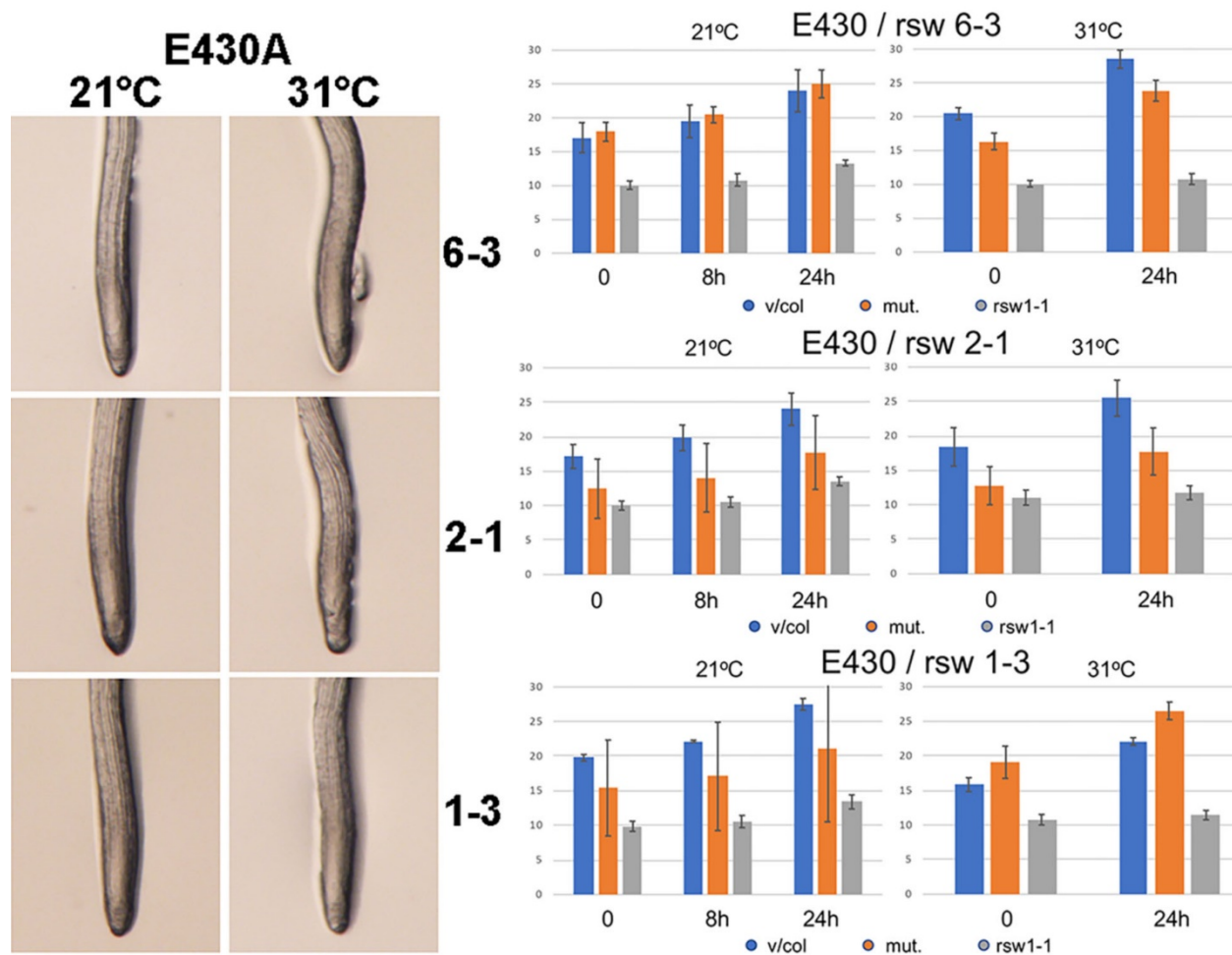


Figure 5.16 Primary root growth of the Glu430Ala *Atcesal* transgene in *rsw1-1* after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures. Three homozygous lines (6-3, 2-1 and 1-3) were tested. Primary root length mean values \pm standard deviation (error bars) (n=4) were measured from scanned images in mm. The pFGC5941 vector in Col-0, Glu430Ala *Atcesal* transgenes and *rsw1-1* control are shown as blue, orange and grey, respectively. (Experiments by Anna T. Olek)

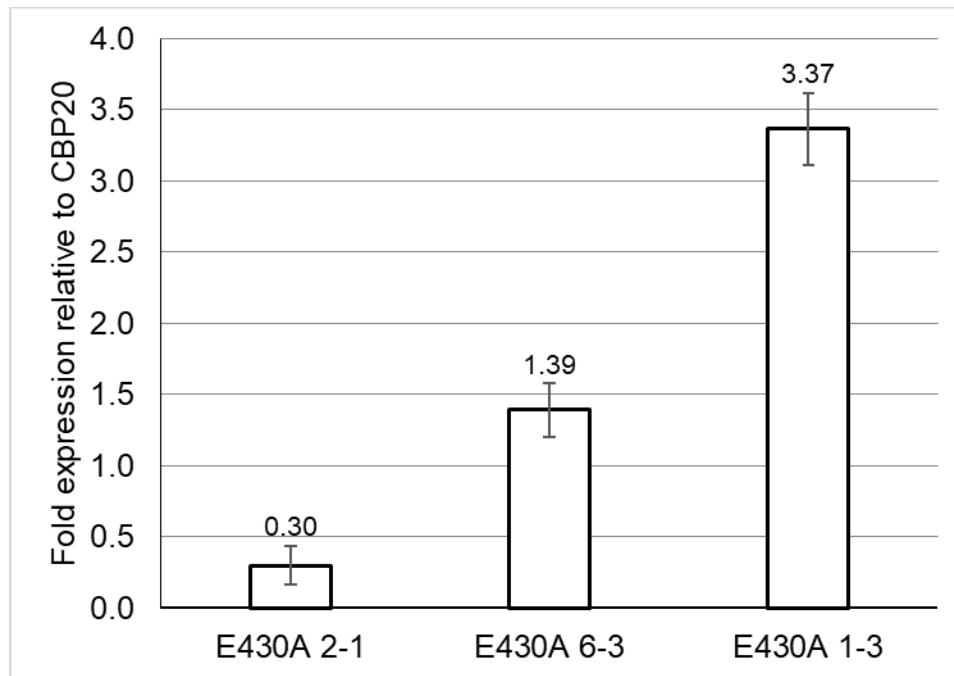


Figure 5.17 Glu430Ala *Atcesal* transgene expression in the *rsw1-1* mutant background relative to *CBP20*. Independent lines are shown as mean values \pm standard deviation (error bars) above their bar graphs (n=3).

The Phe431Ala independent mutant lines 3-5 and 6-4 complemented the radially swollen root tip phenotype of *rsw1-1* at the restrictive (31°C), but not line 2-3 in short-term growth experiments (Figure 5.18). Phe431Ala lines 3-5 and 6-4 showed slight root growth inhibition, while line 2-3 showed severe growth inhibition. Phe431Ala independent lines showed average expression levels of the transgene between ~2.4- and 7.9-fold greater than *CBP20*, with the two highest expressing lines showing better complementation (Figure 5.19). Based on the high variability observed in the expression of the transgene in Phe431Ala line 2-3, it might be expressing the transgene lower than currently estimated. Thus, one possibility could be insufficient expression to complement *rsw1-1* in line 2-3. Like Arg452Ala and Glu430Ala, Phe431Ala is best interpreted as having slight growth inhibition but no radial-swelling, with one independent line behaving as an outlier likely due to transgene expression.

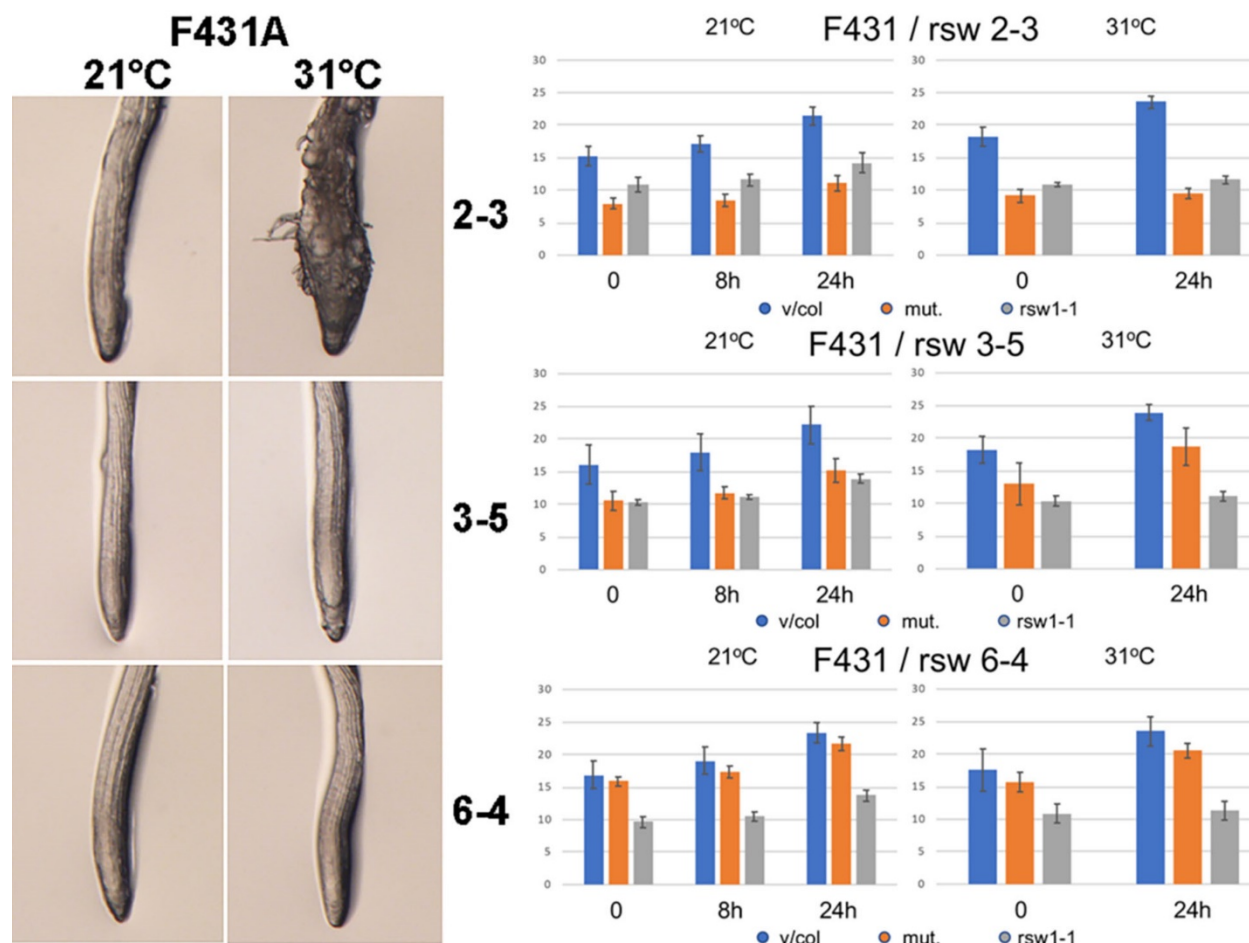


Figure 5.18 Primary root growth of the Phe431Ala *Atcesal* transgene in *rsw1-1* after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures. Three homozygous lines (2-3, 3-5 and 6-4) were tested. Primary root length mean values \pm standard deviation (error bars) (n=4) were measured from scanned images in mm. The pFGC5941 vector in Col-0, Phe431Ala *Atcesal* transgenes and *rsw1-1* control are shown as blue, orange and grey, respectively. (Experiments by Anna T. Olek)

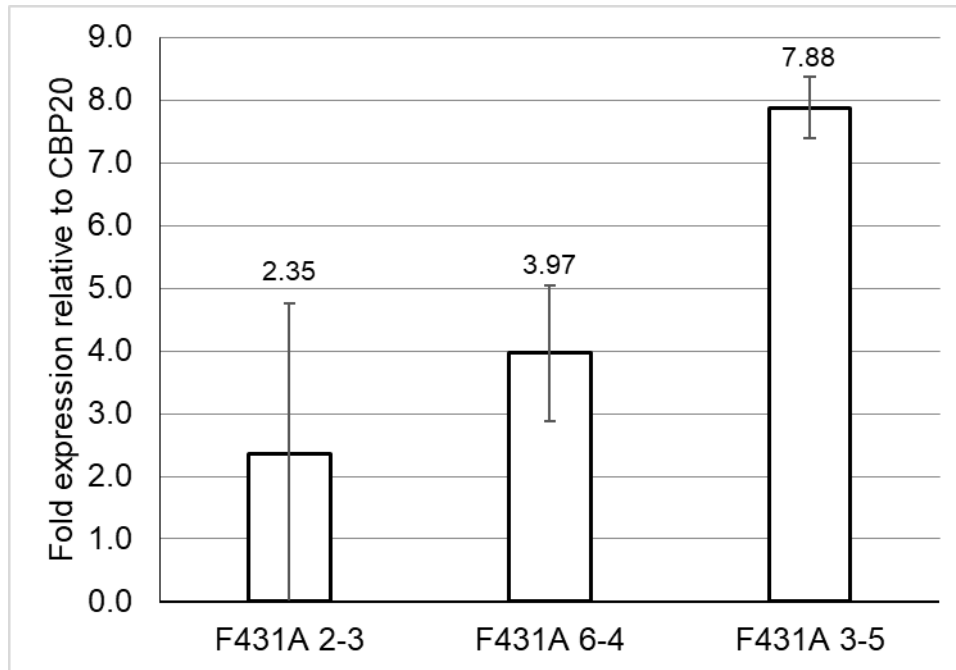


Figure 5.19 Phe431Ala *Atcesa1* transgene expression in the *rswl-1* mutant background relative to *CBP20*. Independent lines are shown as mean values \pm standard deviation (error bars) above their bar graphs (n=3).

Leu812Arg partially complemented the *rswl-1* phenotype at the restrictive temperature (31°C), lacking root swelling but with root lengths that were statistically different from and intermediate to the *rswl-1/AtCESA1 2.2* and Col-0 positive controls and *rswl-1* negative control (Figure 5.8 and 5.20). Partial complementation might indicate that the mutants alter normal AtCESA1 function within the CSC but do not interfere with overall AtCESA1 structure as the latter would likely result in a phenotype as severe as *rswl-1*. Strangely, proximate expression levels resulted in significantly less complementation of the *rswl-1* growth phenotype compared to a line with reduced levels of gene expression (ANOVA $P < 0.05$) (Figure 5.8). More independent lines will need to be tested to establish this difference. Morphologically, the Leu812Arg mutant differed from Col-0 and *rswl-1/AtCESA1 2.2* positive controls not just in length but also appeared to have more root hairs, similar to *rswl-1* (Figure 5.9). Unlike *rswl-1*, however, the root hairs appear shorter and roots do not initially appear to show the same root swollen phenotype as *rswl-1*, despite the reduction in growth. In this way the Leu812Arg mutant has a phenotype that is different than *rswl-1*. Short-term growth experiments had similar results for root growth inhibition at the restrictive temperature (31°C) and showed that root growth was also inhibited at the

permissive temperature (21°C) (Figure 5.20). In both independent lines the primary root length was not significantly different from *rswl-1* at the restrictive (31°C) or permissive temperature (21°C), unlike the initial experiments, regardless of expression levels (Figure 5.21). Photos illustrate the lack of radially swollen root tips at both permissive (21°C) and restrictive (31°C) temperatures (Figure 5.20).

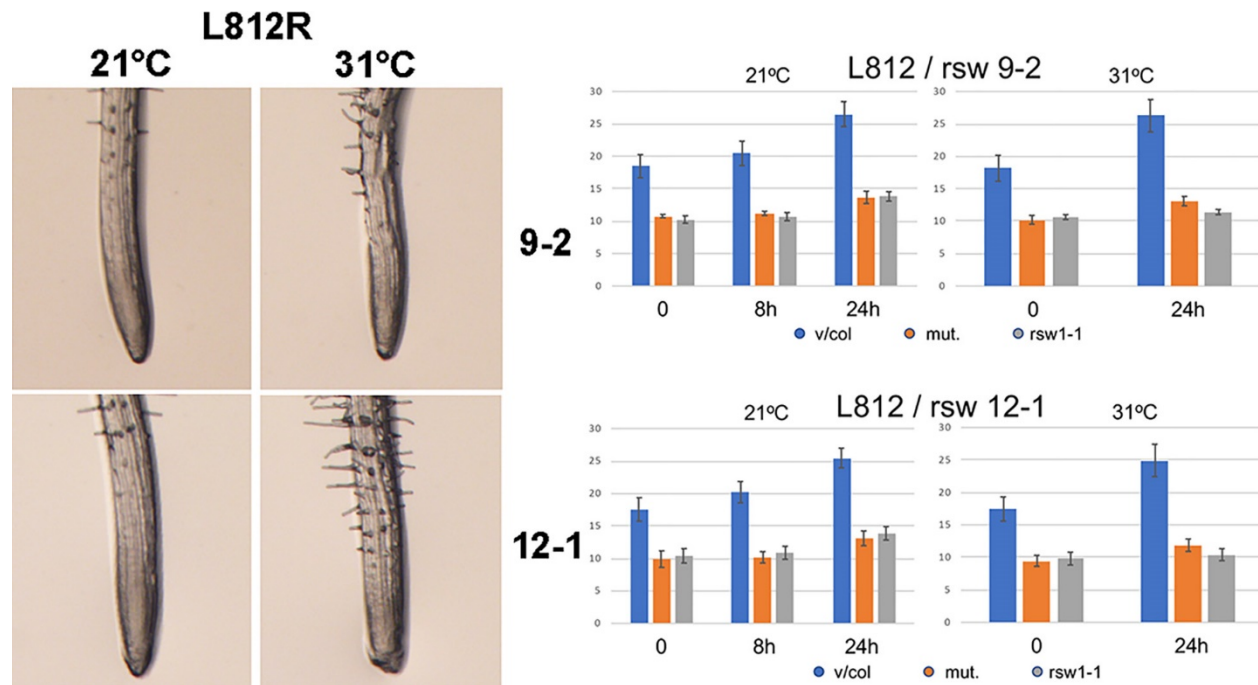


Figure 5.20 Primary root growth of the Leu812Arg *Atcesal* transgene in *rswl-1* after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures. Two homozygous lines (9-2 and 12-1) were tested. Primary root length mean values \pm standard deviation (error bars) (n=4) were measured from scanned images in mm. The pFGC5941 vector in Col-0, Leu812Arg *Atcesal* transgenes and *rswl-1* control are shown as blue, orange and grey, respectively. (Experiments by Anna T. Olek)

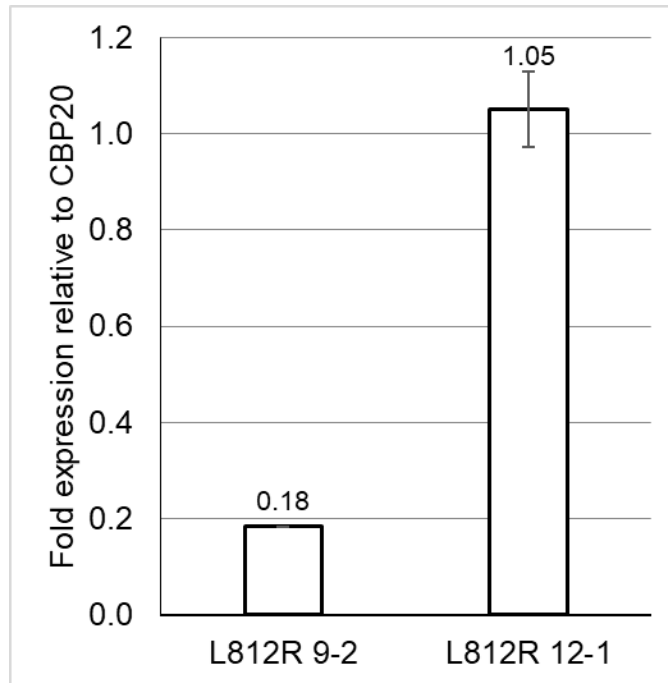


Figure 5.21 Leu812Arg *Atcesa1* transgene expression in the *rswl-1* mutant background relative to *CBP20*. Independent lines are shown as mean values \pm standard deviation (error bars) above their bar graphs (n=3).

The Asp438Ala/Tyr439Ala double mutant showed little complementation of root-tip growth at the restrictive temperature (31°C) (Figure 5.8). All of the lines were significantly different in root length compared to Col-0, *rswl-1/AtCESA1* 2.2 and *rswl-1* but were not significantly different from one another (ANOVA $P < 0.05$). Growth rates for the Asp438Ala/Tyr439Ala 7.6 independent line before and after the transition to the restrictive temperature (31°C) are significantly greater than *rswl-1* alone and less than Col-0 and *rswl-1/AtCESA1* 2.2 positive controls (ANOVA $P < 0.05$) (Figure 5.22). In general, the Asp438Ala/Tyr439Ala double mutant was morphologically most similar to the *rswl-1* mutant but the mutant appears less radially swollen, although still thicker than Col-0 and *rswl-1/AtCESA1* 2.2 positive controls, and the radial swelling is spread more evenly across the root (Figure 5.9). Short-term growth experiments had similar results at the restrictive temperature (31°C), showing growth inhibition and root-tip radial swelling (Figure 5.23). At the permissive temperature (21°C), the Asp438Ala/Tyr439Ala mutant lines showed increased growth relative to *rswl-1*, but lower than the positive control (pFGC5941 vector in Col-0) indicating a degree of complementation of the *rswl-1* growth inhibition phenotype. No significant correlation between the level of expression,

ranging from ~0.2- to 8.9-fold relative to *CBP20*, and complementation were observed, with all three independent lines behaving similarly (Figure 5.24).

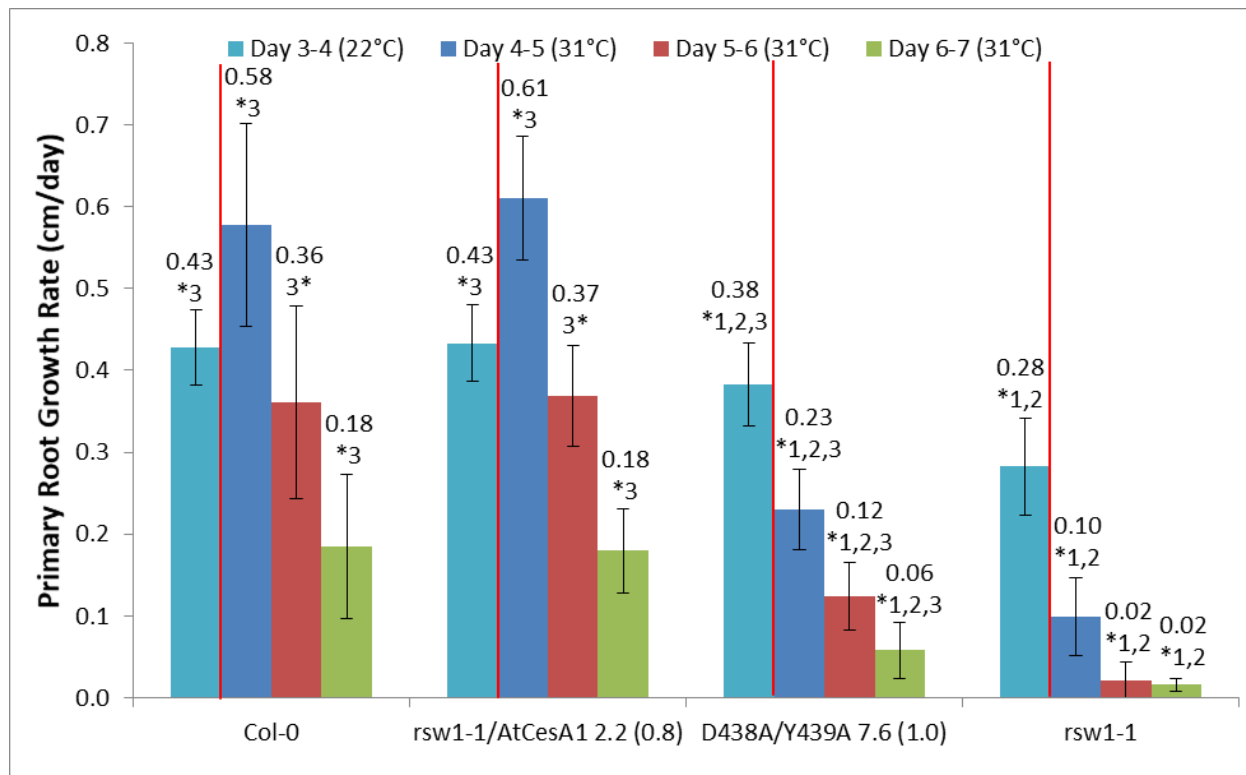


Figure 5.22 Growth rate comparison between controls and the poorly complementing Asp438Ala/Tyr439Ala 7.6 mutant, expressing near native levels, before and after the temperature was increased to the *rsw1-1* restrictive temperature (31°C) (red line). Primary root mean growth rates (cm/day, labeled above each graph) \pm standard deviation (error bars) (n=17-24) were calculated from measurements of scanned images. Statistically significant differences (ANOVA $P < 0.05$) are labeled with an asterisk and number corresponding to differences between Col-0 (1), *rsw1-1/AtCESA1 2.2* (2) and *rsw1-1* (3). Mean expression levels relative to *CBP20* are labeled in parenthesis for *rsw1-1/AtCESA1 2.2* and Asp438Ala/Tyr439Ala 7.6 independent lines.

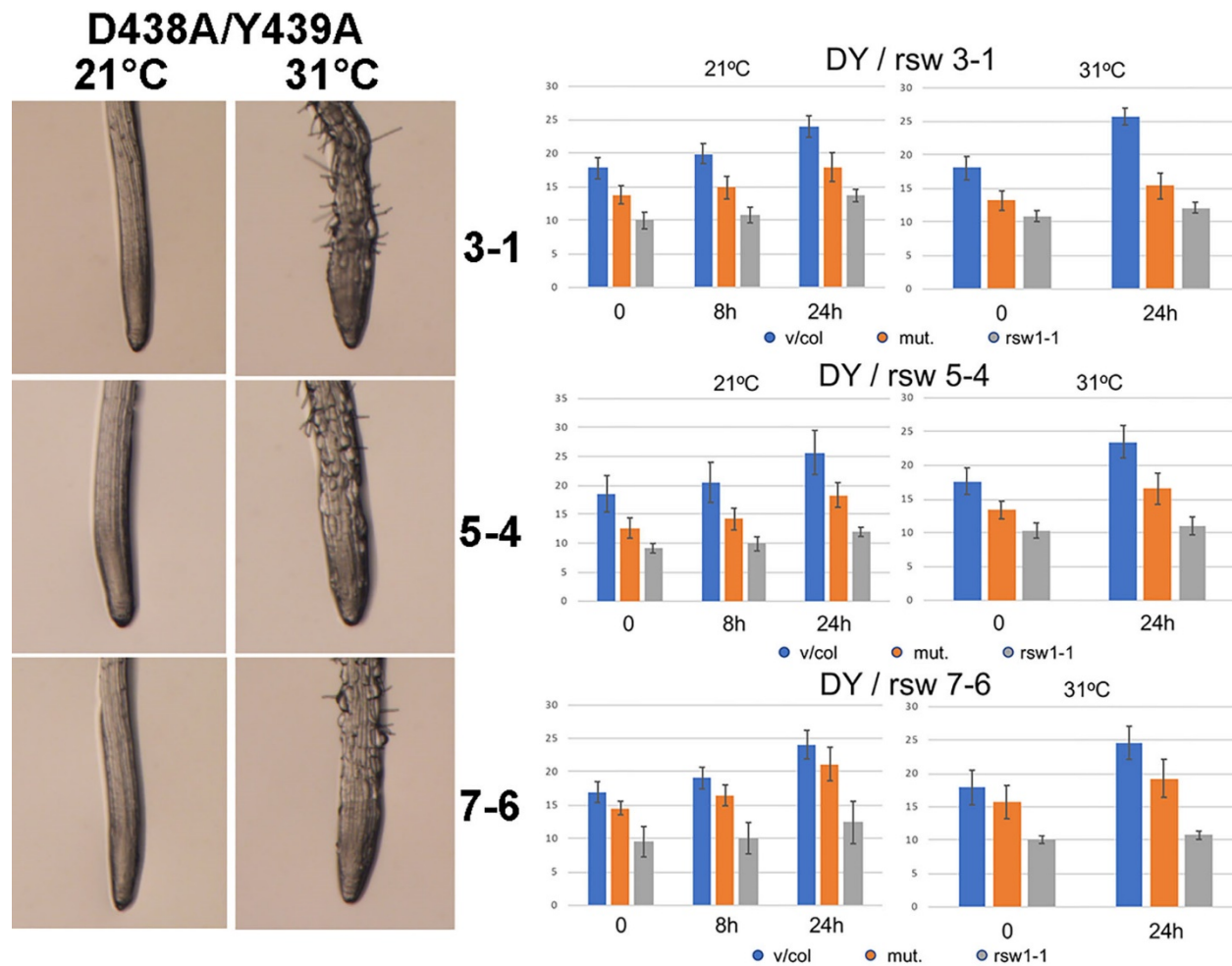


Figure 5.23 Primary root growth of the Asp438Ala/Tyr439Ala *Atcesal* transgene in *rsw1-1* after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures. Three homozygous lines (3-1, 5-4 and 7-6) were tested. Primary root length mean values \pm standard deviation (error bars) (n=4) were measured from scanned images in mm. The pFGC5941 vector in Col-0, Asp438Ala/Tyr439Ala *Atcesal* transgenes and *rsw1-1* control are shown as blue, orange and grey, respectively. (Experiments by Anna T. Olek)

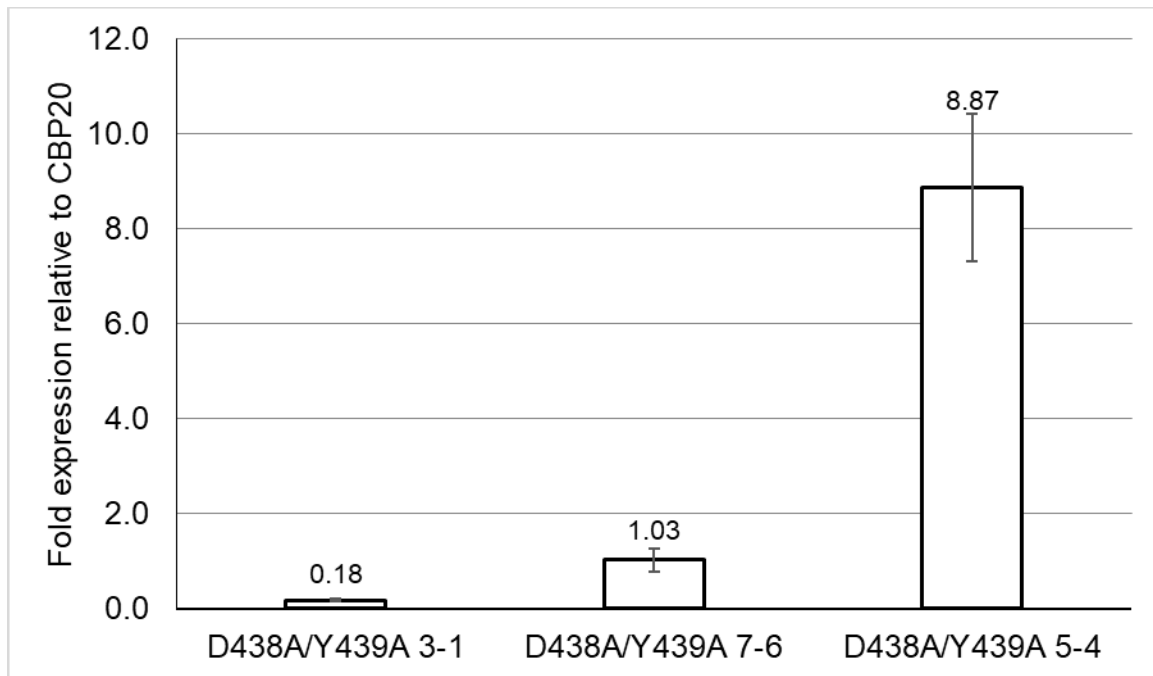


Figure 5.24 Asp438Ala/Tyr439Ala *Atcesa1* transgene expression in the *rsw1-1* mutant background relative to *CBP20*. Independent lines are shown as mean values \pm standard deviation (error bars) above their bar graphs (n=3).

The Arg453Lys (equivalent to *fra6* in *Atcesa8* (Zhong et al., 2003)) and Pro417Ala mutants showed the poorest complementation in short-term growth experiments, despite significant transgene expression (Figure 5.25-5.28). For both mutants, the *rsw1-1* radial swelling and growth inhibition phenotypes were observed at the restrictive temperature (31°C). Unlike the Asp438Ala/Tyr439Ala mutants, Arg453Lys and Pro418Ala also showed increased growth inhibition at the permissive temperature (21°C), leading to shorter roots compared to *rsw1-1*, but no radially-swollen root tip. Expression levels were similar for the three Arg453Lys independent lines, ranging between ~1.6- and 2.6-fold expression relative to *CBP20* (Figure 5.26), but varied more for Pro417Ala mutants that ranged between ~1.2- and 15.5-fold expression relative to *CBP20* (Figure 5.28). In either case, there was no correlation between expression levels and complementation, indicating these mutants do not functionally complement *rsw1-1*. Significant increases in growth inhibition at both restrictive (31°C) and permissive (21°C) temperatures compared to *rsw1-1* also indicates that these mutations might be more severe than *rsw1-1*, reducing root-tip growth more than *rsw1-1* or stacking with *rsw1-1* to an increased effect.

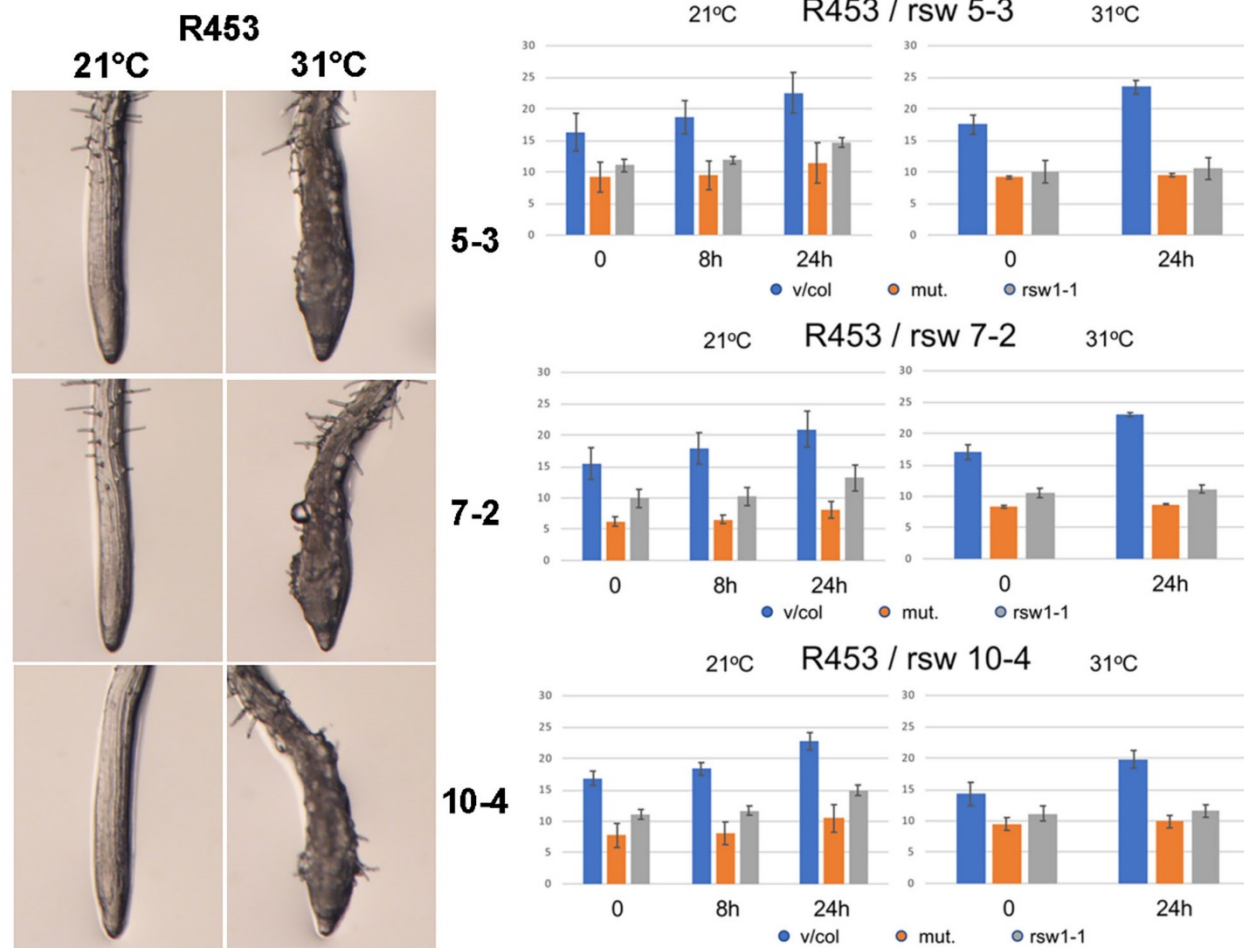


Figure 5.25 Primary root growth of the Arg453Lys (equivalent to *fra6* in *Atcesa8* (Zhong et al., 2003)) *Atcesa1* transgene in *rsw1-1* after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures. Three homozygous lines (5-3, 7-2 and 10-4) were tested. Primary root length mean values \pm standard deviation (error bars) (n=4) were measured from scanned images in mm. The pFGC5941 vector in Col-0, Arg453Lys *Atcesa1* transgenes and *rsw1-1* control are shown as blue, orange and grey, respectively. (Experiments by Anna T. Olek)

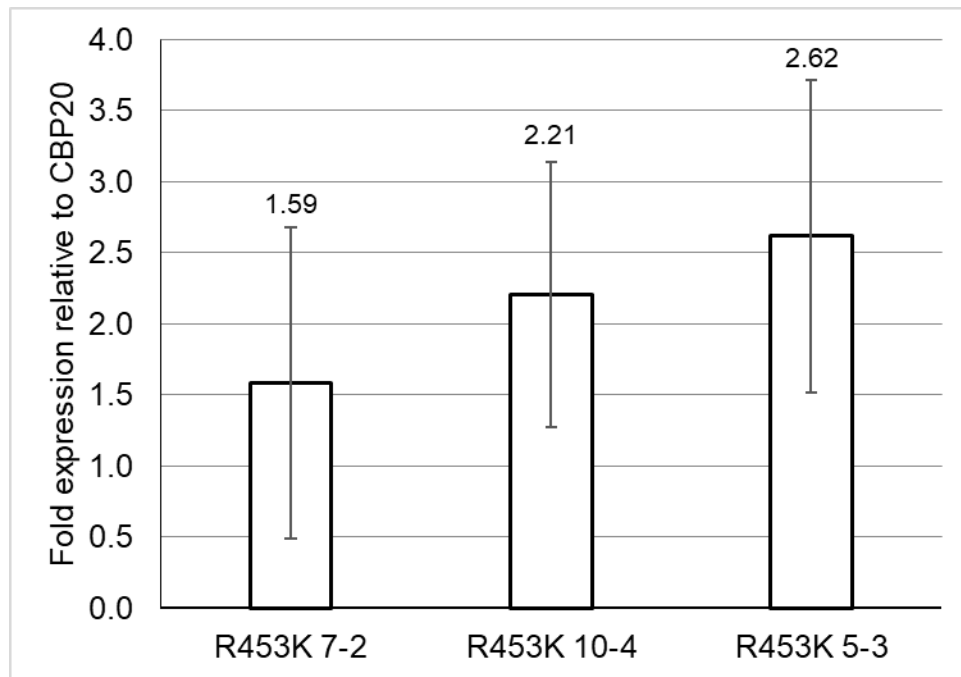


Figure 5.26 Arg453Lys *Atcesal* transgene expression in the *rsw1-1* mutant background relative to *CBP20*. Independent lines are shown as mean values \pm standard deviation (error bars) above their bar graphs (n=3).

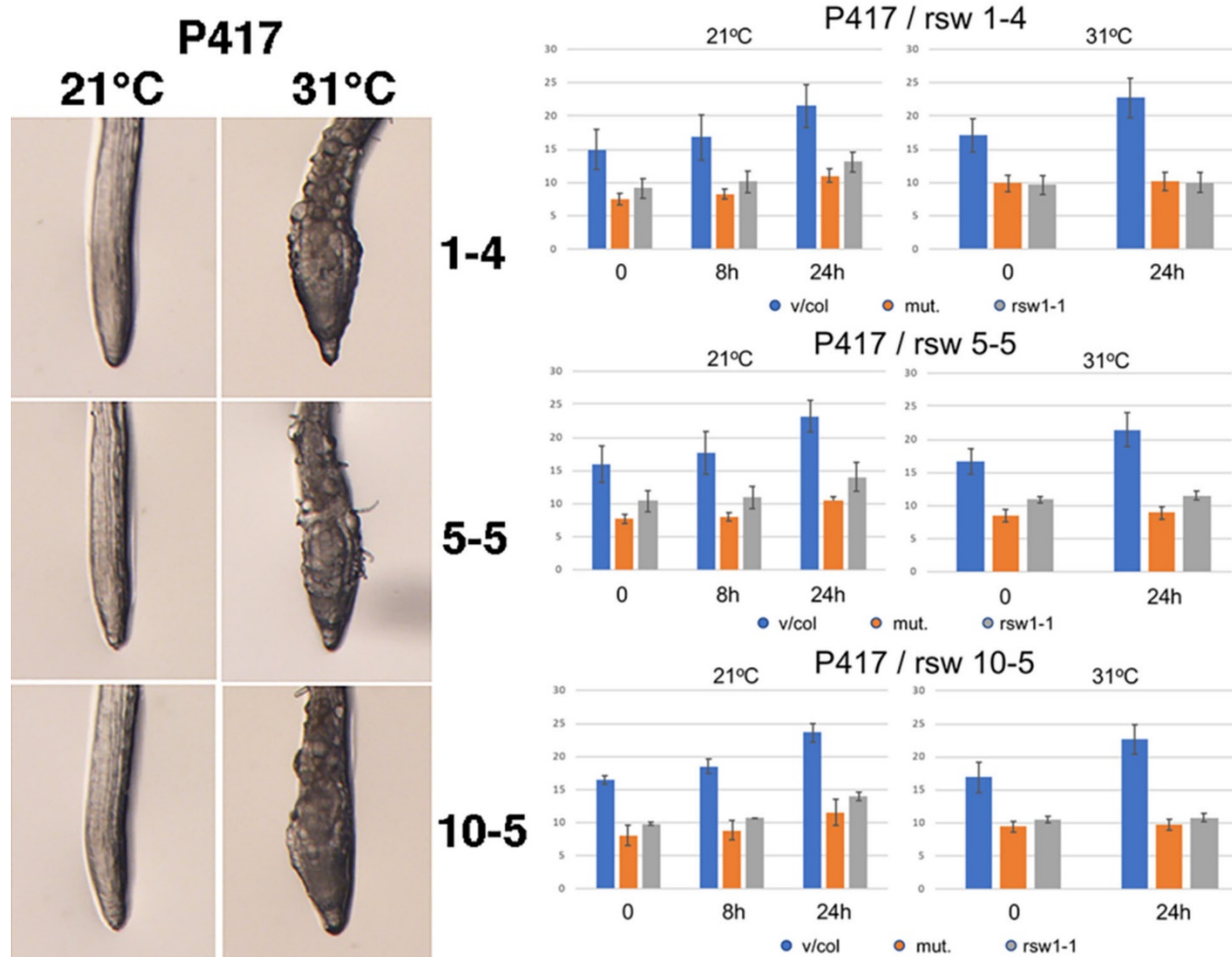


Figure 5.27 Primary root growth of the Pro417Ala *Atcesal* transgene in *rsw1-1* after five days at the permissive temperature (21°C) followed by one day at the permissive (21°C) or restrictive (31°C) temperatures. Three homozygous lines (1-4, 5-5 and 10-5) were tested. Primary root length mean values \pm standard deviation (error bars) (n=4) were measured from scanned images in mm. The pFGC5941 vector in Col-0, Pro417Ala *Atcesal* transgenes and *rsw1-1* control are shown as blue, orange and grey, respectively. (Experiments by Anna T. Olek)

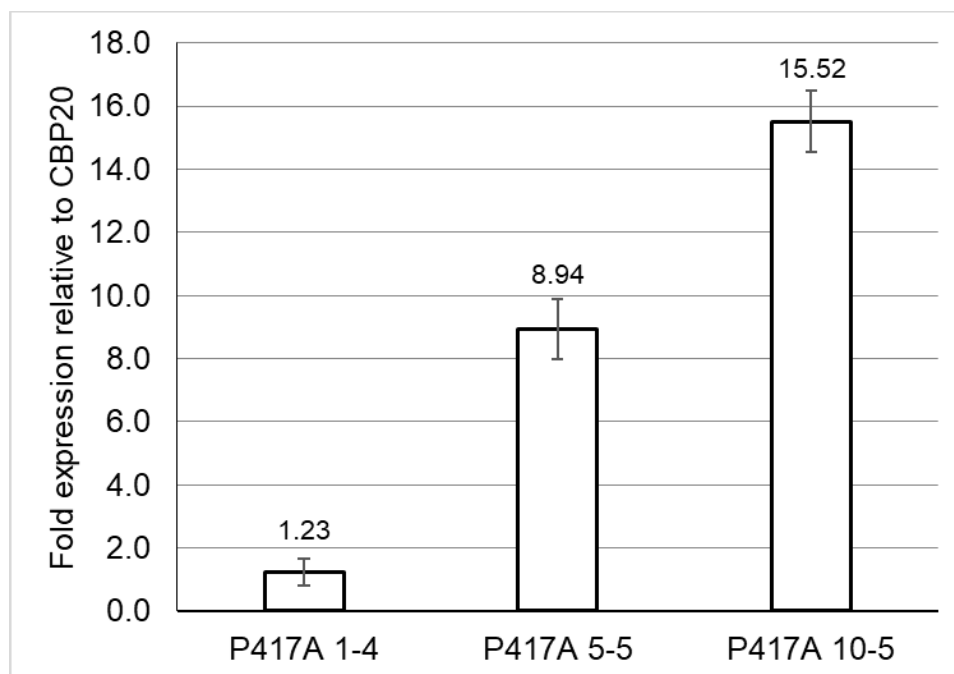


Figure 5.28 Pro417Ala *Atcesa1* transgene expression in the *rsw1-1* mutant background relative to *CBP20*. Independent lines are shown as mean values \pm standard deviation (error bars) above their bar graphs (n=3).

5.4 Discussion

Evaluating the atomic interactions and molecular surface of the P-CR as it relates to the CatD and the catalytic core model has identified regions and features that might be important for hypothetical interactions near the substrate entry portal. Other hypotheses for P-CR function are to facilitate oligomeric interactions between CESAs in the CSC or with a CESA-interacting protein. From analysis of the models generated in previous chapters, several mutants were designed to probe the highly conserved residues of the sequences identified in these models and to determine if they are important for CESA or CSC. In order to evaluate their importance in an *in vivo* assay, it was necessary to apply the knowledge gained from the X-ray crystal and SAXS structures of OsCESA8 to the homologous AtCESA1 P-CR. Due to the high sequence conservation of the P-CR, particularly in the structurally important residues, homology modeling based on the OsCESA8 P-CR creates a nearly identical model for the AtCESA1 P-CR, and likely any class of P-CR from plants. Thus, mutants and potential interactions informed by the OsCESA8 structures can be

directly applied to *AtCESA1* *in vivo*. Preliminary testing on control and homozygous transformants indicates that the *AtCESA1* transgene can complement the *rsw1-1* mutant with some mutants showing complete complementation and others showing partial or poor complementation (Table 5.1 and Figure 5.29). Based on the location of these mutated residues relative to hypothetically important regions in and around the P-CR, it is possible to increase the understanding of the roles structural elements of the P-CR might play.

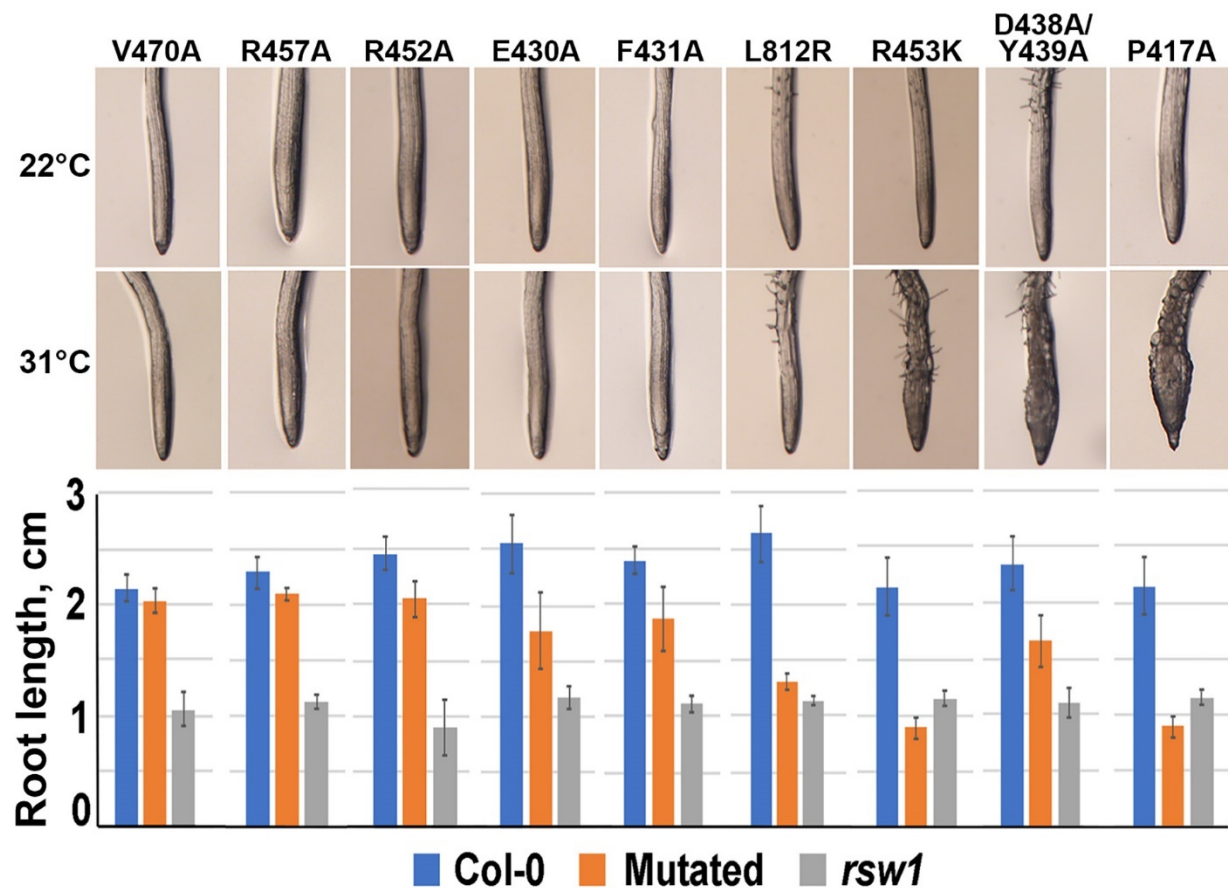


Figure 5.29 Root morphology and Primary root length of P-CR mutants grown for five days at the permissive temperature (22°C) followed by one day at the permissive (22°C) or restrictive (31°C) temperatures. Primary root length mean values \pm standard deviation (error bars) (n=4). The pFGC5941 vector in Col-0, *Atcesa1* transgenes and *rsw1-1* control are shown as blue, orange and grey, respectively. (Experiments by Anna T. Olek).

5.4.1 Mutations that Fully or Partially Complement

Several of the mutants fully complemented the phenotypes of *rsw1-1*, indicating a functional AtCESA1 can be produced even with these changes. Preliminary inferences on the oligomeric state of CESA1 can be drawn from some of these results. For example, in AtCESA1, Arg457 does not appear to be critical to function, as the Arg457Ala mutant completely complemented the *rsw1-1* radially swollen root tip and growth inhibition phenotypes (Figure 5.29). This indicates the P-CR might not form a trimer using the contacts in the P-CR crystal (Figure 2.12). As the OsCESA8 Arg458 was only circumstantially associated with the VGIC-based P-CR tetramer models, the complete complementation of the *Atcesal* Arg457Ala mutant does not support or refute this model but would indicate Arg457 is not important to this tetramer if it exists. A follow up mutant to probe this question could be a mutation of the AtCESA1 Arg457 to Pro, which could interfere with α -helical interactions in HX2. An *Atcesal* Arg457Pro mutant could interfere greatly with the folding around Cav 2 or the P-CR, which could be more impactful on function of the P-CR.

The *Atcesal* Val470Ala mutant also complemented both the *rsw1-1* radially swollen root tip and growth inhibition phenotypes (Figure 5.29). This does not support the hypothesis that the P-CR interacts with the gating loop near the FxVTxK motif, but alone cannot refute this hypothesis. In BCSA, interactions between the gating loop and the IF2 and IF3 helices are thought to be maintained largely by weak hydrophobic interactions between the gating loop Phe503 and Val505 residues and catalytic core Ile377, Leu487 and Ile520 residues (Morgan et al., 2016). Thus, it is reasonable that P-CR interaction with this loop might also involve weak hydrophobic interactions that allow for a strong enough interaction to achieve one position state, and weak enough that it can transition to another state, as in BCSA. Hypothetically, altering the hydrophobic contributions from the P-CR could disrupt this interaction. These considerations led to the creation of the *Atcesal* Val470Ala mutant, which slightly lowers the hydrophobicity of the R-group and creates a smaller surface for the BCSA-homologous gating loop to “grind” against. The mutation from Val to Ala is only a slight change in structure and hydrophobicity, which might not have been significant enough to observably alter function. Mutating AtCESA1 Val470 to a larger hydrophobic or to a larger charged residue, such as *Atcesal* Val470Phe or Val470Asp mutants, might be more impactful. Alternatively, other residues near the BCSA gating loop in this model, such as AtCESA1 Thr408, Ile466, Asn467 or Ala471, could be mutated to further test this hypothesis.

AtCESA1 Arg452 does not appear to be critical for function as the *Atcesa1* Arg452Ala mutants showed complete complementation of the *rsw1-1* radially swollen root tip phenotype and only slight growth inhibition at the restrictive temperature (Figure 5.29). The potential interaction at Cav 2 in which Arg452 participates could depend more on its other features whose mutation, either combined with the AtCESA1 Arg452 or alone, could be more informative to an interaction here. An additional mutant could be a mutation of the AtCESA1 Arg452 to Glu, which would introduce a residue with a more similar size but would alter the charge and remove the potential for a π - π stacking interaction with AtCESA1 Tyr439. This would theoretically be a much larger structural and functional disturbance, which might subsequently alter the AtCESA1 Tyr439 orientation in a way that disturbs its function.

The *Atcesa1* Glu430Ala mutant also complemented the *rsw1-1* radially swollen root tip phenotype with slight growth inhibition in some lines (Figure 5.29). Overexpression, however, in one line produced increased growth and might indicate transcript abundance and/or currently unknown phenotypes associated with the Glu430Ala mutation that are capable of slightly increasing growth relative to positive controls. This indicates that the hydrogen bonds between the AtCESA1 Glu430 and the P-CR coiled-coil helix HX2 are not critical to maintain the connector loop position, as a loss of connector loop interactions with the coiled-coil would be predicted to dramatically alter the P-CR structure and presumably its function. Furthermore, the P-CR tetramer model or potential interactions with the catalytic core are less likely but could also theoretically occur without the OsCESA8 Glu431 R-group in those models. Some of the other mutants, like the *Atcesa1* Leu812Arg and Phe431Ala mutants, might also provide information on the hypothetical interactions near the AtCESA1 Glu430. An additional Glu430Arg mutant could also be utilized to test whether a charge change and increase in residue size potentially clashes with an interaction near the catalytic core IF2 helix, which the *Atcesa1* Glu430Ala mutant might have been tolerated due to its smaller size.

Like the *Atcesa1* Glu430Ala mutant, the *Atcesa1* Phe431Ala mutant (aligned to OsCESA8 Trp432) complemented the *rsw1-1* phenotypes, with only slight growth inhibition in some lines (Figure 5.29). One line had radially swollen root tips and significant growth inhibition, but also poorly expressed the transgene. Together this implies the AtCESA1 P-CR does not make a critical interaction with the catalytic core dependent on Glu430 or Phe431. This also does not support the VGIC-based P-CR tetramer hypothesis, as this is the most prominent residue in that interaction.

The slight growth inhibition observed in some lines potentially links these nearby residues and could be worth investigating further. However, given the lack of a clear phenotype tied to these mutants or consistent evidence for growth inhibition, which could be variation between lines, other mutants and hypotheses are more likely to report on the function of the P-CR and structure of CESAs and the CSC than mutants Arg452, Glu430, Phe431 or other nearby residues.

5.4.2 Mutations that Poorly Complement

The *Atcesal* Leu812Arg transgene poorly complemented the *rsw1-1* phenotype, producing a different phenotype (Figure 5.29). The mutant complements *rsw1-1* root swelling, but significantly inhibits growth at restrictive and permissive temperatures. Thus, this mutant supports a model where the P-CR interacts with the catalytic core near AtCESA1 Leu812 in a way that is important for catalytic function (Figure 5.30), which could be related to substrate selectivity or interactions with the FxVTxK motif loop. However, the complementation of the *Atcesal* Glu430Ala and Phe431Ala casts doubt on the hypothesis that the small α -helix on the connector loop interacts with this region through Phe431 and/or Glu430, as they would be expected to elicit a similarly severe phenotype as Leu812Arg if this interaction were disturbed from the P-CR side.

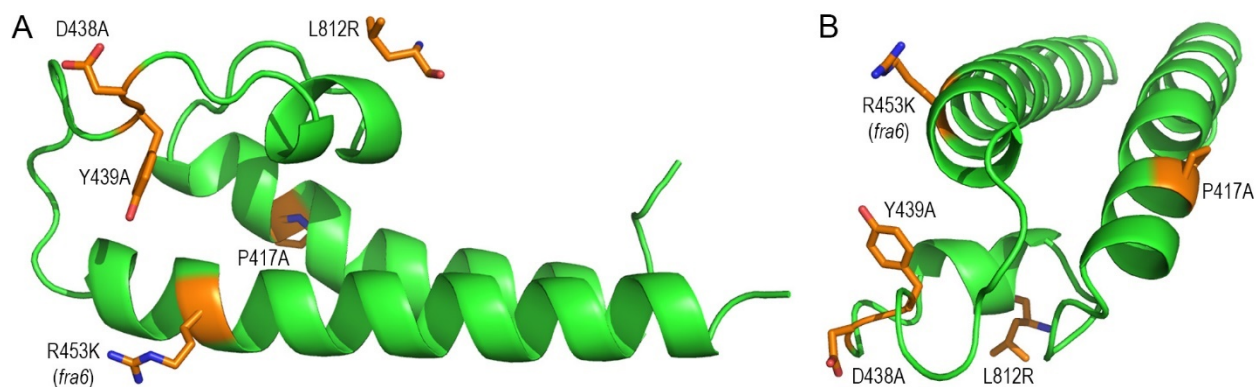


Figure 5.30 *Atcesal* mutants that poorly complemented *Arabidopsis thaliana* *rsw1-1* mutant (orange, stick models), shown on the best homology model of the AtCESA1 P-CR domain using RaptorX (Kallberg et al., 2012). L812R is part of the BCSA homologous catalytic core, modeled relative to the P-CR from the best fit of the P-CR to the CatD (other catalytic core residues not shown).

Alternatively, the *Atcesal* Leu812Arg mutant could interfere with catalytic activity as part of the catalytic core of CESA. Without the extended C-terminus of BCSA the N-terminus of the

IF2 helix appears solvent exposed, which is then covered by the P-CR in the best fit model of the P-CR with the CatD, as mentioned above. For this reason Arg was chosen to replace Leu812 as it might interfere with the hypothetical P-CR interaction, but if not interacting with the P-CR might be solvent exposed and unlikely to interfere with the catalytic core, as it is similar in size to the buried Phe it aligns to in BCSA. The results indicate neither an interaction with the P-CR nor is CESA functional with the Arg mutation as hypothesized. This produces four hypotheses from the models explored in this work, CESA misfolding, interference with the CSC or CESA-interacting proteins, an inactivated catalytic core or disturbed interactions with the loop containing the FxVTxK motif. The latter can be modeled through the homology between BCSA and CESA as similarity in both the catalytic core and FxVTxK motif loop exists. However, due to uncertainty in the structural homology between BCSA and CESA in the FxVTxK loop, the N-terminus of the IF2 helix and the C-terminus following the final TMH, direct molecular interpretations of the *Atcesal* Leu812Arg mutant and other nearby mutants should be approached with caution. Additional mutations at this site could be the use of different mutants at Leu812 to modify the properties of the end of the IF2 helix such as *Atcesal* Leu812Ser/Trp/Glu mutants.

The *Atcesal* Asp438Ala/Tyr439Ala double mutant transgene does not complement the *rswl-1* radially swollen root tip and growth inhibition phenotypes, but showed primary root growth significantly greater than *rswl-1* (Figure 5.29). Based on the complementation observed in mutating AtCESA1 Arg452, it is also less likely that the AtCESA1 Tyr439 is potentially important through the π - π stacking interaction with AtCESA1 Arg452 (Figure 5.30). This means the presumably more exposed hydrophobic surface of the AtCESA1 Tyr439 in the hypothetically unstacked *Atcesal* Arg452Ala mutant must be accommodated by the P-CR and Cav 2, perhaps through reorganized but functional packing. This indicates the electrostatic interactions of the AtCESA1 Asp438 and Tyr439 might be more important than the π - π stacking interaction of AtCESA1 Tyr439 with AtCESA1 Arg452, perhaps in a general binding site, hypothetical tetrameric interaction or through a P-CR crystallographic trimer. Single mutants of the AtCESA1 Asp438 and Tyr439 could provide information on whether just one of these residues could be responsible for the loss of complementation. Future experiments on the *Atcesal* Arg453Lys (equivalent to *fra6* in *Atcesa8* (Zhong et al., 2003)) and Arg457Ala mutants might also be informative to hypothetical interactions at Cav 2 as they form a prominent feature of Cav 2 with distinct interactions observed in the P-CR trimeric interaction. Additional mutants exploring

whether the highly conserved OsCESA8 Glu461 that hydrogen bonds to the OsCESA8 Tyr440, the less conserved OsCESA8 Ser435 hypothetically hydrogen bonding to the OsCESA8 Asp439, or the highly conserved OsCESA8 Lys464 near Asp439 could indicate whether the reciprocal loss of hydrogen bonding causes a similar lack of complementation.

The *Atcesa1* Arg453Lys mutant fails to complement both *rsw1-1* radially swollen root tip and growth inhibition phenotypes, with growth inhibition as bad or worse in some lines compared to *rsw1-1* at both the permissive and restrictive temperatures (Figure 5.29). The *rsw1-1* radially swollen root tip did not appear at permissive temperatures, however, most likely indicating this is not a phenotype associated with the *Atcesa1* Arg453Lys mutant, but most likely *rsw1-1*. This demonstrates the *fra6* mutation causes phenotypes resulting from both primary and secondary cell wall isoforms. Although AtCESA1 Arg453 appears critical to the homologous OsCESA8 trimer model, complementation of the *Atcesa1* Arg457Ala mutant, aligned to apparently critical residue OsCESA8 Arg458 in this interaction, does not support this. Thus, the source of the *Atcesa1* Arg453Lys mutant's failure to complement does not appear to be due to interference with a P-CR or CESA trimer. Similarly, the VGIC-based tetramer model is not supported by other residues that appeared more important, such as the *Atcesa1* Phe431Ala mutant. Thus, while the *Atcesa1* Arg453Lys and *Atcesa8 fra6* Arg362Lys mutants are critical, the functional source and molecular mechanism of the interference with CESAs remains unclear (Figure 5.30). The severity of the phenotypes associated with *fragile fiber* mutations in secondary cell walls leads to the assumption that normal CESA functions are severely disturbed. In primary cell walls, *rsw* mutations are often associated with similarly severe disruptions of normal CESA functions. The lack of any phenotype at permissive temperatures for this mutant indicates the *Atcesa1* Arg453Lys does not create a *rsw* phenotype, assuming it is not temperature sensitive like *rsw1-1*, which is reasonable as the *Atcesa8 fra6* Arg362Lys is not temperature sensitive. This supports the idea that severe secondary cell wall mutations do not exactly correlate with severe primary cell wall mutations. More likely, the radially swollen root tip and growth inhibition phenotypes of *Atcesa1* mutants, the latter being a phenotype of *Atcesa1* Arg453Lys, are independent of one another. The independence of these phenotypes is also observed in the *Atcesa1* Leu812Arg mutant and might be an important clue to how these mutants alter function. Finally, the *Atcesa1* Asp438Ala/Tyr439Ala double mutant did not complement the *rsw1-1* phenotype; these residues also surround Cav 2 in the OsCESA8 P-CR crystal structure. One hypothesis could be that *Atcesa1* Arg453Lys, Leu812Arg and

Asp438Ala/Tyr439Ala double mutant all contribute to a similar functional interference near Cav 2. However, some Cav 2 mutants in this study complemented *rsw1-1*, such the *Atcesa1* Arg452Ala and Arg457Ala mutants. Thus, determining whether mutant phenotypes observed for these Cav 2 mutants are independent and how specific features of Cav 2 alter function in the P-CR, CESA and CSC should contribute greatly to the understanding of plant cellulose synthesis.

The OsCESA8 Pro418 creates a well-resolved and prominent bend in the P-CR HX1 α -helix near the aromatic hydrophobic core formed between HX1, HX2 and the connector loop (Figure 5.30). This is due to the inability of Pro to form hydrogen bonds in the backbone α -helix of HX1. In Chapter 2, I hypothesized that this bend is necessary to accommodate the packing of several aromatic residues arranged in a coiled-coil knob-in-hole (KIH) heptad repeat, which in most coiled-coils consists of smaller aliphatic residues. This hypothesis is supported by the failure of the *Atcesa1* Pro417Ala mutant to complement the radially swollen root tip phenotype at restrictive temperatures and a strong growth inhibition phenotype at both permissive and restrictive temperatures (Figure 5.29). This might indicate misfolding of the P-CR, if the hydrophobic core becomes unstable without this bend. This could lead to a failure to form the hydrophobic core or exposed hydrophobic residues that could lead to misfolding or degradation of the entire CESA. Alternatively, the *Atcesa1* Pro417Ala mutant could interfere with a P-CR function specifically, which could be tied to the bend in HX1 it creates. The contribution of the aromatic residues is likely to create a strong hydrophobic core and stabilize the cavities identified in this study between α -helices and random coil structures. This conserved Pro residue might be a critical part of forming the P-CR structure and could be essential for folding or function. Future experiments could investigate whether this mutant folds properly *in vitro* and in plants. If that is so, this mutant could be characteristic for the loss of a more discrete loss of function in CESA.

5.4.3 Future Work

Based on the results of these initial experiments, the hypothetical trimer, tetramer and interactions between the P-CR connector loop (through Trp432 or other aromatics) and the catalytic core are not supported as mutants to hypothetically critical residues do not compromise CESA function. Other hypotheses, such as interactions between the CESA loop containing the FxVTxK motif and catalytic core near OsCESA8 Leu813 in the OsCESA8 homology model were supported but require further experimentation. This also includes the hypothesis that the OsCESA8 P-CR bend

in the HX1 α -helix at Pro418 is critical, as well as the hypothesis that P-CR Cav 2 is critical. Results for Cav 2 were mixed with some mutants complementing (*Atcesa1* Arg452Ala and Arg457Ala) and others not complementing *rsw1-1* (*Atcesa1* Arg453Lys (*fra6*) and Asp438Ala/Tyr439Ala double mutant) making it unclear what the function of Cav 2 might be.

In order to understand complementation analyses of these mutants in the *rsw1-1* background, several experiments need to be completed. The first consideration is whether the experiments can be done with dark-grown rather than light-grown seedlings. Dark-grown etiolated seedlings are advantageous as a system because root growth occurs at a higher rate. A more thorough analysis of the phenotypes of these mutants is needed, including replicates, and measurement of changes in root tip diameter, root morphology, as well as root hairs. Another experiment is to determine whether a dominant negative phenotype is observed in the Leu812Arg, Pro417Ala or Asp438Ala/Tyr439Ala double mutant at or below the permissive temperature (23°C) for *rsw1-1*. The *Atcesa8 fra6* mutant was reported to not show a dominant negative in the heterozygous form. The *Atcesa1* Arg453Lys mutant could be a useful tool in transheterozygote mutant experiments. Light microscopy experiments could also employ selective staining of the samples to look for differences in sugar deposition in the PCW, such as Calcofluor-white staining of weakly reduced β -glucans (Maeda et al., 1967) compared to the Congo red staining of reducing sugars (Wood, 1980). This could be accompanied by field emission scanning electron microscopy images of mutant roots and hypocotyl stems to look for alterations in cell morphology compared to controls, which might take the form of anisotropic growth and disturbed cell expansion/elongation as is sometimes observed in CESA mutants (Fagard et al., 2000; Daras et al., 2009; Fujita et al., 2013).

Growth phenotypes in *cesa* mutants are not always directly correlated with cellulose synthesis rates, with some mutants showing wild-type levels of cellulose, such as in the *any1* mutant (Fujita et al., 2013). Thus, the most critical experiments are to determine the amount of cellulose produced by mutant and control plants at or below the permissive temperature (23°C) and restrictive temperature (31°C) for *rsw1-1*. Cellulose deposition can be measured during growth at restrictive temperatures as a percentage of ^{14}C incorporation into cellulose using an ectopic addition of D-[U- ^{14}C]-Glucose to growing hypocotyl and roots, which can be done similar to methods in (Daras et al., 2009).

A series of biophysical experiments could then be performed to look at overall changes in cell wall structure, cellulose microfibril crystallinity and microfibril structure. Fourier-transform infrared spectroscopy has been used to look at changes in cell wall composition through changes in characteristic peaks of cellulose, other sugars and lignin when comparing the mutants to either an *rsw1-1/AtCESA1* positive control or *rsw1-1* negative control (Sene et al., 1994). Cellulose microfibril crystallinity is traditionally evaluated by the quality of the X-ray fiber diffraction pattern but can also be evaluated by ssNMR techniques. ssNMR can also be used to look at the polarization of cellulose microfibrils, which combined with either field emission scanning electron microscopy or atomic force microscopy imaging can reveal mesoscale structures of the cell wall.

The *AtCESA1* transgene used in this work is capable of complementing the *rsw1-1* phenotype, implying that it replaces the deficient *rsw1-1 Atcesal* in the CSC. The *Atcesal* Leu812Arg, Pro417Ala or Asp438Ala/Tyr439Ala double mutants that did not complement the *rsw1-1* phenotype could result from failure to produce the transgenic *Atcesal* mutant protein, misfolding of this protein, failure to properly traffic to the plasma membrane, failure to properly integrate into a CSC, disrupted catalysis or improper recycling from the plasma membrane. Synthesis of the mutant protein could be established by Western blot detection using antibodies capable of distinguishing between the mutant protein, the *rsw1-1* protein and other CESA isoforms. Alternatively, it might be possible to take advantage of the *rsw1-1* phenotype in that it leaves the plasma membrane at the restrictive temperature (Arioli et al., 1998; S. Chen et al., 2010; Fujita et al., 2013), while another mutant protein might not. However, there might still be *AtCESA1* epitopes in other membranes, in particular the Golgi apparatus membranes. Thus, separating tissues might be necessary or histological staining might be preferred. Quantitative proteomics using label-free mass spectrometry could also be a good method for detecting mutant proteins versus *rsw1-1* protein, similar to previous isoform stoichiometry experiments (X. Zhang et al., 2018).

If mutant proteins are produced, the next question becomes whether they are being properly trafficked to the plasma membrane and integrating into the CSC. One approach is to take advantage of the well-established fluorescently tagged CESAs, such as the YFP-*AtCESA6* transgene to monitor trafficking, motility and co-localization using confocal microscopy (Paradez et al., 2006; S. Chen et al., 2010; Lei et al., 2012; Fujita et al., 2013; Slabaugh et al., 2014b, 2016; W. Zhang et al., 2019). Known to complement the *Atcesa6 procuste1 (prc1-1)* mutant (Fagard et al., 2000),

the YFP-AtCESA6 could act as a reporter on the proper trafficking and formation of the CSC when a mutant protein is expressed. Furthermore, fluorescently tagging an *AtCESA1* transgene, including some of these mutants, could allow for experiments that test for co-localization with the YFP-AtCESA6 or other fluorescently tagged proteins related to cellulose synthesis. Other known mutant proteins or CESA-interacting proteins related to cellulose synthesis could also be integrated in complementation experiments in order to look for phenotypic dominance or similarity, potentially informing on a role for a *cesa* mutant(s) related to these other proteins. Finally, while establishing cellulose synthesis quantitatively is critical, another method for assessing catalysis and cellulose microfibril synthesis is through the motility of the fluorescent proteins integrated into the CSC, as CSCs are known to move through the membrane along microtubule tracks when actively synthesizing cellulose microfibrils.

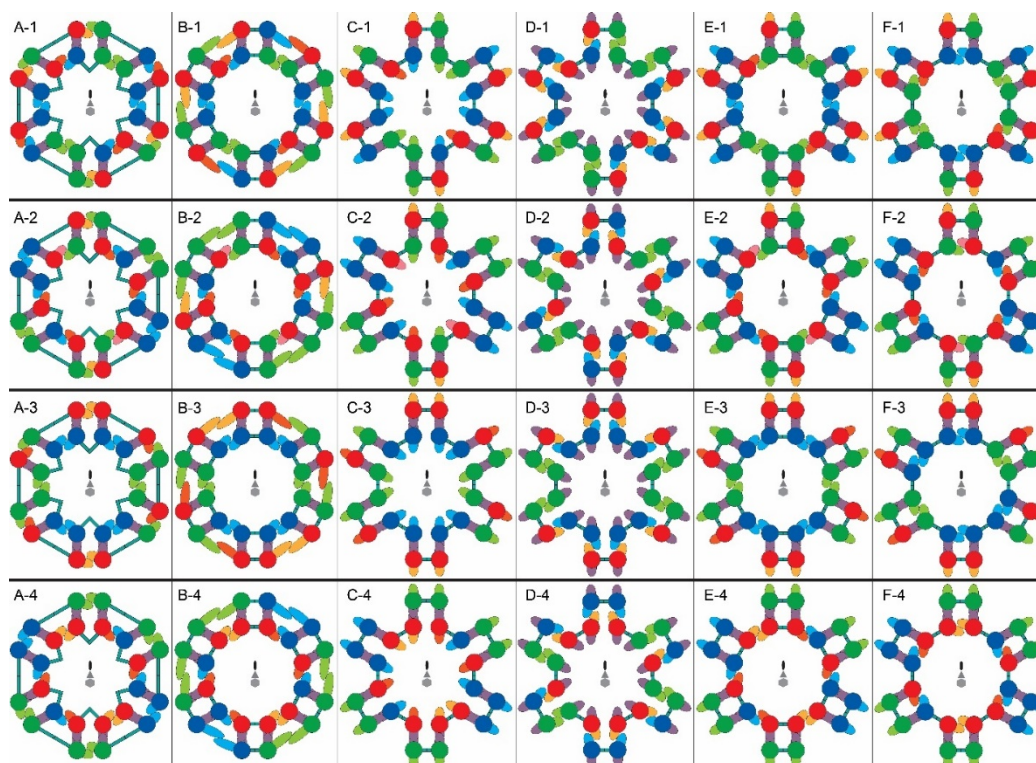
5.5 Acknowledgments

Phillip S. Rushton designed and cloned the P-CR mutants, selected for homozygosity, measured expression for a subset of mutants and performed preliminary complementation experiments. Anna T. Olek cloned *AtCESA1* into Columbia (Col-0) and performed complementation experiments. Some methods and figures were created by Anna T. Olek, as indicated in the text.

APPENDIX

Appendix 1.1 Supplemental Figures

Appendix 1.1 Figure 1 Additional hypothetical models of the CSC using P-CR dimers (light purple ellipse) to form the CSC particle. The circular catalytic core of a single CESA is colored by its class for the critical isoform class (red) or less critical isoforms (green and blue). The CSR (ellipse) is colored light green or blue for the less critical isoforms and differentially for the critical isoform to reflect different positions and/or function within a model (light/dark orange and pink). Possible models for completely filled interactions (A,B), minimally connected (C,D) and intermediate connectivity (E,F) are shown. Four fundamental symmetries describe the critical isoforms position; a 3:1 outer ring to inner ring ratio (A,B,C,D,E,F-1), a 1:3 outer ring to inner ring ratio (A,B,C,D,E,F-2), all critical isoforms in the outer ring (A,B,C,D,E,F-3) or all in the inner ring (A,B,C,D,E,F-4). CSR dimers of the less critical isoform are shown as homodimers but can also be heterodimers (not shown) or polarized across the CSC (not shown). ZnF (narrow teal boxes) homo- and heterodimeric interactions are shown within and between particles. Completely filled dimeric interactions of both the CSR and ZnF leads to long range interactions, assumed to be the ZnF (A) or potentially different variable length CSRs dimers (B). Assuming minimal dimeric interactions linking the isoforms but including the functionally critical and known dimeric interaction of the ZnF, implies the CSR (C) or P-CR (D) could have another role outside of core CSC oligomerization or be unbound. When intermediate connectivity of the CSR and ZnF is considered, models could have inner isoforms make CSR dimers within the CSC, while peripheral isoforms could have another role outside of core CSC oligomerization or be unbound, with CSR dimers between (E) or within (F) particles.



208

| | 1 | 23 | 4 | 56 | 7 | 8 | 9 | Rhodobacter_sphaeroides_3719093 | 370 | PETFSFQGRW | 383 |
|-----------------------------------|-----|---------------------|--------------|-------------------------|-------------------|--|-----|---------------------------------|---------------|------------|-----|
| Amborella_trichopoda_18432708 | 404 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 529 | 813 | PINLSRLNQVLRW | 826 | |
| Phoenix_dactylifera_103703779 | 394 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 519 | 805 | PINLSRLNQVLRW | 818 | |
| Elaeis_guineensis_105040177 | 394 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 519 | 805 | PINLSRLNQVLRW | 818 | |
| Oryza_sativa_4337958 | 397 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 520 | 804 | PINLSRLNQVLRW | 817 | |
| Brachypodium_distachyon_100821105 | 395 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 522 | 803 | PINLSRLNQVLRW | 819 | |
| Sorghum_bicolor_8080953 | 392 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 517 | 801 | PINLSRLNQVLRW | 816 | |
| Zea_mays_541802 | 394 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 519 | 803 | PINLSRLNQVLRW | 816 | |
| Musa_acuminata_103976459 | 394 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 519 | 806 | PINLSRLNQVLRW | 819 | |
| Nelumbo_nucifera_104592390 | 403 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 528 | 813 | PINLSRLNQVLRW | 826 | |
| Vitis_vinifera_100254404 | 402 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 527 | 813 | PINLSRLNQVLRW | 826 | |
| Ricinus_communis_8259478 | 402 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 527 | 811 | PINLSRLNQVLRW | 824 | |
| Jatropha_curcas_105637690 | 402 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 527 | 813 | PINLSRLNQVLRW | 826 | |
| Populus_trichocarpa_7459022 | 401 | AMLTFEALSETSEFAKWKV | PCKKHSIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 526 | 812 | PINLSRLNQVLRW | 825 | |
| Glycine_max_100810567 | 402 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 527 | 813 | PINLSRLNQVLRW | 826 | |
| Phaseolus_vulgaris_18619630 | 402 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 527 | 813 | PINLSRLNQVLRW | 826 | |
| Vigna_radiata_107616937 | 402 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 527 | 813 | PINLSRLNQVLRW | 826 | |
| Medicago_truncatula_11426766 | 404 | AMLSFALSETSEFAKWKV | PCKKHSIEPRAP | EYFLQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 529 | 815 | PINLSRLNQVLRW | 828 | |
| Fragaria_vesca_101314532 | 386 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 511 | 797 | PINLSRLNQVLRW | 810 | |
| Prunus_persica_18793862 | 389 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 514 | 800 | PINLSRLNQVLRW | 813 | |
| Dalus_domestica_103430335 | 389 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 514 | 800 | PINLSRLNQVLRW | 813 | |
| Cucumis_sativus_101205724 | 401 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 526 | 810 | PINLSRLNQVLRW | 823 | |
| Cucumis_melo_103503814 | 401 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 526 | 810 | PINLSRLNQVLRW | 823 | |
| Eucalyptus_Grandis_104437051 | 399 | AMLTFEALSETSEFAKWKV | PCKKHNIEPRAP | EYFAQKIDYLDKDKIQPSFVKER | RAMKREYEEFKVRINAL | VAKAKTPEEGWTMDGTPWGNPNRDPHGMQVFLHSGGLTDGNEPLRV | 524 | 810 | PINLSRLNQVLRW | 823 | |
| Citrus_sinensis_102614 | | | | | | | | | | | |

| | 1 | 23 | 4 | 56 | 7 | 8 | 9 | | | | | | |
|--|-----|---------------------|--------------|------------|-----|------|--------|---|---|-----|-----------------|-----------------|-----|
| <i>Amborella trichopoda</i> 18439664 | 405 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFAQKIDY | LKD | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGQSGGHDTEGNELPRIV | 530 | 824 | PINLSRDLNQLVRLW | 837 | |
| <i>Phoenix dactylifera</i> 103711978 | 401 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFQKRI | DY | LKD | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGQSGGHDTEGNELPRIV | 526 | 818 | PINLSRDLNQLVRLW | 831 |
| <i>Elaeis guineensis</i> 105052190 | 401 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFQKRI | DY | LKD | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGQSGGHDTEGNELPRIV | 526 | 819 | PINLSRDLNQLVRLW | 832 |
| <i>Oryza sativa</i> 4343710 | 401 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFQKRI | DY | LKD | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGQSGGHDTEGNELPRIV | 526 | 821 | PINLSRDLNQLVRLW | 834 |
| <i>Brachypodium distachyon</i> 100833360 | 401 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFQKRI | DY | LKD | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGQSGGHDTEGNELPRIV | 526 | 821 | PINLSRDLNQLVRLW | 834 |
| <i>Sorghum bicolor</i> 8080185 | 398 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFQKRI | DY | LKD | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGQSGGHDTEGNELPRIV | 523 | 819 | PINLSRDLNQLVRLW | 834 |
| <i>Zea mays</i> 541805 | 396 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFQKRI | DY | LKD | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGQSGGHDTEGNELPRIV | 521 | 815 | PINLSRDLNQLVRLW | 828 |
| <i>Musa acuminata</i> 103973182 | 402 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFQKRI | DY | LKD | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGQSGGHDTEGNELPRIV | 527 | 820 | PINLSRDLNQLVRLW | 830 |
| <i>Nelumbo nucifera</i> 104606920 | 402 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFQKRI | DY | LKD | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGQSGGHDTEGNELPRIV | 527 | 817 | PINLSRDLNQLVRLW | 833 |
| <i>Vitis vinifera</i> 100255463 | 412 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFAQKIDY | LKD | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGHNGVRDVEGNELPRIV | 537 | 826 | PINLSRDLNQLVRLW | 839 | |
| <i>Ricinus communis</i> 8273696 | 414 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFQKRI | DY | LKN | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGHSGRVDRVEGNELPRIV | 539 | 826 | PINLSRDLNQLVRLW | 839 |
| <i>Jatropha curcas</i> 105636590 | 414 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFQKRI | DY | LKN | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGHSGRVDRVEGNELPRIV | 539 | 827 | PINLSRDLNQLVRLW | 840 |
| <i>Populus trichocarpa</i> 7457156 | 413 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFQKRI | DY | LKN | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGHSGRVDRVEGNELPRIV | 538 | 826 | PINLSRDLNQLVRLW | 839 |
| <i>Glycine max</i> 100806786 | 412 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFAQKIDY | LKD | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGHSGRVDRVEGNELPRIV | 537 | 825 | PINLSRDLNQLVRLW | 838 | |
| <i>Phaseolus vulgaris</i> 18625246 | 410 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFAQKIDY | LKD | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGHSGRVDRVEGNELPRIV | 535 | 823 | PINLSRDLNQLVRLW | 836 | |
| <i>Vigna radiata</i> 106770139 | 405 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFAQKIDY | LKD | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGHSGRVDRVEGNELPRIV | 530 | 818 | PINLSRDLNQLVRLW | 831 | |
| <i>Medicago truncatula</i> 11428008 | 410 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFAQKIDY | LKD | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGHSGRVDRVEGNELPRIV | 535 | 823 | PINLSRDLNQLVRLW | 836 | |
| <i>Fragaria vesca</i> 101351050 | 411 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFALKIDY | LKD | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGHSGRVDRVEGNELPRIV | 536 | 823 | PINLSRDLNQLVRLW | 836 | |
| <i>Prunus persica</i> 18767664 | 414 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFAQKIDY | LKD | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGHSGRVDRVEGNELPRIV | 539 | 826 | PINLSRDLNQLVRLW | 839 | |
| <i>Malus domestica</i> 103432738 | 413 | AMLTFEALSETSEFAKRWV | PCKKFNIEPRAP | EWYFAQKIDY | LKD | KVLP | SVFKER | RAMKREYEEFKVIRINALVAKAKVPEEGWTMQDGTWPWGNVNRDHPGMIQVFLGHSGRVDRVEGNELPRIV | 538 | 825 | PINLSRDLNQLVRLW | 838 | |
| <i>Cucumis sativus</i> 101214 | | | | | | | | | | | | | |

Appendix 1.2 Figure 7 Clustal Omega HHM multiple sequence alignment (Sievers et al., 2011) of the *Selaginella moellendorffii* and *Physcomitrella patens* CESA P-CRs and BCSA homologous IF2 helices. This region of the IF2 helix is predicted to be potential catalytic core P-CR interaction based on the best fit model of the P-CR to the CatD SAXS density (Olek et al., 2014; Rushton et al., 2017). Sequences, labeled with species and NCBI GeneID, were retrieved from the KEGG orthology databases for CESAs (KO ID: K10999) (Kanehisa et al., 2016). Mutations designed in this study are shown in and numbered at the top. Sequence conservation and similarity defined by Clustal Omega as identical residues (labeled [*]), strongly similar properties (labeled [:], >0.5 in the Gonnet PAM 250 matrix), weakly similar (labeled [.] , ≤ 0.5 and > 0 in the Gonnet PAM 250 matrix) and dissimilar (blank, 0 in the Gonnet PAM 250 matrix) is shown below the alignments for the CESA1 class, a combination of all classes and *Rhodobacter sphaeroides* BCSA for the homologous IF2 helix.

```

                                1          23      4          56      7          8      ↓          Rhodobacter_sphaeroides_3719093_370_PETFASFIQQRGW_383
Selaginella_moellendorffii_9643759 408_AMLTFEALSETSEFARKWVFCKKFNIEPRAPEMYFAQKIDYLDKDKVQPSFVKERRAMKREYEEFKVRINALVAKAQKVPEEGWTMQDGTWPWPGNNTRDHPGMIQVFLGHSGGHDTEGNELPRLV_533|822_PINLSDRLNQVLRW_835
Selaginella_moellendorffii_9644533 409_AMLTFEGLSETSEFARKWVFFVKYKYNIEPRAPEMYFAQKIDYLDKDKVQPSFVKERRAMKREYEEFKVRVNMVAKAQKVPEEGWTMQDGTWPWPGNNTRDHPGMIQVFLGHSGGHDTEGNELPRLV_534|801_PINLSDRLHQVLRW_814
Selaginella_moellendorffii_9641937 404_AMLTFEALSETSEFARKWVFFCKKFSIEPRAPEMYFAQKIDYLDKDKVQPSFVKERRAMKREYEEFKVRMNALVAKAQKIPPEEGWTMQDGTWPWPGNNTRDHPGMIQVFLGHSGGHDTEGNELPRLV_529|809_PINLSDRLHQVLRW_822
Selaginella_moellendorffii_9656492 393_AMLTFEALSETSEFARKWVFFCKKFSIEPRAPEMYFAQKIDYLDKDKVQPTFVKERRAMKREYEEFKVRINALVAKAHKMPPEEGWTMQDGTWPWPGNNTRDHPGMIQVFLGHSGGHDTEGNELPRLV_518|819_PINLSDRLNQVLRW_832
Physcomitrella_patens_112285828 401_AMLTFEALSETSEFARKWVFFCKKFLIEPRAPEMYFAQKIDYLDKDKVQATFVKERRAMKREYEEFKVRVNALVAKAMKVPEEGWTMQDGTWPWPGNNRSDHPGMIQVFLGHSGGHDTGDELPRLV_526|810_PINLSDRLNQVLRW_823
Physcomitrella_patens_112276250 398_AMLTFCISSETSEFARKWVFFCKKFSIEPRAPEMYFAQKIDYLDKDKVQPTFVKERRAMKREYEEFKVRVNALVAKAQKVPEEGWTMQDGTWPWPGNNRSDHPGMIQVFLGHSGGHDTGDELPRLV_523|821_PINLSDRLNQVLRW_834
Physcomitrella_patens_112280291 400_AMLTFEALSETSEFARKWVFFCKKFTIEPRAPEMYFAQKIDYLDKDKVQPTFVKERRAMKREYEEFKVRVNALVAKALKVPEEGWTMQDGTWPWPGNNKSDHPGMIQVFLGHSGGHDTGDELPRLV_525|822_PINLSDRLNQVLRW_835
***** :***** ** :***** *****:***** :***:*****:*****:***** A*:***** ***** *****:***** ** :*****
All Classes      :: *: : :*:::. *:** :. :***:* ** *.**::*: *:::*** :** ** *: :* *: * * * * * * * * * * * * * * * * * * * * * * * *
                                Rhodobacter_sphaeroides_3719093 * :. :.* **

```

REFERENCES

- Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., . . . Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 2), 213-221.
- Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., . . . Ecker, J. R. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, 301(5633), 653-657.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25(17), 3389-3402.
- Amor, Y., Haigler, C. H., Johnson, S., Wainscott, M., & Delmer, D. P. (1995). A membrane-associated form of sucrose synthase and its potential role in synthesis of cellulose and callose in plants. *Proc Natl Acad Sci USA*, 92(20), 9353-9357.
- Anderson, C. T., Carroll, A., Akhmetova, L., & Somerville, C. R. (2010). Real-time imaging of cellulose reorientation during cell wall expansion in *Arabidopsis* roots. *Plant Physiol*, 152(2), 787-796.
- Andrade, M. A., Chacon, P., Merele, J. J., & Moran, F. (1993). Evaluation of secondary structure of proteins from UV circular-dichroism spectra using an unsupervised learning neural-Network. *Protein Eng*, 6(4), 383-390.
- Appenzeller, L., Doblin, M. S., Barreiro, R., Wang, H. Y., Niu, X. M., Kollipara, K., . . . Dhugga, K. S. (2004). Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (CesA) gene family. *Cellulose*, 11(3-4), 287-299.
- Arioli, T., Peng, L., Betzner, A. S., Burn, J., Wittke, W., Herth, W., . . . Williamson, R. E. (1998). Molecular analysis of cellulose biosynthesis in *Arabidopsis*. *Science*, 279(5351), 717-720.
- Aslanidis, C., & de Jong, P. J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res*, 18(20), 6069-6074.
- Atanassov, II, Pittman, J. K., & Turner, S. R. (2009). Elucidating the mechanisms of assembly and subunit interaction of the cellulose synthase complex of *Arabidopsis* secondary cell walls. *J Biol Chem*, 284(6), 3833-3841.

- Bannigan, A., Wiedemeier, A. M., Williamson, R. E., Overall, R. L., & Baskin, T. I. (2006). Cortical microtubule arrays lose uniform alignment between cells and are oryzalin resistant in the Arabidopsis mutant, radially swollen 6. *Plant Cell Physiol*, 47(7), 949-958.
- Bashline, L., Li, S., Anderson, C. T., Lei, L., & Gu, Y. (2013). The endocytosis of cellulose synthase in Arabidopsis is dependent on $\mu 2$, a clathrin-mediated endocytosis adaptin. *Plant Physiol*, 163(1), 150-160.
- Bashline, L., Li, S., & Gu, Y. (2014). The trafficking of the cellulose synthase complex in higher plants. *Ann Bot*, 114(6), 1059-1067.
- Bashline, L., Li, S., Zhu, X., & Gu, Y. (2015). The TWD40-2 protein and the AP2 complex cooperate in the clathrin-mediated endocytosis of cellulose synthase to regulate cellulose biosynthesis. *Proc Natl Acad Sci USA*, 112(41), 12870-12875.
- Baskin, T. I., Betzner, A. S., Hoggart, R., Cork, A. H., & Williamson, R. E. (1992). Root morphology mutants in Arabidopsis thaliana. *Aust J Plant Physiol*, 19(4), 427-437.
- Basu, D., Tian, L., Debrosse, T., Poirier, E., Emch, K., Herock, H., . . . Showalter, A. M. (2016). Glycosylation of a fasciclin-like arabinogalactan-protein (SOS5) mediates root growth and seed mucilage adherence via a cell wall receptor-like kinase (FEI1/FEI2) pathway in Arabidopsis. *PLOS ONE*, 11(1), e0145092.
- Beeckman, T., Przemeck, G. K., Stamatiou, G., Lau, R., Terryn, N., De Rycke, R., . . . Berleth, T. (2002). Genetic complexity of cellulose synthase a gene function in Arabidopsis embryogenesis. *Plant Physiol*, 130(4), 1883-1893.
- Benfey, P. N., Linstead, P. J., Roberts, K., Schiefelbein, J. W., Hauser, M. T., & Aeschbacher, R. A. (1993). Root development in Arabidopsis: four mutants with dramatically altered root morphogenesis. *Development*, 119(1), 57-70.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., . . . Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Res*, 28(1), 235-242.
- Bichet, A., Desnos, T., Turner, S. R., Grandjean, O., & Hofte, H. (2001). BOTERO1 is required for normal orientation of cortical microtubules and anisotropic cell expansion in Arabidopsis. *Plant J*, 25(2), 137-148.
- Bischoff, V., Desprez, T., Mouille, G., Vernhettes, S., Gonneau, M., & Hofte, H. (2011). Phytochrome regulation of cellulose synthesis in Arabidopsis. *Curr Biol*, 21(21), 1822-1827.

- Boex-Fontvieille, E., Davanture, M., Jossier, M., Zivy, M., Hodges, M., & Tcherkez, G. (2014). Photosynthetic activity influences cellulose biosynthesis and phosphorylation of proteins involved therein in *Arabidopsis* leaves. *J Exp Bot*, 65(17), 4997-5010.
- Boisson-Dernier, A., Lituiev, D. S., Nestorova, A., Franck, C. M., Thirugnanarajah, S., & Grossniklaus, U. (2013). ANXUR receptor-like kinases coordinate cell wall integrity with growth at the pollen tube tip via NADPH oxidases. *PLOS Biol*, 11(11), e1001719.
- Bosca, S., Barton, C. J., Taylor, N. G., Ryden, P., Neumetzler, L., Pauly, M., . . . Seifert, G. J. (2006). Interactions between MUR10/CesA7-dependent secondary cellulose biosynthesis and primary cell wall structure. *Plant Physiol*, 142(4), 1353-1363.
- Bowling, A. J., & Brown, R. M., Jr. (2008). The cytoplasmic domain of the cellulose-synthesizing complex in vascular plants. *Protoplasma*, 233(1-2), 115-127.
- Bringmann, M., Li, E., Sampathkumar, A., Kocabek, T., Hauser, M. T., & Persson, S. (2012). POM-POM2/cellulose synthase interacting1 is essential for the functional association of cellulose synthase and microtubules in *Arabidopsis*. *Plant Cell*, 24(1), 163-177.
- Brown, D. M., Zeef, L. A., Ellis, J., Goodacre, R., & Turner, S. R. (2005). Identification of novel genes in *Arabidopsis* involved in secondary cell wall formation using expression profiling and reverse genetics. *Plant Cell*, 17(8), 2281-2295.
- Burk, D. H., Liu, B., Zhong, R., Morrison, W. H., & Ye, Z. H. (2001). A katanin-like protein regulates normal cell wall biosynthesis and cell elongation. *Plant Cell*, 13(4), 807-827.
- Burn, J. E., Hurley, U. A., Birch, R. J., Arioli, T., Cork, A. H., & Williamson, R. E. (2002). The cellulose-deficient *Arabidopsis* mutant *rsw3* is defective in a gene encoding a putative glucosidase II, an enzyme processing N-glycans during ER quality control. *Plant J*, 32(6), 949-960.
- Burton, R. A., Shirley, N. J., King, B. J., Harvey, A. J., & Fincher, G. B. (2004). The *CesA* gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. *Plant Physiol*, 134(1), 224-236.
- Callis, J. (2014). The ubiquitination machinery of the ubiquitin system. *The Arabidopsis Book*, 12, e0174.
- Cano-Delgado, A. I., Metzlauff, K., & Bevan, M. W. (2000). The *eli1* mutation reveals a link between cell expansion and secondary cell wall formation in *Arabidopsis thaliana*. *Development*, 127(15), 3395-3405.

- Cano-Delgado, A. I., Penfield, S., Smith, C., Catley, M., & Bevan, M. (2003). Reduced cellulose synthesis invokes lignification and defense responses in *Arabidopsis thaliana*. *Plant J*, 34(3), 351-362.
- Carpita, N. C. (2011). Update on mechanisms of plant cell wall biosynthesis: how plants make cellulose and other (1->4)-beta-D-glycans. *Plant Physiol*, 155(1), 171-184.
- Carpita, N. C., & McCann, M. C. (2015). Characterizing visible and invisible cell wall mutant phenotypes. *J Exp Bot*, 66(14), 4145-4163.
- Carpita, N. C., Ralph, J., & McCann, M. C. (2015). The cell wall. In *Biochemistry & Molecular Biology of Plants*, eds B. Buchanan, W. Gruissem, R. Jones, & K. Vickers. Hoboken, NJ: Wiley Blackwell, 52-108.
- Carroll, A., Mansoori, N., Li, S., Lei, L., Vernhettes, S., Visser, R. G., . . . Trindade, L. M. (2012). Complexes with mixed primary and secondary cellulose synthases are functional in *Arabidopsis* plants. *Plant Physiol*, 160(2), 726-737.
- Carroll, A., & Specht, C. D. (2011). Understanding plant cellulose synthases through a comprehensive investigation of the cellulose synthase family sequences. *Front Plant Sci*, 2, 5.
- Catterall, W. A. (2014). Structure and function of voltage-gated sodium channels at atomic resolution. *Exp Physiol*, 99(1), 35-51.
- Chen, J., Sawyer, N., & Regan, L. (2013). Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci*, 22(4), 510-515.
- Chen, L., Wu, L. Y., Wang, Y., Zhang, S., & Zhang, X. S. (2006). Revealing divergent evolution, identifying circular permutations and detecting active-sites by protein structure comparison. *BMC Struct Biol*, 6, 18.
- Chen, S., Ehrhardt, D. W., & Somerville, C. R. (2010). Mutations of cellulose synthase (CESA1) phosphorylation sites modulate anisotropic cell expansion and bidirectional mobility of cellulose synthase. *Proc Natl Acad Sci USA*, 107(40), 17188-17193.
- Chen, S., Jia, H., Zhao, H., Liu, D., Liu, Y., Liu, B., . . . Somerville, C. R. (2016). Anisotropic cell expansion is affected through the bidirectional mobility of cellulose synthase complexes and phosphorylation at two critical residues on CESA3. *Plant Physiol*, 171(1), 242-250.

- Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., . . . Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, 66(1), 12-21.
- Chen, W., Li, Q., Cao, J., Liu, Y., Li, J., Zhang, J., . . . Yu, H. (2015). Revealing the structures of cellulose nanofiber bundles obtained by mechanical nanofibrillation via TEM observation. *Carbohydr Polym*, 117, 950-956.
- Chen, Z., Hong, X., Zhang, H., Wang, Y., Li, X., Zhu, J. K., & Gong, Z. (2005). Disruption of the cellulose synthase gene, AtCesA8/IRX1, enhances drought and osmotic stress tolerance in Arabidopsis. *Plant J*, 43(2), 273-283.
- Cheung, A. Y., & Wu, H. M. (2011). THESEUS 1, FERONIA and relatives: a family of cell wall-sensing receptor kinases? *Curr Opin Plant Biol*, 14(6), 632-641.
- Cho, S. H., Du, J., Sines, I., Poosarla, V. G., Vepachedu, V., Kafle, K., . . . Nixon, B. T. (2015). In vitro synthesis of cellulose microfibrils by a membrane protein from protoplasts of the non-vascular plant *Physcomitrella patens*. *Biochem J*, 470(2), 195-205.
- Cho, S. H., Purushotham, P., Fang, C., Maranas, C., Díaz-Moreno, S. M., Bulone, V., . . . Nixon, B. T. (2017). Synthesis and self-assembly of cellulose microfibrils from reconstituted cellulose synthase. *Plant Physiol*, 175(1), 146-156.
- Clough, S. J., & Bent, A. F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J*, 16(6), 735-743.
- Collings, D. A., Gebbie, L. K., Howles, P. A., Hurley, U. A., Birch, R. J., Cork, A. H., . . . Williamson, R. E. (2008). Arabidopsis dynamin-like protein DRP1A: a null mutant with widespread defects in endocytosis, cellulose synthesis, cytokinesis, and cell expansion. *J Exp Bot*, 59(2), 361-376.
- Connolly, M. L. (1993). The molecular surface package. *J Mol Graph*, 11(2), 139-141.
- Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res*, 14(6), 1188-1190.
- Crowell, E. F., Bischoff, V., Desprez, T., Rolland, A., Stierhof, Y. D., Schumacher, K., . . . Vernhettes, S. (2009). Pausing of Golgi bodies on microtubules regulates secretion of cellulose synthase complexes in Arabidopsis. *Plant Cell*, 21(4), 1141-1154.

- Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., & Scheible, W. R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol*, *139*(1), 5-17.
- Daras, G., Rigas, S., Penning, B., Milioni, D., McCann, M. C., Carpita, N. C., . . . Hatzopoulos, P. (2009). The thanatos mutation in *Arabidopsis thaliana* cellulose synthase 3 (AtCesA3) has a dominant-negative effect on cellulose synthesis and plant growth. *New Phytol*, *184*(1), 114-126.
- Dauter, Z., Dauter, M., & Dodson, E. (2002). Jolly SAD. *Acta Crystallogr D Biol Crystallogr*, *58*(Pt 3), 494-506.
- Delmer, D. P. (1999). CELLULOSE BIOSYNTHESIS: Exciting times for a difficult field of study. *Annu Rev Plant Physiol Plant Mol Biol*, *50*, 245-276.
- Desnos, T., Orbovic, V., Bellini, C., Kronenberger, J., Caboche, M., Traas, J., & Hofte, H. (1996). Procuste1 mutants identify two distinct genetic pathways controlling hypocotyl cell elongation, respectively in dark- and light-grown *Arabidopsis* seedlings. *Development*, *122*(2), 683-693.
- Desprez, T., Juraniec, M., Crowell, E. F., Jouy, H., Pochylova, Z., Parcy, F., . . . Vernhettes, S. (2007). Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA*, *104*(39), 15572-15577.
- Desprez, T., Vernhettes, S., Fagard, M., Refregier, G., Desnos, T., Aletti, E., . . . Hofte, H. (2002). Resistance against herbicide isoxaben and cellulose deficiency caused by distinct mutations in same cellulose synthase isoform CESA6. *Plant Physiol*, *128*(2), 482-490.
- Dhugga, K. S. (2001). Building the wall: genes and enzyme complexes for polysaccharide synthases. *Curr Opin Plant Biol*, *4*(6), 488-493.
- DiMaio, F., Terwilliger, T. C., Read, R. J., Wlodawer, A., Oberdorfer, G., Wagner, U., . . . Baker, D. (2011). Improved molecular replacement by density- and energy-guided protein structure optimization. *Nature*, *473*(7348), 540-543.
- Dimitroff, G., Little, A., Lahnstein, J., Schwerdt, J. G., Srivastava, V., Bulone, V., . . . Fincher, G. B. (2016). (1,3;1,4)-beta-glucan biosynthesis by the CSLF6 enzyme: position and flexibility of catalytic residues influence product fine structure. *Biochemistry*, *55*(13), 2054-2061.

- Ding, S. Y., & Himmel, M. E. (2006). The maize primary cell wall microfibril: a new model derived from direct visualization. *J Agric Food Chem*, *54*(3), 597-606.
- Ellis, C., Karafyllidis, I., Wasternack, C., & Turner, J. G. (2002). The Arabidopsis mutant *cev1* links cell wall signaling to jasmonate and ethylene responses. *Plant Cell*, *14*(7), 1557-1566.
- Ellis, C., & Turner, J. G. (2001). The Arabidopsis mutant *cev1* has constitutively active jasmonate and ethylene signal pathways and enhanced resistance to pathogens. *Plant Cell*, *13*(5), 1025-1033.
- Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, *60*(Pt 12 Pt 1), 2126-2132.
- Endler, A., Kesten, C., Schneider, R., Zhang, Y., Ivakov, A., Froehlich, A., . . . Persson, S. (2015). A mechanism for sustained cellulose synthesis during salt stress. *Cell*, *162*(6), 1353-1364.
- Endler, A., Schneider, R., Kesten, C., Lampugnani, E. R., & Persson, S. (2016). The cellulose synthase companion proteins act non-redundantly with CELLULOSE SYNTHASE INTERACTING1/POM2 and CELLULOSE SYNTHASE 6. *Plant Signal Behav*, *11*(4), e1135281.
- Facette, M. R., Shen, Z., Björnsdóttir, F. R., Briggs, S. P., & Smith, L. G. (2013). Parallel proteomic and phosphoproteomic analyses of successive stages of maize leaf development. *Plant Cell*, *25*(8), 2798-2812.
- Fagard, M., Desnos, T., Desprez, T., Goubet, F., Refregier, G., Mouille, G., . . . Hofte, H. (2000). PROCUSTE1 encodes a cellulose synthase required for normal cell elongation specifically in roots and dark-grown hypocotyls of Arabidopsis. *Plant Cell*, *12*(12), 2409-2424.
- Feng, W., Kita, D., Peaucelle, A., Cartwright, H. N., Doan, V., Duan, Q., . . . Dinneny, J. R. (2018). The FERONIA receptor kinase maintains cell-wall integrity during salt stress through Ca(2+) signaling. *Curr Biol*, *28*(5), 666-675.
- Feraru, E., Feraru, M. I., Kleine-Vehn, J., Martinière, A., Mouille, G., Vanneste, S., . . . Friml, J. (2011). PIN polarity maintenance by the cell wall in Arabidopsis. *Curr Biol*, *21*(4), 338-343.
- Fernandes, A. N., Thomas, L. H., Altaner, C. M., Callow, P., Forsyth, V. T., Apperley, D. C., . . . Jarvis, M. C. (2011). Nanostructure of cellulose microfibrils in spruce wood. *Proc Natl Acad Sci USA*, *108*(47), E1195-E1203.

- Fischetti, R., Stepanov, S., Rosenbaum, G., Barrea, R., Black, E., Gore, D., . . . Bunker, G. B. (2004). The BioCAT undulator beamline 18ID: a facility for biological non-crystalline diffraction and X-ray absorption spectroscopy at the Advanced Photon Source. *J Synchrotron Radiat*, 11(Pt 5), 399-405.
- Fox, N. K., Brenner, S. E., & Chandonia, J. M. (2014). SCOPe: structural classification of proteins-extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Res*, 42(Database issue), D304-D309.
- Fujimoto, M., Suda, Y., Vernhettes, S., Nakano, A., & Ueda, T. (2014). Phosphatidylinositol 3-kinase and 4-kinase have distinct roles in intracellular trafficking of cellulose synthase complexes in *Arabidopsis thaliana*. *Plant Cell Physiol*, 56(2), 287-298.
- Fujita, M., Himmelsbach, R., Ward, J., Whittington, A., Hasenbein, N., Liu, C., . . . Wasteneys, G. O. (2013). The anisotropy1 D604N mutation in the *Arabidopsis* cellulose synthase1 catalytic domain reduces cell wall crystallinity and the velocity of cellulose synthase complexes. *Plant Physiol*, 162(1), 74-85.
- Gardiner, J. C., Taylor, N. G., & Turner, S. R. (2003). Control of cellulose synthase complex localization in developing xylem. *Plant Cell*, 15(8), 1740-1748.
- Giddings, T. H., Jr., Brower, D. L., & Staehelin, L. A. (1980). Visualization of particle complexes in the plasma membrane of *Micrasterias denticulata* associated with the formation of cellulose fibrils in primary and secondary cell walls. *J Cell Biol*, 84(2), 327-339.
- Gillmor, C. S., Poindexter, P., Lorieau, J., Palcic, M. M., & Somerville, C. R. (2002). Alpha-glucosidase I is required for cellulose biosynthesis and morphogenesis in *Arabidopsis*. *J Cell Biol*, 156(6), 1003-1013.
- Gonneau, M., Desprez, T., Guillot, A., Vernhettes, S., & Hofte, H. (2014). Catalytic subunit stoichiometry within the cellulose synthase complex. *Plant Physiol*, 166(4), 1709-1712.
- Goss, C. A., Brockmann, D. J., Bushoven, J. T., & Roberts, A. W. (2012). A CELLULOSE SYNTHASE (CESA) gene essential for gametophore morphogenesis in the moss *Physcomitrella patens*. *Planta*, 235(6), 1355-1367.
- Grigoryan, G., & Keating, A. E. (2008). Structural specificity in coiled-coil interactions. *Curr Opin Struct Biol*, 18(4), 477-483.

- Gu, F., Bringmann, M., Combs, J. R., Yang, J., Bergmann, D. C., & Nielsen, E. (2016). Arabidopsis CSLD5 functions in cell plate formation in a cell cycle-dependent manner. *Plant Cell*, 28(7), 1722-1737.
- Gu, Y., Kaplinsky, N., Bringmann, M., Cobb, A., Carroll, A., Sampathkumar, A., . . . Somerville, C. R. (2010). Identification of a cellulose synthase-associated protein required for cellulose biosynthesis. *Proc Natl Acad Sci USA*, 107(29), 12866-12871.
- Guo, H., Li, L., Ye, H., Yu, X., Algreen, A., & Yin, Y. (2009). Three related receptor-like kinases are required for optimal cell elongation in Arabidopsis thaliana. *Proc Natl Acad Sci USA*, 106(18), 7648-7653.
- Gutierrez, R., Lindeboom, J. J., Paredez, A. R., Emons, A. M., & Ehrhardt, D. W. (2009). Arabidopsis cortical microtubules position cellulose synthase delivery to the plasma membrane and interact with cellulose synthase trafficking compartments. *Nat Cell Biol*, 11(7), 797-806.
- Haigler, C. H., & Brown, R. M., Jr. (1986). Transport of rosettes from the golgi apparatus to the plasma membrane in isolated mesophyll cells of Zinnia elegans during differentiation to tracheary elements in suspension culture. *Protoplasma*, 134(2-3), 111-120.
- Haigler, C. H., & Roberts, A. W. (2019). Structure/function relationships in the rosette cellulose synthesis complex illuminated by an evolutionary perspective. *Cellulose*, 26(1), 227-247.
- Harholt, J., Sorensen, I., Fangel, J., Roberts, A. W., Willats, W. G., Scheller, H. V., . . . Ulvskov, P. (2012). The glycosyltransferase repertoire of the spikemoss Selaginella moellendorffii and a comparative study of its cell wall. *PLOS ONE*, 7(5), e35846.
- Harris, D. M., Corbin, K., Wang, T., Gutierrez, R., Bertolo, A. L., Petti, C., . . . Debolt, S. (2012). Cellulose microfibril crystallinity is reduced by mutating C-terminal transmembrane region residues CESA1A903V and CESA3T942I of cellulose synthase. *Proc Natl Acad Sci USA*, 109(11), 4098-4103.
- Hazen, S. P., Hawley, R. M., Davis, G. L., Henrissat, B., & Walton, J. D. (2003). Quantitative trait loci and comparative genomics of cereal cell wall composition. *Plant Physiol*, 132(1), 263-271.
- Heim, D. R., Roberts, J. L., Pike, P. D., & Larrinua, I. M. (1989). Mutation of a locus of Arabidopsis thaliana confers resistance to the herbicide isoxaben. *Plant Physiol*, 90(1), 146-150.

- Hematy, K., Sado, P. E., Van Tuinen, A., Rochange, S., Desnos, T., Balzergue, S., . . . Hofte, H. (2007). A receptor-like kinase mediates the response of Arabidopsis cells to the inhibition of cellulose synthesis. *Curr Biol*, 17(11), 922-931.
- Hemsley, P. A., Weimar, T., Lilley, K. S., Dupree, P., & Grierson, C. S. (2013). A proteomic approach identifies many novel palmitoylated proteins in Arabidopsis. *New Phytol*, 197(3), 805-814.
- Hernández-Blanco, C., Feng, D., Hu, J., Sanchez-Vallet, A., Deslandes, L., Llorente, F., . . . Molina, A. (2007). Impairment of cellulose synthases required for Arabidopsis secondary cell wall formation enhances disease resistance. *Plant Cell*, 19(3), 890-903.
- Herth, W. (1985). Plasma-membrane rosettes involved in localized wall thickening during xylem vessel formation of *Lepidium sativum* L. *Planta*, 164(1), 12-21.
- Hill, J. L., Jr., Hammudi, M. B., & Tien, M. (2014). The Arabidopsis cellulose synthase complex: a proposed hexamer of CESA trimers in an equimolar stoichiometry. *Plant Cell*, 26(12), 4834-4842.
- Hill, J. L., Jr., Hill, A. N., Roberts, A. W., Haigler, C. H., & Tien, M. (2018b). Domain swaps of Arabidopsis secondary wall cellulose synthases to elucidate their class specificity. *Plant Direct*, 2(7), e00061.
- Hill, J. L., Jr., Josephs, C., Barnes, W. J., Anderson, C. T., & Tien, M. (2018a). Longevity in vivo of primary cell wall cellulose synthases. *Plant Mol Biol*, 96(3), 279-289.
- Holland, N., Holland, D., Helentjaris, T., Dhugga, K. S., Xoconostle-Cazares, B., & Delmer, D. P. (2000). A comparative analysis of the plant cellulose synthase (CesA) gene family. *Plant Physiol*, 123(4), 1313-1324.
- Holm, L., Kääriäinen, S., Rosenström, P., & Schenkel, A. (2008). Searching protein structure databases with DaliLite v.3. *Bioinformatics*, 24(23), 2780-2781.
- Holm, L., & Laakso, L. M. (2016). Dali server update. *Nucleic Acids Res*, 44(W1), W351-355.
- Holm, L., & Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res*, 38(Web Server issue), W545-W549.
- Howles, P. A., Birch, R. J., Collings, D. A., Gebbie, L. K., Hurley, U. A., Hocart, C. H., . . . Williamson, R. E. (2006). A mutation in an Arabidopsis ribose 5-phosphate isomerase reduces cellulose synthesis and is rescued by exogenous uridine. *Plant J*, 48(4), 606-618.

- Howles, P. A., Gebbie, L. K., Collings, D. A., Varsani, A., Broad, R. C., Ohms, S., . . . Williamson, R. E. (2016). A temperature-sensitive allele of a putative mRNA splicing helicase down-regulates many cell wall genes and causes radial swelling in *Arabidopsis thaliana*. *Plant Mol Biol*, 91(1-2), 1-13.
- Hu, Y., Zhong, R., Morrison, W. H., 3rd, & Ye, Z. H. (2003). The *Arabidopsis* RHD3 gene is required for cell wall biosynthesis and actin organization. *Planta*, 217(6), 912-921.
- Hu, Z., Vanderhaeghen, R., Cools, T., Wang, Y., De Clercq, I., Leroux, O., . . . De Veylder, L. (2016). Mitochondrial defects confer tolerance against cellulose deficiency. *Plant Cell*, 28(9), 2276-2290.
- Jacob-Wilk, D., Kurek, I., Hogan, P., & Delmer, D. P. (2006). The cotton fiber zinc-binding domain of cellulose synthase A1 from *Gossypium hirsutum* displays rapid turnover in vitro and in vivo. *Proc Natl Acad Sci USA*, 103(32), 12191-12196.
- Jones, D. M., Murray, C. M., Ketelaar, K. J., Thomas, J. J., Villalobos, J. A., & Wallace, I. S. (2016). The emerging role of protein phosphorylation as a critical regulatory mechanism controlling cellulose biosynthesis. *Front Plant Sci*, 7, 684.
- Kallberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., & Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nat Protoc*, 7(8), 1511-1522.
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*, 44(D1), D457-D462.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*, 10(6), 845-858.
- Knott, B. C., Crowley, M. F., Himmel, M. E., Zimmer, J., & Beckham, G. T. (2016). Simulations of cellulose translocation in the bacterial cellulose synthase suggest a regulatory mechanism for the dimeric structure of cellulose. *Chem Sci*, 7(5), 3108-3116.
- Konarev, P. V., Petoukhov, M. V., Volkov, V. V., & Svergun, D. I. (2006). ATSAS 2.1, a program package for small-angle scattering data analysis. *J Appl Crystallogr*, 39, 277-286.
- Kondo, T., Nojiri, M., Hishikawa, Y., Togawa, E., Romanovitch, D., & Brown, R. M., Jr. (2002). Biodirected epitaxial nanodeposition of polymers on oriented macromolecular templates. *Proc Natl Acad Sci USA*, 99(22), 14008-14013.

- Kotake, T., Aohara, T., Hirano, K., Sato, A., Kaneko, Y., Tsumuraya, Y., . . . Kawasaki, S. (2011). Rice Brittle culm 6 encodes a dominant-negative form of CesaA protein that perturbs cellulose synthesis in secondary cell walls. *J Exp Bot*, 62(6), 2053-2062.
- Kozin, M. B., & Svergun, D. I. (2001). Automated matching of high- and low-resolution structural models. *J Appl Crystallogr*, 34(1), 33-41.
- Krissinel, E. (2007). On the relationship between sequence and structure similarities in proteomics. *Bioinformatics*, 23(6), 717-723.
- Krissinel, E., & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr*, 60(Pt 12 Pt 1), 2256-2268.
- Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol*, 372(3), 774-797.
- Kubicki, J. D., Yang, H., Sawada, D., O'Neill, H., Oehme, D. P., & Cosgrove, D. J. (2018). The shape of native plant cellulose microfibrils. *Sci Rep*, 8(1), 13983.
- Kumar, M., Atanassov, I., & Turner, S. R. (2017). Functional analysis of cellulose synthase (CESA) protein class specificity. *Plant Physiol*, 173(2), 970-983.
- Kumar, M., Mishra, L., Carr, P., Pilling, M., Gardner, P., Mansfield, S. D., & Turner, S. R. (2018). Exploiting CELLULOSE SYNTHASE (CESA) class-specificity to probe cellulose microfibril biosynthesis. *Plant Physiol*, 177(1), 151-167.
- Kumar, M., & Turner, S. R. (2015a). Plant cellulose synthesis: CESA proteins crossing kingdoms. *Phytochemistry*, 112, 91-99.
- Kumar, M., & Turner, S. R. (2015b). Protocol: a medium-throughput method for determination of cellulose content from single stem pieces of *Arabidopsis thaliana*. *Plant Methods*, 11, 46.
- Kumar, M., Wightman, R., Atanassov, I., Gupta, A., Hurst, C. H., Hemsley, P. A., & Turner, S. R. (2016). S-Acylation of the cellulose synthase complex is essential for its plasma membrane localization. *Science*, 353(6295), 166-169.
- Kurek, I., Kawagoe, Y., Jacob-Wilk, D., Doblin, M. S., & Delmer, D. P. (2002). Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domains. *Proc Natl Acad Sci USA*, 99(17), 11109-11114.

- La, D., Esquivel-Rodriguez, J., Venkatraman, V., Li, B., Sael, L., Ueng, S., . . . Kihara, D. (2009). 3D-SURFER: software for high-throughput protein surface comparison and analysis. *Bioinformatics*, 25(21), 2843-2844.
- Lairson, L. L., Henrissat, B., Davies, G. J., & Withers, S. G. (2008). Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem*, 77, 521-555.
- Lane, D. R., Wiedemeier, A., Peng, L., Hofte, H., Vernhettes, S., Desprez, T., . . . Williamson, R. E. (2001). Temperature-sensitive alleles of RSW2 link the KORRIGAN endo-1,4-beta-glucanase to cellulose synthesis and cytokinesis in Arabidopsis. *Plant Physiol*, 126(1), 278-288.
- Lei, L., Li, S., Du, J., Bashline, L., & Gu, Y. (2013). Cellulose synthase INTERACTIVE3 regulates cellulose biosynthesis in both a microtubule-dependent and microtubule-independent manner in Arabidopsis. *Plant Cell*, 25(12), 4912-4923.
- Lei, L., Li, S., & Gu, Y. (2012). Cellulose synthase complexes: composition and regulation. *Front Plant Sci*, 3, 75.
- Lei, L., Singh, A., Bashline, L., Li, S., Yingling, Y. G., & Gu, Y. (2015). CELLULOSE SYNTHASE INTERACTIVE1 is required for fast recycling of cellulose synthase complexes to the plasma membrane in Arabidopsis. *Plant Cell*, 27(10), 2926-2940.
- Lei, L., Zhang, T., Strasser, R., Lee, C. M., Gonneau, M., Mach, L., . . . Gu, Y. (2014). The jiaoyao1 mutant is an allele of korrigan1 that abolishes endoglucanase activity and affects the organization of both cellulose microfibrils and microtubules in Arabidopsis. *Plant Cell*, 26(6), 2601-2616.
- Li, A., Xia, T., Xu, W., Chen, T., Li, X., Fan, J., . . . Peng, L. (2013). An integrative analysis of four CESA isoforms specific for fiber cellulose production between *Gossypium hirsutum* and *Gossypium barbadense*. *Planta*, 237(6), 1585-1597.
- Li, B., Turuvekere, S., Agrawal, M., La, D., Ramani, K., & Kihara, D. (2008). Characterization of local geometry of protein surfaces with the visibility criterion. *Proteins*, 71(2), 670-683.
- Li, S., Bashline, L., Lei, L., & Gu, Y. (2014). Cellulose synthesis and its regulation. *Arabidopsis Book*, 12, e0169.
- Li, S., Lei, L., Somerville, C. R., & Gu, Y. (2012). Cellulose synthase interactive protein 1 (CS11) links microtubules and cellulose synthase complexes. *Proc Natl Acad Sci USA*, 109(1), 185-190.

- Liebminger, E., Grass, J., Altmann, F., Mach, L., & Strasser, R. (2013). Characterizing the link between glycosylation state and enzymatic activity of the endo-beta1,4-glucanase KORRIGAN1 from *Arabidopsis thaliana*. *J Biol Chem*, 288(31), 22270-22280.
- Linding, R., Russell, R. B., Neduva, V., & Gibson, T. J. (2003). GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res*, 31(13), 3701-3708.
- Liu, L., Shang-Guan, K., Zhang, B., Liu, X., Yan, M., Zhang, L., . . . Zhou, Y. (2013). Brittle Culm1, a COBRA-like protein, functions in cellulose assembly through binding cellulose microfibrils. *PLOS Genet*, 9(8), e1003704.
- Liu, Z., Schneider, R., Kesten, C., Zhang, Y., Somssich, M., Zhang, Y., . . . Persson, S. (2016). Cellulose-microtubule uncoupling proteins prevent lateral displacement of microtubules during cellulose synthesis in *Arabidopsis*. *Dev Cell*, 38(3), 305-315.
- Lo, W. C., Huang, P. J., Chang, C. H., & Lyu, P. C. (2007). Protein structural similarity search by Ramachandran codes. *BMC Bioinformatics*, 8, 307.
- Lo, W. C., Lee, C. Y., Lee, C. C., & Lyu, P. C. (2009). iSARST: an integrated SARST web server for rapid protein structural similarity searches. *Nucleic Acids Res*, 37(Web Server issue), W545-W551.
- Lukowitz, W., Nickle, T. C., Meinke, D. W., Last, R. L., Conklin, P. L., & Somerville, C. R. (2001). *Arabidopsis* *cyt1* mutants are deficient in a mannose-1-phosphate guanylyltransferase and point to a requirement of N-linked glycosylation for cellulose biosynthesis. *Proc Natl Acad Sci USA*, 98(5), 2262-2267.
- MacKinnon, I. M., Sturcova, A., Sugimoto-Shirasu, K., His, I., McCann, M. C., & Jarvis, M. C. (2006). Cell-wall structure and anisotropy in *procuste*, a cellulose synthase mutant of *Arabidopsis thaliana*. *Planta*, 224(2), 438-448.
- Maeda, H., & Ishida, N. (1967). Specificity of binding of hexopyranosyl polysaccharides with fluorescent brightener. *J Biochem*, 62(2), 276-278.
- Mansoori, N., Timmers, J., Desprez, T., Alvim-Kamei, C. L., Dees, D. C., Vincken, J. P., . . . Trindade, L. M. (2014). KORRIGAN1 interacts specifically with integral components of the cellulose synthase machinery. *PLOS ONE*, 9(11), e112387.
- Manzano, C., Abraham, Z., López-Torrejón, G., & Del Pozo, J. C. (2008). Identification of ubiquitinated proteins in *Arabidopsis*. *Plant Mol Biol*, 68(1), 145-158.

- Matthews, J. F., Beckham, G. T., Bergenstr hle-Wohlert, M., Brady, J. W., Himmel, M. E., & Crowley, M. F. (2012). Comparison of cellulose I β simulations with three carbohydrate force fields. *J Chem Theory Comput*, 8(2), 735-748.
- McFarlane, H. E., Doring, A., & Persson, S. (2014). The cell biology of cellulose synthesis. *Annu Rev Plant Biol*, 65, 69-94.
- Merz, D., Richter, J., Gonneau, M., Sanchez-Rodriguez, C., Eder, T., Sormani, R., . . . Hauser, M. T. (2017). T-DNA alleles of the receptor kinase THESEUS1 with opposing effects on cell wall integrity signaling. *J Exp Bot*, 68(16), 4583-4593.
- Miart, F., Desprez, T., Biot, E., Morin, H., Belcram, K., Hofte, H., . . . Vernhettes, S. (2014). Spatio-temporal analysis of cellulose synthesis during cell plate formation in Arabidopsis. *Plant J*, 77(1), 71-84.
- Michael, T. P., Mockler, T. C., Breton, G., McEntee, C., Byer, A., Trout, J. D., . . . Chory, J. (2008). Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLOS Genet*, 4(2), e14.
- Mitchell, A., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., . . . Finn, R. D. (2015). The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res*, 43(Database issue), D213-D221.
- Morgan, J. L., McNamara, J. T., Fischer, M., Rich, J., Chen, H. M., Withers, S. G., & Zimmer, J. (2016). Observing cellulose biosynthesis and membrane translocation in crystallo. *Nature*, 531(7594), 329-334.
- Morgan, J. L., McNamara, J. T., & Zimmer, J. (2014). Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. *Nat Struct Mol Biol*, 21(5), 489-496.
- Morgan, J. L., Strumillo, J., & Zimmer, J. (2013). Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature*, 493(7431), 181-186.
- Mueller, S. C., & Brown, R. M., Jr. (1980). Evidence for an intramembrane component associated with a cellulose microfibril-synthesizing complex in higher plants. *J Cell Biol*, 84(2), 315-326.
- Murzin, A. G., Brenner, S. E., Hubbard, T., & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol*, 247(4), 536-540.

- Nakagami, H., Sugiyama, N., Mochida, K., Daudi, A., Yoshida, Y., Toyoda, T., . . . Shirasu, K. (2010). Large-scale comparative phosphoproteomics identifies conserved phosphorylation sites in plants. *Plant Physiol*, 153(3), 1161-1174.
- Neumann, U., Brandizzi, F., & Hawes, C. (2003). Protein transport in plant cells: in and out of the Golgi. *Ann Bot*, 92(2), 167-180.
- Newman, R. H., Hill, S. J., & Harris, P. J. (2013). Wide-angle X-ray scattering and solid-state nuclear magnetic resonance data combined to test models for cellulose microfibrils in mung bean cell walls. *Plant Physiol*, 163(4), 1558-1567.
- Nickle, T. C., & Meinke, D. W. (1998). A cytokinesis-defective mutant of Arabidopsis (cyt1) characterized by embryonic lethality, incomplete cell walls, and excessive callose accumulation. *Plant J*, 15(3), 321-332.
- Nixon, B. T., Mansouri, K., Singh, A., Du, J., Davis, J. K., Lee, J. G., . . . Haigler, C. H. (2016). Comparative structural and computational analysis supports eighteen cellulose synthases in the plant cellulose synthesis complex. *Sci Rep*, 6, 28696.
- Nobles, D. R., Romanovicz, D. K., & Brown, R. M., Jr. (2001). Cellulose in cyanobacteria. Origin of vascular plant cellulose synthase? *Plant Physiol*, 127(2), 529-542.
- Norris, J. H., Li, X., Huang, S., Van de Meene, A. M. L., Tran, M. L., Killeavy, E., . . . Roberts, A. W. (2017). Functional specialization of cellulose synthase isoforms in a moss shows parallels with seed plants. *Plant Physiol*, 175(1), 210-222.
- Nuhse, T. S., Stensballe, A., Jensen, O. N., & Peck, S. C. (2004). Phosphoproteomics of the Arabidopsis plasma membrane and a new phosphorylation site database. *Plant Cell*, 16(9), 2394-2405.
- Oehme, D. P., Downton, M. T., Doblin, M. S., Wagner, J., Gidley, M. J., & Bacic, A. (2015). Unique aspects of the structure and dynamics of elementary Ibeta cellulose microfibrils revealed by computational simulations. *Plant Physiol*, 168(1), 3-17.
- Okekeogbu, I. O., Pattathil, S., González Fernández-Niño, S. M., Aryal, U. K., Penning, B. W., Lao, J., . . . Carpita, N. C. (2019). Glycome and proteome components of Golgi membranes are common between two angiosperms with distinct cell-wall structures. *Plant Cell*, 31(5), 1094-1112.

- Olek, A. T., Rayon, C., Makowski, L., Kim, H. R., Ciesielski, P., Badger, J., . . . Carpita, N. C. (2014). The structure of the catalytic domain of a plant cellulose synthase and its assembly into dimers. *Plant Cell*, 26(7), 2996-3009.
- Otwinowski, Z., & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*, 276, 307-326.
- Paradez, A. R., Wright, A., & Ehrhardt, D. W. (2006). Microtubule cortical array organization and plant cell morphogenesis. *Curr Opin Plant Biol*, 9(6), 571-578.
- Paredez, A. R., Somerville, C. R., & Ehrhardt, D. W. (2006). Visualization of cellulose synthase demonstrates functional association with microtubules. *Science*, 312(5779), 1491-1495.
- Park, S., Song, B., Shen, W., & Ding, S. Y. (2019). A mutation in the catalytic domain of cellulose synthase 6 halts its transport to the Golgi apparatus. *J Exp Bot*, 70(21), 6071-6083.
- Pear, J. R., Kawagoe, Y., Schreckengost, W. E., Delmer, D. P., & Stalker, D. M. (1996). Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. *Proc Natl Acad Sci USA*, 93(22), 12637-12642.
- Peng, L., Hocart, C. H., Redmond, J. W., & Williamson, R. E. (2000). Fractionation of carbohydrates in Arabidopsis root cell walls shows that three radial swelling loci are specifically involved in cellulose production. *Planta*, 211(3), 406-414.
- Penning, B. W., Hunter, C. T., 3rd, Tayengwa, R., Eveland, A. L., Dugard, C. K., Olek, A. T., . . . Carpita, N. C. (2009). Genetic resources for maize cell wall biology. *Plant Physiol*, 151(4), 1703-1728.
- Persson, S., Paredez, A., Carroll, A., Palsdottir, H., Doblin, M. S., Poindexter, P., . . . Somerville, C. R. (2007). Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in Arabidopsis. *Proc Natl Acad Sci USA*, 104(39), 15566-15571.
- Peterson, L. X., Shin, W. H., Kim, H., & Kihara, D. (2018). Improved performance in CAPRI round 37 using LZerD docking and template-based modeling with combined scoring functions. *Proteins*, 86(Suppl 1), 311-320.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, 25(13), 1605-1612.

- Polko, J. K., Barnes, W. J., Voiniciuc, C., Doctor, S., Steinwand, B., Hill, J. L., Jr., . . . Kieber, J. J. (2018). SHOU4 proteins regulate trafficking of cellulose synthase complexes to the plasma membrane. *Curr Biol*, 28(19), 3174-3182.
- Polko, J. K., & Kieber, J. J. (2019). The regulation of cellulose biosynthesis in plants. *Plant Cell*, 31(2), 282-296.
- Purushotham, P., Cho, S. H., Diaz-Moreno, S. M., Kumar, M., Nixon, B. T., Bulone, V., & Zimmer, J. (2016). A single heterologously expressed plant cellulose synthase isoform is sufficient for cellulose microfibril formation in vitro. *Proc Natl Acad Sci USA*, 113(40), 11360-11365.
- Pysh, L. D. (2015). Two alleles of the AtCesA3 gene in Arabidopsis thaliana display intragenic complementation. *Am J Bot*, 102(9), 1434-1441.
- Pysh, L. D., Alexander, N., Swatzyna, L., & Harbert, R. (2012). Four alleles of AtCESA3 form an allelic series with respect to root phenotype in Arabidopsis thaliana. *Physiol Plant*, 144(4), 369-381.
- Ramakrishnan, V., & Biou, V. (1997). Treatment of multiwavelength anomalous diffraction data as a special case of multiple isomorphous replacement. *Methods Enzymol*, 276, 538-557.
- Rao, Y., Yang, Y., Xin, D., Li, X., Zhai, K., Ma, B., . . . Zeng, D. (2013). Characterization and cloning of a brittle culm mutant (bc88) in rice (Oryza sativa L.). *Chin Sci Bull*, 58(24), 3000-3006.
- Reiter, W. D., Chapple, C., & Somerville, C. R. (1997). Mutants of Arabidopsis thaliana with altered cell wall polysaccharide composition. *Plant J*, 12(2), 335-345.
- Richmond, T. A., & Somerville, C. R. (2000). The cellulose synthase superfamily. *Plant Physiol*, 124(2), 495-498.
- Richmond, T. A., & Somerville, C. R. (2001). Integrative approaches to determining Csl function. *Plant Mol Biol*, 47(1-2), 131-143.
- Robert, S., Bichet, A., Grandjean, O., Kierzkowski, D., Satiat-Jeunemaitre, B., Pelletier, S., . . . Vernhettes, S. (2005). An Arabidopsis endo-1,4-beta-D-glucanase involved in cellulose synthesis undergoes regulated intracellular cycling. *Plant Cell*, 17(12), 3378-3389.
- Roberts, A. W., & Bushoven, J. T. (2007). The cellulose synthase (CESA) gene superfamily of the moss Physcomitrella patens. *Plant Mol Biol*, 63(2), 207-219.

- Robinson, D. G., & Quader, H. (1980). Structure, synthesis and orientation of microfibrils. VII. Microtubule reassembly in vivo after cold treatment in *Oocystis* and its relevance to microfibril orientation. *Eur J Cell Biol*, 21(2), 229-230.
- Rohl, C. A., Strauss, C. E., Misura, K. M., & Baker, D. (2004). Protein structure prediction using Rosetta. *Methods Enzymol*, 383, 66-93.
- Rossmann, M. G., Moras, D., & Olsen, K. W. (1974). Chemical and biological evolution of nucleotide-binding protein. *Nature*, 250(463), 194-199.
- Roudier, F., Fernandez, A. G., Fujita, M., Himmelsbach, R., Borner, G. H., Schindelman, G., . . . Benfey, P. N. (2005). COBRA, an Arabidopsis extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation. *Plant Cell*, 17(6), 1749-1763.
- Rubio-Díaz, S., Pérez-Pérez, J. M., González-Bayón, R., Muñoz-Viana, R., Borrega, N., Mouille, G., . . . Micol, J. L. (2012). Cell expansion-mediated organ growth is affected by mutations in three EXIGUA genes. *PLOS ONE*, 7(5), e36500.
- Rudolph, U., Gross, H., & Schnepf, E. (1989). Investigations of the turnover of the putative cellulose-synthesizing particle “rosettes” within the plasma membrane of *Funaria hygrometrica* protonema cells. *Protoplasma*, 148(2), 57-69.
- Rushton, P. S., Olek, A. T., Makowski, L., Badger, J., Steussy, C. N., Carpita, N. C., & Stauffacher, C. V. (2017). Rice cellulose synthaseA8 plant-conserved region is a coiled-coil at the catalytic core entrance. *Plant Physiol*, 173(1), 482-494.
- Sael, L., Li, B., La, D., Fang, Y., Ramani, K., Rustamov, R., & Kihara, D. (2008). Fast protein tertiary structure retrieval based on global surface shape similarity. *Proteins*, 72(4), 1259-1273.
- Sampathkumar, A., Gutierrez, R., McFarlane, H. E., Bringmann, M., Lindeboom, J., Emons, A. M., . . . Persson, S. (2013). Patterning and lifetime of plasma membrane-localized cellulose synthase is dependent on actin organization in Arabidopsis interphase cells. *Plant Physiol*, 162(2), 675-688.
- Sato, S., Kato, T., Kakegawa, K., Ishii, T., Liu, Y. G., Awano, T., . . . Shibata, D. (2001). Role of the putative membrane-bound endo-1,4-beta-glucanase KORRIGAN in cell elongation and cellulose synthesis in Arabidopsis thaliana. *Plant Cell Physiol*, 42(3), 251-263.

- Saxena, I. M., Brown, R. M., Jr., Fevre, M., Geremia, R. A., & Henrissat, B. (1995). Multidomain architecture of beta-glycosyl transferases: implications for mechanism of action. *J Bacteriol*, 177(6), 1419-1424.
- Saxena, I. M., Lin, F. C., & Brown, R. M., Jr. (1990). Cloning and sequencing of the cellulose synthase catalytic subunit gene of *Acetobacter xylinum*. *Plant Mol Biol*, 15(5), 673-683.
- Scavuzzo-Duggan, T. R., Chaves, A. M., Singh, A., Sethaphong, L., Slabaugh, E., Yingling, Y. G., . . . Roberts, A. W. (2018). Cellulose synthase 'class specific regions' are intrinsically disordered and functionally undifferentiated. *J Integr Plant Biol*, 60(6), 481-497.
- Scheible, W. R., Eshed, R., Richmond, T., Delmer, D. P., & Somerville, C. R. (2001). Modifications of cellulose synthase confer resistance to isoxaben and thiazolidinone herbicides in *Arabidopsis* Ixr1 mutants. *Proc Natl Acad Sci USA*, 98(18), 10079-10084.
- Schiefelbein, J. W., & Somerville, C. R. (1990). Genetic Control of Root Hair Development in *Arabidopsis thaliana*. *Plant Cell*, 2(3), 235-243.
- Schindelman, G., Morikami, A., Jung, J., Baskin, T. I., Carpita, N. C., Derbyshire, P., . . . Benfey, P. N. (2001). COBRA encodes a putative GPI-anchored protein, which is polarly localized and necessary for oriented cell expansion in *Arabidopsis*. *Genes Dev*, 15(9), 1115-1127.
- Schmidtke, P., Le Guilloux, V., Maupetit, J., & Tufféry, P. (2010). fpocket: online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Res*, 38(Web Server issue), W582-W589.
- Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*, 9(7), 671-675.
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005a). Geometry-based flexible and symmetric protein docking. *Proteins*, 60(2), 224-231.
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005b). PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res*, 33(Web Server issue), W363-W367.
- Schrack, K., Debolt, S., & Bulone, V. (2012). Deciphering the molecular functions of sterols in cellulose biosynthesis. *Front Plant Sci*, 3, 84.
- Schulze, W. X., Yao, Q., & Xu, D. (2015). Databases for plant phosphoproteomics, in *Plant Phosphoproteomics: Methods and Protocols*. New York, NY: Springer, 207-216.

- Sechet, J., Marion-Poll, A., & North, M. H. (2018). Emerging functions for cell wall polysaccharides accumulated during eudicot seed development. *Plants*, 7(4), 81.
- Sene, C. F. B., McCann, M. C., Wilson, R. H., & Grinter, R. (1994). Fourier-transform raman and fourier-transform infrared spectroscopy (an investigation of five higher plant cell walls and their components). *Plant Physiol*, 106(4), 1623-1631.
- Sethaphong, L., Davis, J. K., Slabaugh, E., Singh, A., Haigler, C. H., & Yingling, Y. G. (2016). Prediction of the structures of the plant-specific regions of vascular plant cellulose synthases and correlated functional analysis. *Cellulose*, 23(1), 145-161.
- Sethaphong, L., Haigler, C. H., Kubicki, J. D., Zimmer, J., Bonetta, D., DeBolt, S., & Yingling, Y. G. (2013). Tertiary model of a plant cellulose synthase. *Proc Natl Acad Sci USA*, 110(18), 7512-7517.
- Shih, H. W., Miller, N. D., Dai, C., Spalding, E. P., & Monshausen, G. B. (2014). The receptor-like kinase FERONIA is required for mechanical signal transduction in Arabidopsis seedlings. *Curr Biol*, 24(16), 1887-1892.
- Shim, I., Law, R., Kileeg, Z., Stronghill, P., Northey, J. G. B., Strap, J. L., & Bonetta, D. T. (2018). Alleles causing resistance to isoxaben and flupoxam highlight the significance of transmembrane domains for CESA protein function. *Front Plant Sci*, 9, 1152.
- Shindyalov, I. N., & Bourne, P. E. (1998). Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng*, 11(9), 739-747.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., . . . Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, 7, 539.
- Sillitoe, I., Lewis, T. E., Cuff, A., Das, S., Ashford, P., Dawson, N. L., . . . Orengo, C. A. (2015). CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res*, 43(Database issue), D376-D381.
- Slabaugh, E., Davis, J. K., Haigler, C. H., Yingling, Y. G., & Zimmer, J. (2014a). Cellulose synthases: new insights from crystallography and modeling. *Trends Plant Sci*, 19(2), 99-106.

- Slabaugh, E., Scavuzzo-Duggan, T. R., Chaves, A., Wilson, L., Wilson, C., Davis, J. K., . . . Haigler, C. H. (2016). The valine and lysine residues in the conserved FxVTxK motif are important for the function of phylogenetically distant plant cellulose synthases. *Glycobiology*, 26(5), 509-519.
- Slabaugh, E., Sethaphong, L., Xiao, C., Amick, J., Anderson, C. T., Haigler, C. H., & Yingling, Y. G. (2014b). Computational and genetic evidence that different structural conformations of a non-catalytic region affect the function of plant cellulose synthase. *J Exp Bot*, 65(22), 6645-6653.
- Somerville, C. R. (2006). Cellulose synthesis in higher plants. *Annu Rev Cell Dev Biol*, 22, 53-78.
- Sorek, N., Sorek, H., Kijac, A., Szemenyei, H. J., Bauer, S., Hematy, K., . . . Somerville, C. R. (2014). The Arabidopsis COBRA protein facilitates cellulose crystallization at the plasma membrane. *J Biol Chem*, 289(50), 34911-34920.
- Speicher, T. L., Li, P. Z., & Wallace, I. S. (2018). Phosphoregulation of the plant cellulose synthase complex and cellulose synthase-like proteins. *Plants*, 7(3), 52.
- Sugimoto, K., Williamson, R. E., & Wasteneys, G. O. (2001). Wall architecture in the cellulose-deficient *rswl* mutant of *Arabidopsis thaliana*: microfibrils but not microtubules lose their transverse alignment before microfibrils become unrecognizable in the mitotic and elongation zones of roots. *Protoplasma*, 215(1-4), 172-183.
- Svergun, D. I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr*, 25, 495-503.
- Svergun, D. I., Barberato, C., & Koch, M. H. J. (1995). CRY SOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr*, 28, 768-773.
- Svergun, D. I., Petoukhov, M. V., & Koch, M. H. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophys J*, 80(6), 2946-2953.
- Sánchez-Rodríguez, C., Ketelaar, K., Schneider, R., Villalobos, J. A., Somerville, C. R., Persson, S., & Wallace, I. S. (2017). BRASSINOSTEROID INSENSITIVE2 negatively regulates cellulose synthesis in *Arabidopsis* by phosphorylating cellulose synthase 1. *Proc Natl Acad Sci USA*, 114(13), 3533-3538.

- Sánchez-Rodríguez, C., Shi, Y., Kesten, C., Zhang, D., Sancho-Andrés, G., Ivakov, A., . . . Persson, S. (2018). The cellulose synthases are cargo of the TPLATE adaptor complex. *Mol Plant*, 11(2), 346-349.
- Tanaka, K., Murata, K., Yamazaki, M., Onosato, K., Miyao, A., & Hirochika, H. (2003). Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall. *Plant Physiol*, 133(1), 73-83.
- Tateno, M., Brabham, C., & DeBolt, S. (2016). Cellulose biosynthesis inhibitors - a multifunctional toolbox. *J Exp Bot*, 67(2), 533-542.
- Taylor, N. G. (2007). Identification of cellulose synthase AtCesA7 (IRX3) in vivo phosphorylation sites--a potential role in regulating protein degradation. *Plant Mol Biol*, 64(1-2), 161-171.
- Taylor, N. G. (2008). Cellulose biosynthesis and deposition in higher plants. *New Phytol*, 178(2), 239-252.
- Taylor, N. G., Gardiner, J. C., Whiteman, R., & Turner, S. R. (2004). Cellulose synthesis in the Arabidopsis secondary cell wall. *Cellulose*, 11(3-4), 329-338.
- Taylor, N. G., Howells, R. M., Huttly, A. K., Vickers, K., & Turner, S. R. (2003). Interactions among three distinct CesA proteins essential for cellulose synthesis. *Proc Natl Acad Sci USA*, 100(3), 1450-1455.
- Taylor, N. G., Laurie, S., & Turner, S. R. (2000). Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in Arabidopsis. *Plant Cell*, 12(12), 2529-2540.
- Taylor, N. G., Scheible, W. R., Cutler, S., Somerville, C. R., & Turner, S. R. (1999). The irregular xylem3 locus of Arabidopsis encodes a cellulose synthase required for secondary cell wall synthesis. *Plant Cell*, 11(5), 769-780.
- Terwilliger, T. C., Dimaio, F., Read, R. J., Baker, D., Bunkoczi, G., Adams, P. D., . . . Echols, N. (2012). phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. *J Struct Funct Genomics*, 13(2), 81-90.
- The PyMOL Molecular Graphics System, Version 2.2. (2019).
- Thomas, L. H., Forsyth, V. T., Sturcova, A., Kennedy, C. J., May, R. P., Altaner, C. M., . . . Jarvis, M. C. (2013). Structure of cellulose microfibrils in primary cell walls from collenchyma. *Plant Physiol*, 161(1), 465-476.

- Timmers, J., Vernhettes, S., Desprez, T., Vincken, J. P., Visser, R. G., & Trindade, L. M. (2009). Interactions between membrane-bound cellulose synthases involved in the synthesis of the secondary cell wall. *FEBS Lett*, 583(6), 978-982.
- Tran, M. L., & Roberts, A. W. (2016). Cellulose synthase gene expression profiling of *Physcomitrella patens*. *Plant Biol*, 18(3), 362-368.
- Tung, C. H., Huang, J. W., & Yang, J. M. (2007). Kappa-alpha plot derived structural alphabet and BLOSUM-like substitution matrix for rapid search of protein structure database. *Genome Biol*, 8(3), R31.
- Turner, S. R., & Kumar, M. (2018). Cellulose synthase complex organization and cellulose microfibril structure. *Philos Trans A Math Phys Eng Sci*, 376(2112), 20170048.
- Turner, S. R., & Somerville, C. R. (1997). Collapsed xylem phenotype of *Arabidopsis* identifies mutants deficient in cellulose deposition in the secondary cell wall. *Plant Cell*, 9(5), 689-701.
- Uversky, V. N. (2019). Intrinsically disordered proteins and their “mysterious” (meta)physics. *Front Phys*, 7(10).
- Vain, T., Crowell, E. F., Timpano, H., Biot, E., Desprez, T., Mansoori, N., . . . Vernhettes, S. (2014). The cellulase KORRIGAN is part of the cellulose synthase complex. *Plant Physiol*, 165(4), 521-1532
- Van der Does, D., Boutrot, F., Engelsdorf, T., Rhodes, J., McKenna, J. F., Vernhettes, S., . . . Zipfel, C. (2017). The *Arabidopsis* leucine-rich repeat receptor kinase MIK2/LRR-KISS connects cell wall integrity sensing, root growth and response to abiotic and biotic stresses. *PLOS Genet*, 13(6), e1006832.
- Vandavasi, V. G., Putnam, D. K., Zhang, Q., Petridis, L., Heller, W. T., Nixon, B. T., . . . O'Neill, H. (2016). A structural study of CESA1 catalytic domain of *Arabidopsis* cellulose synthesis complex: evidence for CESA trimers. *Plant Physiol*, 170(1), 123-135.
- Vergara, C. E., & Carpita, N. C. (2001). Beta-D-glycan synthases and the CesaA gene family: lessons to be learned from the mixed-linkage (1-->3),(1-->4)beta-D-glucan synthase. *Plant Mol Biol*, 47(1-2), 145-160.
- Voss, N. R., & Gerstein, M. (2010). 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res*, 38(Web Server issue), W555-W562.

- Wagner, T. A., & Kohorn, B. D. (2001). Wall-associated kinases are expressed throughout plant development and are required for cell expansion. *Plant Cell*, 13(2), 303-318.
- Wang, D., Qin, Y., Fang, J., Yuan, S., Peng, L., Zhao, J., & Li, X. (2016). A missense mutation in the zinc finger domain of OsCESA7 deleteriously affects cellulose biosynthesis and plant growth in rice. *PLOS ONE*, 11(4), e0153993.
- Wang, H., Lockwood, S. K., Hoeltzel, M. F., & Schiefelbein, J. W. (1997). The ROOT HAIR DEFECTIVE3 gene encodes an evolutionarily conserved protein with GTP-binding motifs and is required for regulated cell enlargement in Arabidopsis. *Genes Dev*, 11(6), 799-811.
- Wang, J., Elliott, J. E., & Williamson, R. E. (2008). Features of the primary wall CESA complex in wild type and cellulose-deficient mutants of Arabidopsis thaliana. *J Exp Bot*, 59(10), 2627-2637.
- Wang, J., Howles, P. A., Cork, A. H., Birch, R. J., & Williamson, R. E. (2006). Chimeric proteins suggest that the catalytic and/or C-terminal domains give Cesa1 and Cesa3 access to their specific sites in the cellulose synthase of primary walls. *Plant Physiol*, 142(2), 685-695.
- Wang, T., & Hong, M. (2016). Solid-state NMR investigations of cellulose structure and interactions with matrix polysaccharides in plant primary cell walls. *J Exp Bot*, 67(2), 503-514.
- Watanabe, Y., Meents, M. J., McDonnell, L. M., Barkwill, S., Sampathkumar, A., Cartwright, H. N., . . . Mansfield, S. D. (2015). Visualization of cellulose synthases in Arabidopsis secondary cell walls. *Science*, 350(6257), 198-203.
- Watanabe, Y., Schneider, R., Barkwill, S., Gonzales-Vigil, E., Hill, J. L., Jr., Samuels, A. L., . . . Mansfield, S. D. (2018). Cellulose synthase complexes display distinct dynamic behaviors during xylem transdifferentiation. *Proc Natl Acad Sci USA*, 115(27), E6366-E6374.
- Whitmore, L., & Wallace, B. A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res*, 32(Web Server issue), W668-W673.
- Wiedemeier, A. M. (1998). Analysis of Arabidopsis thaliana growth anisotropy mutants: genetic, physiological, and cytological characterization. (Ph. D.), University of Missouri, Columbia.

- Wiedemeier, A. M., Judy-March, J. E., Hocart, C. H., Wasteneys, G. O., Williamson, R. E., & Baskin, T. I. (2002). Mutant alleles of Arabidopsis RADIALLY SWOLLEN 4 and 7 reduce growth anisotropy without altering the transverse orientation of cortical microtubules or cellulose microfibrils. *Development*, 129(20), 4821-4830.
- Williamson, R. E., Burn, J. E., Birch, R. J., Baskin, T. I., Arioli, T., Betzner, A. S., & Cork, A. H. (2001). Morphology of *rsw1*, a cellulose-deficient mutant of *Arabidopsis thaliana*. *Protoplasma*, 215(1-4), 116-127.
- Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., & Provart, N. J. (2007). An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. *PLOS ONE*, 2(8), e718.
- Wong, H. C., Fear, A. L., Calhoon, R. D., Eichinger, G. H., Mayer, R., Amikam, D., . . . et al. (1990). Genetic organization of the cellulose synthase operon in *Acetobacter xylinum*. *Proc Natl Acad Sci USA*, 87(20), 8130-8134.
- Wood, P. J. (1980). Specificity in the interaction of direct dyes with polysaccharides. *Carbohydr Res*, 85(2), 271-287.
- Wu, S., Scheible, W. R., Schindelasch, D., Van Den Daele, H., De Veylder, L., & Baskin, T. I. (2010). A conditional mutation in *Arabidopsis thaliana* separase induces chromosome non-disjunction, aberrant morphogenesis and cyclin B1;1 stability. *Development*, 137(6), 953-961.
- Wu, S., & Zhang, Y. (2008). MUSTER: improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins*, 72(2), 547-556.
- Xia, Y., Lei, L., Brabham, C., Stork, J., Strickland, J., Ladak, A., . . . DeBolt, S. (2014). Acetobixan, an inhibitor of cellulose synthesis identified by microbial bioprospecting. *PLOS ONE*, 9(4), e95245.
- Xu, S. L., Rahman, A., Baskin, T. I., & Kieber, J. J. (2008). Two leucine-rich repeat receptor kinases mediate signaling, linking cell wall biosynthesis and ACC synthase in *Arabidopsis*. *Plant Cell*, 20(11), 3065-3079.
- Yan, C., Yan, S., Zeng, X., Zhang, Z., & Gu, M. (2007). Fine mapping and isolation of Bc7(t), allelic to OsCesA4. *J Genet Genom*, 34(11), 1019-1027.

- Yang, H., Zimmer, J., Yingling, Y. G., & Kubicki, J. D. (2015). How cellulose elongates-a QM/MM study of the molecular mechanism of cellulose polymerization in bacterial CESA. *J Phys Chem B*, 119(22), 6525-6535.
- Ye, D., Kiemle, S. N., Rongpipi, S., Wang, X., Wang, C., Cosgrove, D. J., . . . Gomez, E. D. (2018). Resonant soft X-ray scattering reveals cellulose microfibril spacing in plant primary cell walls. *Sci Rep*, 8(1), 12449.
- Ye, Y., & Godzik, A. (2004). FATCAT: a web server for flexible structure comparison and structure similarity searching. *Nucleic Acids Res*, 32(Web Server issue), W582-W585.
- Yu, Y. (2018). Cellulose synthase stoichiometry varies among species and tissues. *Plant Physiol*, 177(3), 873-874.
- Zhang, B., Deng, L., Qian, Q., Xiong, G., Zeng, D., Li, R., . . . Zhou, Y. (2009). A missense mutation in the transmembrane domain of CESA4 affects protein abundance in the plasma membrane and results in abnormal cell wall biosynthesis in rice. *Plant Mol Biol*, 71(4-5), 509-524.
- Zhang, M., Henquet, M., Chen, Z., Zhang, H., Zhang, Y., Ren, X., . . . Gong, Z. (2009). LEW3, encoding a putative α -1,2-mannosyltransferase (ALG11) in N-linked glycoprotein, plays vital roles in cell-wall biosynthesis and the abiotic stress response in *Arabidopsis thaliana*. *Plant J*, 60(6), 983-999.
- Zhang, T., Zheng, Y., & Cosgrove, D. J. (2016). Spatial organization of cellulose microfibrils and matrix polysaccharides in primary plant cell walls as imaged by multichannel atomic force microscopy. *Plant J*, 85(2), 179-192.
- Zhang, W., Cai, C., & Staiger, C. J. (2019). Myosins XI are involved in exocytosis of cellulose synthase complexes. *Plant Physiol*, 179(4), 1537-1555.
- Zhang, X., Dominguez, P. G., Kumar, M., Bygdell, J., Miroshnichenko, S., Sundberg, B., . . . Niittylä, T. (2018). Cellulose synthase stoichiometry in aspen differs from *Arabidopsis* and Norway spruce. *Plant Physiol*, 177(3), 1096-1107.
- Zhang, X., Henriques, R., Lin, S. S., Niu, Q. W., & Chua, N. H. (2006). Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat Protoc*, 1(2), 641-646.

- Zhang, Y., Inouye, H., Yang, L., Himmel, M. E., Tucker, M., & Makowski, L. (2015). Breakdown of hierarchical architecture in cellulose during dilute acid pretreatments. *Cellulose*, 22(3), 1495-1504.
- Zhang, Y., Nikolovski, N., Sorieul, M., Vellosillo, T., McFarlane, H. E., Dupree, R., . . . Dupree, P. (2016). Golgi-localized STELLO proteins regulate the assembly and trafficking of cellulose synthase complexes in Arabidopsis. *Nat Commun*, 7, 11656.
- Zhang, Y., & Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*, 33(7), 2302-2309.
- Zhang, Z., Bharatham, K., Sherman, W. A., & Mihalek, I. (2010). deconSTRUCT: general purpose protein database search on the substructure level. *Nucleic Acids Res*, 38(Web Server issue), W590-W594.
- Zhong, R., Burk, D. H., Morrison, W. H., 3rd, & Ye, Z. H. (2002). A kinesin-like protein is essential for oriented deposition of cellulose microfibrils and cell wall strength. *Plant Cell*, 14(12), 3101-3117.
- Zhong, R., Burk, D. H., Morrison, W. H., 3rd, & Ye, Z. H. (2004). FRAGILE FIBER3, an Arabidopsis gene encoding a type II inositol polyphosphate 5-phosphatase, is required for secondary wall synthesis and actin organization in fiber cells. *Plant Cell*, 16(12), 3242-3259.
- Zhong, R., Burk, D. H., Nairn, C. J., Wood-Jones, A., Morrison, W. H., 3rd, & Ye, Z. H. (2005). Mutation of SAC1, an Arabidopsis SAC domain phosphoinositide phosphatase, causes alterations in cell morphogenesis, cell wall synthesis, and actin organization. *Plant Cell*, 17(5), 1449-1466.
- Zhong, R., Morrison, W. H., 3rd, Freshour, G. D., Hahn, M. G., & Ye, Z. H. (2003). Expression of a mutant form of cellulose synthase AtCesA7 causes dominant negative effect on cellulose biosynthesis. *Plant Physiol*, 132(2), 786-795.
- Zhong, R., Pena, M. J., Zhou, G. K., Nairn, C. J., Wood-Jones, A., Richardson, E. A., . . . Ye, Z. H. (2005). Arabidopsis fragile fiber8, which encodes a putative glucuronyltransferase, is essential for normal secondary wall synthesis. *Plant Cell*, 17(12), 3390-3408.
- Zhong, R., Taylor, J. J., & Ye, Z. H. (1997). Disruption of interfascicular fiber differentiation in an Arabidopsis mutant. *Plant Cell*, 9(12), 2159-2170.

- Zhu, J., & Weng, Z. (2005). FAST: a novel protein structure alignment algorithm. *Proteins*, 58(3), 618-627.
- Zhu, X., Li, S., Pan, S., Xin, X., & Gu, Y. (2018). CSI1, PATROL1, and exocyst complex cooperate in delivery of cellulose synthase complexes to the plasma membrane. *Proc Natl Acad Sci USA*, 115(15), E3578-E3587.
- Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kubler, J., Lozajic, M., . . . Alva, V. (2018). A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol*, 430(15), 2237-2243.
- Zulawski, M., Braginets, R., & Schulze, W. X. (2013). PhosPhAt goes kinases--searchable protein kinase target information in the plant phosphorylation site database PhosPhAt. *Nucleic Acids Res*, 41(Database issue), D1176-D1184.