Structural study of plant cellulose synthase and atypical profilin3

Qiao, Zhu

2019

Qiao, Z. (2019). Structural study of plant cellulose synthase and atypical profilin3. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/137087>

<https://doi.org/10.32657/10356/137087>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 18:09:20 SGT

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

Structural study of plant cellulose synthase and atypical profilin3

Zhu Qiao

SCHOOL OF BIOLOGICAL SCIENCES

2019

Structural study of plant cellulose synthase and atypical profilin3

Zhu Qiao

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological
University in partial fulfilment of the requirement for the
degree of Doctor of Philosophy

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me except where otherwise stated in this thesis. The thesis work has not been submitted for a degree or professional qualification to any other university or institution. I declare that this thesis is written by myself and is free of plagiarism and of sufficient grammatical clarity to be examined. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

16/12/19
.....
Date

.....
Zhu Qiao

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it of sufficient grammatical clarity to be examined. To the best of my knowledge, the thesis is free of plagiarism and the research and writing are those of the candidate's except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

23/12/2019

.....
Date

.....
Assoc Prof Gao Yonggui

Authorship Attribution Statement

This thesis contains material from 1 paper published in the following peer-reviewed journal in which I am listed as an author.

Chapter 2 is published as

Qiao, Zhu*, Sun, He*, Ng, Justin Tze Yang, Ma, Qianqian, Koh, Si Hui, Mu, Yuguang, . . . Gao, Yong-Gui. (2019). Structural and computational examination of the Arabidopsis profilin-Poly-P complex reveals mechanistic details in profilin-regulated actin assembly. *Journal of Biological Chemistry*. doi:10.1074/jbc.RA119.011307. (* equal contribution)

The contributions of the co-authors are as follows:

- Assoc Prof Gao Yonggui and Asst Prof Miao Yansong supervised the project.
- I purified the proteins used in the publication and performed all the crystal related work at the School of Biological Science, NTU.
- Sun He, Ma Qianqian, and Koh Si Hui did the yeast-based *in vivo* work.
- Ng Justin Tze Yang did the molecular dynamics simulation under the supervision of Assoc Prof Mu Yuguang.
- Sun He, Ng Justing Tze Yang and I did the data analysis, figure preparation, and wrote the manuscript; Assoc Prof Mu Yuguang, Assoc Prof Gao Yonggui and Asst Prof Miao Yansong revised the manuscript.

16/12/19
.....
Date

.....
Zhu Qiao

Acknowledgement

I would like to express my sincere gratitude to my supervisor Assoc Prof Gao Yonggui for accepting me as his Ph. D student. His mentoring, continuous support, patience, and trust during my Ph. D study mature me to be a research scientist. His suggestions can always guide me pave the way to progress the research when I am faced with difficulties. His broad knowledge scope, specification on structural biology, and enthusiasm to scientific research are all very inspiring and influencing for me. He is such an excellent mentor that I am deeply grateful for working with him for the past four years.

I also want to thank my thesis advisory committee: Assoc Prof. Liang Zhao-Xun and Assnt Prof. Miao Yansong for their insightful comments and encouragement during the thesis advisory committee meeting. Their expertise with different background guide me and broaden my research from various perspectives. Moreover, I learned a lot from Prof Miao while we drafted and revised the Profilin manuscript. His meticulous attitude and passion for interesting biological questions are truly what I should learn from. I also thank Assoc Prof Mu Yuguang who guides the molecular dynamics simulation of Profilin project.

I want to thank my important collaborator Sun He, Justin Tze-Yang Ng. With Sun He's specialization in yeast, plant genetics, biochemistry, and cell biology and Justin's expertise in molecular simulation and computation, we finished the profilin project together by such an interdisciplinary collaboration. It's such a pleasure to

work with them. Moreover, I thank Dr. Basu Shibom, Dr. Wang Meitian from Paul Scherrer Institute (PSI, Switzerland) for collecting the Selenium anomalous signal dataset of AtCesA3^{CatD} crystals and diffraction datasets of AtPRF3Δ22. I also would like to thank Dr. Andrew Wong for teaching me the negative staining and Cryo grids preparation, operation of T12 electron microscopy and Arctica electron microscopy.

I would like to thank all Gao lab members for creating such an enjoyable and friendly environment. Dr. Veerendra Kumar and Dr. Rya Ero are always willing to help me and give me suggestions. The discussions with Dr. Yan Xinfu, Dr. Bu Wenting, Jin Shengyang are also of great value and inspiration. I also want to express my thanks to entire Gao lab members, Goh Kwok Jian, Su Weixin, Jackie Tan Yen, and Loh Zhi Yang.

Most importantly, I want to express my gratitude and love to my parents and my family. From the primary school to the Ph. D study, it is such a long journey and they always encourage me and finally I am going to finish my study as a student. And most of all, my loved wife, Liu Ping whose faithful support and companion to me is truly appreciated.

Lastly, I would present my thanks to Nanyang Technological University for providing me the research scholarship, as well as to Nanyang Institute of Structural Biology and School of Biological Sciences for the excellent research facility and environment.

List of Tables

Table 1: Recipe of 40x amino acid mixture.....	52
Table 2. Mutation sites in CesAs with annotation.	90
Table 3: Full-length CesA constructs summary.....	92
Table 4. The proteins used for calibration.	125
Table 5. The calculated molecular weight of AtPRF3 variants.	127

List of Figures

Figure 1. Chemical structure of cellulose	16
Figure 2. The synthesis and functions of UDP-Glc in plants	18
Figure 3. Structure of typical GT-A type and GT-B type glycosyltransferase.....	19
Figure 4. Diversity of Bacterial cellulose synthase operons.....	21
Figure 5. Structure of bacterial BcsA and BcsB complex	22
Figure 6. The activation of BcsA by c-di-GMP.....	24
Figure 7. Catalytic mechanism of BcsA	26
Figure 8. Crystal structures of BcsC	27
Figure 9. Structure of <i>Acetobacter xylinum</i> BcsD.....	28
Figure 10. Crystal structure of BcsG	29
Figure 11. Crystal structure of BcsZ.....	30
Figure 12. The schematic diagram of plant cell wall.....	32
Figure 13. Model of plant CSC and CSC related proteins.....	34
Figure 14. Domain organization of Arabidopsis AtCesA3.....	35
Figure 15. <i>In vivo</i> trafficking of CSC	40
Figure 16. Chemical structures of cellulose synthesis inhibitors.....	42
Figure 17. Characterization of cellulose synthesis inhibitors	43
Figure 18. Dimer model of Rice CesA8 ^{CatD} calculated by SAXS	45
Figure 19. SAXS model of AtCesA1 ^{CatD}	46
Figure 20. Crystal structure of OsCesA8 ^{P-CR} dimer.....	47
Figure 21. Flow-chart of experimental approaches	64
Figure 22. Small-scale test of AtCesA3 ^{CatD} constructs.....	64
Figure 23. Nickel-affinity purification of AtCesA3 ^{CatD} VC11	65
Figure 24. Purified CesA ^{CatD} homologs.....	66
Figure 25. Disordered region prediction and small-scale expression test of C-SR loop deletions	68
Figure 26. Purification of AtCesA3 ^{CatD} LD0	69
Figure 27. The initial crystal screening hits.....	70
Figure 28. The optimized crystals of OsCesA8 ^{CatD} LD0	71
Figure 29. Crystal optimization of AtCesA3 ^{CatD} LD0	72

Figure 30. Purification of Se-Met substituted AtCesA3 ^{CatD} LD0	73
Figure 31. Optimized crystal of Se-Met substituted AtCesA3 ^{CatD} LD0	74
Figure 32. Purification of AtCesA3 ³¹⁷⁻⁸¹⁰	75
Figure 33. The crystals and diffraction image of AtCesA3 ³¹⁷⁻⁸¹⁰	77
Figure 34. Calibration of AtCesA3 ³¹⁷⁻⁷⁹² and AtCesA3 ³¹⁷⁻⁸¹⁰	78
Figure 35. Crystal structure of AtCesA3 ³¹⁷⁻⁸¹⁰	80
Figure 36. Dimer interface of AtCesA3 ³¹⁷⁻⁸¹⁰	83
Figure 37. UDP-Glc bound structure of AtCesA3 ³¹⁷⁻⁷⁹²	85
Figure 38. Comparison of AtCesA3 ^{CatD} to BcsA GT domain	87
Figure 39. Possible trimers of AtCesA3 ^{CatD}	89
Figure 40. Mutation sites mapped on AtCesA3	91
Figure 41. Small-scale purification results of AtCesAs	93
Figure 42. Small-scale expression test of AtCesAs isoforms	94
Figure 43. Detergent testing on AtCesA1 constructs	95
Figure 44. Purification of AtCesA1	96
Figure 45. Negative staining of AtCesA1	97
Figure 46. Grafix of full-length AtCesA1	99
Figure 47. The typical micrograph of cross-linked AtCesA1	100
Figure 48. 2D average results of AtCesA1	101
Figure 49. The CTF fitting and auto-picked particles	103
Figure 50. The 2D-classification of AtCesA1	104
Figure 51. The 3D classification of AtCesA1	105
Figure 52. Fitting BcsA to AtCesA1 Cryo-EM map	106
Figure 53. Edge-to-edge interaction of AtCesA	110
Figure 54. CSC model	111
Figure 55. Cytoskeleton in eukaryotic cells	114
Figure 56. Structure of tubulin and tubulin-like protein FtsZ	115
Figure 57. Structure of Actin and actin like protein	116
Figure 58. Overview of proteins regulating actin filaments	117
Figure 59. Multiple sequence alignment of AtPRF1-5	119
Figure 60. Purification of AtPRF3 variants	125

Figure 61. Calibration size-exclusion of four standard proteins.....	126
Figure 62. The calibration of AtPRF3 variants.....	127
Figure 63. The optimized crystals of AtPRF3 Δ 37.....	128
Figure 64. The representative crystals of AtPRF3 Δ 22 variants.....	129
Figure 65. The purification of AtPRF2.....	130
Figure 66. Structure of AtPRF3	131
Figure 67. Crystal structure of apo and Poly-P bound AtPRF2.....	133
Figure 68. The steric hindrance of Poly-P binding pocket with AtPRF3 ^{NTE}	134

Abbreviations

<i>A. thaliana</i>	<i>Arabidopsis thaliana</i>
ARM	Armadillo
BiFC	Bimolecular fluorescence
BIN2	Brassinosteroid Insensitive
C-SR	Class-specific region
CatD	Catalytic domain
CBI	Cellulose biosynthesis inhibitor
CBM	Carbohydrate binding module
CBM49	Carbohydrate binding module family 49
CD	Circular dichroism
CesA	Cellulose synthase
Cryo-EM	Cryo electron microscopy
CSC	Cellulose synthase complex
DCB	2,6-dichlorobenzonitrile
DDM	n-Dodecyl- β -D-Maltopyranoside
<i>E. coli</i>	<i>Escherichia coli</i>
FC-12	Fos-Choline-12
GA	Glutaraldehyde
GDN	Glyco-diosgenin
GFP	Green fluorescent protein
GH	Glycosylhydrolase
GPI	Glycosylphosphatidylinositol
GT	Glycosyltransferase
<i>H. vulgare</i>	<i>Hordeum vulgare</i>
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K _d	dissociation constant
LMNG	Lauryl maltose neopentyl glycol
MASC	Microtubule- associated CesA compartment

MST	Microscale thermophoresis
NTE	N terminal extension
<i>O. sativa</i>	<i>Oryza sativa</i>
OD	Optical density
P-CR	Plant-Conserved Region
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PIPES	1,4-Piperazinediethanesulfonic acid
Poly-P	Polyproline region
PPP	Protein Production Platform
RMSD	Root mean square deviation
SAXS	Small angle x-ray scattering
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide electrophoresis
Se-Met	Selenium methionine
SEC	Size-exlcusion chromatography
SmaCC	Small CesA compartment
SuSy	Sucrose synthases
TEV	Tobacco etch virus
T _m	Half-melting temperature
TMW	Theoretical molecular weight
TPR	Tetratricopeptide repeat
UDP	Uridine-5'-diphosphate
UDP-Glc	Uridine-5'-diphosphate glucose
UGPase	UDP-Glc pyrophosphorylase
USPase	UDP-sugar pyrophosphorylase
UTP	Uridine-5'-triphosphate
VPP	Voltage phase plate
WT	Wild type
β-ME	β-mercaptoethanol

Table of Contents

Acknowledgement	1
List of Tables	3
List of Figures.....	4
Abbreviations.....	7
Summary.....	14
Part 1. Structural study of plant cellulose synthase.....	16
1. Introduction.....	16
1.1 Introduction to cellulose.....	16
1.2 Introduction to UDP-Glc.....	17
1.3 Introduction to glycosyltransferase	18
1.4 Introduction to bacterial cellulose synthase	19
1.4.1 Function of bacterial cellulose.....	19
1.4.2 Diversity of bacterial cellulose synthase operon	20
1.4.3 Bacterial cellulose synthase BcsA and BcsB.....	21
1.4.3.1 Structure of BcsA and BcsB.....	21
1.4.3.2 The activation of BcsA by c-di-GMP.....	23
1.4.3.3 The catalytic mechanism of BcsA.....	24
1.4.4 Bacterial cellulose synthase C (BcsC).....	26
1.4.5 Bacterial cellulose synthase D (BcsD).....	27
1.4.6 Bacterial cellulose synthase G (BcsG).....	28
1.4.7 Bacterial cellulose synthase Z (BcsZ)	29
1.5 Introduction to plant cell wall	31
1.5.1 The chemical nature of primary carbohydrates in plant cell wall	32
1.5.2 Introduction of plant CSC.....	33
1.5.3 Introduction of plant CesA	34
1.5.4 Korrigan1	36
1.5.5 Cobra.....	37
1.5.6 Cellulose synthase-interaction protein 1	37
1.5.7 Cellulose synthase companion proteins.....	38
1.5.8 The assembly and trafficking of CSC.....	39

1.5.9 The regulation of CSC	40
1.6 The similarity and interactions among CesAs	41
1.7 Inhibition of cellulose biosynthesis.....	42
1.8 Previous structural study of plant cellulose synthase.....	44
1.8.1 Cryo-tomography model of CSC.....	44
1.8.2 Dimeric model of rice CesA ^{CatD}	45
1.8.3 Trimeric model of AtCesA1 ^{CatD}	45
1.8.4 Crystal structure of rice CesA8 ^{P-CR}	46
2. Research objectives.....	48
3. Materials and methods	49
3.1 Molecular cloning	49
3.2 Protein expression and purification in <i>E. coli</i>	49
3.2.1 Small-scale expression test.....	49
3.2.2 Large-scale protein purification.....	50
3.3 Molecular weight determination	52
3.4 Se-Met substituted protein expression and purification.....	52
3.5 Full-length cellulose synthase expression and purification in insect cell	53
3.5.1 Full-length cellulose synthase expression in insect cell	53
3.5.2 Small-scale purification of full-length cellulose synthase.....	55
3.5.3 Large-scale purification of full-length cellulose synthase.....	56
3.6 Crystal screening, optimization, and structure determination.....	57
3.6.1 Crystal screening	57
3.6.2 Crystal optimization and data collection	57
3.6.3 Structural determination	57
3.7 Cross-linking of full-length cellulose synthase.....	58
3.8 Negative staining.....	59
3.8.1 Negative staining grid preparation.....	59
3.8.2 Negative staining data processing	60
3.9 Cryo-EM grid preparation and data processing	60
3.9.1 Cryo-EM grid preparation	60
3.9.2 Motion correction and CTF estimation.....	61
3.9.3 Particle picking and 2D classification	61
3.9.4 3D Initial model building, 3D classification, and refinement.....	62

4. Results	63
4.1 Protein preparation	64
4.1.1 Soluble constructs screening on AtCesA3 ^{CatD}	64
4.1.2 Purification of AtCesA3 ^{CatD} VC11	64
4.1.3 Screening of cellulose synthase isoforms	65
4.1.4 Design of C-SR loop deletions and small-scale expression test.....	67
4.1.5 Purification of CesA ^{CatD} loop deleted constructs	69
4.2 Crystal screening, optimization, and structure determination.....	70
4.2.1 Crystal screening of purified CesA ^{CatD}	70
4.2.2 Crystal optimization of OsCesA8 ^{CatD} LD0	70
4.2.3 Crystal optimization of AtCesA3 ^{CatD} LD0.....	71
4.2.4 Se-Met protein purification and crystal optimization	73
4.2.5 Purification of AtCesA3 ³¹⁷⁻⁸¹⁰	74
4.2.6 Crystal screening and optimization of AtCesA3 ³¹⁷⁻⁸¹⁰	76
4.2.7 Molecular weight determination by SEC	77
4.3 Structure of AtCesA3 ^{CatD}	78
4.3.1 Structure of AtCesA3 ³¹⁷⁻⁸¹⁰	78
4.3.2 Dimer interface of AtCesA3 ³¹⁷⁻⁸¹⁰	81
4.3.3 UDP-Glc and Mn ²⁺ binding of AtCesA3 ³¹⁷⁻⁷⁹²	83
4.3.4 Structural comparison with BcsA GT domain.....	85
4.3.5 Proposed trimer models of AtCesA3 ^{CatD}	87
4.3.6 Mutation sites mapping on AtCesA3 ^{CatD}	89
4.4 Full-length cellulose synthase expression and purification	92
4.4.1 Small scale expression of AtCesA1 and AtCesA3 constructs.....	92
4.4.2 Detergent screening of AtCesA1	94
4.4.3 Purification of AtCesA1	95
4.5 Negative staining of AtCesA1	96
4.5.1 negative staining of AtCesA1	96
4.5.2 Cross-linking of purified AtCesA1 by Grafix	98
4.5.3 Negative staining data processing of Grafixed AtCesA1	99
4.6 Cryo-EM data processing of cross linked AtCesA1	101
4.6.1 Motion correction, CTF estimation and particle picking.....	102
4.6.2 2D classification	103
4.6.3 3D model generation and model fitting	104

5. Discussion	107
5.1 Function of plant P-CR and C-SR domains	107
5.2 Mechanism of plant cellulose synthesis	108
5.3 Oligomerization of plant cellulose synthase	108
5.4 Sample preparation for Cryo-EM.....	112
Chapter 2: Structural study of Arabidopsis Profilin3.....	113
1 introduction	113
1.1 Introduction to cytoskeleton.....	113
1.2 Introduction to eukaryotic actin network	116
1.3 Introduction to Profilin.....	117
2. Research objectives.....	120
3. Materials and method.....	121
3.1 Molecule cloning.....	121
3.2 Protein expression and purification	121
3.3 Crystallization, optimization, and structure determination	122
3.3.1 Crystal screening of AtPRFs	122
3.3.2 Crystal optimization of AtPRFs.....	122
3.3.3 Data collection and structure determination	123
4. Results	124
4.1 Purification of AtPRF3Δ37, AtPRF3Δ22, and AtPRF3.....	124
4.2 Calibration size-exclusion chromatography.....	125
4.2.1 Standard protein calibration.....	125
4.2.2 Calibration of AtPRFs	126
4.3 Crystal optimization of AtPRF proteins.....	127
4.3.1 Crystal optimization of AtPRF3Δ37.....	127
4.3.2 Crystal optimization of AtPRF3Δ22.....	128
4.4 Purification, crystallization of AtPRF2, and Poly-P bound AtPRF2	129
4.5 Structure of AtPRF3Δ37 and AtPRF3Δ22	130
4.6 Structural analysis of AtPRF2 and Poly-P bound AtPRF2.....	132
4.7 N terminus of AtPRF3Δ22 occupied the Poly-P binding groove	133
5. Discussion	135

5.1 Profilin is evolutionally conserved but AtPRF3 is unique.....	135
5.2 Molecular dynamics simulation aids in understanding AtPRF3 ^{NTE} dynamics.....	135
5. References.....	137
6. Appendix.....	149
1. Genes encoded by the bacterial cellulose synthase operon.....	149
2. Constructs summary of CesaA for X-ray crystallography study	150
3. Data collection and refinement statistics of structures in Chapter 1	151
4. Secondary prediction of full-length AtCesA3.....	152
5. Multiple sequence alignment of AtCesAs.....	154
6. Secondary structure prediction of AtPRF3	156
7. Data collection and refinement statistics of structures in Chapter 2	157
7. Publication	158

Summary

There are in total of two chapters included in this thesis. The first chapter focuses on the structural study of Arabidopsis cellulose synthase and the second chapter focuses on the structural study of Arabidopsis atypical profilin3.

Plant cellulose synthase catalytic domains are cloned, expressed in *E. coli*, purified, and crystallized. The crystal structure of AtCesA3^{CatD} is solved by experimental method utilizing the selenium anomalous signal. Four structures that are AtCesA3³¹⁷⁻⁷⁹², UDP-AtCesA3³¹⁷⁻⁷⁹², UDP-Glc-AtCesA3³¹⁷⁻⁷⁹², and AtCesA3³¹⁷⁻⁸¹⁰ are reported in chapter 1. The structure of AtCesA3³¹⁷⁻⁸¹⁰ shows a conserved inner core GT domain surrounded by P-CR and C-SR domains. Residues that are important for UDP-Glc and Mn²⁺ binding are identified. The AtCesA3^{D765} that is the catalytic residue resides in the flexible loop and the AtCesA3^{QxxRW} motif is inserted into the catalytic region that is different to bacterial cellulose synthase BcsA, which suggests a different catalytic and translocation mechanism for plant cellulose synthesis. Moreover, the AtCesA3³¹⁷⁻⁸¹⁰ is a homo-dimer in solution and the dimer interfaces are mapped to β 6 and α 9. Notably, the β 6 from one protomer interacts with the other by main chain oxygen and nitrogen forming in a total of four hydrogen bonds, which is classified as “edge to edge interaction” mode. Mutation of AtCesA3^{C573P} on β 6 disrupts the dimer formation along β 6- β 6' resulting in a monomeric state in solution. Furthermore, the full-length AtCesA1 is also successfully expressed and purified from insect cells. Preliminary Cryo-EM map at low resolution demonstrated trimeric assembly, which paves the way to better elucidate the organization of CSC.

Arabidopsis profilin3 (AtPRF3) encodes 37 amino acids long N terminal extension, which is atypical among the other four profilins. AtPRF3 having such an N terminal extension binds formin1 Poly-P region stronger than other profilin isoforms and is more prone to aggregate, which results in negative regulation on actin polymerization. The crystal structure of AtPRF3 Δ 37, AtPRF3 Δ 22, AtPRF2, and Poly-P bound AtPRF2 are determined and discussed in chapter 2. The profilin core domain of AtPRF1, AtPRF2, and AtPRF3 are structurally conserved with an RMSD of ~ 0.4 . The N terminal residues K31 to N37 of AtPRF3 are resolved which protrude along with the first α helix. Three residues (W3, Y6, and Y125) of AtPRF2 are identified to be directly involved in Poly-P binding. The N terminal extension of AtPRF3 Δ 22 is found to occupy the Poly-P binding pocket compared to Poly-P bound AtPRF2, which suggests the flexibility of N terminal extension towards Poly-P binding.

Part 1. Structural study of plant cellulose synthase

1. Introduction

1.1 Introduction to cellulose

Cellulose is the most abundant biopolymer on earth with a formula of $(C_6H_{10}O_5)_n$ in which hundreds to thousands of D-glucose are β (1 $\rightarrow$ 4) linked with glycosidic bonds (Figure 1) [1]. Because of the polymerization, the two ends of cellulose are chemically different. One end contains the D-glucopyranose unit in equilibrium with the aldehyde group connecting the C1 and C5 atom (reducing end), while the other end has an anomeric C atom linked with an glycosidic bond (non-reducing end) (Figure 1) [2]. Cellulose is a vital component of cell wall in a variety of organisms, ranging from bacteria, and algae to green plants. Cellulose has a wide range of industrial applications, such as raw material for paper and fibers, as well as serving the source of renewable biofuels [1]. Cellulose is synthesized by cellulose synthase (CesA) that adds UDP-glucose (UDP-Glc) to elongating non-reducing end of cellulose chain and translocated to the apoplast concurrently.

Figure 1. Chemical structure of cellulose

The minimum repeating unit of cellulose, cellobiose is shown in parentheses. The Non-reducing and reducing ends are shown. The figure is adapted [3].

1.2 Introduction to UDP-Glc

The UDP-Glc is not only the key substrate for sucrose and polysaccharides synthesis, but it also serves as the sugar donor for many glycosylation processes on proteins, lipids, and hormones [4]. In animals, the UDP-Glc can serve as an extracellular signaling molecule that activates its receptor P2Y₁₄ which is a broadly expressed G protein-coupled receptor thus triggers several downstream signaling responses like the activation of MAP kinase signaling pathway [5]. However, there is no such candidate UDP-Glc receptors have been identified on the plant plasma membrane yet.

Three distinctive enzyme classes in plants, UDP-Glc pyrophosphorylase (UGPase), UDP-sugar pyrophosphorylase (USPase) and sucrose synthases (SuSy) are responsible for the synthesis the UDP-Glc *in vivo* (Figure 2) [4]. The UTP-glucose-1-phosphate uridylyltransferase which is a member of UGPase family catalyzes the reversible replacement of phosphoanhydride bond on UTP by phosphate group on glucose-1-phosphate to produce UDP-Glc [6]. USPase uses a broad range of substrates including glucose-1-phosphate, galactose-1-phosphate, xylose-1-phosphate, arabinose-1-phosphate and glucuronate-1-phosphate to synthesize the UDP-monosaccharide in the presence of UTP [7]. SuSy can break down sucrose with UTP to UDP-Glc and D-fructose [8]. UDP-Glc is mainly produced by UGPase in soft tissues and the degradation of sucrose by SuSy in sink tissues in plants [4]. The expression of UGPase, USPase, and SuSy are fine regulated to sustain the UDP-Glc concentration in plants [4].

Figure 2. The synthesis and functions of UDP-Glc in plants

The three reversible synthesis pathways of UDP-Glc are shown with the enzymes labeled. The functions of UDP-Glc are shown on the right. The figure is adapted [4].

1.3 Introduction to glycosyltransferase

Glycosyltransferase (GT) transfers the activated sugar donor in the form of nucleotide diphosphate sugars to acceptor substrates, such as lipid, protein, nucleic acid, antibiotics, and even small molecules [9]. There are over 90 GT families classified with the canonical D, D, TED, Q/RxxRW GT signature and it is estimated that GT genes account for 1% to 2% in total of the genes in an organism including both prokaryotes and eukaryotes [9]. However, they generally fall into two folds (GT-A and GT-B folds) based on structural analysis (Figure 3). The GT-A fold shows central twisted β sheets which are surrounded by several α helices on both sides, which resembles a Rossmann-like fold [10]. Like GT-A fold, the GT-B fold has two Rossmann-like domains that are loosely associated and face each other with the catalytic pocket in the groove. The structures of SpsA from *Bacillus subtilis* and T4 phage β -glucosyltransferase are examples of GT-A and GT-B fold, respectively [11]. Cellulose synthase belongs to GT-A type GT2 family glycosyltransferase.

Figure 3. Structure of typical GT-A type and GT-B type glycosyltransferase

(A) Structure of SpsA from *Bacillus subtilis* are shown in cartoon and colored in rainbow from blue to red indicating the N and C termini (PDB code: 1QGQ). The UDP and catalytic residues are shown in sticks. The metal ion is shown in spheres. (B) Structure of T4 β -glycosyltransferase (PDB code: 1JG7). The bound UDP is shown in sticks. The metal ion is shown in spheres. Figures are prepared by PyMol [12].

1.4 Introduction to bacterial cellulose synthase

1.4.1 Function of bacterial cellulose

In addition to plants that are commonly known to produce cellulose for cell wall formation, cellulose is also produced by a wide variety of bacteria. Bacteria secrete cellulose and form the biofilm along with other components, like proteins, lipids, and DNA [13]. Biofilm enables bacteria to adhere together to form the local community in which bacterial cells share nutrients and get protected from harmful factors in the environment such as antibiotics and the host body's immune system [14]. Moreover, cellulose production enables the bacterium to adhere to its host. For example, *Rhizobium leguminosarum*, the nitrogen-fixing endosymbiont of the economically important soybean, produces cellulose that aids in adhesion and infection in root hair [14].

1.4.2 Diversity of bacterial cellulose synthase operon

Bacterial cellulose synthase complex (CSC) with four-gene *bcsABCD* forming an operon is first discovered in *Gluconacetobacter xylinus* [15, 16]. Genome-wide analysis reveals the diversity of cellulose synthase operon. The CSC operons are classified into three major types with *Komagataeibacter xylinus*, *Escherichia coli* (*E. coli*), and *Agrobacterium tumefaciens* operons as the respective representatives. Each of the three major operon types is further divided into several subtypes based on the specific gene presence and arrangement (Figure 4) [14]. The characteristic feature of the first type operon is the presence of *bcsD* gene. The second *E. coli*-like type *bcs* operon is the presence of both *bcsE* and *bcsG* gene but without the *bcsD* gene. The third type operon generally includes *bcsZ* and additional gene *bcsK*, but not *bcsD*, *bcsE* or *bcsG* [14]. Surprisingly, there is one class of *bcs* like operon that includes *bcsA* and most other *bcs* genes, but lacks of *bcsB* gene. Such *bcsB* absence operon is present in cellulose-producing cyanobacteria *Anabaena* sp. *PCC7120* and *Thermosynechococcus vulcanus*. Possibly, some potential candidates like HlyD and BcsW could function to translocate the cellulose chain to the outer membrane [14]. All the associated and identified cellulose synthase genes are summarized in Appendix 1. BcsA and BcsB constitute the core bacterial CSC that synthesizes cellulose. Other components of bacterial CSC vary in organism and generally function in three aspects that are modulating the expression of bacterial cellulose synthase, exporting the newly elongated chain to the outer membrane, and helping cellulose microfibril crystallization. These components are vital for *in vivo* cellulose

yield but not *in vitro*. In the below section, the detailed introduction of the core genes in the bcs operon and structures of solved cellulose synthase genes will be discussed.

Figure 4. Diversity of Bacterial cellulose synthase operons

Gene designations are shown in Appendix 1. The displayed operons are from *Komagataeibacter xylinus* E25 (Ia), *Dickeya dadantii* Ech703 (Ib), *Burkholderia phymatum* STM815 (Ic), *Salmonella enterica* serovar Typhimurium (IIa), *Pseudomonas putida* KT2440 (IIb), *Burkholderia mallei* ATCC 23344 (IIc), *Chromobacterium violaceum* ATCC 12472 (IId), *Agrobacterium fabrum* C58 (IIIa), *Methylobacterium extorquens* PA1 (IIIb), *Azospirillum lipoferum* 4B (IIIc), *Acidiphilium cryptum* JF-5 (IIId), *Nostoc punctiforme* PCC 73102 (IVa), and *Nostoc* sp. PCC 7120 (IVb7). The gene symbols are drawn approximately to scale. Empty shapes indicate open reading frames of unknown function. The checkered pattern indicates the tetratricopeptide repeat (TPR) domain. The figure is adapted from Trends in microbiology review [14].

1.4.3 Bacterial cellulose synthase BcsA and BcsB

1.4.3.1 Structure of BcsA and BcsB

The crystal structure of BcsA and BcsB complex from *Rhodobacter sphaeroides* is a breakthrough in understanding the molecular mechanism of bacterial cellulose

synthesis [17]. The BcsA is an inner membrane protein with 8 transmembrane helices, with each 4 transmembrane helices located in N and C termini, respectively. The GT domain that is responsible for the catalysis of cellulose chain is inserted between the fourth and fifth transmembrane helix. Besides, BcsA has a C-terminal PilZ domain that binds to bacterial secondary messenger c-di-GMP to regulate the cellulose synthesis activity. BcsB is a periplasmic protein with one C terminal transmembrane helix directly interacting with BcsA in the inner membrane. BcsB is primarily composed of four domains that are all majorly β sheet fold. The domain 3 of BcsB helps coordinate cellulose chain (Figure 5). The GT domain contains D, Dx D, D and Q/RxxRW residues that are universally conserved among GT family. Notably, the first D and second D (namely D179 and D246) bind to the uracil and diphosphate group of UDP, whereas the third D (D343) is the catalytic base facing to the non-reducing end of the cellulose chain. The QxxRW motif help bind to the terminal disaccharide of the glucan [17].

Figure 5. Structure of bacterial BcsA and BcsB complex

The BcsA is colored from N terminal GT domain in lime, the transmembrane domain in cyan to C terminal PilZ domain in red. The rest connecting-regions are in light blue. The BcsB is colored in rainbow from blue to red indicating the N and C termini. The cellulose is shown in sticks and colored in green. (PDB code: 4HG6). The figures are generated by PyMol [12].

1.4.3.2 The activation of BcsA by c-di-GMP

It is well known that the activity of bacterial cellulose synthase is regulated by c-di-GMP through BcsA's PilZ domain. Three important features are identified to be important for the regulation of cellulose synthesis activity of BcsA which are "gating loop" from residue 499 to 517, TM8-PilZ linker which connects the transmembrane helix 8 and the C terminal PilZ domain, and the "finger helix" which includes the "TED" motif [18]. The c-di-GMP bound BcsA is primarily auto-inhibited by the salt bridge formed between R580 in the gating loop and E371 in the GT domain compared to the BcsA structure in the absence of c-di-GMP which is referred to as an inactive state [17, 19]. Binding of c-di-GMP to BcsA induces the conformational change of TM8-PilZ linker. The R580 is flipped to the c-di-GMP thus forms hydrogen bonds with c-di-GMP's guanidine group. Such a drastic conformational change of R580 disrupts the salt bridge formed by R580 and E371 and liberates the flexibility of the gating loop thus activates the BcsA (Figure 6).

Figure 6. The activation of BcsA by c-di-GMP

Three structures (PDB code: 4HG6, 4P00, 4P02) are superimposed based on GT domain. The inactivated BcsA is colored in cyan, the activated BcsA with UDP is colored in lime, the activated BcsA without UDP is colored in yellow. The bound cellulose is shown in sticks. The PilZ domain is colored in cartoon. GT domain is shown in surface. The gating loop and the linker are labeled. The conformational change of R580 is indicated by the arrow. The figure is generated by PyMol [12].

1.4.3.3 The catalytic mechanism of BcsA

The initiation of cellulose synthesis needs the sitosterol- β -glucoside serving as the primer, after which the cellulose synthesis enters the elongation stage and the sitosterol- β -glucoside is cleaved [20]. However, there is a lack of further evidence of such a primer requirement for cellulose synthesis initiation. By combining the *in crystallo* enzymology, the catalytic mechanism of BcsA is well characterized through various conformational states [17-19]. The synthesis of cellulose starts from the

substrate binding. Then BcsA positions the substrate, transfers the glucose to the non-reducing end of cellulose, translocates the newly added cellulose, and finally exchanges the UDP with UDP-Glc for the next cycle (Figure 7) [19]. After the BcsA is activated by c-di-GMP by binding to the C terminal PilZ domain, the substrate UDP-Glc enters the substrate-binding pocket and the gating loop is retracted to interact with UDP-Glc to stabilize it. The donor glucose is pushed in the vicinity of acceptor glucose and TED motif in finger helix. The D343 in the TED motif serves as the base catalyst which deprotonates the C4 hydroxyl group of the glucose, thus facilitates the direct S_N2 -like displacement of the activated phosphate leaving group [9]. The newly added glucose is stabilized by the phosphate group of the UDP and conserved residue D246, which is the product-bound state. In such a state, the gating loop remains inserted to the catalytic pocket and the finger helix keeps pointing up as in the substrate-bound state. Then the gating loop is retracted to release the UDP to exchange it to UDP-Glc and the finger helix is reset to the down position to interact with the terminal glucose in the cellulose chain. Then upon the new UDP-Glc entering the substrate-binding pocket, a new cycle synthesis begins [19].

Figure 7. Catalytic mechanism of BcsA

(A) The four states of BcsA (substrate-bound, product-bound, pre-translocation, and translocation-intermediate with PDB codes 4HG6, 5EIY, 5EJ1, 5EJZ, respectively) are superimposed together by GT domain. The catalytic residues are labeled. the finger helix and gating loop are shown in cartoon. The cellulose glucan and the UDP are shown in sticks. The Mg^{2+} is shown in spheres. (B-E) The four states of cellulose processing, substrate-bound, product-bound, pre-translocation, translocation-intermediate are colored in green gradient from light green to dark green, respectively. The figures are generated by PyMol [12].

1.4.4 Bacterial cellulose synthase C (BcsC)

BcsC is an outer membrane protein with over 1000 amino acids which functions to translocate cellulose chain from periplasm to the outside membrane [21]. The N terminal part is classified as tetratricopeptide repeat (TRP) domain located in the periplasm. The C terminal part is the outer transmembrane domain with a β -barrel shape (Figure 8. A) [22]. The typical TRP motif is composed of two anti-parallel α helices linked by a β turn. Multiple TPR motifs can be assembled to form a superhelix structure. The structure of N terminal six TPR motifs from *Enterobacter CjF-002* is

determined in 2017 (Figure 8. B) [21]. Interestingly, one non-TPR α helix is inserted in the middle of TPR motifs in the crystal structure. Because of the inserted helix, the flexibility of the TPR motif gets increased, which could be important for cellulose exporting. The C terminal TPR repeat and β barrel domain are also solved recently, which shows the β barrel pore is composed of 16 β strands and the C terminus is deeply buried into the pore (Figure 8. C) [22]. Currently, there is still no evidence to indicate whether BcsC could directly interact with BcsA or BcsB.

Figure 8. Crystal structures of BcsC

(A) A schematic view of BcsC. (B) N terminal TRP repeats of BcsC from *Enterobacter* CJE-002 are colored in blue to green gradient indicating the N and C termini (PDB code: 5XW7). (C) The C terminal TRP repeat and β -barrel domain of BcsC from *E. coli* are colored in yellow to red gradient indicating the N and C termini (PDB code: 6TZK). The figures are generated by PyMol [12].

1.4.5 Bacterial cellulose synthase D (BcsD)

Bacterial cellulose synthase D (BcsD), a periplasmic protein with molecular weight around 17 kDa, is not necessary for *in vitro* cellulose production. But the cellulose

production is drastically decreased *in vivo* while the *bcsd* gene is disrupted, indicating its importance in cellulose synthesis in bacterium [23]. The crystal structure of BcsD from *Acetobacter xylinum* is determined and the BcsD is proposed to be important in cellulose chain extrusion and polymerization [24]. The AxBcsD is composed of five α helices and four β strands which form an anti-parallel β sheet surrounded by the five α helices with four α helices on one side and one α helix on the other side (Figure 9. A). The AxBcsD forms an octamer that is confirmed by size-exclusion chromatography (Figure 9. B). The complex structure of AxBcsD with cellopentaose is also determined with relatively low occupancy of cellopentaose, which indicates the binding affinity between AxBcsD and cellopentaose is low [24].

Figure 9. Structure of *Acetobacter xylinum* BcsD

(A) The structure of AxBcsD protomer is shown in cartoon and colored in rainbow from blue to red indicating the N and C termini (PDB code: 3A8E). (B) The structure of AxBcsD octamer is shown in cartoon with protomers colored respectively. The figures are generated by PyMol [12].

1.4.6 Bacterial cellulose synthase G (BcsG)

BcsG is an inner membrane protein with four transmembrane helices in N terminus and a periplasmic AlkP domain in C terminus [25]. Mutating *bcsG* gene significantly decreases the secreted cellulose level, indicating the importance of BcsG *in vivo* [26]. The crystal structure of the BcsG C terminal periplasmic domain shows seven β sheets sandwiched by several α helices and another four β sheets in C terminus (Figure 10) [25]. BcsG can chemically modify the growing glucan chain in *E. coli* and *S. typhimurium* by adding the phosphoethanolamine group to every other glucosyl residue to produce the modified cellulose, which is important for extracellular matrix assembly and biofilm formation [27]. The N terminal membrane helix is important for the direct interaction with BcsA and C terminal AlkP domain transfers the phosphoethanolamine group onto the growing cellulose molecule in the presence of Zn^{2+} [25].

Figure 10. Crystal structure of BcsG

(A) Side view of BcsG (PDB code: 5OLT) is shown in cartoon and colored in rainbow from blue to red indicating the N and C terminus. The bound Zn^{2+} is shown as spheres. (B) Top view of BcsG is shown in cartoon. The figures are generated by PyMol [12].

1.4.7 Bacterial cellulose synthase Z (BcsZ)

BcsZ is a periplasmic protein classified to be glycoside hydrolases (GH-8) family with a typical $(\alpha/\alpha)_6$ fold that six pairs of α helices form the inner and outer rings (Figure 11) [28, 29]. BcsZ is important for optimal cellulose synthesis *in vivo* by studying the *Acetobacter xylinum* and *Agrobacterium tumefaciens* homologs of BcsZ (CMCax and CelC, respectively). Reduced rate of cellulose secretion is observed while the glucanase activity of BcsZ is depleted, whereas the cellulose formation is stimulated with the increased concentration of BcsZ homolog CMCax [30, 31]. BcsZ can hardly degrade cellulose and purified BcsZ can efficiently degrade carboxymethyl-cellulose embedded in agar plates, it is hypothesized that BcsZ functions before the cellulose microfibril formation [29]. Even though the cellulose yield is decreased while the *bcsz* gene is knocked out *in vivo*, BcsZ does not affect the *in vitro* cellulose synthase activity [29].

Figure 11. Crystal structure of BcsZ

(A) The side view of BcsZ (PDB code: 3QXF) is shown in cartoon and colored in rainbow from blue to red indicating the N and C termini. the cellopentaose is shown in sticks. (B) The top view of BcsZ is shown in cartoon. The cellopentaose is shown in sticks. The figures are generated by PyMol [12].

1.5 Introduction to plant cell wall

The plant cell wall is one of the major differences between plant cells and mammalian cells [32]. It performs a large variety of biological functions such as providing mechanical strength to plant cells, protecting the plant cells against the osmotic stress by allowing cells to develop turgor pressure, and serving the immune defense for pathogen infection. Because the plant cell wall is formed between adjacent cells, it is also vital for intercellular communication [33]. Plant cell walls can have up to three layers that are middle lamella, primary cell wall, and secondary cell wall. The middle lamella rich in pectin is the most outside layer forming the interface among surrounding plant cells to glue them together. The primary cell wall is a generally thin, flexible, and extensive layer formed during cell growth. The major component of the primary cell wall is carbohydrates that include cellulose, hemicellulose, and pectin. The cellulose microfibrils are synthesized by six lobed rosette-shaped cellulose synthase complex (CSC) and glued together by hemicellulose to form the cellulose-hemicellulose network that is surrounded by pectin matrix. The thick secondary cell wall is formed inside the primary cell wall after the cell is mature (Figure 12). The secondary cell wall is not necessarily formed in all plant cell types. The composition of the secondary cell wall could differ significantly. It includes a wide variety of additional compounds that can modify its mechanical property. The secondary cell wall that makes up wood generally includes 35-50% cellulose, 20-35% hemicellulose, and 10-25% lignin that is a complex phenolic polymer driving water out to strengthen the cell wall. Proteins are also present in the plant cell wall as important components.

Figure 12. The schematic diagram of plant cell wall

The major components of plant cell walls that are cellulose microfibril, hemicellulose, and pectin are shown. The Primary wall resides between plasma membrane and middle lamella. The figure is adapted [34] .

1.5.1 The chemical nature of primary carbohydrates in plant cell wall

Different from the strong crystalline biopolymer cellulose in plant secondary cell wall that is composed of 7, 000 - 15, 000 glucose units, hemicellulose is a generally short (around 500 - 3, 000 sugar length) heterogeneous biopolymer which can be easily hydrolyzed by the diluted acid or base [35]. According to different compositions, there are many kinds of hemicellulose, such as xylan, glucuronoxylan, arabinoxylan, glucomannan, and xyloglucan that are cross-linked to pectin and cellulose to form fibers network [36]. The hemicellulose is synthesized in the Golgi apparatus and transported to the cell matrix by the trans-Golgi network [36]. Pectin is a structurally heterogeneous biopolymer that is present in the primary cell wall of terrestrial plants non-woody part [37]. Pectin is the major component of middle lamella that can bind

cells together and be deposited into the cell wall by the trans-Golgi network during exocytosis [37]. It is rich in galacturonic acid and often serves as a gelling agent in the food industry [38]. Lignin is chemically cross-linked phenolic polymers synthesized from glycosylated monolignols in the cytosol [39]. The monolignols are linked to glucose to increase the water solubility and transported into the apoplast through the cell membrane, after which the glucose group is cut off and the monolignols are polymerized into lignin [40]. Because of its hydrophobic nature, lignin is important in conducting water in plant tissues [41]. Lignin is covalently linked to hemicellulose to cross-linker various plant cell wall polymers, providing the mechanical strength to the cell wall especially in vascular and support tissues [40].

1.5.2 Introduction of plant CSC

Plant cellulose is synthesized by rosette terminal complex (RTC) at the plasma membrane that is later identified as plant cellulose synthase complex (CSC) [42, 43]. The CSC contains most likely 18 or 36 subunits and each subunit is responsible for synthesizing the corresponding cellulose chain [1, 44]. Recently the 18-CesA model has become more compelling through electron microscopy and computational modeling study of the CSC [45-47]. The detailed introduction of CSC components and CSC associated protein will be discussed below (Figure 13).

Figure 13. Model of plant CSC and CSC related proteins

Six-lobed CSC is shown with the elongating cellulose. CSC-associated proteins include cellulose synthase interactive protein 1 (CS1), companion of cellulose synthase 1/2 (CC1/2), KORRIGAN1 (KOR1), and COBRA. The figure is adapted [48].

1.5.3 Introduction of plant Cesa

Plant Cesa is first identified based on the canonical D, D, D, QxxRW motif and the sequence similarity to bacterial Cesa [49]. Similar to bacterial Cesa, plant Cesa also belongs to GT-2 superfamily [9]. The total number of *cesa* genes varies in species [50]. For *A. thaliana*, there are in total of 10 *cesa* genes identified. Each CSC contains three different cellulose synthase subunits in a 1:1:1 stoichiometry [51]. The AtCesA1, AtCesA3, AtCesA6, AtCesA6-like cellulose synthase (AtCesA2, AtCesA5, and AtCesA9) are involved in primary cell wall formation. The AtCesA4, AtCesA7, and AtCesA8 are involved in secondary cell wall formation [52]. The function of AtCesA10 remains unknown so far. Plant Cesa is predicted to contain one N terminal zinc finger domain, 8 transmembrane helices that are inserted by one conserved cytosolic GT domain between the second and third transmembrane helix (Figure 14) [53]. The zinc finger domain is possibly implicated in homo-dimer or hetero-dimer

(with other CesA) formation [54, 55]. The GT domain catalyzes the cellulose synthesis and translocates the elongating cellulose to the apoplast together with the transmembrane channel formed by the eight transmembrane helices. The plant GT domain contains two more insertions compared to bacterial CesA. One is the plant-conserved region (P-CR) and the other one is the class-specific region (C-SR) [1]. The exact functions of P-CR and C-SR have not been identified, though it is presumed that they may be involved in homo-oligomers or hetero-oligomers formation.

Figure 14. Domain organization of Arabidopsis AtCesA3

The domain organization of Arabidopsis cellulose synthase 3 (AtCesA3) is shown. TM: transmembrane helix, GT: glycosyltransferase domain. The domain boundaries are labeled.

Both *Populus tremula x tremuloides* PttCesA8 and *Physcomitrella patens* PpCesA5 heterologously expressed in yeast are both functional when purified and reconstituted in liposome [56, 57]. Plant CesA prefers to using Mn^{2+} as the cofactor for catalysis which is different from bacterial cellulose synthase BcsA utilizing the Mg^{2+} as the cofactor [56]. The enriched PttCesA8 is further characterized to catalyze the cellulose synthesis with a K_m of 30 μM at a constant Mn^{2+} concentration and the protein is only functional while reconstituted into the liposome, but not in detergent micelle

[56]. Notably, truncating the N terminus before the first transmembrane helix doesn't significantly reduce the PttCesA8 catalytic activity [56].

1.5.4 Korrigan1

In Arabidopsis, Korrigan1 (Kor1) belongs to membrane-anchored type II integral membrane glycosyl hydrolase with a cytosolic domain, a single transmembrane helix and an extracellular catalytic domain [58]. Kor1 is the plant counterpart of bacterial cellulose synthase BcsZ with unknown specific function in plants [59]. The *kor1-1* mutant showed dwarf, reduced growth, short hypocotyl, irregular surface, and altered cell wall architecture compared to the wild type [60]. Kor1 is required for normal CSC activity as the CSC velocity in *kor1-1* mutant is drastically decreased from 253 nm/min to 124 nm/min [61]. The movement of GFP-AtKor1 is similar to GFP-AtCesA3 and GFP-AtCesA6 in Arabidopsis. In epidermal cells, GFP-Kor1 substantially overlaps with mCherry-AtCesA1 in the plasma membrane [61]. The interaction of AtKor1 and AtCesAs is also confirmed by the split-ubiquitin method. The yeast grows when AtKor1 is used as the bait by fusing to C terminal part of ubiquitin and transcription factor while AtCesA1, AtCesA3, or AtCesA6 are used as prey but not the unrelated protein [62]. *In vitro* pull-down assay of the AtCesAs catalytic domain with the cytosolic domain of AtKor1 also corroborated the interaction [63]. Bimolecular fluorescence (BiFC) analysis also indicated that AtKor1 interacts with various AtCesAs [61]. The association of AtKor1 and AtCesAs supports that it is an integral component of CSC and is closely involved in cellulose synthesis. Notably, PtKor1 is also involved in secondary wall formation demonstrated by RNAi silencing study in *Populus* [64].

1.5.5 Cobra

In Arabidopsis, Cobra encodes a glycosylphosphatidylinositol-anchored (GPI) protein with 456 amino acids localized primarily in the plasma membrane and is necessary for oriented cell expansion both in the primary and secondary cell wall formation [65]. Cobra mutants have abnormal roots that expand radially rather than longitudinally under certain growth conditions [66, 67]. Brittle Culm1 (BC1), a COBRA-like protein in rice, also belongs to GPI anchored protein which can be released to cell wall after the removal of GPI anchor. BC1 includes one carbohydrate-binding module in the N terminus which enables it to bind crystalline cellulose thus regulates the crystallinity of cellulose [68].

1.5.6 Cellulose synthase-interaction protein 1

In Arabidopsis, Cellulose synthase-interaction protein 1 (CSI1) encodes a 2,050-residue protein that is co-expressed and regulated with primary cell wall CesAs [69]. The CSI1 is majorly composed of armadillo (ARM) repeat with the C terminal C2 domain, indicating it is likely a membrane-anchored protein [70]. CSI1 directly interacts with microtubule, as well as the catalytic domain of CesAs (AtCesA1, AtCesA3, and AtCesA6) [69, 70]. CSI1 could be co-aligned with microtubule and CSC revealing that CSI1 could function as the bridge that links CSC and microtubule [69, 70]. The *csi1* mutant shows altered microtubule-dependent dynamics of CSC and reduced cellulose content [69-71]. In Arabidopsis, CSI2 and CSI3 are another two CSI1 like proteins [70, 72]. CSI2 is identified as a pseudo-gene since no protein expression is detected in any of the tissues such as rosette leaves, light or dark-grown

whole seedling, flowers, siliques, stamen, and pollen. The expression of CSI3 is detected in many tissues, like CSI1. The *csi3* mutant has no obvious defects, but *csi1/csi3* double mutant showed more severe cell expansion defects. CSI3 also associates with primary CSC and co-aligned with microtubule [72].

1.5.7 Cellulose synthase companion proteins

When the plant is under adverse environments, like salt stress, the cortical microtubule providing plant cell morphology is quickly disrupted and gradually re-assembled. Companion of cellulose synthase proteins (CCs) are identified to be constitutional components of CSC protecting the plants from salt stress by promoting the formation of the salt-tolerant microtubule. The *cc1/cc2* double mutant shows reduced growth and cell swelling when cellulose synthase inhibitors, microtubule inhibitors, or salt stress is present, whereas there is no obvious phenotype change under normal growth condition [73]. CC1 and CC2 protect the microtubule and keep it integrates with CSC during adverse environments. The *cc1/cc2* mutant leads to disrupted microtubule array during adverse environments. CC proteins are type I transmembrane protein with around 320 amino acids in Arabidopsis. The N terminal part of CC protein is in cytosol involved in interaction with microtubule and the C terminal part is in apoplast [73]. Further dissection of the CC1 N terminus reveals that it bundles microtubule together thus increases microtubule stability. Y26/Y27 of CC1 are important for interacting with microtubule demonstrated by saturation transfer difference NMR measurements. Mutation of Y26A/Y27A of CC1 causes reduced hypocotyl length and cellulose content for the *cc1/cc2* mutant with low salt tolerance, indicating the phenotypes are not restored in *cc1/cc2* mutant [74]. CC proteins and

CsII perform distinct roles during cellulose synthesis, and specifically, CC proteins are necessary when the components of CSC get impaired [75].

1.5.8 The assembly and trafficking of CSC

Fluorescently tagged Cesa has tremendously accelerated the study of trafficking and characterization of CSC since 2006 [76]. The trans-Golgi network functions as a hub for protein sorting and transportation. CSC is imaged to be present in several compartments which include the Golgi apparatus, small post-Golgi-related compartments (SmaCCs), and microtubule-associated compartments (MASCs) in the cell cortex and cell membrane [77]. The Golgi apparatus localized STELL1 and 2 (STL1/2) proteins are likely to control the assembly of CSC in the Golgi apparatus since the *stl1/stl2* mutant has reduced insertion of CSC into the membrane and decreased primary and secondary cell wall cellulose compared to the wild type [78]. The STL protein that is a GT75 family homolog interacts with CesAs [78]. SHOU4/SHOU4L are identified to be novel plant-specific membrane proteins with N terminus interacting strongly with AtCesA1, weakly with AtCesA3, and not with AtCesA6 by yeast two-hybrid system [79]. The interaction of N terminus of SHOU4 and Cesa1 is further confirmed by co-immunoprecipitation assay [79]. The *shou4/shou4l* mutant shows an increased level of CSC and non-crystalline cellulose, which suggests *shou4* mediates the suppression of cell growths (Figure 15) [79].

Figure 15. *In vivo* trafficking of CSC

The assembly of CSC occurs in the Golgi apparatus where STL proteins interact with CSC and thus help its assembly and transportation. CSC is secreted by the trans Golgi network to the plasma membrane. POM2/CSI1, the exocyst complex, and PTL are required for CSC delivery. SHOU4/4L interacts with CesA and negatively affects CSC exocytosis. The CC1 and CC2 are integrated components of CSC, associate with SmaCCs/MASCs under salt stress and mediate the cortical microtubule and CSC dynamics, allowing for their recovery. Kor is a component of CSC required for optimal cellulose biosynthesis and COB is necessary for proper microfibril orientation. The endocytosed CSCs may undergo recycling back to the PM or might be targeted for degradation. The figure is adapted [80].

1.5.9 The regulation of CSC

Posttranslational modifications (PTMs) play an important role in regulating CSC function thus affects cellulose synthesis. The most extensively studied one is the phosphorylation. CesA is phosphorylated in the N terminal hyper-variable region and central cytosolic GT domain [81]. Phosphorylation mimicking of AtCesA3^{S211E} and AtCesA3^{T212E} revealed that the phosphorylation status of AtCesA3^{S211} and

AtCesA3^{T212} affects the anisotropic expansion and cellulose synthesis [82]. Moreover, mimicking the phosphorylation of AtCesA1^{T166E}, AtCesA1^{S686E}, or AtCesA1^{S688E} can complement the temperature-sensitive *cesa1-1* allele while abolishing the phosphorylation by mutating Thr/Ser to Ala cannot [81]. Notably, many CSC interacting proteins, such as CSI1, Kor1, CC1, and CC2, are also regulated by phosphorylation [48]. Brassinosteroid Insensitive (BIN2) protein kinase is the only one currently has been identified to be involved in phosphorylating CesAs [83]. Another posttranslational modification is the S-acylation of secondary cell wall CSC. For instance, secondary cell wall CesAs are S-acylated in the GT domain and C terminus. Abolishing the S-acylation by cysteine mutation does not affect the assembling and trafficking of CSC, but does affect the localization of CSC to plasma membrane [84].

1.6 The similarity and interactions among CesAs

The sequences of plant CesAs share over 70% identities among each other, and the sequence alignment is shown in Appendix 5. Multiple CesAs interact and thus assemble into the CSC. However, how the individual CesA interacts with each other remains unknown [85]. Results of *in vitro* pull-down assay, yeast two-hybrid assay and *in vivo* bimolecular fluorescence complementation assay (BiFC) have all revealed that AtCesA1, AtCesA3, and AtCesA6 can form homo-oligomers or hetero-oligomers [86-88]. An 840-kDa protein complex that represents an oligomer of several CesAs is detected by blue-native polyacrylamide gel electrophoresis (blue-native PAGE) [89]. Similar to primary cell wall CSC, the secondary cell wall component AtCesA4, AtCesA7 and AtCesA8 can all form hetero-dimers [90]. But, only AtCesA4 can

strongly interact with itself based on pull-down assay and yeast two-hybrid assay. These data suggest a different model of the secondary cell wall compared to the primary cell wall [1].

1.7 Inhibition of cellulose biosynthesis

Herbicides are classified by their mode of actions, like abolishing light processing, cell metabolism, growth, and division. The cellulose synthesis inhibitor (CBI) is classified as group L herbicides which affect cell growth [91]. Based on their chemical structures, the CBIs are further classified to Benzamide, Alkylazines, Nitriles, and Triazolocarbonoxamide (Figure 16).

Figure 16. Chemical structures of cellulose synthesis inhibitors

CBIs inhibit the cellulose synthesis. The figure is adapted from Herbicide Resistant Action Committee (www.hracglobal.com).

Several requirements must be fulfilled to identify a compound as a cellulose synthesis inhibitor. First, the plant seeds show dwarf growth and the rapidly growing tissue

show radial swelling like hypocotyl while treated (Figure 17. A and B). Second, the reduced cellulose content is proportional to the dose of inhibitors applied. Third, a drastically decreased intake of C^{14} glucose into the cell wall occurs in less than 2 hours (Figure 17 C). Despite the complexity of CSC assembly, trafficking, and organization, the fluorescent protein labeled CesaA like YFP-AtCesA6 with live-cell imaging greatly aids in characterizing the mode of actions of cellulose synthesis inhibitors (Figure 17 D)[76].

Figure 17. Characterization of cellulose synthesis inhibitors

(A) The effects of CBI on monocots and dicots as compared to untreated counterparts that include dwarfed phenotype and ectopic lignification, which are similar to CesaA deficient mutants. (B) CBI treated plant cells become radially swollen and irregularly shaped, which is visualized microscopically using a PIP2:GFP (plasma membrane reporter) and laser scanning confocal microscopy. Scale bar = 10 μm [92]. (C) Tracing of C^{14} radio-labeled glucose uptake and incorporation into cellulose. (D) CBI change in populations the population of CesaA

particles at the plasma membrane, cortex, or Golgi using the plants expressing RFP:TUA (red) and YFP:AtCesA6 (yellow) in CBI treated and untreated plants. The figure is adapted from a published review [91].

Therefore, cellulose synthesis inhibitors are also classified based on their mode of actions into three groups. Cellulose synthesis inhibitors in group one cause clearance of CSC from the plasma membrane with Isoxaben as an example [76]. The second group of cellulose synthesis inhibitors leads to increased CSC while the movement of CSC is stalled, for example, Indaziflam or DCB (2,6-dichlorobenzonitrile) [93]. Those in the third group, such as morlin (7-ethoxy-4-methyl chromen-2-one), change the trajectory of CSC to the plasma membrane and the speed of CSC at the plasma membrane focal plane [94].

1.8 Previous structural study of plant cellulose synthase

1.8.1 Cryo-tomography model of CSC

Although the plant cellulose synthase has been identified for over 20 years, the high-resolution structure of plant cellulose synthase is still not available. The catalytic domain of cellulose synthases, like cotton CesaA, rice Cesa8, and Arabidopsis AtCesA1, are computationally modeled [53, 87, 88]. But they are not experimentally verified and are hardly correct because the models are either *ab initio* generated or homologous modeled based on bacterial BcsA. *Physcomitrella patens protonemal* plasma membrane is frozen and the rosette-shaped six lobed CSC particles are manually picked from the reconstructed tomography model. After data processing, it reveals the diameter of CSC is around 23.6 nm. The average measurement of

individual lobes is around 17.0 ± 0.7 nm, supporting the 18-subunit CSC model. Notably, the adjacent lobes are quite separated with distinctive boundaries [95].

1.8.2 Dimeric model of rice CesA^{CatD}

The rice CesA8^{CatD} is purified, characterized by size-exclusion chromatography, modeled by low-resolution small-angle X-ray scattering (SAXS) [87]. The protein binds the ligands (UDP and UDP-Glc) and exhibits two oligomer states (monomer and dimer). The dimer is sensitive to reducing agents, like DTT. The dimer model calculated by SAXS is around 160 Å at the longest dimension showing a two-fold symmetry (Figure 18).

Figure 18. Dimer model of Rice CesA8^{CatD} calculated by SAXS

The longest dimension is labeled and monomer is mapped to the dimer model

1.8.3 Trimeric model of AtCesA1^{CatD}

The catalytic domain of Arabidopsis thaliana CesA1 (AtCesA1^{CatD}) is expressed in *E. coli*, solubilized with detergent sodium lauroyl sarcosine, and purified as mainly in monomeric state. Interestingly, the AtCesA1^{CatD} monomer would be transformed to trimer after removing the detergent as SAXS calculation gave a maximum dimension of 11.3 nm in terms of R_g value (Figure 19) [88]. Circular dichroism (CD) thermal

melting study results showed that the detergent-solubilized protein is more thermally stable in the presence of UDP or UDP-Glc with increased half-melting temperature (T_m) from 69.5 °C to 72.4°C and 74.5 °C upon adding of UDP-Glc and UDP, respectively. However, it's skeptical that such a protein can be thermally stable with T_m above 70°C. Also, the trimeric protein is resistant to reducing agents like DTT, which rules out the possibility of specific or non-specific disulfide bond formation.

Figure 19. SAXS model of AtCesA1^{CatD}

The fitting of *ab initio* computational model built by ROSETTA (shown in orange and surface in dark gray) of AtCesA1^{CatD} to SAXS model (light gray).

1.8.4 Crystal structure of rice CesA8^{P-CR}

The crystal structure of rice CesA8^{P-CR} trespassing the residues 399-523 reveals that the rice CesA8^{P-CR} is composed of two α helices connected by a 25 amino acids long loop (Figure 20) [96]. However, only residues 399-475 can be modeled in the electron density map, which indicates residues 476-523 is flexible in such a construct. The protein is a dimer in both solution and crystal structure. Therefore, the author suggests P-CR could be the region responsible for CesA oligomerization [96].

Figure 20. Crystal structure of OsCesA8^{P-CR} dimer

(A) Side view of OsCesA8^{P-CR} dimer (PDB ID: 5JNP). Chain A is shown in cartoon and colored in rainbow from blue to red indicating the N and C termini. The chain B is colored in cyan. (B) Top view of OsCesA8^{P-CR}. The figures are generated by PyMol [12].

2. Research objectives

Considering plant cellulose's dominant abundance on earth and many potential applications especially being the source of renewable biofuel, understanding the mechanism of plant cellulose synthesis is of great importance. In contrast to tremendous progress in understanding bacterial cellulose biosynthesis, limited information is available for structural basis and molecular mechanism of plant cellulose biosynthesis due to its unique features. First, the plant cellulose synthase contains an N terminal zinc finger domain and two additional motifs (P-CR and C-SR) in GT domain that are not present in bacterial cellulose synthase. Secondly, the plant CSC is composed of multiple cellulose synthase subunits, while bacterial CSC only contains one cellulose synthase. Thirdly, many other proteins are associated with CSC and possibly regulate the CSC function. How plant cellulose is synthesized and how individual cellulose synthase interacts with each other to assemble into CSC largely remain elusive. To address these questions, we investigate the structures of the catalytic domain of plant cellulose synthase and full-length plant cellulose synthase through X-ray crystallography and Cryo-Electron Microscopy.

3. Materials and methods

3.1 Molecular cloning

The cellulose synthase genes from *Arabidopsis thaliana* were purchased from TAIR (<https://www.arabidopsis.org/>) and the cellulose synthase genes from *Hordeum vulgare* were purchased from Genetic Resource Center (NARO, Japan). The cellulose synthase genes from *Oryza sativa* were amplified from the rice cDNA library. The catalytic domains of cellulose synthase were amplified by Q5 DNA polymerase (New England Biolabs), digested by restriction enzymes (New England Biolabs), and ligated to pET-28a (+) (Novagen) derived vector pDBHT-2-TEV with an N terminal cleavable hexahistidine tag or pET-28a (+) (Novagen) with C terminal non-cleavable hexahistidine tag. The C-SR region loop deletions were completed by site-directed mutagenesis. Full-length cellulose synthase genes were cloned into pFastbac1 (Invitrogen) with a C terminal hexahistidine tag. All the plasmids were sequenced and confirmed by ^{1st}Base (Singapore). The constructs used in this study were summarized in Appendix 2.

3.2 Protein expression and purification in *E. coli*

The sequenced plasmids expressing CesA^{CatD} were transformed into *E. coli* BL21 (DE3). A single colony was picked and inoculated into 5 mL 2x YT medium and cultured overnight.

3.2.1 Small-scale expression test

For small-scale protein purification, 10 mL 2x YT medium was used for cell culture. The cells were harvested, resuspended in buffer-HisA (25 mM Na-HEPES, pH 7.5, 150 mM NaCl, 20 mM Imidazole), sonicated for 1 minute (4 seconds on, 8 seconds off), and clarified by centrifugation. The supernatant was collected and incubated with 15 µl Nickel NTA beads (Biobasic, Singapore). After extensively washing out with the buffer-HisA, the protein was eluted in the buffer-HisB (25 mM Na-HEPES, pH 7.5, 150 mM NaCl, 250 mM Imidazole). The eluted protein was loaded to SDS-PAGE to check purity.

3.2.2 Large-scale protein purification

For large-scale protein purification, the cells were inoculated into 5-liter 2x YT medium in six flasks. When the O.D reached 0.8, the cells were cooled down in cold room for half an hour then induced by 0.1 mM Isopropyl β -D-1-thiogalactopyranoside (IPTG) at 16°C for 16 hours. The cells were harvested by centrifugation at 4,000 rpm (JLA-8.1, Beckman, USA) for 10 minutes and the pellet was resuspended in buffer-HisA. The cells were lysed by LM20 Microfluidizer and clarified by centrifugation at 20,000 rpm (JA-25.5, Beckman, USA) at 4°C for 1 hour. The supernatant was collected and filtered through 0.22 µm filter and was further loaded onto a HisTrap FF 5 mL column (GE Healthcare) pre-equilibrated with buffer-HisA. After extensively washing, the target protein was eluted by the imidazole gradient. The samples were checked by SDS-PAGE.

For the constructs having an N terminal cleavable hexahistidine tag, the fractions with less impurity were pooled and dialyzed overnight against buffer-GF (25 mM Na-HEPES, pH 7.5, 150 mM NaCl) by adding Tobacco etch virus (TEV) protease for hexahistidine tag cleavage. The overnight-dialyzed protein sample was load onto HisTrap FF 5 mL column (GE healthcare) again to remove the uncut protein and TEV protease. The protein without tag was then concentrated to 5 mL and loaded to HiLoad Superdex 200 16/60 (GE Healthcare) column pre-equilibrated with buffer-GF. The protein fractions were pooled, concentrated and snap-frozen in liquid nitrogen and stored at -80 °C for later use.

For the constructs with C terminal non-cleavable hexahistidine tag, the protein samples eluted from nickel affinity purification were pooled and loaded onto HiPrep 26/60 desalting column (GE Healthcare) pre-equilibrated with buffer-QA (25 mM Na-HEPES, pH 7.5, 50 mM NaCl) to remove the imidazole and decrease the salt concentration. Subsequently, the desalted protein fraction was loaded to HiTrap Q HP 5 mL (GE Healthcare) pre-equilibrated with buffer-QA. The protein was eluted by increasing NaCl gradient in buffer-QB (25 mM Na-HEPES, pH 7.5, 1 M NaCl). The peaks containing target protein were collected, concentrated, and loaded to HiLoad Superdex 200 16/60 (GE Healthcare) pre-equilibrated with buffer-GF. The protein fractions were pooled, concentrated and snap-frozen in liquid nitrogen and stored at -80 °C for later use.

3.3 Molecular weight determination

Superdex 200 increase 10/300 (GE Healthcare) column was used for molecular weight determination. 100 µL protein at a concentration of 2 mg/mL was loaded onto the column pre-equilibrated with buffer-GF. The flow rate was set as 0.4 mL/min and the fractions were collected. The theoretical molecular weight (TMW) was calculated based on a standard curve that was plotted with standard protein in calibration kit (GE healthcare).

3.4 Se-Met substituted protein expression and purification

One single colony of freshly transformed *E. coli* BL21 (DE3) with the plasmid expressing AtCesA3⁽³¹⁷⁻⁷⁹²⁾ was picked into 50 mL 2x YT medium supplemented with 30 µg/mL kanamycin and cultured overnight at 37°C with 200 rpm shaking. The over-night culture was pelleted down by centrifugation at 4,000 g for 5 minutes. The pellet was washed twice with M9 medium and re-suspended in 50 mL M9 medium. Then 6 mL of resuspended bacteria were inoculated into 1-liter M9 medium supplemented with 30 µg/mL kanamycin and cultured at 37°C with 200 rpm shaking till OD600 was around 0.8. The cell was cooled down to 16°C in cold room, added by 25 mL 40x amino acid mixture (Table 1) and 1 mL 25 mg/mL selenium methionine (Se-Met) each liter of bacterial culture, and grown at 16°C for 1 hour before induced by IPTG at a final concentration of 0.2 mM. The Se-Met protein purification methods used were the same as native protein purification methods described in method section 3.2.2.

Table 1: Recipe of 40x amino acid mixture

Amino acids/250 mL	Weight
L-Lysine	1 g
L-Phenylalanine	1 g
L-Threonine	1 g
L-Isoleucine	0.5 g
L-Leucine	0.5 g
L-Valine	0.5 g

Note: the total volume is 250 mL.

3.5 Full-length cellulose synthase expression and purification in insect cell

3.5.1 Full-length cellulose synthase expression in insect cell

2 µl of the recombinant plasmid encoding Arabidopsis CesA was added into 100 µl ice thawed *E. coli* DH10Bac competent cells. The mixture was incubated on ice for 10 minutes, heat-shocked at 42°C for 1 min, and chilled on ice for 5 minutes. 900 µl of SOC medium was added to the cells. The cells were incubated at 37°C with shaking at 200 rpm for 5 hours till the suspension culture became cloudy. 10 µl cell culture was mixed with 700 µl 2xYT medium and plated onto the plates (containing 50 µg/mL Kanamycin, 7 µg/mL Gentamycin, 10 µg/mL Tetracycline, 100 µg/mL Blue-gal and 40 µg/mL IPTG). The plates were inverted and incubated for 48 hours at 37°C until the blue and white colonies can be identified. Colony polymerase chain reaction (PCR) was performed with M13 forward primer (5'-CCCAGTCACGACGTTGTAAAACG-3') and reverse primer (5'-AGCGGATAACAATTTCACACAGG-3') to verify the white colonies are indeed the ones finished the transposition. Only the colonies with correct insertion were selected for culturing and bacmid extraction.

The verified colonies were picked and inoculated into 5 mL of 2xYT medium (containing 50 µg/mL Kanamycin, 7 µg/mL Gentamycin and 10 µg/mL Tetracycline) and cultured overnight at 37°C with shaking at 200 rpm. Cells were spun down and the pellets were re-suspended with 250 µL P1 Buffer (Biobasic plasmid miniprep kit) followed by 250 µL P2 Buffer. 350 µL P3 Buffer was slowly added into the tube with gentle mixing. The total mixture was subjected to centrifugation at 15,000 g for 10 minutes. The supernatant was gently transferred to an ice-cooled 1.5 mL Eppendorf tube containing 800 µL of isopropanol, mixed gently by inverting the tube several times and placed on ice for 20 minutes. The mixture was spun down at 15,000 g for 15 minutes at 4 °C, and the pellet was washed with 1 mL 70% ethanol by inverting the tube several times, then spun down at 15,000 g for 5 minutes. The wash step was repeated once. The pellet was air-dried in the hood.

Spodoptera frugiperda Sf9 insect cells were maintained in Sf 900 III medium (Gibco™) supplemented with 1% FBS, 50 U/mL penicillin and 50 µg/mL streptomycin at 27°C with shaking at 130 rpm until the population reached 2×10^6 cells/mL. The 6-well tissue culture plate was used for the transfection. Each well was dispensed by 1 mL cells and incubated at 27°C for 1 hour to allow cells to attach to the surface. Meanwhile, 10 µL of recombinant bacmid DNA was mixed with 100 µL of Grace insect medium (Gibco™) in a 1.5 mL Eppendorf tube. 6 µL of Cellfectin (Gibco™) was mixed with 100 µL of Grace's insect medium in 1.5 mL Eppendorf tube. Then two tubes were mixed and incubated for 30 minutes. After cell attachment, the supernatant medium was carefully removed. 2 mL of fresh Grace's insect medium

was then gently dispensed to wash the cell by shaking the plate slowly. The supernatant medium was carefully removed again before overlaying the cells with 0.2 mL of liposome embedded DNA. Another 1 mL of Grace's insect medium was added into the individual wells and the plate was incubated at 27°C for 4 h. The transfection mixture within Grace's medium was carefully removed and 2 mL Sf 900 III medium (containing 1% FBS, 50 U/mL penicillin and 50 µg/mL streptomycin) was dispensed into the wells. The plate was sealed and incubated at 27°C.

The remaining medium in the plate was carefully transferred into 2 mL Eppendorf tube and centrifuged at 500 g for 5 minutes at room temperature 72 hours post-transfection. The supernatant, which was P0 virus, was gently transferred into a new 1.5 mL Eppendorf tube and stored at 4°C protected from light. 3 mL Sf9 cells at 2×10^6 cells/mL cultured in Sf 900 III medium were added to 24-well deep cell culture plates (Thomas Scientific) with 30 µL P0 virus. Cells were grown with shaking at 450 rpm at 27°C for 48 hours. The supernatant was collected by centrifuging the cell culture at 1,000 g for 5 minutes and stored at 4°C. That is P1 virus stock. P2 virus was obtained by repeating the procedure depending on the volume of virus needed.

3.5.2 Small-scale purification of full-length cellulose synthase

50 mL Sf9 cells at 2×10^6 cells/mL were used for small-scale expression with 0.5 mL P2 virus for infection. The cells were harvested by centrifugation at 500 g 48 hours post-transfection, resuspended in 10 mL ice-cooled buffer-QA (25 mM Na-HEPES pH, 7.5, 50 mM NaCl) with protease inhibitor cocktail (Thermal Fisher). The cells

were disrupted by glass homogenizer. The membrane pellet was collected by centrifugation at 36, 000 rpm for 45 minutes (SW41, Beckman, USA). The membrane pellet was washed one time using the buffer-QA, and washed two more times by high salt buffer-QB (25 mM Na-HEPES, pH 7.5, 1 M NaCl). The washed membrane pellet was then resuspended in 10 mL buffer-solubilization (25 mM Na-HEPES, pH 7.5, 150 mM NaCl, 10% glycerol, 1% detergent) with protease inhibitor cocktail and rotated at cold room for 2 hours. Then the soluble fraction was collected by centrifugation at 36, 000 rpm for 45 minutes (SW41, Beckman, USA). The soluble fraction containing the detergent-solubilized membrane protein was subjected to standard small-scale nickel-affinity purification described in method section 3.2.1.

3.5.3 Large-scale purification of full-length cellulose synthase

200 mL insect cells at 2×10^6 cells/mL were used for expression with 2 mL P2 virus infected. The cells were harvested by centrifugation at 500 rpm (JLA-8.1, Beckman, USA) at 4°C 48 hours post-transfection, resuspended in 50 mL ice-cooled buffer-QA with protease inhibitor cocktail. The cells were disrupted by glass homogenizer. The membrane pellet was collected by spinning at 36, 000 rpm for 45 minutes (Ti45, Beckman, USA). The membrane pellet was washed twice in buffer-QA, and washed another two times in buffer-QB. The washed membrane pellet was then resuspended in 20 mL buffer-solubilization with protease inhibitor and rotated in cold room for 2 hours. Then the soluble fraction was collected by centrifugation at 36, 000 rpm for 45 minutes (SW41, Beckman, USA). The soluble fraction, which contained the detergent-solubilized membrane protein, was subjected to standard nickel-affinity purification described in method section 3.2.2.

3.6 Crystal screening, optimization, and structure determination

3.6.1 Crystal screening

The commercial kits (Crystal screening, Index, Salt Rx, PEG Ion, and PEG Rx) from Hampton research and the Intelli 96 wells sitting drop plates (Art Robbins Instruments) were used for crystal screening which was set up by mosquito (TTP Labtech) with varied protein and reservoir ratio at 100 nL: 50 nL, 100 nL: 100 nL and 50 nL: 100 nL, respectively. The plates were sealed by clear tape and stored in rock imager (Formulatrix®) for automatic imaging at 20°C.

3.6.2 Crystal optimization and data collection

The crystals of AtCesA3⁽³¹⁷⁻⁷⁹²⁾ were grown in 0.1 M magnesium formate dehydrate with 15% w/v Polyethylene glycol (PEG) 3, 350. After several rounds of optimization, the final crystal growth condition is 5% w/v PEG 3, 350, 100 mM 1,4-Piperazinediethanesulfonic acid (PIPES), pH 7.5, and 10 mM MnCl₂. The crystals were cryo-protected into mother liquor with 25% glycerol. The diffraction datasets were collected from Australian light source MAX2. The Se-Met substituted AtCesA3⁽³¹⁷⁻⁷⁹²⁾ crystals were grown in the same condition with WT and handled similarly. The Selenium anomalous signal dataset was collected in Paul Scherrer Institute (PSI, Switzerland) at 12.698 keV.

3.6.3 Structural determination

The Se-Met crystal dataset was processed by XDS with Friedel's law is false [97]. The dataset was indexed in P2₁ space group with unit cell parameter (54.80 Å, 146.76

Å, 91.37 Å, 90°, 102.33°, 90°) and the resolution was cut off at 2.5 Å. There were two molecules in one asymmetrical unit indicated by the Matthews coefficient in the CCP4 suite [98]. The Phenix Autosol program was used for automatic experimental phasing [99]. In a total of 20 Selenium atoms were found with high fidelity and the initial protein model of AtCesA3⁽³¹⁷⁻⁷⁹²⁾ was built based on the substructure of Selenium atoms after density modification. The model was further manually built by COOT and refined by Phenix refinement [100]. The diffraction datasets of AtCesA3⁽³¹⁷⁻⁸¹⁰⁾ crystals were collected in Taiwan synchrotron and Australia synchrotron. Because the crystals of AtCesA3⁽³¹⁷⁻⁸¹⁰⁾ were less than 100 µm and very thin, they could only tolerate roughly 90-degree rotation of data collection. Multiple datasets were scaled together to finally get 2.9 Å resolution. The space group was P2₁2₁2₁ with unit cell parameter (49.14 Å, 58.43 Å, 321.89 Å, 90.00°, 90.00°, 90.00°). There were two molecules in one asymmetrical unit indicated by the Matthews coefficient. The model of AtCesA3⁽³¹⁷⁻⁷⁹²⁾ was used as template for the structure determination of AtCesA3⁽³¹⁷⁻⁸¹⁰⁾. The structure of AtCesA3⁽³¹⁷⁻⁸¹⁰⁾ was manually built by Coot and refined by Phenix Refinement.

3.7 Cross-linking of full-length cellulose synthase

To prepare the Grafix gradient, two buffer solutions were prepared separately based on the buffer (25 mM Na-HEPES, pH7.5, 100mM NaCl, 0.05% FC-12): the top buffer contained 10% Sucrose, and the bottom buffer contained 30% Sucrose and 0.05% Glutaraldehyde (GA) [101]. Around 2.1 mL top solution was first added to a 4.4 mL centrifuge tube (Beckman, USA) and the bottom solution was carefully

injected by blunt-end syringe into the centrifuge tube bottom by pushing up the top solution. The continuous gradient was formed by executing the standard 10% - 30% sucrose gradient program in Gradient Master 108 (Biocomp) and cooled down in cold room for one hour before using. 100 μ L 1 mg/mL protein was loaded on the top of the gradient and subjected to ultracentrifugation with a speed of 50, 000 rpm (SW60, Beckman, USA) for 16 hours at 4°C. Following the ultracentrifugation, the peristaltic pump was used for fractionation at 4°C from the bottom to top and every four drops were collected as one fraction. All the fractions were collected and subjected to 7.5% SDS-PAGE analysis.

3.8 Negative staining

3.8.1 Negative staining grid preparation

The suitable fractions from Grafix were concentrated by Vivaspın® 500 100k MWCO concentrator (Sartorius) with buffer exchanged to 20 mM Na-HEPES, pH 7.5, 100 mM NaCl, 0.05% FC-12. Continuous carbon film-coated copper grids with 200 mesh (Electron Microscopy Sciences, CF200-CU) were glow-discharged for 35 seconds and 4 μ L of protein at 0.04 mg/mL was applied to the grids. After 1 minute incubation, the protein was blotted away by filter paper (Whatman). 4 μ L of 2% uranyl acetate was immediately added to the grid, incubated for 1 minute, and gently blotted away by filter paper (Whatman). The grids were air-dried for 10 minutes before loaded to FEI Tecnai T12 spirit equipped with a 4K Eagle CCD camera. The images were collected at 49,000 nominal magnification with a pixel size of 2.11 Å/pixel. Micrographs were recorded at an electron dose around 20 e-/Å².

3.8.2 Negative staining data processing

Negative staining micrographs were imported and auto CTF corrected in EMAN2 [102]. The particles were picked using EMAN2 particle picker in SWARM mode. Particle coordinate files were saved for each micrograph as box files. For the initial 2D classification, all the particles were used as input in 50 classes with a default particle mask applied.

3.9 Cryo-EM grid preparation and data processing

3.9.1 Cryo-EM grid preparation

The suitable fractions from Grafix were collected and the buffer was exchanged to final buffer (20 mM Na-HEPES, pH 7.5, 100 mM NaCl, 0.05% Fc-12) by VivoSpin 500 100 kDa MWCO concentrator two times to remove the sucrose. The final concentration of AtCesA1 used for Cryo-EM was 0.2 mg/mL. The Quantifoil R2/1 2nm Carbon 300 mesh Copper grids were glow-discharged for 22s before mounted to Vitrobot tweezers. The blotting was performed under 4°C with over 90% humidity. The blotting protocol is 1s blot time, 0 blot force, 30s waiting time, 1 blot total, and 0s drain time. 4 µL of protein sample was applied to the cryo grids and plunged into liquid ethane container and carefully transferred to a grid box in liquid nitrogen. The cryo grids were clipped before loading to FEI Arctica electron microscopy. Digital micrographs in movie mode were recorded at 200 kV 78, 000 magnification with defocus range between 0 and -0.7 µm under low-dose mode with voltage phase plate

(VPP) with falcon direct detector. The beam was aligned and eucentric height and astigmatism were adjusted before imaging.

3.9.2 Motion correction and CTF estimation

All the recorded micrographs were imported into Relion 3.0 and the data processing was following the Relion 3.0 manual [103]. All micrographs were subjected to the motion correction by MotionCor2. The motion-corrected micrographs were auto fitted to CTF estimation with CTFFIND-4.1. Because the Cryo-EM data was collected with the voltage phase plate, the initial phase shift was applied between 0 and 180 degrees with a step size of 10 degrees. After the CTF estimation, all the CTF fitted micrographs were manually checked and micrographs with bad CTF estimation were removed from further processing.

3.9.3 Particle picking and 2D classification

2 CTF estimated micrographs were select for manual particle picking. The particle size was set to 200 Å. In a total of around 1000 particles were manually picked and 2D class averaged in 20 classes. Only the classes with distinct features comparing to the background were selected as good classes. The good classes were then used as the 2-D reference for auto-picking against all good CTF estimated micrographs. The picking threshold was set to 0.9 and the minimum inter-particle distance was set as 180 Å. In a total of 186, 897 particles were auto-picked from 604 micrographs with an average of 309 particles per micrograph. All the particles were subsequently extracted and 2D class averaged into 100 classes. Only the classes with distinct features comparing to the background were selected as good classes.

3.9.4 3D Initial model building, 3D classification, and refinement

The particles belonging to good 2D class averages were used as input, CTF corrected and only one class was used with mask diameter as 280 Å for initial model. The default parameters for 3D classification and refinement were used. The resolution was estimated by the golden standard FSC curve.

4. Results

Obtaining protein crystals with decent diffraction capacities under X-ray illumination is always the bottleneck in most of the protein crystallography work, which results in applying multiple strategies to engineer the protein target, such as fusion solubility tags to improve the protein target solubility and expression level, removing the flexible region by truncating the terminus or internal unstructured regions, reducing the surface entropy by lysine methylation and surface-entropy reduction mutagenesis [104-106]. For successful structure determination, the protein target must undergo the construct design, protein expression and purification, crystal screening and optimization, X-ray diffraction data collection, and structure determination [107]. The protein may fail in any step and the project may need to start from the beginning to re-design the protein construct. To make the experimental steps clearer and the thesis easy to read, one flow-chart describing the experimental approach to structure determination was provided with the individual steps described in detail (Figure 21).

Figure 21. Flow-chart of experimental approaches

The experimental steps were described in the hierarchy from the construct design (top) to structure determination (below). The failed steps are labeled in red and the structures were labeled in green. The flow-chart was generated online (<https://online.visual-paradigm.com/>)

4.1 Protein preparation

4.1.1 Soluble constructs screening on AtCesA3^{CatD}

The cDNA of AtCesA3^{CatD} covering the amino acids 305 - 843 was codon-optimized and synthesized by GenScript (USA). Various N terminal and C terminal truncations of AtCesA3^{CatD} were designed and expressed with the hexahistidine tag. From the small-scale expression test with the help of the Protein Production Platform (PPP, NTU), only two constructs were soluble (Figure 22). Interestingly, the constructs need to be specifically around T13 (AtCesA3^{T317}) and S488 (AtCesA3^{S792}) in order to be soluble.

Figure 22. Small-scale test of AtCesA3^{CatD} constructs

Red bar indicated the construct was either no expression or insoluble. The bar in green indicated the construct was soluble.

4.1.2 Purification of AtCesA3^{CatD}VC11

The soluble construct of AtCesA3^{CatD} (AtCesA3^{CatD}VC11) was purified by one-step nickel-affinity purification (Figure 23). The theoretical protein size was 53.3 kDa and the protein ran between 50 kDa and 60 kDa molecular weight protein markers judging from the SDS-PAGE. The protein was very pure from SDS-PAGE without any obvious contaminant protein, though the protein elution peak was not symmetrical. The nickel elution fractions were collected. However, the protein precipitated heavily during concentration and it could only be concentrated to ~2 mg/mL. Because of the precipitation issue, the protein was not loaded to the size-exclusion column for molecular weight and homogeneity characterization. The protein continued to precipitate even after removing the precipitant, which suggested the protein was unstable and not suitable for subsequent crystal screening.

Figure 23. Nickel-affinity purification of AtCesA3^{CatD}VC11

The left panel showed the nickel gradient elution profile of AtCesA3^{CatD}VC11. The fractions collected and load for SDS-PAGE were labeled by black line. The right panel showed the SDS-PAGE result of corresponding nickel-affinity purification fractions.

4.1.3 Screening of cellulose synthase isoforms

The cellulose synthases could have similar three-dimensional structures across isoforms and organisms considering the high sequence similarity (> 60%). The *Arabidopsis thaliana* (AtCesA1, AtCesA3, AtCesA6, AtCesA7, AtCesA8, and

AtCesA10), *Hordeum vulgare* (HvCesA1-HvCesA6 and HvCesA8), and *Oryza sativa* (OsCesA4 and OsCesA8) cellulose synthases were cloned for protein expression and purification with the same design as AtCesA3^{CatD}VC11 which was cloned into pET28-a (+) expression vector with an N terminal hexahistidine tag. Among all the cellulose synthases, all of them were soluble after small-scale expression tests except for AtCesA7^{CatD} and OsCesA4^{CatD}. All the soluble CesA^{CatD} isoforms were purified by nickel-affinity purification. AtCesA6^{CatD} dominantly aggregated during purification. The HvCesA2^{CatD}, HvCesA4^{CatD}, HvCesA5^{CatD}, and OsCesA8^{CatD} were excluded for further purification and characterization because of low yield. The rest of CesA^{CatD} was purified and loaded to SDS-PAGE (Figure 24). The result demonstrated that these proteins were generally pure after nickel-affinity purification. Notably, HvCesA8^{CatD} had the highest yield, and it could also be concentrated to around 8 mg/mL for crystal screening indicating its stability during purification.

Figure 24. Purified CesA^{CatD} homologs.

Purified catalytic domain of cellulose synthase from *Arabidopsis thaliana* and *Hordeum vulgare*. The protein markers were labeled. The input amount for each lane varied.

4.1.4 Design of C-SR loop deletions and small-scale expression test

Multiple strategies were employed to crystallize the catalytic domain of cellulose synthase including terminal truncations, homolog screening, and surface entropy reduction (data not shown). However, none of them gave any crystal hits. HvCesA8^{CatD} was observed to form phase separation after setting up for crystal screening, which was a good sign indicating the protein was close to being crystallized. One structurally disordered region composed of 47 amino acids from K318 to M364 was found in the C-SR region based on secondary structure prediction of HvCesA8^{CatD} by Psipred (Figure 25. A) [108]. Given the fact that the crystallization process is packing the ordered molecules but the large disordered region increases the molecule's flexibility thus decreases the propensity to crystallization, deleting the predicted disordered region could help crystallization [109]. Therefore, HvCesA8^{CatD} was selected as the starting construct because it had the highest soluble expression level. Two C-SR loop deletion versions were initially designed with the varied glycine-serine linkers to compensate for the unknown distance between the two ends generated by the disordered region deletion. The amino acids from HvCesA8^{P639} to HvCesA8^{P670} were deleted and named as C-SR version 1 (C-SR-1). For C-SR-1 constructs, LD0 means complete deletion, LD8 means the region was substituted with 8 amino acids long linker ((GGGGS)G₃) and likewise, LD16 ((GGGGS)₃G), and LD18 (GG(GGGGS)₃G). The amino acids from HvCesA8^{C627} to HvCesA8^{R662} were also deleted and named as C-SR version 2 (C-SR-

2). For C-SR-2 constructs, LD4 (GGGS), LD6 (GGSGG), and LD10 (GGGS)₂ were designed. All the loop deletion constructs were tested by small-scale expression. It turned out all four C-SR-1 constructs were highly soluble, which indicated the deleted region was indeed disordered and the deletion didn't disrupt the protein structure (Figure 25. B). However, all C-SR-2 constructs were insoluble (Figure 25. C). Therefore, the C-SR-1 LD0 construct was selected for large-scale purification, as direct deletion was fine and a flexible linker was not necessary.

Figure 25. Disordered region prediction and small-scale expression test of C-SR loop deletions

A. Secondary structure prediction of HvCesA8⁽³²⁶⁻⁸⁴⁸⁾ by PSIPRED. The α helices were highlighted in purple and the β sheets were highlighted in yellow. The predicted disordered region was highlighted by red squares. B. Small-scale expression test of HvCesA8 C-SR-1 loop deletion constructs. C. Small-scale expression test of HvCesA8 C-SR-2 loop deletion constructs.

4.1.5 Purification of Cesa^{CatD} loop deleted constructs

Likewise, all the previous mentioned constructs with soluble expression were loop deleted, among which the soluble constructs were used for large-scale purifications. The purification of AtCesA3^{CatD}LD0 was shown in below (Figure 26). The purification of AtCesA1^{CatD}LD0, AtCesA6^{CatD}LD0, AtCesA8^{CatD}LD0, HvCesA1^{CatD}LD0, HvCesA3^{CatD}LD0, HvCesA6^{CatD}LD0, HvCesA8^{CatD}LD0, OsCesA8^{CatD}LD0 were not shown because the protein preparations were roughly the same. The AtCesA3^{CatD}LD0 was quite pure after the first step nickel-affinity purification. The hexahistidine tag was efficiently removed by the TEV protease and the tag-free protein was collected in the flow-through fractions by a 2nd round nickel-affinity purification. Subsequently, the protein sample was applied to a size-exclusion column. The target protein was eluted at 75.97 mL from HiLoad Superdex 200 16/60 column, which indicated the protein was monomeric in solution. The protein fractions were rather pure after the size-exclusion chromatography judging from the SDS-PAGE result. The fractions were collected, concentrated to 10 mg/mL, and snap-frozen in liquid nitrogen for further use.

Figure 26. Purification of AtCesA3^{CatD}LD0

The size-exclusion chromatography was shown on the left with retention volume labeled. the SDS-PAGE on the right showed the overall purification. S: supernatant after cell disruption

and centrifugation. FT: the nickel-affinity purification flow through. E: the elution of nickel-affinity purification. AC: after TEV protease cleavage. FT: flow-through of samples collected from 2nd time nickel-affinity purification. 20: wash buffer containing 20 mM Imidazole. Size: fractions corresponding to the left chromatography.

4.2 Crystal screening, optimization, and structure determination

4.2.1 Crystal screening of purified Cesa^{CatD}

All the purified C-SR loop deletion proteins were subjected to crystal screening at a concentration of ~10 mg/mL. Serendipitously, two proteins were successfully crystallized which were OsCesA8^{CatD}LD0 and AtCesA3^{CatD}LD0. There were a lot of crystallization conditions from Index screening kit that yielded crystals for OsCesA8^{CatD}LD0. However, only one crystallization condition that was Index H8 (0.1 M magnesium formate dehydrate, 15% w/v Polyethylene glycol 3, 350) yielded crystals for AtCesA3^{CatD}LD0 (Figure 27). The crystals were verified to be protein crystals by UV illumination in Rock Imager, though the crystal size was very small.

Figure 27. The initial crystal screening hits

A. Crystals of OsCesA8^{CatD}LD0 grown in Index H10. The left image showed the visible light imaging of the drop and the right image showed the UV illumination of the crystals. B.

Crystals of AtCesA3^{CatD}LD0 grown in Index H8. The left image showed the visible light imaging of the drop and the right figure showed the UV illumination of the crystals.

4.2.2 Crystal optimization of OsCesA8^{CatD}LD0

The OsCesA8^{CatD}LD0 had a lot of crystallization conditions using PEG 3, 350 as the precipitant. However, most crystals were stacked together to form round shaped cluster after several rounds of optimization, which was far from satisfactory (Figure 28). Concurrently, the optimization of AtCesA3^{CatD}LD0 was also ongoing and single decent crystals can be obtained. It was quite surprising that OsCesA8^{CatD}LD0 can be crystallized many conditions, but it was hard to optimize the condition to grow decent crystals. However, AtCesA3^{CatD}LD0 could only be crystallized in one condition that could be easily optimized. Thus, the optimization of OsCesA8^{CatD}LD0 suspended.

Figure 28. The optimized crystals of OsCesA8^{CatD}LD0

One optimization drop was shown here and the scale bar was 100 μm .

4.2.3 Crystal optimization of AtCesA3^{CatD}LD0

The crystal optimization was performed by varying the PEG concentration from 3% to 15% and the pH range from 6.0 to 8.2. The best crystal growth condition was 100 mM PIPES, pH 7.5, 5% PEG 3, 350, 10 mM MnCl₂ (Figure 29. A). However, the protein precipitation still occurred in the drop. Considering the protein precipitant impeded the crystal harvesting, the additive screening was performed to get rid of the

protein precipitant. The additive number 67 (NDSB-256) successfully helped to grow crystals without heavy protein precipitation (Figure 29 B). The ligands (UDP, UDP-Glc, cellobiose, and the mixture) at a concentration of 5 mM were incubated with the protein before the optimization setup to get the ligand-bound crystal structures. The crystals with the ligands looked no much difference with the apo ones (Figure 29 C-E).

Figure 29. Crystal optimization of AtCesA3^{CatD}LD0

A. The Apo crystals grew in 5% PEG 3, 350, 100 mM PIPES, pH 7.5, 10 mM MnCl₂. B. The Apo crystals grew in 5% PEG 3, 350, 100 mM PIPES, pH 7.5, 10 mM MnCl₂ with 3% NDSB-256. C. The crystals grew in 5% PEG 3, 350, 100 mM PIPES, pH 7.5, 10 mM MnCl₂, 3% NDSB-256, 5 mM UDP-Glc. D. The crystals grew in 5% PEG 3, 350, 100 mM PIPES, pH 7.5, 10 mM MnCl₂, 3% NDSB-256, 5 mM cellobiose. E. The crystals grew in 5% PEG 3, 350, 100 mM PIPES, pH 7.5, 10 mM MnCl₂, 3% NDSB-256, 5 mM UDP-Glc and cellobiose.

The crystals with or without the ligands were harvested by adding 25% glycerol into the mother liquor. The diffraction datasets were collected in Australian synchrotron and processed with XDS package [97]. The molecular replacement using bacterial cellulose synthase BcsA catalytic domain and the computational models predicted by Robetta (University of Washington, USA) both failed to solve the AtCesA3^{CatD}LD0 structure. Therefore, the Selenium-Methionine (Se-Met) substitution method was used to experimentally solve the protein structure.

4.2.4 Se-Met protein purification and crystal optimization

The protein purification procedure was the same with the WT protein purification. The Se-Met protein was pure after the size-exclusion chromatography (Figure 30). The protein fractions were collected and concentrated to ~10 mg/mL, and snap-frozen in liquid nitrogen for later use.

Figure 30. Purification of Se-Met substituted AtCesA3^{CatD}LD0

Left. The size-exclusion chromatography of Se-Met substituted AtCesA3^{CatD}LD0. Right. The SDS-PAGE of the corresponding fractions. E: nickel affinity purification elution, AC: after

TEV protease cleavage, FT: the 2nd nickel column follow through, and 20: 20 mM imidazole wash.

The Se-Met substituted crystals of AtCesA3^{CatD}LD0 grew under the same crystallization condition as WT protein. The crystal shape was also similar to WT protein crystals (Figure 31). The presence of selenium was confirmed by Se K edge scanning.

Figure 31. Optimized crystal of Se-Met substituted AtCesA3^{CatD}LD0

The single crystal of Se-Met substituted AtCesA3^{CatD}LD0 was shown with the scale bar as 100 μm .

4.2.5 Purification of AtCesA3³¹⁷⁻⁸¹⁰

After the structure determination of AtCesA3^{CatD}LD0, it revealed that the loop deletion could be further optimized to delete the residues AtCesA3⁶³⁰⁻⁶⁹⁸ that were not seen from the electron density map indicating the flexibility. Therefore, the AtCesA3⁶³⁰⁻⁶⁹⁸ was substituted by a GGGGS linker, which resulted in the construct named AtCesA3³¹⁷⁻⁷⁹². Still, the construct of AtCesA3³¹⁷⁻⁷⁹² didn't include the QxxRW motif for GT family in its C terminus which could be important for cellulose binding. One longer construct AtCesA3CatD³¹⁷⁻⁸¹⁰ that had extended C terminus thus included the QxxRW motif was generated with C-terminal hexahistidine tag. The

purification of AtCesA3³¹⁷⁻⁸¹⁰ was shown in Figure 32. The nickel-affinity purification fractions were pooled together, desalted to 50 mM NaCl and subjected to anion ion exchange. The first peak of gradient elution from the HiTrap Q HP column contained most of the target protein. The second peak contained relatively more contaminant protein. So only the first peak fractions were collected, concentrated, and loaded to size-exclusion column. The protein eluted at 70.75 mL with a symmetrical peak, indicating the protein could be dimer in solution. The SDS-PAGE of the fractions showed the protein was quite pure after the size-exclusion purification. The size-exclusion fractions with less contaminant protein were concentrated to 10 mg/mL and snap-frozen by liquid nitrogen for future use.

Figure 32. Purification of AtCesA3³¹⁷⁻⁸¹⁰

(A) and (B). Ion exchange chromatography and the corresponding SDS-PAGE of AtCesA3³¹⁷⁻⁸¹⁰. The fractions were indicated by black line. (C) and (D). The size-exclusion chromatography and the corresponding SDS-PAGE. The fractions were indicated by black line.

4.2.6 Crystal screening and optimization of AtCesA3³¹⁷⁻⁸¹⁰

The crystals of AtCesA3³¹⁷⁻⁸¹⁰ appeared in many conditions using PEG 3, 350 as precipitant. Multiple strategies were used to optimize the crystals by varying the molecular weight and concentration of PEG, as well as the pH range. Tiny crystals appeared almost immediately after the optimization setup at both room temperature and 4°C. Some other optimization kits from Hampton research, like additive screening, silver bullet, and detergent screening were further utilized for crystal optimization, but none of them improved the crystal size much. Separating the protein drop with reservoir drop and linking the two drops using a cat beard also failed to slow down the crystal growth. The final optimized crystals used for data collection were grown in 100 mM MES, pH 6.5, 0.2 M potassium sodium tartrate, 5% PEG 3, 350. The crystals were in good shape with around 50 µm in length, 40 µm in width, but only less than 10 µm thick (Figure 33. A). The small crystal size made it extremely difficult to harvest the crystals. Besides, an unusually long exposure time (30 seconds) for synchrotron light source was applied for data collection due to the weak diffraction spots (Figure 33. B). The crystals could only tolerate less than 90-degree rotation data collection. Therefore, multiple crystal diffraction datasets were indexed into the same space group with P2₁2₁2₁, scaled, and merged to finally reach the resolution at 2.9 Å.

Figure 33. The crystals and diffraction image of AtCesA3³¹⁷⁻⁸¹⁰

A. Typical crystals image of AtCesA3³¹⁷⁻⁸¹⁰. Scale bar 100 μm. B. The diffraction image of AtCesA3³¹⁷⁻⁸¹⁰ with the resolution ring shown.

4.2.7 Molecular weight determination by SEC

Based on size-exclusion chromatography profiles of AtCesA3³¹⁷⁻⁷⁹² and AtCesA3³¹⁷⁻⁸¹⁰ from large-scale purification, they were eluted at around 76 mL and 70 mL respectively, which indicated these two constructs had a discrepancy regarding the oligomeric state. The molecular weight determinations for both proteins as well as standard proteins were performed by size-exclusion chromatography (Figure 34). From the results, the AtCesA3³¹⁷⁻⁸¹⁰ was eluted at 12.83 mL and AtCesA3³¹⁷⁻⁷⁹² was eluted at 12.66 mL and 13.57 mL. According to the standard curve, it was quite obvious that AtCesA3³¹⁷⁻⁷⁹² formed both monomer and dimer in solution, with the monomer as the majority. In contrast, AtCesA3³¹⁷⁻⁸¹⁰ was only a dimer in solution. Inclusion of the C terminal QxxRW motif could help to stabilize the protein in a dimer state, suggesting the importance of the C terminal QxxRW motif in protein function.

Figure 34. Calibration of AtCesA3³¹⁷⁻⁷⁹² and AtCesA3³¹⁷⁻⁸¹⁰

Three standard proteins (Aldolase 158 kDa, Conalbumin 75 kDa, and Ovalbumin 44kDa) were used in the calibration with elution volume indicated. The profile of AtCesA3³¹⁷⁻⁷⁹² was colored in red and the profile of AtCesA3³¹⁷⁻⁸¹⁰ was colored in green.

4.3 Structure of AtCesA3^{CatD}

Two constructs with the different C terminus of AtCesA3CatD have been solved by X-ray crystallography. One is the AtCesA3³¹⁷⁻⁷⁹² and the other one is the AtCesA3³¹⁷⁻⁸¹⁰ which has the canonical QxxRW motif for glycosyltransferase. UDP and UDP-Glc bound structures of AtCesA3³¹⁷⁻⁷⁹² in the presence of Mn²⁺ have also obtained by co-crystallization. In the below section, the overall structure of AtCesA3³¹⁷⁻⁸¹⁰ and the ligand-bound structure of AtCesA3³¹⁷⁻⁷⁹² will be discussed. The data collection and refinement statistics of all the structures of AtCesA3 have been summarized in Appendix 3.

4.3.1 Structure of AtCesA3³¹⁷⁻⁸¹⁰

The crystal structure of AtCesA3³¹⁷⁻⁸¹⁰ was solved in P2₁2₁2₁ space group at 2.95 Å with two molecules in one asymmetrical unit (Figure 35. A). The protein structure

was built with residues ranging from 334-575, 593-600, 610-630, 704-785, 796-810. Despite a few flexible regions missing due to poor electron density, the C terminal QxxRW motif was built in the crystal structure that covered the entire GT domain and provided the insights about the catalytic mechanism of cellulose synthesis.

The structure of AtCesA3³¹⁷⁻⁸¹⁰ can be divided into three parts that are central GT domain, P-CR domain, and C-SR domain (Figure 35. A). The GT domain occupying the inner layer is mainly composed of 11 β strands surrounded by three α helices. The N terminal part of the core GT domain including the α 1, α 5, α 6, and β 1 to β 5 aligns quite good with the typical GT-2 glycosyltransferase SpsA from *Streptococcus pneumoniae*, indicating the similar nucleotide-binding mechanism (Figure 35.D). The remaining C terminal part of the core GT domain includes the β 6 to β 12 and the α 9. Interestingly, the remaining C terminal part of the core GT domain is composed of two layers of β sheets, with β 6, β 7, β 9, β 12 in one layer, and β 1- β 4, β 8, β 10, and β 11 in the other layer. The C-terminal part of the core GT domain drastically differs from SpsA, implying the uniqueness of plant cellulose synthase (14). The QxxRW motif which is the signature of GT domain and predicted to be the cellulose-binding helix is assigned as α 9 in AtCesA3³¹⁷⁻⁸¹⁰ and inserted into the catalytic pocket. Notably, there is one disulfide bond formed by C618 and C782, though the catalytic domain resides in the cytosol.

Besides the core GT domain, the two unique domains of plant cellulose synthase (P-CR and C-SR) can be unambiguously resolved. The P-CR and C-SR regions are both in the outside layer surrounding the core GT domain (Figure 35. A). The P-CR

domain consists of $\alpha 2$, $\alpha 3$, $\alpha 4$, and $\beta 3$ (Figure 35. B). The previously reported dimer structure of the rice Cesa8 P-CR is definitely crystallography-artifact as the interaction partner occupies the core GT domain when the P-CR dimer structure is aligned to AtCesA3 structure [96] (Figure 35. E). The C-SR domain consists of about 70 amino acids long (with dominating positively charged amino acids, like lysine and arginine) intrinsically disordered region by secondary structure prediction and mathematical modeling [110]. The resolved part of C-SR⁷⁰⁴⁻⁷⁵⁰ is composed of two helices named $\alpha 7$ and $\alpha 8$ that reside outside of the core GT domain (Figure 35.C). Considering that C-SR contains a large cluster of positively charged amino acids, it is speculated that C-SR lies under the membrane lipid to anchor the GT domain to the membrane [111].

Figure 35. Crystal structure of AtCesA3³¹⁷⁻⁸¹⁰

(A) Structure of AtCesA3³¹⁷⁻⁸¹⁰ was shown in cartoon. The secondary structure elements were labeled from N terminus to C terminus. The protein was colored in lime, the P-CR was colored in wheat and the C-SR was colored in yellow. The un-modeled regions were linked by dashed lines (B) An enlarged view of the P-CR region colored in wheat. (C) An enlarged view of the C-SR region colored in yellow and the distance between the loop deletions termini was showed. (D) The alignment of SpsA colored in grey (PDB code: 1H7L) to AtCesA3³¹⁷⁻⁸¹⁰. (E) The alignment of Rice dimer OsCesA8 P-CR (PDB code: 5JNP) to the structure of AtCesA3³¹⁷⁻⁸¹⁰. The OsCesA8 P-CR structure was colored in yellow and salmon. The structure of AtCesA3³¹⁷⁻⁸¹⁰ was colored in lime.

4.3.2 Dimer interface of AtCesA3³¹⁷⁻⁸¹⁰

Oligomerization of cellulose synthase is very important as the individual cellulose synthase interacts with each other to form the six lobes CSC. The purified AtCesA3³¹⁷⁻⁸¹⁰ is a dimer in solution according to the calibration result as mentioned before. The configuration of AtCesA3³¹⁷⁻⁸¹⁰ dimer can be revealed by the crystal structure, as each protomer is symmetrically two-fold rotated, with a total of 711.3 Å² surface area buried by PISA calculation (Figure 36. A) [112]. Both β6 and α9 are involved in dimer formation. AtCesA3^{Q571/V572/C573/Y574} from β6 of two protomers form four pairs of hydrogen bonds by backbone interaction to generate an anti-parallel β sheet (Figure 36.B). Moreover, the AtCesA3^{R800} from α9 also forms a hydrogen bond with the other molecule's AtCesA3^{Q803} (Figure 36.C). Due to the lack of a clear electron density map for the side chains of the amino acids in α9, there might be some other potential hydrogen bond pairs missing in our structure. Interestingly, AtCesA3³¹⁷⁻⁷⁹² that truncated the C terminal extended QxxRW motif from AtCesA3^{A793} to AtCesA3^{S811} is a dominant monomer, which suggests a crucial role of the helix α9 in dimer formation.

To validate the importance of AtCesA3 $\beta 6$ in dimer formation, the AtCesA3^{Q571} and AtCesA3^{C573} that are involved in hydrogen bond pairs were mutated to proline to disrupt the dimer interface since proline lacks backbone amide proton [113]. These two protein mutants were purified on a large scale following the similar protocol for WT protein purification. The protein samples of AtCesA3³¹⁷⁻⁸¹⁰, AtCesA3³¹⁷⁻⁸¹⁰Q571P, AtCesA3³¹⁷⁻⁸¹⁰C573P, and AtCesA3³¹⁷⁻⁷⁹² were loaded to calibration size-exclusion column to characterize the oligomeric state (Figure 36. D). Surprisingly, the effects of the two mutations Q571P and C573P were quite different. Q571P mutation did not affect the oligomeric state of AtCesA3³¹⁷⁻⁸¹⁰ at all since it had almost the same retention volume with WT protein. However, the C573P mutation can completely abolish the dimer formation of AtCesA3³¹⁷⁻⁸¹⁰ because the C573P mutant protein had only one retention volume at 13.52 mL. It is concluded that the C573P mutation can abolish the dimer formation compared to the profiles of AtCesA3³¹⁷⁻⁸¹⁰ and AtCesA3³¹⁷⁻⁷⁹². Taken together, the $\beta 6$ is indeed important for the dimerization of plant cellulose synthase.

Figure 36. Dimer interface of AtCesA3³¹⁷⁻⁸¹⁰

(A). The dimer structure of AtCesA3³¹⁷⁻⁸¹⁰ with one molecule colored in lime and the other one colored in cyan. The P-CR and C-SR domains are colored in wheat and yellow. (B) and (C). The $\beta 6$ and $\alpha 9$ interface with hydrogen bond pairs are shown in sticks. (D) Calibration of AtCesA3 variants. AtCesA3³¹⁷⁻⁸¹⁰ is shown in cyan, AtCesA3³¹⁷⁻⁸¹⁰Q571P is shown in red, AtCesA3³¹⁷⁻⁸¹⁰C573P is shown in green, and AtCesA3³¹⁷⁻⁷⁹² is shown in blue.

4.3.3 UDP-Glc and Mn²⁺ binding of AtCesA3³¹⁷⁻⁷⁹²

The UDP and UDP-Glc bound AtCesA3³¹⁷⁻⁷⁹² structures were finally obtained by collecting more than twenty datasets while only two datasets having the respective UDP and UDP-Glc. The UDP-Glc bound structure of AtCesA3³¹⁷⁻⁸¹⁰ was difficult to get because the crystals of AtCesA3³¹⁷⁻⁸¹⁰ were extremely tiny thus difficult to handle. Therefore, the UDP-Glc bound AtCesA3⁽³¹⁷⁻⁷⁹²⁾ was superimposed to the crystal structure of AtCesA3³¹⁷⁻⁸¹⁰ to get the UDP-Glc bound AtCesA3³¹⁷⁻⁸¹⁰ (Figure 37. A). Although the crystals were co-crystallized with the UDP, cellobiose, and Mn²⁺, both of the UDP and UDP-Glc bound structures don't have any electron density of cellobiose, which suggests the lack of coordination of cellobiose. Besides, the glucose

group from UDP-Glc also has no solid electron density, so the UDP-Glc bound AtCesA3³¹⁷⁻⁷⁹² structure is virtually the same as UDP bound AtCesA3⁽³¹⁷⁻⁷⁹²⁾ (Figure 37. B). The UDP-Glc lies in the groove formed by β 1, the loop connecting β 1 and α 1, and α 5 (Figure 37. A). The catalytic residue D765 resides in the loop that is flexible compared to the AtCesA3³¹⁷⁻⁷⁹² and AtCesA3³¹⁷⁻⁸¹⁰ structures. It is postulated that the cellulose-binding region falls into the area between the α 9 and the catalytic loop, which is also just facing the glucose group of UDP-Glc. The Mn²⁺ is coordinated by the oxygen atom in the carboxylic acid group of D545, the C2 hydroxyl group from the ribose group in UDP-Glc, the amino group from K521 and three water molecules of which two are well resolved in the crystal structure (Figure 37. C). The presence of Mn²⁺ is confirmed by the fluorescence scan in Australian synchrotron. For the UDP-Glc binding, D379 forms a hydrogen bond with uracil O4 of UDP-Glc. S343 and E350 form hydrogen bonds with the C3 and C2 hydroxyl group of UDP-Glc, respectively. K349 and K520 form ionic bonds with α and β phosphate of UDP-Glc (Figure 37. D). There is no obvious conformational change observed between apo and UDP bound structure of AtCesA3³¹⁷⁻⁷⁹² except the K349, K520, and K521 are flexible in the absence of UDP-Glc (Figure 37. D).

Figure 37. UDP-Glc bound structure of AtCesA3³¹⁷⁻⁷⁹²

(A) The UDP-Glc bound structure of AtCesA3³¹⁷⁻⁷⁹² in lime is aligned to AtCesA3³¹⁷⁻⁸¹⁰ in cyan. The UDP-Glc is shown in sticks and colored in salmon. The catalytic loop with residue D765 residues is outlined by the blue dashed eclipse. The putative cellulose-binding region is labeled in grey eclipse. (B) The unbiased Fo-Fc electron density map contoured at 2.2 σ of Mn^{2+} and UDP-Glc is shown. (C) The binding of Mn^{2+} . The D545, K521, and UPP-Glc are shown in sticks. The interaction distances are labeled. (D) The binding of UDP-Glc to AtCesA3³¹⁷⁻⁷⁹². The apo structure of AtCesA3³¹⁷⁻⁷⁹² is colored in grey and the UDP-Glc bound structure of AtCesA3³¹⁷⁻⁷⁹² is colored in lime. The residues are shown in sticks.

4.3.4 Structural comparison with BcsA GT domain

The catalytic mechanism of bacterial cellulose synthase has been well characterized by the structural study of BcsA-BcsB complex in various states [17-19]. In particular, the BcsA-BcsB co-purified with the short cellulose chain catalyzes the UDP-Glc

hydrolysis to transfer the glucose to the non-reducing end of the elongating cellulose chain and translocate the newly elongated chain and release the UDP [17-19]. The bacterial BcsA¹²⁷⁻⁴⁰¹ was aligned to the UDP-Glc bound structure of AtCesA3³¹⁷⁻⁸¹⁰ with a root mean square deviation (RMSD) about 2.7 by DALI server (Figure 38. A) [114]. As the AtCesA3 also has the canonical D, D, D, QxxRW motif, the AtCesA3^{D379} and AtCesA3^{D545} can be aligned quite good with BcsA^{D179} and BcsA^{D246}, respectively (Figure 38. B). Similar to BcsA, the function of AtCesA3^{D379} is to bind the uridine group of UDP-Glc and the AtCesA3^{D545} coordinates the Mn²⁺. In Arabidopsis, the AtCesA3^{S343} and AtCesA3^{E350} bind to the ribose group in UDP-Glc, while in bacterium, BcsA^{Q229} and BcsA^{E151} bind to the ribose group in UDP-Glc. Besides, the AtCesA3^{K349} and AtCesAK⁵²⁰ forms ion bonds with the phosphate group in UDP-Glc, while in bacterium BcsA^{R382} in the QxxRW motif helps coordinate the phosphate group (Figure 38. B). Two major structural differences are observed for the third D and QxxRW motif between BcsA and AtCesA3. First, the BcsA^{D343} which resides in the “finger helix” from the “TED” motif serves the general base to attack the C4 hydroxyl group of acceptor cellulose whereas its counterpart in AtCesA3^{D765} lies in a loop connecting β 10 and β 11 with different positioning, suggesting its drastic flexibility [19]. Second, the BcsA^{QxxRW} helps stabilize the phosphate group of UDP-Glc by BcsA^{R382} and the cellulose by BcsA^{Q379}, BcsA^{R380}, and BcsA^{W383}. In contrast, the AtCesA3^{QxxRW} which is located in the α 9 helix is inserted toward the catalytic pocket (Figure 38. B). Taken together, it is likely that the overall structural arrangement for the catalysis and cellulose translocation of plant cellulose synthase is different from bacterium cellulose synthase. Unlike bacterial cellulose synthase BcsA

that is activated by its C terminal PilZ domain by binding to c-di-GMP, it remains unclear whether AtCesA3 needs activation to initiate the cellulose synthesis.

Figure 38. Comparison of AtCesA3^{CatD} to BcsA GT domain

(A) Structural comparison of AtCesA3³¹⁷⁻⁸¹⁰ to BcsA CatD (PDB code: 4HG6). The AtCesA3³¹⁷⁻⁸¹⁰ is colored in lime, the BcsA is colored in grey. The bound cellulose and UDP-Glc are shown in sticks. (B) Structural comparison of catalytic residues and QxxRW motif from AtCesA3³¹⁷⁻⁷⁹² and BcsA CatD, respectively. The AtCesA3³¹⁷⁻⁷⁹² is colored in lime with the bound UDP-Glc colored in salmon. The BcsA is colored in grey. The non-conserved residues are outlined by red dash eclipses.

4.3.5 Proposed trimer models of AtCesA3^{CatD}

The cellulose synthase isoforms share an overall 70% similarity in terms of the GT domain (Appendix 3). It is highly likely that the crystal structure of AtCesA3^{CatD} can serve as a structural model for the other isoforms. The homo-trimer models of AtCesA3^{CatD} were generated using online server GalaxyHomomer (Figure 39) [115]. Two rules were used to evaluate if the trimer model is physiologically relevant based on the crystal structure of AtCesA3CatD: 1) The C-SR which is largely positively charged and presumably beneath the membrane lipid shall not be the primary

interface among the three protomers; 2) the catalytic pocket shall be solvent accessible. Based on the two rules, model 2 and 4 are ruled out as their interface involves the C-SR region. The C-SR in model 1 is presumably located beneath the membrane lipid and the catalytic pocket is solvent accessible. The interface is mostly contributed by the core GT domain. The C-SR region in model 3 is located in the outer shell, which makes it still possible to interact with the membrane lipid. However, the catalytic pocket is deeply buried into the interface, makes it less efficient for substrate entering and exit. Similar to model 3, the catalytic pocket in model 5 is also buried inside the trimer model, making it unlikely to be physiologically possible. Taken together, it seems that the model 1 is the most likely model for the trimeric AtCesA. Model 1 has a length of 94 Å, which is consistent with the reported 10.6 nm diameter of each CSC lobe. From the negative stain model of CSC, it seems each lobe of CSC is quite loosely connected, which raises the question about how the individual lobes interact with each to resemble the CSC.

Figure 39. Possible trimers of AtCesA3^{CatD}

(A-E). Trimer models of AtCesA3. The P-CR is colored in wheat and the C-SR is colored in yellow. The rest region is colored by chain. The structures were shown in cartoon.

4.3.6 Mutation sites mapping on AtCesA3^{CatD}

Mutants of cellulose synthases greatly aided in understanding the functions of CesAs and CSC. The plant CesA was identified based on the mutants that had the dwarf phenotype, reduced growth, and decreased cellulose content [116]. The crystal structure of AtCesA3^{CatD} can help understand how these mutations affect the CesA and even CSC function. The mutants of CesAs are adapted from the review and summarized in Table 2 [1]. The mutants are classified into three groups. The first group leads to the premature termination of cellulose synthase, thus affecting the CSC assembly and cellulose synthesis. The second group occurs in the cytosolic GT

domain. The third group is single amino acid substitution while it does not occur in the GT domain. So only group 2 mutations will be discussed below.

Table 2. Mutation sites in CesAs with annotation

	Mutation	Original CesA	Disrupttion	Sovent accessible
Group 1	Y258*	CesA6		
	Q306*	CesA4		
	W465*	CesA7		
	Q700*	CesA6		
	S722*	CesA6		
	W757*	CesA6		
	W993*	CesA5		
Group 2	S377F	CesA3	Y	N
	R437K	CesA8	N	Y
	A522V	CesA3	N	Y
	A533V	CesA1	N	N
	P578L/S/T ¹	CesA3/CesA7	X	X
	D588N	CesA1	X	X
	E603N	CesA1	X	X
	G615S	CesA1	N	Y
	G617E	CesA3	Y	N
	S761L	CesA8	N	Y
	E764K	CesA1	N	Y
	D765N	CesA1/CesA8	N	Y
	H773Y	CesA1	N	Y
Group 3	S301F	CesA3		
	L/I872F ²	CesA8		
	A/F888V ³	CesA1		
	G916E	CesA3		
	T942I	CesA3		
	G998D	CesA3		
	R1045W	CesA6		
	P1056S	CesA3		

* means premature termination of translation, Y means yes, N means No, and X means not applicable. 1: Multiple mutation sites were reported. 2: L is the WT amino acid in CesA8 and the corresponding amino acid in CesA3 is I. 3: A is the WT amino acid in CesA1 and the corresponding amino acid in CesA3 is F. The table is summarized from [63, 113, 116-135] and adapted from a published review [1].

In group 2 mutations, P578, D588, and E603 are not built because of lacking electron density, indicating they may reside in the flexible loops. S377F and G617E disrupt the internal structure of the GT domain after mapping to the crystal structure of AtCesA3^{CatD}. A522, G615, S761, E764, and D765 are in the close vicinity of catalytic pocket, of which D765 is the catalytic residue. These mutations could affect substrate binding, catalysis, or cellulose elongation. Remarkably, A533, R437, and H773 are all solvent accessible, and these mutations appear to be not so detrimental for the enzyme integrity and activity. R437K is even the same type of amino acid substitution, possibly the R437K affects the cellulose synthase assembly of CSC (Figure 40).

Figure 40. Mutation sites mapped on AtCesA3

The protein is shown in cartoon and colored in rainbow from blue to red indicating the N and C termini. The mutation sites are shown in spheres.

4.4 Full-length cellulose synthase expression and purification

4.4.1 Small scale expression of AtCesA1 and AtCesA3 constructs

Full-length AtCesAs constructs were tested for expression in insect cells. The membrane was harvested, washed by low salt and high salt buffer, solubilized in 1% Fos-Choline-12 (FC-12). After the centrifugation to remove insoluble material, the solubilized membrane was undergone small-scale nickel-affinity purification and eluted in buffer containing 250 mM imidazole and 0.05% FC12. The solubilized membrane and nickel-affinity purification elution were loaded onto SDS-PAGE for analysis (Figure 41). The constructs were summarized in Table 3.

Table 3: Full-length CesA constructs summary

Number	CesA	Amino Acids
41	AtCesA1	1-1081
42	AtCesA1NT1	211-1081
45	AtCesA1NT1LD2	211-1081(Δ 630-698::GGGGS)
47	AtCesA3	1-1065
52	AtCesA1NTLD	1-1081(Δ 120-204)
53	AtCesA1NTLD-LD2	1-1081(Δ 120-204+ Δ 646-714::GGGGS)
54	AtCesA3NTLD	1-1065(Δ 100-187)

Note: Δ means the designated region deletion

The CesA variants were generally expressed very good and can be efficiently extracted by 1% FC-12 except for construct 42. The proteins can also be purified in reasonable purity after one-step nickel-affinity purification. However, double bands were observed in 52, 53, 54 and both bands were verified to be CesA by mass spectrum, possibly because of degradation.

Figure 41. Small-scale purification results of AtCesAs

The constructs numbers were indicated on top of the gel. SM: Solubilized membrane, E: nickel beads elution.

Considering that there are ten isoforms of AtCesAs in *Arabidopsis thaliana* and full-length AtCesA1's expression was very good, other isoforms were also tested in insect cell expression system following the same strategy (Figure 42). However, overall the expression of all other isoforms was not as good as AtCesA1. Only AtCesA8 can be enriched after the nickel purification with the very faint band presenting in SDS-PAGE. So the full-length AtCesA1 was selected for further purification and subsequent characterization.

Figure 42. Small-scale expression test of AtCesAs isoforms

The numbers indicated the AtCesAs in Arabidopsis. SM: solubilized membrane.

4.4.2 Detergent screening of AtCesA1

Detergents are important factors in solubilizing membrane protein and even some detergents are preferred for Cryo-EM study depending on the physiochemical characters [136]. The commonly used detergents for membrane protein Cryo-EM structure determination are lauryl maltose neopentyl glycol (LMNG), glyco-diosgenin (GDN), and n-Dodecyl- β -D-Maltopyranoside (DDM), which are summarized by Anatrace (<https://www.anatrace.com/Landing/2018/Cryo-EM-Update-Oct18>). Hence, LMNG and DDM were used for small-scale purification of the AtCesAs (Figure 43). Compared to the small-scale purification by FC-12, LMNG and DDM solubilized the membrane less efficiently than FC12. The target bands were very faint with many contaminant bands present. Nevertheless, construct 52 was the best among the constructs purified by LMNG and DDM. Thus construct 52 was further large-scale purified and extracted by either LMNG or DDM. The large-scale purification of

construct 52 in LMNG and DDM were both very impure and unstable (data not show).

Figure 43. Detergent testing on AtCesA1 constructs

The constructs were small-scale purified by LMNG shown on the left and DDM shown on the right. The target protein bands were outlined by green rectangle.

4.4.3 Purification of AtCesA1

Because of the relatively high expression level, the AtCesA1 was purified using 200 mL insect cells that were harvested 48 hours post-transfection. The harvested membrane was solubilized in 1% FC-12 and undergone nickel-affinity purification. The protein was further purified by size-exclusion chromatography (Figure 44). The protein was eluted with two peaks, with retention volumes at ~ 9.4 mL and 10.6 mL, respectively. Based on the calibration size-exclusion chromatography, the former could be trimeric CesA and the latter dimeric CesA. The protein identity was also confirmed by mass spectrometry.

Figure 44. Purification of AtCesA1

Left: Size-exclusion chromatography of purified AtCesA1. The profile of AtCesA1 was shown on the left with peaks labeled with elution volume. Right: The elution fractions were loaded to SDS-PAGE on the right.

4.5 Negative staining of AtCesA1

4.5.1 negative staining of AtCesA1

Since the rapid advancement of Cryo-EM technology, the full-length AtCesA1 could be suitable for structural investigation by Cryo-EM. As the general practice, negative staining electron microscopy was used to characterize the AtCesA's homogeneity. A total of 4 μ L protein samples from nickel-affinity purification, potential trimer peak, and potential dimer peak with a concentration of 0.03 mg/mL were applied to glow-discharged copper grids and visualized under Tecan Spirit 120 kV electron microscopy at a magnification of 49,000 (Figure 45).

Figure 45. Negative staining of AtCesA1

(A) The protein sample eluted from nickel-affinity purification. (B) The 9.4 mL suspected AtCesA1 trimmer. (C) The 10.6 mL suspected AtCesA1 dimer. The typical particles were labeled in green sphere. The typical fibril-shaped particles were labeled in green rectangle.

Even though the eluted protein from nickel-affinity purification was quite pure and had a nice symmetrical peak on size-exclusion chromatography, the protein was not so homogeneous judging from the negative staining. There were two distinctive particle classes. One was the round-shaped particle; the other one was the fibril-shaped particles. For the former, the size was also not very uniform, ranging from 10 + nm to 20 + nm, with particles of 17 nm as the major population. For the fibril-

shaped particles, it was very unlikely the AtCesA1 could have this kind of conformation. And the lengths of fibrils were also not uniform.

4.5.2 Cross-linking of purified AtCesA1 by Grafix

Because the AtCesA1 was not homogeneous enough for Cryo-EM, AtCesA1 was cross-linked with glutaraldehyde (GA) following the GraDeR protocol except for changing the detergent from LMNG to FC-12 [101, 137]. The concept of Grafix is to improve the sample quality especially for large protein complex by a glycerol gradient centrifugation in which a gradient of chemical fixation reagent is also applied (Figure 46. A)[137]. The AtCesA1 had the propensity to form oligomers that migrated above the 200-kDa protein marker. The fractions 1-4 were very heterogeneous under the electron microscope and the particle sizes were over 20 nm. Fractions 5, 6, and 7 had the strongest band and the particles were more homogeneous than the fractions 1-4 (Figure 46. B). The Fractions eluted later than fraction 8 tended to be smeary on SDS-PAGE, possibly due to ununiformed cross-linking. The fractions 5, 6, and 7 were thus collected and loaded to size-exclusion column, showing one sharp and symmetrical peak which was eluted at ~8.7 mL with an estimated molecular weight of ~ 400 kDa (Figure 46. C). The Fraction 6 was stained by uranium acetate and visualized under electron microscopy at a concentration of 0.04 mg/mL. The particles were majorly 17 nm in the longest dimension (Figure 46. D). There were very few fibril-shaped particles and aggregated particles. In conclusion, the AtCesA1 homogeneity was greatly improved by Grafix compared to not cross-linked protein sample.

Figure 46. Grafix of full-length AtCesA1

(A) Schematic view of the principle of Grafix [137]. (B) The SDS-PAGE of gradix fractions. (C) The Size-exclusion chromatography of the Grafixed fractions 5, 6, and 7. The elution volume was labeled. (D) The negative staining of Grafixed sample of faction 6 with typical particles labeled by green star, the scale bar was 100 nm.

4.5.3 Negative staining data processing of Grafixed AtCesA1

In a total of 170 micrographs of the Grafixed AtCesA1 were collected by the Tecnai 120k Spirit electron microscope and processed by EMAN2 [102]. The typical micrograph of the negative staining data was shown in Figure 47. The sample was still not very homogeneous, though particles were majorly 17 nm. Particles larger than 20 nm that were possibly aggregation were also presented in the micrograph. The 17 nm particles were picked and resulted in around 40, 000 particles.

Figure 47. The typical micrograph of cross-linked AtCesA1

Fraction 6 from Grafix was used for negative staining data collection. The scale bar was 100 nm. The 17 nm typical particles were labeled with green circle and typical bad particles were labeled with red circle.

All the particles were 2D class averaged into 50 classes that were shown in Figure 48. Some good classes showed the presence of an internal channel surrounded by the trimer-like feature of AtCesA1, which was consistent with the current model that each lobe of CSC contained three isoforms of CesA. Since the cross-linked AtCesA1 was the best protein sample can be obtained so far and trimer features can be achieved by negative staining, the AtCesA1 was subjected to Cryo-EM with the advancement of Cryo-EM [138].

Figure 48. 2D average results of AtCesA1.

The green square on the upper right of each image indicated “good” class and the blue square on the upper right of each image indicated “bad” class. The green circle outlined the trimer like feature 2D class averages.

4.6 Cryo-EM data processing of cross linked AtCesA1

Cryo-EM has undergone rapid advancement and many high-resolution structures have been published in recent years [139]. Cryo-EM structure of glutamate dehydrogenase with a molecular weight of 334 kDa was solved at 1.8 Å [140]. The smallest protein solved by cryo-EM was 64 kDa methemoglobin at 2.9 Å resolution by traditional Cryo-EM [141]. The smallest proteins solved by the Volta phase plate (VPP) Cryo-EM were 3.2 Å resolution structure of 64 kDa hemoglobin and 52 kDa streptavidin [142, 143]. Even smaller proteins like green fluorescent protein (GFP) with molecular weight as small as 26 kDa was engineered to bind a protein scaffold, thus the scaffold bound GFP was solved to 3.8 Å by Cryo-EM [144].

Because the biological sample is sensitive to the radiation damage, unlike the negative-staining electron microscopy, the electron dose used for Cryo-EM data collection is usually quite low, which results in the low contrast. VPP has been developed to improve the poor contrast of biological samples [145]. VPP is a very thin layer (around 10 nm) amorphous carbon which is implemented in the back focal plane of the electron microscope. The insertion of the VPP exposes amorphous carbon under the electron beam and thus keeps it heated. The irradiation of the VPP results in the negative potential which produces the phase shift between the scattered and unscattered electrons [145]. However, VPP can't improve the final resolution, though the data collection is under low defocus range comparing to the traditional data collection which utilizes the defocus range around -1 μm to -2 μm .

4.6.1 Motion correction, CTF estimation and particle picking

The Cryo-EM recorded micrographs were motion-corrected, CTF estimated and corrected during processing by Relion3 [103]. A typical CTF fitting micrograph and auto-picked particles in CTF corrected micrographs were shown in Figure 49. The protein concentration seemed a bit high, thus particles were too dense on the micrographs. The picking threshold was set to 0.9 in order not to pick too many noise particles.

Figure 49. The CTF fitting and auto-picked particles

Left: the auto fitted CTF estimation. Right: auto-picking of CTF corrected micrograph.

4.6.2 2D classification

All the auto-picked particles were extracted with a box size of 240 pixels and masked with 28 nm, then as input for 2D class average in 100 classes. 7 out of 100 classes were selected as good classes. The rest class averages were discarded because either the class averages were too big (~25 nm), or not distinctive from the background (Figure 50). Only the rough shape can be obtained from the 2D classification and no secondary structure information can be seen from 2D class averages.

Figure 50. The 2D-classification of AtCesA1

The selected “good” classes were shown.

4.6.3 3D model generation and model fitting

7 “good” class averages including 15, 284 particles were used to build the initial model and subsequently, the 3D classification was performed (Figure 51). Class 1 showed a 3-fold like symmetry. Class 2 and class 3 seem to be a degraded form of class 1. Therefore, class 1 was selected and used for 3D auto-refinement. Finally, there were only 5, 168 particles in class 1 and used for auto refinement.

Figure 51. The 3D classification of AtCesA1.

(A), (B), and (C) each showed the Class 1, 2, 3, respectively. The 180° rotation was shown below.

Because the 3D class averages showed the trimer-like feature, the auto-refinement was done by two approaches. One was keeping the C1 symmetry and the other one was applying the 3-fold symmetry against the Class 1 that contained only 5,168 particles. Both the resolutions of 3D auto-refined structures of AtCesA1 were only around 26 Å according to the golden standard FSC resolution estimation. The C1 symmetry map was quite flat with one channel in the central part. However, the C3 symmetry map showed an additional blob on top of the C1 symmetry map central part (Figure 52). Because of the low resolution, currently, it's difficult to validate which map is more physiologically possible. The I-TASSER protein structure prediction server was used to predict the full-length model of AtCesA3 by providing the crystal

structure of AtCesA3³¹⁷⁻⁸¹⁰ [146]. However, all the predicted models had a loosely disordered region in the N terminus and C-SR region, which can hardly be used for Cryo-EM model fitting. Since the crystal structure of bacterial cellulose synthase BcsA has already been solved and the central GT domains of BcsA and AtCesA3CatD aligns well [17], the BcsA was used as the model to fit into the 3D Cryo-EM map. Three molecules of BcsA can be well fitted into the map either C1 symmetry or C3 applied symmetry.

Figure 52. Fitting BcsA to AtCesA1 Cryo-EM map

Top: C1 symmetry map fitting (Left: side view. Right: top view). Below: C3 symmetry fitting (Left: side view. Right: top view).

5. Discussion

5.1 Function of plant P-CR and C-SR domains

The first crystal structure of the plant cellulose synthase catalytic domain is solved and discussed in this thesis. AtCesA3 has an overall conserved GT domain with GT-A fold, though the sequence similarity to other GT family members is very low. The P-CR and C-SR domains are sandwiching the core GT domain. One R362K substitution on AtCesA8 was reported to cause reduced fiber wall thickness and cellulose content [129]. However, the R362K substitution is very unlikely to affect the catalytic activity of full-length cellulose synthase because the R362 is far apart from the catalysis region and solvent accessible judging from the crystal structure of AtCesA3^{CatD}. Since the R362 is located in the P-CR region, it is possible the P-CR may be involved in oligomerization of cellulose synthase or the interaction with CSC interacting -partners that help its assembly and transportation to the plasma membrane.

The C-SR domain is largely disordered by secondary structure prediction [108]. It is composed of long positively charged amino acids clusters in the first 30 amino acids and is rich in negatively charged amino acids in the following 20 amino acids. Such an arrangement of positive and negative amino acids is shared by all the cellulose synthases, though the C-SR region is the least conserved region (Appendix 5). The C-SR domain is very flexible demonstrated by molecular dynamics simulation, though the analysis is based on *Physcomitrella patens* cellulose synthase which shares a similar feature of Arabidopsis cellulose synthase [147]. It is also reported that the

disordered C-SR could be functionally interchangeable among different classes in spite of the class-specific feature [110, 148]. Such disordered nature of C-SR combining with N terminal flexible region between the zinc finger domain and the first transmembrane helix makes plant cellulose synthase notoriously difficult for purification and characterization.

5.2 Mechanism of plant cellulose synthesis

Although the plant CesA shares the conserved D, D, D, QxxRW motif with bacterial BcsA, the catalytic arrangement of plant CesA could be quite different from bacterial cellulose synthase for two aspects from the structural comparison. The first one is the QxxRW motif from the CesA is inserted towards the catalytic pocket and the other one is the catalytic residue D765 lies in the flexible loop connecting the β 10 and β 11 strand. Although the co-crystal structure of UDP and UDP-Glc bound AtCesA3^{CatD} have been determined, the structures failed to show the binding to the cellobiose, which could be explained by the importance of the transmembrane channel to coordinate the cellulose chain. Therefore, purified AtCesA3^{CatD} has no catalytic activity. To completely elucidate the catalytic mechanism of plant cellulose synthesis, the structure of full-length plant cellulose synthase is required.

5.3 Oligomerization of plant cellulose synthase

The edge-to-edge interaction is very common and important for protein functions [149]. For example, the tetrameric domain of human tumor suppressor gene p53, mutating the amino acids involved in the dimer formation completely abolish its

function, the amyloid-like fibrils from amyloidogenic peptide fragments which cause the neurodegenerative Alzheimer disease [150, 151]. AtCesA3^{CatD} forms dimer by β - β strand backbone interaction that is also edge-to-edge interaction, in addition to the α 9 helix interaction. Mutation of AtCesA3^{C573P} transforms the AtCesA3^{CatD} dimer to monomer further validates the importance of β - β strand backbone interaction for dimerization (Figure 53. A) [113]. Analysis of the β 6 sequence in Arabidopsis cellulose synthases shows the position of Q571 is more favorable for lysine, but all are positively charged residues. The position of V572 is all hydrophobic, and uncharged residues (V, L, and C). C573 is conserved among all the AtCesA1, 3, 4, 6, 7, 8. Y574 is almost conserved in all AtCesAs except for AtCesA8 with F (Figure 53. B). Such amino acid conservation and edge-to-edge interaction perhaps enable the AtCesA isoforms to form homo-oligomers or hetero-oligomers without the specific amino acid requirement (Figure 53. C). The dimer structure of AtCesA3CatD is consistent with the phenomenon that different isoforms of cellulose synthases interact with each other, which could be the earliest stage of CSC assembly. The dimer structure of AtCesA3^{CatD} is in agreement with the dimer shape of cellulose synthase determined by SAXS [87]. Because of the edge-to-edge interaction, such a feature also enables the cellulose synthase interacting with other protein classes with solvent exposing β strand for its assembly and transportation. However, most importantly, it needs to be validated *in vivo* whether such a dimer configuration of cellulose synthase catalytic domain is physiologically relevant, which is currently ongoing.

Figure 53. Edge-to-edge interaction of AtCesA

(A) Schematic view of edge-to-edge interaction and the proline mutation disruption. The figure was adapted [113]. (B) The sequence alignment of $\beta 6$ of Arabidopsis cellulose synthases. (C) Possible interaction among Arabidopsis cellulose synthase.

Considering the plant cellulose synthase isoforms are quite conserved in sequence and high-resolution structure of CSC is still the main puzzle, AtCesA3CatD structure was used to generate the homotrimer model which is consistent with the diameter of the negative stain result of each lobe of CSC [95]. The trimer model of full-length AtCesA1 further demonstrates that the six lobes CSC has only three subunits in each lobe, which is in agreement with current prevailing CSC models. Moreover, from the negative stain structure of CSC, it showed that the CSC is quite loosely connected regarding the six lobes. The CSC is assembled, then transported to the plasma

membrane by Golgi vesicles. The interactions that keep the six lobes together have not been identified. It is possible by the N terminal zinc finger domain, GT domain, or transmembrane domain. How the plant cellulose synthase is assembled into the CSC and how the subunits interact with each other require further study.

Two CSC models (“compact” and “loose”) are shown in Figure 54. For a compact CSC model, each cellulose synthase needs at least three interaction interfaces to interact with other isoforms for CSC formation. However, if the CSC model is loosely connected, then each cellulose synthase only needs to interact with the other two CesA isoforms in the same lobe. Whichever model of the CSC is, the dimer structure of AtCesA3^{CatD} showing the bipartite edge-to-edge interaction can only be one of the interaction interfaces. Future structural work of the entire CSC complex is required to answer how the CSC is assembled.

Figure 54. CSC model

(A) and (B): the “compact” and “loose” model of 18 subunits CSC. Each color represents one CesA subunit.

5.4 Sample preparation for Cryo-EM

The purification and *in vitro* characterization of plant cellulose synthase is extremely difficult, and currently, there is still a lack of experimental high-resolution CSC structure. The purification of full-length AtCesA1 and the low-resolution Cryo-EM reconstruction was carried out in the thesis. Although the rapid development of techniques, like direct electron detectors with improved quantum efficiency, automated data collection, and robust data analysis software, the sample preparation remains the bottleneck of Cryo-EM [136, 152]. The main obstacle for membrane protein Cryo-EM is the determination of the chemical conditions which can solubilize the membrane protein and retain its native state [136]. The most widely and successfully used detergents, polymers from the Anatrace are GDN, DDM, LMNG, and amphipols like Amphipol A8-35, Amphipol PMAL-C8, as well as the Nanodisc [153]. Different from most membrane proteins showing a sharp contrast between 2D class averages and background, the 2D classification of AtCesA1 only shows rough shape, which is possibly resulted from heterogeneous and bad quality particles. Therefore, after the model building and auto-refinement, only a small fraction of the total particles is used and the model resolution is only around 26 Å. Even though the Cryo-EM model is at low resolution, it demonstrates the feature with a 3-fold symmetry which is similar to the trimeric model from SAXS analysis, as shown in the introduction section 1.7.3. More construct optimization and protein purification procedures need to be carried out to improve the sample purify for high-resolution Cryo-EM structure determination.

Chapter 2: Structural study of Arabidopsis Profilin3

1 introduction

1.1 Introduction to cytoskeleton

The cytoskeleton is universally present in all kingdoms of lives, from prokaryotes, archaea to eukaryotes, though it was initially believed that only the eukaryotic cells possess the cytoskeleton network [154]. The cytoskeleton provides the cell morphology and mechanical support. It is involved in cell signaling, migration, intracellular transportation, and cell division [154]. Three types of cytoskeletal networks are present in eukaryotic cells that are microtubule, intermediate filament, and microfilament formed by respective proteins (Figure 55). The microtubule is a hollow cylinder, with ~23 nm in diameter formed by 13 protofilaments which are polymerized by α and β tubulin dimers (Figure 55. B) [155]. The intermediate filament, generally 10 nm in diameter, is formed by diverse proteins sharing a common structural feature which is a central coiled-coil domain (Figure 55. C) [156]. Microfilament, also known as actin filament, is composed of monomeric actin, assembled into two chains that intertwined to form filament chains (Figure 55. A) [157].

Figure 55. Cytoskeleton in eukaryotic cells

(A-C): the schematic drawing of the microtubule, microfilament and intermediate filament. The figure is adapted from online source (<https://www.open.edu/openlearn/science-maths-technology/science/tour-the-cell/content-section-4.2>).

It is until in the 1990s that bacteria FtsZ is identified to be a tubulin homolog [158]. FtsZ belongs to the GTPase family and mutation of FtsZ abolishes the cell division [158]. The crystal structure of *Methanococcus jannaschii* FtsZ1 is solved in 1998 which shows the high similarity to tubulin despite low sequence similarity <10% (Figure 56. B) [159]. There are other tubulin homologs in bacteria identified so far, BtubA/B and TubZ proteins [160-162].

Figure 56. Structure of tubulin and tubulin-like protein FtsZ

(A) Crystal structure of *Bos taurus* tubulin colored in rainbow from blue to red indicating the N and C termini (PDB code: 1JFF) (B) Crystal structure of *Methanocaldococcus jannaschii* FtsZ colored in rainbow from blue to red indicating the N and C termini (PDB code: 1W5A).

Bacterial actin homologs (MreB and ParM) are discovered by sequence alignment based on the ATPase motif, after which it is demonstrated that MreB forms the cytoskeletal structure in bacterial cells (Figure 57. B) [163, 164]. Unlike the eukaryotic actin network which shares remarkable conservation among organisms, bacterial actin homologs varies greatly in species and as many as tens of actin homologs have been identified [165]. Both eukaryotic intermediate filament (IF) and bacterial counterparts proteins are quite variable. Crescentin and FilP are identified to be members of the bacterial IF-like protein [166, 167].

Figure 57. Structure of Actin and actin like protein

(A) Crystal structure of rabbit actin colored in rainbow from blue to red indicating the N and C termini (PDB code: 1ATN). The subdomains, Nucleotide binding cleft, and barbed-end groove were marked. (B) Crystal structure of *Thermotoga maritima* MreB colored in rainbow from blue to red indicating the N and C termini (PDB code: 1JCG). (C) Crystal structure of *Escherichia coli* ParM colored in rainbow from blue to red indicating the N and C termini (PDB code: 1MWM). The bound nucleotide is shown in sticks and metal ion in spheres.

1.2 Introduction to eukaryotic actin network

Actin is one of the most abundant proteins expressed in eukaryotic cells with ~ 370 amino acids divided into four subdomains [168]. Because the actin filament is polarized structure, the pointed end and barbed end are labeled (Figure 57. A) [169]. The nucleotide-binding cleft is located at the pointed end and the barbed end groove which is responsible for many proteins binding is important for actin assembly regulation (Figure 57. A). Actin strongly binds to ATP or ADP in the presence of Ca^{2+} or Mg^{2+} . Monomeric actin spontaneously polymerizes under the physiological salt condition and many proteins in the cell are involved in regulating actin polymerization thus affecting the cytoskeleton[168]. These actin-regulating proteins are classified into several classes based on their function. Monomeric actin-binding proteins bind monomeric actin either protecting it from the polymerization or bringing it to barbed end to promote actin filament elongation, such as Profilin, Thymosin- β 4. Proteins like cofilin, gelsolin family cut actin filaments thus disrupt actin filaments are classified as severing proteins. Nucleation proteins like Arp2/3 produce branched actin filaments, formin initiates unbranched actin filaments. Capping proteins bind to the barbed end or pointed end of actin filaments. Tropomodulin is a pointed end-capping protein that wraps around three terminal actins at the pointed end and blocks

monomer actin exchange. Cross-linking proteins which have two actin-binding sites can stabilize the high order structure of actin filaments [168]. The functions of these different actin-binding protein classes are shown in Figure 58.

Figure 58. Overview of proteins regulating actin filaments

Proteins regulating actin filaments include monomer binding, elongation such as formins, capping proteins, severing proteins, cross-linking proteins, and branching protein Arp2/3 complex. This figure is adapted from a review published on cold spring harbor perspective biology [168].

1.3 Introduction to Profilin

Profilins are small but abundant cytosolic proteins with molecular weight approximately 15 kDa and are universally expressed in eukaryotes and regulate actin filament elongation by binding with monomeric actin and formin proteins [170-172]. In *A. thaliana*, there are in total of five Profilin isoforms (AtPRF1-5) that share over 70% sequence identity except for AtPRF3 having an extra N-terminal 37 amino acids (Figure 59). AtPRF1, AtPRF2, and AtPRF3 are ubiquitously expressed in plant tissues, while AtPRF4 and AtPRF5 are only selectively expressed in pollen [173,

174]. Formin-mediated actin assembly is one of the two major assembly machines in plants in generating long actin cables for plant growth and immune defense. During actin elongation, profilin transiently binds with the formin polyproline region (Poly-P: RVPPPPPPPPPLP) by the first and fourth α helix to deliver actin monomer into the filament-growing end and after that released quickly for next rounds of actin monomer recruitment [175]. Moreover, profilin can also directly interact with actin with the surface involving the third and fourth helix. However, with an extended N-terminus, AtPRF3 binds formin1 Poly-P with $K_d \sim 30 \mu M$ which is stronger than other profilin isoforms (usually with $K_d \sim 80 \mu M$) [172]. AtPRF3 strongly inhibits actin spontaneous polymerization by *in vitro* actin polymerization assay and the inhibition effect was even stronger in the presence of formin, which makes it atypical among profilin isoforms in Arabidopsis [172]. Moreover, the AtPRF3 N terminal extension (NTE) contains a highly disordered region that is usually a signaling hub during signal transduction by posttranslational modification [176].

Figure 59. Multiple sequence alignment of AtPRF1-5

The multiple sequence alignment of AtPRF1-5 is generated by ClustalW and the figure is generated by Esprit [177, 178].

2. Research objectives

Arabidopsis AtPRF3 is unique compared to other profilin isoforms by having N terminal 37 amino acids extension. Such a unique feature enables it have a stronger binding affinity to AtFormin1 thus negatively regulates the actin polymerization. This study aims to understand how the AtPRF3^{NTE} can aid to the increased binding affinity of AtPRF3 to AtFormin1 Poly-P by structural study and molecular dynamics simulation.

3. Materials and method

3.1 Molecule cloning

The genes encoding *Arabidopsis thaliana* AtPRF2 (Gene Bank AY114048.1) and AtPRF3 (Gene Bank BT000885.1) were cloned into a modified vector of pET-21d (+) with N-terminal histidine tag and human 3C protease cleavage site. The AtPRF3 variants AtPRF3 Δ 37, AtPRF3 Δ 22, AtPRF3 Δ 22M1, and AtPRF3 Δ 22M2 were generated by site-directed mutagenesis followed by sequencing from 1st base (Singapore).

3.2 Protein expression and purification

All AtPRF variants were transformed into *E. coli* RosettaTM (DE3). Single colonies were picked and cultured for expression. Bacteria were induced by 0.1 mM IPTG at 16 °C for 16 hours for protein expression and harvested by centrifugation at 4, 000 rpm for 10 minutes (JLA-8.1, Beckman, USA). The pellet was resuspended in buffer-HisA (25 mM Na-HEPES, pH 7.5, 150 mM NaCl, 20 mM Imidazole), lysed by LM20 microfluidizer (Microfluidics, USA) and clarified by centrifugation at 20, 000 rpm (JA-25.5, Beckman, USA) at 4 °C for 1 hour. The supernatant was collected, filtered through 0.22 μ m filter and loaded onto a HisTrap FF 5 mL column (GE Healthcare) and the eluted fractions were pooled and dialyzed overnight against buffer-GF (25 mM Na-HEPES, pH 7.5, 150 mM NaCl, 5 mM β -ME) by adding human 3C protease to cleave the histidine tag. The tag cleaved protein was load onto HiLoad Superdex 75 16/60 (GE Healthcare) pre-equilibrated with buffer-GF. The eluted protein fractions

were pooled, concentrated to 20 mg/mL and snap-frozen in liquid nitrogen and stored at -80 °C for later use.

3.3 Crystallization, optimization, and structure determination

3.3.1 Crystal screening of AtPRFs

Purified proteins were subjected to crystal screening against crystallization kits from Hampton research using sitting drop method described in Chapter 1 method section 3.7.1.

3.3.2 Crystal optimization of AtPRFs

The initial crystal screening of AtPRF3Δ37 showed phase separation in a lot of conditions which generally used salt (ammonium sulfate, magnesium sulfate) as the precipitant. After extensively optimization, the protein was finally crystallized in excellent shape and diffraction capacity in 0.1 M sodium acetate pH 5.0, 1.42 M ammonium sulfate with a protein concentration of 10 mg/mL by seeding and adding glycerol into the reservoir at 4 °C.

AtPRF2 was initially crystallized in many conditions also primarily using ammonium phosphate and ammonium sulfate as precipitant. The final optimized condition for AtPRF2 was 0.1 M Tris-HCl, pH 8.5 and 2.2 M Ammonium phosphate dibasic with a protein concentration of 10 mg/mL.

There were only two initial crystallization conditions that Poly-P bound AtPRF2 was crystallized which were 2.5 M ammonium sulfate, 0.1 M Bis-Tris, propane pH 7.0

and 2.0 M Ammonium citrate tribasic pH 7.0, 0.1 M Bis-Tris propane, pH 7.0.

The final optimized condition for Poly-P bound AtPRF2 was 0.1 M Bis-tris pH, 7.0, 2.0 M ammonium phosphate dibasic with a protein concentration of 10 mg/mL.

The initial crystallization condition of AtPRF3 Δ 22M2 was 100 mM Tris-HCl pH 8.5 and 1.2 M sodium citrate. The final optimized condition for AtPRF3 Δ 22M2 was 0.1 M Tris-HCl pH 7.8 and 0.9 M sodium citrate with a protein concentration of 20 mg/mL.

3.3.3 Data collection and structure determination

The crystals were harvested by adding 25% glycerol to the mother liquor. The AtPRF2 and Poly-P bound AtPRF2 diffraction datasets were collected using a Rigaku FR-X X-ray generator with a PILATUS3R 300K detector. The AtPRF3 Δ 37 dataset was collected from the Australian Light Source MX2 beamline. The AtPRF3 Δ 22 diffraction datasets were collected in Swiss Light Source. The datasets were indexed, integrated and scaled by HKL3000 or processed by XDS [97, 179]. The structures were solved by molecular replacement with Phaser in the Phenix suite with the AtPRF1 structure (PDB code: 3NUL) as the template [99]. The manual model building was finished by COOT [100]. All the structural figures were generated by PyMol program [12]. All the data collection and refinement statistics were summarized in Appendix 7. The multiple sequence alignment was done with ESPript [178].

4. Results

4.1 Purification of AtPRF3 Δ 37, AtPRF3 Δ 22, and AtPRF3

In addition to the full-length AtPRF3, two truncated constructs of AtPRF3 were generated which were removing the first 22 amino acids (AtPRF3 Δ 22) and the entire N terminal 37 amino acids (AtPRF3 Δ 37), respectively. AtPRF3, AtPRF3 Δ 22, and AtPRF3 Δ 37 were purified following the similar protocol (Figure 60). More than half of full-length AtPRF3 aggregated which eluted in void volume judging from size-exclusion chromatography. As for AtPRF3 Δ 22, the truncation of the first 22 amino acids, had significantly reduced aggregation. Furthermore, truncating the entire 37 amino acids region which was not present in other forms of AtPRFs completely avoided the protein aggregation, indicating that the inclusion of the disordered region destabilized the protein and caused the protein more prone to aggregate. AtPRF3, AtPRF3 Δ 22, and AtPRF3 Δ 37 were all in good purity after collecting ~80 mL elution peak from the size-exclusion chromatography fractions.

Figure 60. Purification of AtPRF3 variants

(A, B, and C) Size-exclusion chromatography of AtPRF3 variants. (D) The SDS-PAGE of purified AtPRF3, AtPRF3 Δ 22, and AtPRF3 Δ 37.

4.2 Calibration size-exclusion chromatography

4.2.1 Standard protein calibration

The Superdex 200 10/300 increase column was used for calibration size-exclusion chromatography. Four standard proteins were used for standard equation estimation (Table 4). Blue dextran was used to measure the void volume of the column which was 7.42 mL. The standard equation was calculated by following the analytical size-exclusion chromatography manual (GE Healthcare) (Figure 61).

Table 4. The proteins used for calibration.

Standard	Ve	TMW
Ferritin	9.29	440
Aldolase	10.78	158
Conalbumin	12.84	75
Ovalbumin	14.61	44

Note. Ve: Elution volume. TMW: theoretical molecular weight (/kDa)

Figure 61. Calibration size-exclusion of four standard proteins

Four standard proteins were used for standard equation estimation. The profile of calibration size-exclusion chromatography was shown on the left and the estimated equation was shown on the right.

4.2.2 Calibration of AtPRFs

All of the AtPRF3 variants were subjected to the calibration size-exclusion chromatography for characterization (Figure 62). The respective calculated molecular weights of AtPRF3 variants were calculated based on the standard equation in Figure 59 and were shown in Table 5. The full-length AtPRF3, the truncated variants AtPRF3 Δ 22 and AtPRF3 Δ 37 were all monomer in solution.

Figure 62. The calibration of AtPRF3 variants

The size-exclusion chromatography of four AtPRF3 variants were shown with the elution volume indicated.

Table 5. The calculated molecular weight of AtPRF3 variants.

Protein	TMW	CMW
AtPRF3	18.28	22.25
AtPRF3Δ37	14.04	15.14
AtPRF3Δ22	15.87	16.18
AtPRF3Δ22M2	15.71	17.45

Note: TMW, theoretical molecular weight; CMW, calculated molecular weight, the unit, /kDa.

4.3 Crystal optimization of AtPRF proteins

4.3.1 Crystal optimization of AtPRF3Δ37

AtPRF3Δ37 tended to form phase separation in the primary crystal screening, which was worthy of further optimization. The precipitants in these conditions were mostly ammonium sulfate, ammonium phosphate. Very small and thin crystals appeared from the edges of phase separation after optimizing the precipitant concentration and pH range (Figure 63). However, no diffraction spots could be obtained when those crystals were shot on the Australian synchrotron source X-ray. Additive screening from Hampton research also did not help change the crystals to good morphology and diffraction. Luckily, crystals growing under 4 degree supplemented with 1% to 12% glycerol into the reservoir with seeding yielded excellent crystals that diffracted beyond 1.4 Å (Figure 63).

Figure 63. The optimized crystals of AtPRF3Δ37

The initial optimized crystals were shown on the left and the final optimized crystals grown under 4 degree were shown on the right. The scale bar was 100 μm.

4.3.2 Crystal optimization of AtPRF3Δ22

The path of optimizing the crystals of AtPRF3Δ22 was also very tough. After the extensive optimization including additive screening, the crystals remained extreme thin needle clusters that were not suitable for X-ray diffraction data collection (Figure 64. A). The N terminal extension contains five tandem lysines (K31-K35) after analyzing the amino acid sequence of AtPRF3Δ22. Lysine is positively charged and generally solvent accessible. Because of its intrinsic flexibility, the lysine cluster would be an energy barrier which makes protein recalcitrant to be crystallized [105]. Based on the SerP server (<https://services.mbi.ucla.edu/SER/>) prediction, surface entropy reduction mutant AtPRF3Δ22^{K32A/K33A/K34A} name as AtPRF3Δ22M1 was generated. The AtPRF3Δ22M1 protein was purified similarly as AtPRF3Δ22 and subjected to crystal screening and optimization. AtPRF3Δ22M1 crystals were greatly improved after optimization, though the crystals were still ugly-looking and diffracted very limitedly (Figure 64. B). According to the sequence alignment of AtPRFs, the AtPRF3^{L136} was conserved in all the four AtRPFs except for AtPRF3. Thus, L136Q

mutation was generated base on AtPRF3 Δ 22M1, which resulted in AtPRF3 Δ 22^{K32A/K33A/K34A/L136Q} (AtPRF3 Δ 22M2). AtPRF3 Δ 22M2 was purified following the purification procedure of AtPRF3 Δ 22. AtPRF3 Δ 22M2 could be successfully crystallized in various conditions using sodium citrate as the precipitant. After crystal optimization, the single crystals of AtPRF3 Δ 22M2 were obtained with good morphology (Figure 64. C).

Figure 64. The representative crystals of AtPRF3 Δ 22 variants.

(A-C) The optimized crystals of AtPRF3 Δ 22, AtPRF3 Δ 22M1, and AtPRF3 Δ 22M2. The final optimized condition was shown in the bottom. The scale bar was 100 μ m.

4.4 Purification, crystallization of AtPRF2, and Poly-P bound AtPRF2

Extensively crystal-screening aiming for getting the Poly-P bound structure of AtPRF3 failed. Since profilins were highly similar in regards to the sequence and structure, Poly-P bound AtPRF2 was an alternative option. The purification of AtPRF2 was similar to AtPRF3 Δ 37 (Figure 65). AtPRF2 was readily crystallized and the Poly-P bound AtPRF2 crystal was also quickly optimized to decent crystals (data not shown). The X-ray diffraction datasets collection and structure determination of apo and Poly-P bound AtPRF2 were described in the method section.

Figure 65. The purification of AtPRF2

The size-exclusion chromatography of AtPRF2 was shown on left and the SDS-PAGE of respective fractions were shown on the right.

4.5 Structure of AtPRF3 Δ 37 and AtPRF3 Δ 22

Crystal structure of AtPRF3 Δ 37 is determined beyond 1.5 Å resolution in space group C121 with one molecule in one asymmetric unit. All residues are unambiguously built except for the first residue Met in the N-terminus. AtPRF3 Δ 37 is consisted of seven anti-parallel β sheets that are sandwiched by four α -helices demonstrating a canonical profilin fold (Figure 66. A) [180]. The structure of AtPRF3 Δ 22M2 is determined at 3.6 Å resolution with ten molecules in one asymmetrical unit. Residues 23-30 (QRRSRAKV) are unable to be modeled due to lacking electron density. The structure of AtPRF3 Δ 22M2 demonstrates a protruding N-terminus from K31 to N37 (KKAAATN), compared with AtPRF3 Δ 37 (Figure 66. B). K31- N37 does not directly interact with the profilin core domain, appearing to be dynamic. The nearest possible hydrogen bond pair distance between two AtPRF3 Δ 22M2's N terminuses (K31- N37) is 5.0 Å, indicating no strong interaction between paired AtPRF3 Δ 22M2, though we can't rule out the weak interactions (Figure 66. C). The unbiased Fo-Fc map of K31- N37 in chain A is shown (Figure 66. D). The main chains of K31- N37

are well fitted to the Fo-Fc map, indicating that they are not over-fitted, though AtPRF3 Δ 22M2 proteins are paired in crystal packing (Figure 64. B).

Figure 66. Structure of AtPRF3

(A) The structure of AtPRF3 Δ 37 is colored in cyan. The secondary structures are labeled. (B) The structure of AtPRF3 Δ 22M2 is colored in orange. The N terminal extension K31-N37 colored in violet is shown on the right. The secondary structures are labeled in alphabet. (C) The crystal packing of AtPRF3 Δ 22M2 with crystallographic dimer is shown in cartoon. The N terminal extensive is shown in sticks and the distances are labeled. (D) The unbiased Fo-Fc map is contoured at 2.2 σ of chain A from 3.6 Å crystal structure of AtPRF3 Δ 22M2 with the amino acids labeled.

4.6 Structural analysis of AtPRF2 and Poly-P bound AtPRF2

To better understand how Arabidopsis profilin specifically interacts with Poly-P, the synthetic Poly-P (RVPPPPPPPPPLP) from AtFH1 which is involved in actin polymerization is used to get Poly-P bound AtPRF2 structure [181]. Crystal structures of AtPRF2 in both apo and Poly-P-bound forms are determined at 2.4 Å and 1.9 Å resolution respectively (Figure 67. A and B). The structure of the isolated AtPRF2 shares the same fold as that of AtPRF1 and AtPRF3Δ37, with an RMSD of 0.41 and 0.40 for C-α atoms, respectively (Figure 67. A), suggesting that the overall core structures of Arabidopsis profilin isoforms are highly conserved [182]. As for the Poly-P bound AtPRF2, all the 13 amino acids except for the first arginine in Poly-P can be modeled (Figure 67. B). The Poly-P peptide is held in the hydrophobic groove formed by the first and the fourth helices of AtPRF2. Notably, Poly-P binds to AtPRF2 with 1:1 stoichiometry, in a similar way as that of plant pollen allergenic profilin protein [183]. Three key residues (Trp3, Tyr6, and Tyr125) in AtPRF2 and three essential prolines (Pro4, Pro7, and Pro8) in Poly-P are involved in hydrogen bond formation. Other key residues (His10, Trp33, Leu122, Leu126, Ser129, and Leu131) are also involved in the formation of the hydrophobic pocket for Poly-P binding. No significant conformational change of AtPRF2 is observed upon the binding of Poly-P comparing to the apo structure of AtPRF2 except for Tyr125 shifts 0.5 Å toward Poly-P (Figure 67. C). Unlike human platelet profilin-1 which can adopt two distinct Poly-P backbone orientations [184], Poly-P binds to AtPRF2 via only one direction, indicating a more favorable orientation that is observed in other Poly-P bound profilin structures [183, 185, 186] (Figure 67. D)

Figure 67. Crystal structure of apo and Poly-P bound AtPRF2

(A) Structure alignment of AtPRF1 in yellow (PDB ID: 3NUL), AtPRF2 in lime and AtPRF3D37 in cyan. (B) Poly-P bound structure of AtPRF2. The 2Fo-Fc map contoured at 1.6 σ of Poly-P is shown. (C) Amino acids in AtPRF2 and prolines in Poly-P involved in direct interaction are shown in sticks. (D) Poly-P bound structure alignment of AtPRF2 colored in lime while the Poly-P showed in salmon, human palate PRF1 (PDB ID: 1AWI) in grey, and pollen PRF (PDB ID: 5EV0) in slate.

4.7 N terminus of AtPRF3 Δ 22 occupied the Poly-P binding groove

Poly-P bound AtPRF2 structure can serve as the model of Arabidopsis profilin interacting with Poly-P because of the high sequence and structural similarity. The structure of AtPRF3 Δ 22M2 is aligned to the Poly-P bound structure of AtPRF2 using the profilin core domain resulting in an RMSD of 0.347, which suggests similar

overall folding. Unexpectedly, the residues A35/T36/N37 of NTE occupy the Poly-P binding groove (Figure 68). Considering the AtPRF3 Δ 22M2 is still able to bind Poly-P with higher affinity than AtPRF3 Δ 37 (data not shown), the NTE could be flexible that enables it to accommodate multiple dynamic conformations to modulate Poly-P binding.

Figure 68. The steric hindrance of Poly-P binding pocket with AtPRF3^{NTE}

(A) and (B) two mirror views of aligned structures of Poly-P bound AtPRF2 in lime, and AtPRF3 Δ 22 in orange. The Poly-P is shown in both sticks and surface. The residues having steric hindrance with Poly-P binding in NTE of AtPRF3 are shown in sticks.

5. Discussion

5.1 Profilin is evolutionally conserved but AtPRF3 is unique

Profilins are structurally highly conserved among eukaryotes. Even archaea encoded profilins, like Loki profilin-1, Loki profilin-2, and Odin profilin also adopt the typical profilin fold in spite of extremely low sequence similarity which is ~ 11-17% to human profilin1 but still retains the interaction with rabbit actin [187]. Arabidopsis Profilin3 is characterized by harboring N terminal 37 amino acid more than typical profilins in arabidopsis. The N terminal 1-22 amino acids are disordered and 23-31 amino acids have the propensity to form helix from secondary structure prediction (Appendix 6). Peptides with flanked core region could dynamically evolve nonspecific interactions with interacting proteins in the vicinity of the binding areas [188]. Considering that AtPRF3 has flanked intrinsically disordered region and Poly-P has flanked amino acids in addition to the core prolines involved in interaction with AtPRF3, AtPRF3 probably adopts non-specific and transient interactions that are evolved rapidly under different environmental conditions to accommodate the binding to Poly-P.

5.2 Molecular dynamics simulation aids in understanding AtPRF3^{NTE} dynamics

Though the AtPRF3 Δ 22 that includes the amino acids from Q23 to N37 is crystallized, only the N terminal residues from K31 to N37 can be modeled. The residues from Q23 to V30 are not modeled in the crystal structure, which suggests the residues from Q23 to V30 could be very flexible and don't form stable contact with

profilin core domain, otherwise, they can be stabilized by the core domain and thus be modeled. Flexible regions are always the obstacles for X-ray crystallography study that requires the ordered packing of individual molecules, which also explains the difficulty in crystal optimization, low resolution, and high b factor of AtPRF3 Δ 22. Such flexibility in AtPRF3^{NTE} conformation towards Poly-P binding provides an excellent platform to fine-tune the unique negative function of AtPRF3 in balancing actin assembly with other AtPRF isoforms during plant defense responses. Molecular dynamics simulation is an excellent suitable tool to study protein dynamics. It can be combined with the crystal structure of AtPRF3 Δ 22 to study the N terminal extension dynamics and possibly dissect the additional interaction partners between AtPRF3^{NTE} and Poly-P. In collaboration with Assoc Prof Mu Yuguang from our school, molecular dynamic simulation of AtPRF3^{NTE} has been carried out, but it is beyond my own thesis work thus is not included.

5. References

1. McFarlane, H.E., A. Doring, and S. Persson, *The cell biology of cellulose synthesis*. Annu Rev Plant Biol, 2014. **65**: p. 69-94.
2. Heinze, T., *Cellulose: Structure and Properties*, in *Cellulose Chemistry and Properties: Fibers, Nanocelluloses and Advanced Materials*, O.J. Rojas, Editor. 2016, Springer International Publishing: Cham. p. 1-52.
3. Qiu, X. and S. Hu, "Smart" Materials Based on Cellulose: A Review of the Preparations, Properties, and Applications. Materials (Basel), 2013. **6**(3): p. 738-781.
4. Janse van Rensburg, H.C. and W. Van den Ende, *UDP-Glucose: A Potential Signaling Molecule in Plants?* Front Plant Sci, 2017. **8**: p. 2230.
5. Harden, T.K., et al., *Signalling and pharmacological properties of the P2Y receptor*. Acta Physiol (Oxf), 2010. **199**(2): p. 149-60.
6. Sivaraman, J., et al., *Crystal structure of Escherichia coli glucose-1-phosphate thymidyltransferase (RffH) complexed with dTTP and Mg²⁺*. J Biol Chem, 2002. **277**(46): p. 44214-9.
7. Kotake, T., et al., *Properties and physiological functions of UDP-sugar pyrophosphorylase in Arabidopsis*. Biosci Biotechnol Biochem, 2007. **71**(3): p. 761-71.
8. Zheng, Y., et al., *The structure of sucrose synthase-1 from Arabidopsis thaliana and its functional implications*. J Biol Chem, 2011. **286**(41): p. 36108-18.
9. Lairson, L.L., et al., *Glycosyltransferases: structures, functions, and mechanisms*. Annu Rev Biochem, 2008. **77**: p. 521-55.
10. Charnock, S.J. and G.J. Davies, *Structure of the nucleotide-diphospho-sugar transferase, SpsA from Bacillus subtilis, in native and nucleotide-complexed forms*. Biochemistry, 1999. **38**(20): p. 6380-5.
11. Morera, S., et al., *High resolution crystal structures of T4 phage beta-glucosyltransferase: induced fit and effect of substrate and metal binding*. J Mol Biol, 2001. **311**(3): p. 569-77.
12. DeLano, W.L., *PyMOL: An Open-Source Molecular Graphics Tool*. 2002: CCP4 Newsletter On Protein Crystallography.
13. Hall-Stoodley, L., J.W. Costerton, and P. Stoodley, *Bacterial biofilms: from the natural environment to infectious diseases*. Nat Rev Microbiol, 2004. **2**(2): p. 95-108.
14. Romling, U. and M.Y. Galperin, *Bacterial cellulose biosynthesis: diversity of operons, subunits, products, and functions*. Trends Microbiol, 2015. **23**(9): p. 545-57.
15. Saxena, I.M., F.C. Lin, and R.M. Brown, Jr., *Cloning and sequencing of the cellulose synthase catalytic subunit gene of Acetobacter xylinum*. Plant Mol Biol, 1990. **15**(5): p. 673-83.
16. Wong, H.C., et al., *Genetic organization of the cellulose synthase operon in Acetobacter xylinum*. Proc Natl Acad Sci U S A, 1990. **87**(20): p. 8130-4.

17. Morgan, J.L., J. Strumillo, and J. Zimmer, *Crystallographic snapshot of cellulose synthesis and membrane translocation*. Nature, 2013. **493**(7431): p. 181-6.
18. Morgan, J.L., J.T. McNamara, and J. Zimmer, *Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP*. Nat Struct Mol Biol, 2014. **21**(5): p. 489-96.
19. Morgan, J.L., et al., *Observing cellulose biosynthesis and membrane translocation in crystallo*. Nature, 2016. **531**(7594): p. 329-34.
20. Peng, L., et al., *Sitosterol-beta-glucoside as primer for cellulose synthesis in plants*. Science, 2002. **295**(5552): p. 147-50.
21. Nojima, S., et al., *Crystal structure of the flexible tandem repeat domain of bacterial cellulose synthesis subunit C*. Sci Rep, 2017. **7**(1): p. 13018.
22. Acheson, J.F., Z.S. Derewenda, and J. Zimmer, *Architecture of the Cellulose Synthase Outer Membrane Channel and Its Association with the Periplasmic TPR Domain*. Structure, 2019. **27**(12): p. 1855-1861 e3.
23. Saxena, I.M., et al., *Characterization of genes in the cellulose-synthesizing operon (acs operon) of Acetobacter xylinum: implications for cellulose crystallization*. J Bacteriol, 1994. **176**(18): p. 5735-52.
24. Hu, S.Q., et al., *Structure of bacterial cellulose synthase subunit D octamer with four inner passageways*. Proc Natl Acad Sci U S A, 2010. **107**(42): p. 17957-61.
25. Sun, L., et al., *Structural and Functional Characterization of the BcsG Subunit of the Cellulose Synthase in Salmonella typhimurium*. J Mol Biol, 2018.
26. Serra, D.O., A.M. Richter, and R. Hengge, *Cellulose as an architectural element in spatially structured Escherichia coli biofilms*. J Bacteriol, 2013. **195**(24): p. 5540-54.
27. Thongsomboon, W., et al., *Phosphoethanolamine cellulose: A naturally produced chemically modified cellulose*. Science, 2018. **359**(6373): p. 334-338.
28. Alzari, P.M., H. Souchon, and R. Dominguez, *The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from Clostridium thermocellum*. Structure, 1996. **4**(3): p. 265-75.
29. Mazur, O. and J. Zimmer, *Apo- and cellopentaose-bound structures of the bacterial cellulose synthase subunit BcsZ*. J Biol Chem, 2011. **286**(20): p. 17601-6.
30. Koo, H.M., et al., *Evidence that a beta-1,4-endoglucanase secreted by Acetobacter xylinum plays an essential role for the formation of cellulose fiber*. Biosci Biotechnol Biochem, 1998. **62**(11): p. 2257-9.
31. Kawano, S., et al., *Effects of endogenous endo-beta-1,4-glucanase on cellulose biosynthesis in Acetobacter xylinum ATCC23769*. J Biosci Bioeng, 2002. **94**(3): p. 275-81.
32. Keegstra, K., *Plant Cell Walls*. Plant Physiology, 2010. **154**(2): p. 483-486.
33. Keegstra, K., *Plant cell walls*. Plant Physiol, 2010. **154**(2): p. 483-6.
34. Sticklen, M.B., *Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol*. Nat Rev Genet, 2008. **9**(6): p. 433-43.

35. Gibson, L.J., *The hierarchical structure and mechanics of plant materials*. J R Soc Interface, 2012. **9**(76): p. 2749-66.
36. Scheller, H.V. and P. Ulvskov, *Hemicelluloses*. Annu Rev Plant Biol, 2010. **61**: p. 263-89.
37. Braidwood, L., C. Breuer, and K. Sugimoto, *My body is a cage: mechanisms and modulation of plant cell growth*. New Phytol, 2014. **201**(2): p. 388-402.
38. Mohnen, D., *Pectin structure and biosynthesis*. Curr Opin Plant Biol, 2008. **11**(3): p. 266-77.
39. Martone, P.T., et al., *Discovery of lignin in seaweed reveals convergent evolution of cell-wall architecture*. Curr Biol, 2009. **19**(2): p. 169-75.
40. Boerjan, W., J. Ralph, and M. Baucher, *Lignin biosynthesis*. Annu Rev Plant Biol, 2003. **54**: p. 519-46.
41. Hofte, H. and A. Voxeur, *Plant cell walls*. Curr Biol, 2017. **27**(17): p. R865-R870.
42. Mueller, S.C. and R.M. Brown, Jr., *Evidence for an intramembrane component associated with a cellulose microfibril-synthesizing complex in higher plants*. J Cell Biol, 1980. **84**(2): p. 315-26.
43. Kimura, S., et al., *Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant Vigna angularis*. Plant Cell, 1999. **11**(11): p. 2075-2085.
44. Guerriero, G., J. Fugelstad, and V. Bulone, *What do we really know about cellulose biosynthesis in higher plants?* J Integr Plant Biol, 2010. **52**(2): p. 161-75.
45. Desprez, T., et al., *Organization of cellulose synthase complexes involved in primary cell wall synthesis in Arabidopsis thaliana*. Proc Natl Acad Sci U S A, 2007. **104**(39): p. 15572-7.
46. Newman, R.H., S.J. Hill, and P.J. Harris, *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, 2013. **163**(4): p. 1558-67.
47. Hill, J.L., Jr., M.B. Hammudi, and M. Tien, *The Arabidopsis cellulose synthase complex: a proposed hexamer of CESA trimers in an equimolar stoichiometry*. Plant Cell, 2014. **26**(12): p. 4834-42.
48. Speicher, T.L., P.Z. Li, and I.S. Wallace, *Phosphoregulation of the Plant Cellulose Synthase Complex and Cellulose Synthase-Like Proteins*. Plants (Basel), 2018. **7**(3).
49. Pear, J.R., et al., *Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase*. Proc Natl Acad Sci U S A, 1996. **93**(22): p. 12637-42.
50. Popper, Z.A., et al., *Evolution and diversity of plant cell walls: from algae to flowering plants*. Annu Rev Plant Biol, 2011. **62**: p. 567-90.
51. Gonneau, M., et al., *Catalytic subunit stoichiometry within the cellulose synthase complex*. Plant Physiol, 2014. **166**(4): p. 1709-12.

52. Persson, S., et al., *Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in Arabidopsis*. Proc Natl Acad Sci U S A, 2007. **104**(39): p. 15566-71.
53. Sethaphong, L., et al., *Tertiary model of a plant cellulose synthase*. Proc Natl Acad Sci U S A, 2013. **110**(18): p. 7512-7.
54. Jacob-Wilk, D., et al., *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 U S A, 2006. **103**(32): p. 12191-6.
55. Kurek, I., et al., *Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domains*. Proc Natl Acad Sci U S A, 2002. **99**(17): p. 11109-14.
56. Purushotham, P., et al., *A single heterologously expressed plant cellulose synthase isoform is sufficient for cellulose microfibril formation in vitro*. Proc Natl Acad Sci U S A, 2016. **113**(40): p. 11360-11365.
57. Cho, S.H., et al., *Synthesis and Self-Assembly of Cellulose Microfibrils from Reconstituted Cellulose Synthase*. Plant Physiol, 2017. **175**(1): p. 146-156.
58. Zuo, J., et al., *KORRIGAN, an Arabidopsis endo-1,4-beta-glucanase, localizes to the cell plate by polarized targeting and is essential for cytokinesis*. Plant Cell, 2000. **12**(7): p. 1137-52.
59. Nicol, F., et al., *A plasma membrane-bound putative endo-1,4-beta-D-glucanase is required for normal wall assembly and cell elongation in Arabidopsis*. EMBO J, 1998. **17**(19): p. 5563-76.
60. Robert, S., et al., *An Arabidopsis endo-1,4-beta-D-glucanase involved in cellulose synthesis undergoes regulated intracellular cycling*. Plant Cell, 2005. **17**(12): p. 3378-89.
61. Mansoori, N., et al., *KORRIGAN1 interacts specifically with integral components of the cellulose synthase machinery*. PLoS One, 2014. **9**(11): p. e112387.
62. Vain, T., et al., *The Cellulase KORRIGAN Is Part of the Cellulose Synthase Complex*. Plant Physiol, 2014. **165**(4): p. 1521-1532.
63. Lei, L., et al., *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, 2014. **26**(6): p. 2601-2616.
64. Yu, L., et al., *PtrKOR1 is required for secondary cell wall cellulose biosynthesis in Populus*. Tree Physiol, 2014. **34**(11): p. 1289-300.
65. Schindelman, G., et al., *COBRA encodes a putative GPI-anchored protein, which is polarly localized and necessary for oriented cell expansion in Arabidopsis*. Genes Dev, 2001. **15**(9): p. 1115-27.
66. Roudier, F., et al., *COBRA, an Arabidopsis extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation*. Plant Cell, 2005. **17**(6): p. 1749-63.
67. Ko, J.H., et al., *Loss of function of COBRA, a determinant of oriented cell expansion, invokes cellular defence responses in Arabidopsis thaliana*. J Exp Bot, 2006. **57**(12): p. 2923-36.

68. Liu, L., et al., *Brittle Culm1, a COBRA-like protein, functions in cellulose assembly through binding cellulose microfibrils*. PLoS Genet, 2013. **9**(8): p. e1003704.
69. Gu, Y., et al., *Identification of a cellulose synthase-associated protein required for cellulose biosynthesis*. Proc Natl Acad Sci U S A, 2010. **107**(29): p. 12866-71.
70. Lei, L., et al., *CELLULOSE SYNTHASE INTERACTIVE1 Is Required for Fast Recycling of Cellulose Synthase Complexes to the Plasma Membrane in Arabidopsis*. Plant Cell, 2015. **27**(10): p. 2926-40.
71. Li, S., et al., *Cellulose synthase interactive protein 1 (CSI1) links microtubules and cellulose synthase complexes*. Proc Natl Acad Sci U S A, 2012. **109**(1): p. 185-90.
72. Lei, L., et al., *Cellulose synthase INTERACTIVE3 regulates cellulose biosynthesis in both a microtubule-dependent and microtubule-independent manner in Arabidopsis*. Plant Cell, 2013. **25**(12): p. 4912-23.
73. Endler, A., et al., *A Mechanism for Sustained Cellulose Synthesis during Salt Stress*. Cell, 2015. **162**(6): p. 1353-64.
74. Kesten, C., et al., *The companion of cellulose synthase 1 confers salt tolerance through a Tau-like mechanism in plants*. Nat Commun, 2019. **10**(1): p. 857.
75. Endler, A., et al., *The cellulose synthase companion proteins act non-redundantly with CELLULOSE SYNTHASE INTERACTING1/POM2 and CELLULOSE SYNTHASE 6*. Plant Signal Behav, 2016. **11**(4): p. e1135281.
76. Paredez, A.R., C.R. Somerville, and D.W. Ehrhardt, *Visualization of cellulose synthase demonstrates functional association with microtubules*. Science, 2006. **312**(5779): p. 1491-5.
77. Gutierrez, R., et al., *Arabidopsis cortical microtubules position cellulose synthase delivery to the plasma membrane and interact with cellulose synthase trafficking compartments*. Nat Cell Biol, 2009. **11**(7): p. 797-806.
78. Zhang, Y., et al., *Golgi-localized STELLO proteins regulate the assembly and trafficking of cellulose synthase complexes in Arabidopsis*. Nat Commun, 2016. **7**: p. 11656.
79. Polko, J.K., et al., *SHOU4 Proteins Regulate Trafficking of Cellulose Synthase Complexes to the Plasma Membrane*. Curr Biol, 2018.
80. Polko, J.K. and J.J. Kieber, *The Regulation of Cellulose Biosynthesis in Plants*. Plant Cell, 2019. **31**(2): p. 282-296.
81. Chen, S., D.W. Ehrhardt, and C.R. Somerville, *Mutations of cellulose synthase (CESA1) phosphorylation sites modulate anisotropic cell expansion and bidirectional mobility of cellulose synthase*. Proc Natl Acad Sci U S A, 2010. **107**(40): p. 17188-93.
82. Chen, S., et al., *Anisotropic Cell Expansion Is Affected through the Bidirectional Mobility of Cellulose Synthase Complexes and Phosphorylation at Two Critical Residues on CESA3*. Plant Physiol, 2016. **171**(1): p. 242-50.
83. Sanchez-Rodriguez, C., et al., *BRASSINOSTEROID INSENSITIVE2 negatively regulates cellulose synthesis in Arabidopsis by phosphorylating cellulose synthase 1*. Proc Natl Acad Sci U S A, 2017.

84. Kumar, M., et al., *S-Acylation of the cellulose synthase complex is essential for its plasma membrane localization*. Science, 2016. **353**(6295): p. 166-9.
85. Taylor, N.G., et al., *Interactions among three distinct Cesa proteins essential for cellulose synthesis*. Proc Natl Acad Sci U S A, 2003. **100**(3): p. 1450-5.
86. Carroll, A., et al., *Complexes with mixed primary and secondary cellulose synthases are functional in Arabidopsis plants*. Plant Physiol, 2012. **160**(2): p. 726-37.
87. Olek, A.T., et al., *The structure of the catalytic domain of a plant cellulose synthase and its assembly into dimers*. Plant Cell, 2014. **26**(7): p. 2996-3009.
88. Vandavasi, V.G., et al., *A Structural Study of CESA1 Catalytic Domain of Arabidopsis Cellulose Synthesis Complex: Evidence for CESA Trimers*. Plant Physiol, 2016. **170**(1): p. 123-35.
89. Wang, J., et al., *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, 2006. **142**(2): p. 685-95.
90. Timmers, J., et al., *Interactions between membrane-bound cellulose synthases involved in the synthesis of the secondary cell wall*. FEBS Lett, 2009. **583**(6): p. 978-82.
91. Tateno, M., C. Brabham, and S. DeBolt, *Cellulose biosynthesis inhibitors - a multifunctional toolbox*. J Exp Bot, 2016. **67**(2): p. 533-42.
92. Xia, Y., et al., *Experimental approaches to study plant cell walls during plant-microbe interactions*. Front Plant Sci, 2014. **5**: p. 540.
93. DeBolt, S., et al., *Nonmotile cellulose synthase subunits repeatedly accumulate within localized regions at the plasma membrane in Arabidopsis hypocotyl cells following 2,6-dichlorobenzonitrile treatment*. Plant Physiol, 2007. **145**(2): p. 334-8.
94. DeBolt, S., et al., *Morlin, an inhibitor of cortical microtubule dynamics and cellulose synthase movement*. Proc Natl Acad Sci U S A, 2007. **104**(14): p. 5854-9.
95. Nixon, B.T., et al., *Comparative Structural and Computational Analysis Supports Eighteen Cellulose Synthases in the Plant Cellulose Synthesis Complex*. Sci Rep, 2016. **6**: p. 28696.
96. Rushton, P.S., et al., *Rice Cellulose SynthaseA8 Plant-Conserved Region Is a Coiled-Coil at the Catalytic Core Entrance*. Plant Physiol, 2017. **173**(1): p. 482-494.
97. Kabsch, W., *Xds*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.
98. Winn, M.D., et al., *Overview of the CCP4 suite and current developments*. Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 235-42.
99. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
100. Emsley, P., et al., *Features and development of Coot*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.

101. Hauer, F., et al., *GraDeR: Membrane Protein Complex Preparation for Single-Particle Cryo-EM*. Structure, 2015. **23**(9): p. 1769-1775.
102. Tang, G., et al., *EMAN2: an extensible image processing suite for electron microscopy*. J Struct Biol, 2007. **157**(1): p. 38-46.
103. Zivanov, J., et al., *New tools for automated high-resolution cryo-EM structure determination in RELION-3*. Elife, 2018. **7**.
104. Derewenda, Z.S., *Application of protein engineering to enhance crystallizability and improve crystal properties*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 5): p. 604-15.
105. Goldschmidt, L., et al., *Toward rational protein crystallization: A Web server for the design of crystallizable protein variants*. Protein Sci, 2007. **16**(8): p. 1569-76.
106. Walter, T.S., et al., *Lysine methylation as a routine rescue strategy for protein crystallization*. Structure, 2006. **14**(11): p. 1617-22.
107. Chayen, N.E. and E. Saridakis, *Protein crystallization: from purified protein to diffraction-quality crystal*. Nat Methods, 2008. **5**(2): p. 147-53.
108. McGuffin, L.J., K. Bryson, and D.T. Jones, *The PSIPRED protein structure prediction server*. Bioinformatics, 2000. **16**(4): p. 404-5.
109. Holcomb, J., et al., *Protein crystallization: Eluding the bottleneck of X-ray crystallography*. AIMS Biophys, 2017. **4**(4): p. 557-575.
110. Scavuzzo-Duggan, T.R., et al., *Cellulose synthase 'class specific regions' are intrinsically disordered and functionally undifferentiated*. J Integr Plant Biol, 2018. **60**(6): p. 481-497.
111. Baker, J.A., et al., *Charged residues next to transmembrane regions revisited: "Positive-inside rule" is complemented by the "negative inside depletion/outside enrichment rule"*. BMC Biol, 2017. **15**(1): p. 66.
112. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state*. J Mol Biol, 2007. **372**(3): p. 774-97.
113. Joseph, P.R., et al., *Proline substitution of dimer interface beta-strand residues as a strategy for the design of functional monomeric proteins*. Biophys J, 2013. **105**(6): p. 1491-501.
114. Holm, L., *Benchmarking Fold Detection by DaliLite v.5*. Bioinformatics, 2019.
115. Baek, M., et al., *GalaxyHomomer: a web server for protein homo-oligomer structure prediction from a monomer sequence or structure*. Nucleic Acids Res, 2017. **45**(W1): p. W320-W324.
116. Williamson, R.E., et al., *Morphology of rsw1, a cellulose-deficient mutant of Arabidopsis thaliana*. Protoplasma, 2001. **215**(1-4): p. 116-27.
117. Turner, S.R. and C.R. Somerville, *Collapsed xylem phenotype of Arabidopsis identifies mutants deficient in cellulose deposition in the secondary cell wall*. Plant Cell, 1997. **9**(5): p. 689-701.
118. Arioli, T., et al., *Molecular analysis of cellulose biosynthesis in Arabidopsis*. Science, 1998. **279**(5351): p. 717-20.
119. Taylor, N.G., et al., *The irregular xylem3 locus of Arabidopsis encodes a cellulose synthase required for secondary cell wall synthesis*. Plant Cell, 1999. **11**(5): p. 769-80.

120. Fagard, M., et al., *PROCUSTE1 encodes a cellulose synthase required for normal cell elongation specifically in roots and dark-grown hypocotyls of Arabidopsis*. Plant Cell, 2000. **12**(12): p. 2409-2424.
121. Taylor, N.G., S. Laurie, and S.R. Turner, *Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in Arabidopsis*. Plant Cell, 2000. **12**(12): p. 2529-2540.
122. Scheible, W.R., et al., *Modifications of cellulose synthase confer resistance to isoxaben and thiazolidinone herbicides in Arabidopsis lxr1 mutants*. Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10079-84.
123. Vergara, C.E. and N.C. Carpita, *Beta-D-glycan synthases and the Cesa gene family: lessons to be learned from the mixed-linkage (1-->3),(1-->4)beta-D-glucan synthase*. Plant Mol Biol, 2001. **47**(1-2): p. 145-60.
124. Beeckman, T., et al., *Genetic complexity of cellulose synthase a gene function in Arabidopsis embryogenesis*. Plant Physiol, 2002. **130**(4): p. 1883-93.
125. Desprez, T., et al., *Resistance against herbicide isoxaben and cellulose deficiency caused by distinct mutations in same cellulose synthase isoform CESA6*. Plant Physiol, 2002. **128**(2): p. 482-90.
126. Ellis, C., et al., *The Arabidopsis mutant cev1 links cell wall signaling to jasmonate and ethylene responses*. Plant Cell, 2002. **14**(7): p. 1557-66.
127. Gillmor, C.S., et al., *Alpha-glucosidase I is required for cellulose biosynthesis and morphogenesis in Arabidopsis*. J Cell Biol, 2002. **156**(6): p. 1003-13.
128. Cano-Delgado, A., et al., *Reduced cellulose synthesis invokes lignification and defense responses in Arabidopsis thaliana*. Plant J, 2003. **34**(3): p. 351-62.
129. Zhong, R., et al., *Expression of a mutant form of cellulose synthase AtCesA7 causes dominant negative effect on cellulose biosynthesis*. Plant Physiol, 2003. **132**(2): p. 786-95.
130. Chen, Z., et al., *Disruption of the cellulose synthase gene, AtCesA8/IRX1, enhances drought and osmotic stress tolerance in Arabidopsis*. Plant J, 2005. **43**(2): p. 273-83.
131. Bosca, S., et al., *Interactions between MUR10/CesA7-dependent secondary cellulose biosynthesis and primary cell wall structure*. Plant Physiol, 2006. **142**(4): p. 1353-63.
132. Zhang, B., et al., *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, 2009. **71**(4-5): p. 509-24.
133. Harris, D.M., et al., *Cellulose microfibril crystallinity is reduced by mutating C-terminal transmembrane region residues CESA1A903V and CESA3T942I of cellulose synthase*. Proc Natl Acad Sci U S A, 2012. **109**(11): p. 4098-103.
134. Fujita, M., et al., *The anisotropy1 D604N mutation in the Arabidopsis cellulose synthase1 catalytic domain reduces cell wall crystallinity and the velocity of cellulose synthase complexes*. Plant Physiol, 2013. **162**(1): p. 74-85.

135. Li, F., et al., *A novel FC17/CESA4 mutation causes increased biomass saccharification and lodging resistance by remodeling cell wall in rice*. Biotechnol Biofuels, 2018. **11**: p. 298.
136. Sgro, G.G. and T.R.D. Costa, *Cryo-EM Grid Preparation of Membrane Protein Samples for Single Particle Analysis*. Front Mol Biosci, 2018. **5**: p. 74.
137. Kastner, B., et al., *GraFix: sample preparation for single-particle electron cryomicroscopy*. Nat Methods, 2008. **5**(1): p. 53-5.
138. Jonic, S., *Cryo-electron Microscopy Analysis of Structurally Heterogeneous Macromolecular Complexes*. Comput Struct Biotechnol J, 2016. **14**: p. 385-390.
139. Cheng, Y., *Membrane protein structural biology in the era of single particle cryo-EM*. Curr Opin Struct Biol, 2018. **52**: p. 58-63.
140. Merk, A., et al., *Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery*. Cell, 2016. **165**(7): p. 1698-1707.
141. Herzik, M.A., Jr., M. Wu, and G.C. Lander, *High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM*. Nat Commun, 2019. **10**(1): p. 1032.
142. Khoshouei, M., et al., *Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate*. Nat Commun, 2017. **8**: p. 16099.
143. Fan, X., et al., *Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Å resolution*. Nat Commun, 2019. **10**(1): p. 2386.
144. Liu, Y., D.T. Huynh, and T.O. Yeates, *A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold*. Nat Commun, 2019. **10**(1): p. 1864.
145. Danev, R., et al., *Volta potential phase plate for in-focus phase contrast transmission electron microscopy*. Proc Natl Acad Sci U S A, 2014. **111**(44): p. 15635-40.
146. Yang, J., et al., *The I-TASSER Suite: protein structure and function prediction*. Nat Methods, 2015. **12**(1): p. 7-8.
147. Scavuzzo-Duggan, T.R., et al., *Cellulose synthase 'class specific regions' are intrinsically disordered and functionally undifferentiated*. J Integr Plant Biol, 2018.
148. Kumar, M., I. Atanassov, and S. Turner, *Functional Analysis of Cellulose Synthase (CESA) Protein Class Specificity*. Plant Physiol, 2017. **173**(2): p. 970-983.
149. Cheng, P.N., J.D. Pham, and J.S. Nowick, *The supramolecular chemistry of beta-sheets*. J Am Chem Soc, 2013. **135**(15): p. 5477-92.
150. Jeffrey, P.D., S. Gorina, and N.P. Pavletich, *Crystal structure of the tetramerization domain of the p53 tumor suppressor at 1.7 Å*. Science, 1995. **267**(5203): p. 1498-502.
151. Finder, V.H. and R. Glockshuber, *Amyloid-beta aggregation*. Neurodegener Dis, 2007. **4**(1): p. 13-27.
152. Passmore, L.A. and C.J. Russo, *Specimen Preparation for High-Resolution Cryo-EM*. Methods Enzymol, 2016. **579**: p. 51-86.
153. Denisov, I.G. and S.G. Sligar, *Nanodiscs for structural and functional studies of membrane proteins*. Nat Struct Mol Biol, 2016. **23**(6): p. 481-6.

154. Wickstead, B. and K. Gull, *The evolution of the cytoskeleton*. J Cell Biol, 2011. **194**(4): p. 513-25.
155. Desai, A. and T.J. Mitchison, *Microtubule polymerization dynamics*. Annu Rev Cell Dev Biol, 1997. **13**: p. 83-117.
156. Herrmann, H., et al., *Intermediate filaments: from cell architecture to nanomechanics*. Nat Rev Mol Cell Biol, 2007. **8**(7): p. 562-73.
157. Fletcher, D.A. and R.D. Mullins, *Cell mechanics and the cytoskeleton*. Nature, 2010. **463**(7280): p. 485-92.
158. Lutkenhaus, J.F., H. Wolf-Watz, and W.D. Donachie, *Organization of genes in the ftsA-envA region of the Escherichia coli genetic map and identification of a new fts locus (ftsZ)*. J Bacteriol, 1980. **142**(2): p. 615-20.
159. Lowe, J. and L.A. Amos, *Crystal structure of the bacterial cell-division protein FtsZ*. Nature, 1998. **391**(6663): p. 203-6.
160. Jenkins, C., et al., *Genes for the cytoskeletal protein tubulin in the bacterial genus Prosthecobacter*. Proc Natl Acad Sci U S A, 2002. **99**(26): p. 17049-54.
161. Larsen, R.A., et al., *Treadmilling of a prokaryotic tubulin-like protein, TubZ, required for plasmid stability in Bacillus thuringiensis*. Genes Dev, 2007. **21**(11): p. 1340-52.
162. Tinsley, E. and S.A. Khan, *A novel FtsZ-like protein is involved in replication of the anthrax toxin-encoding pXO1 plasmid in Bacillus anthracis*. J Bacteriol, 2006. **188**(8): p. 2829-35.
163. Bork, P., C. Sander, and A. Valencia, *An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins*. Proc Natl Acad Sci U S A, 1992. **89**(16): p. 7290-4.
164. Jones, L.J., R. Carballido-Lopez, and J. Errington, *Control of cell shape in bacteria: helical, actin-like filaments in Bacillus subtilis*. Cell, 2001. **104**(6): p. 913-22.
165. Derman, A.I., et al., *Phylogenetic analysis identifies many uncharacterized actin-like proteins (Alps) in bacteria: regulated polymerization, dynamic instability and treadmilling in Alp7A*. Mol Microbiol, 2009. **73**(4): p. 534-52.
166. Ausmees, N., J.R. Kuhn, and C. Jacobs-Wagner, *The bacterial cytoskeleton: an intermediate filament-like function in cell shape*. Cell, 2003. **115**(6): p. 705-13.
167. Bagchi, S., et al., *Intermediate filament-like proteins in bacteria and a cytoskeletal function in Streptomyces*. Mol Microbiol, 2008. **70**(4): p. 1037-50.
168. Pollard, T.D., *Actin and Actin-Binding Proteins*. Cold Spring Harb Perspect Biol, 2016. **8**(8).
169. Begg, D.A., R. Rodewald, and L.I. Rebhun, *The visualization of actin filament polarity in thin sections. Evidence for the uniform polarity of membrane-associated filaments*. J Cell Biol, 1978. **79**(3): p. 846-52.
170. Moens, P.D. and L.A. Bagatolli, *Profilin binding to sub-micellar concentrations of phosphatidylinositol (4,5) bisphosphate and*

- phosphatidylinositol (3,4,5) trisphosphate*. Biochim Biophys Acta, 2007. **1768**(3): p. 439-49.
171. Akil, C. and R.C. Robinson, *Genomes of Asgard archaea encode profilins that regulate actin*. Nature, 2018.
 172. Sun, H., et al., *Profilin Negatively Regulates Formin-Mediated Actin Assembly to Modulate PAMP-Triggered Plant Immunity*. Curr Biol, 2018. **28**(12): p. 1882-1895 e7.
 173. Huang, S., et al., *The Arabidopsis profilin gene family. Evidence for an ancient split between constitutive and pollen-specific profilin genes*. Plant Physiol, 1996. **111**(1): p. 115-26.
 174. Kandasamy, M.K., E.C. McKinney, and R.B. Meagher, *Plant profilin isoforms are distinctly regulated in vegetative and reproductive tissues*. Cell Motil Cytoskeleton, 2002. **52**(1): p. 22-32.
 175. Romero, S., et al., *How ATP hydrolysis controls filament assembly from profilin-actin: implication for formin processivity*. J Biol Chem, 2007. **282**(11): p. 8435-45.
 176. Miao, Y., et al., *Phospho-regulation of intrinsically disordered proteins for actin assembly and endocytosis*. FEBS J, 2018.
 177. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega*. Mol Syst Biol, 2011. **7**: p. 539.
 178. Robert, X. and P. Gouet, *Deciphering key features in protein structures with the new ENDscript server*. Nucleic Acids Res, 2014. **42**(Web Server issue): p. W320-4.
 179. Minor, W., et al., *HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes*. Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 8): p. 859-66.
 180. Krishnan, K. and P.D.J. Moens, *Structure and functions of profilins*. Biophys Rev, 2009. **1**(2): p. 71-81.
 181. Michelot, A., et al., *The formin homology 1 domain modulates the actin nucleation and bundling activity of Arabidopsis FORMIN1*. Plant Cell, 2005. **17**(8): p. 2296-313.
 182. Thorn, K.S., et al., *The crystal structure of a major allergen from plants*. Structure, 1997. **5**(1): p. 19-32.
 183. Offermann, L.R., et al., *Structural, Functional, and Immunological Characterization of Profilin Panallergens Amb a 8, Art v 4, and Bet v 2*. J Biol Chem, 2016. **291**(30): p. 15447-59.
 184. Mahoney, N.M., et al., *Profilin binds proline-rich ligands in two distinct amide backbone orientations*. Nat Struct Biol, 1999. **6**(7): p. 666-71.
 185. Ferron, F., et al., *Structural basis for the recruitment of profilin-actin complexes during filament elongation by Ena/VASP*. EMBO J, 2007. **26**(21): p. 4597-606.
 186. Baek, K., et al., *Modulation of actin structure and function by phosphorylation of Tyr-53 and profilin binding*. Proc Natl Acad Sci U S A, 2008. **105**(33): p. 11748-53.
 187. Akil, C. and R.C. Robinson, *Genomes of Asgard archaea encode profilins that regulate actin*. Nature, 2018. **562**(7727): p. 439-443.

188. Frappier, V., M. Duran, and A.E. Keating, *PixelDB: Protein-peptide complexes annotated with structural conservation of the peptide binding mode*. Protein Sci, 2018. **27**(1): p. 276-285.

6. Appendix

1. Genes encoded by the bacterial cellulose synthase operon

	Protein	Length	Operon type	UniProt	Structure	Functional annotation
Core genes	BcsA	750–870	I, II, III, IV	P37653	4HG6	Cellulose synthase catalytic subunit A
	BcsB	770–800	I, II, III	P37652	4HG6	Cellulose synthase subunit B c (periplasmic)
	BcsC	1150	I, II	P37650	5XW7	Cellulose synthase subunit C, spans periplasm and outer membrane
	BcsD	155	I	P19451	3A8E	Cellulose synthase subunit D (periplasmic)
	BcsE	500–750	II	Q8ZLB5		Cellulose synthase cytoplasmic subunit E, binds c-di-GMP [42] d
	BcsF	60	II	Q7CPI8		Membrane-anchored subunit (1 TM segment)
	BcsG	550	II	Q7CPI7	50LT	Contains 4 TM segments and a periplasmic AlkP domain
	BcsQ	250	I, II	Q8ZLB6		ParA/MinD-related NTPase, may localize BCS at the cell pole e
	BcsR	65	II	P0ADJ3		Likely regulatory subunit
	BcsZ	370	I, II, III	Q8ZLB7	1WZZ, 3QXQ	Endo- β -1,4-glucanase (cellulase), periplasmic
lineage specific genes	BcsH	340	Ia	Q76KK0		Cellulose-complementing protein A, specific for acetobacteria b
	BglX	730	Ia	Q8RTY7		β -Glucosidase, glycosyl hydrolase family 3, secreted
	BcsO	200	Ib	D4IFF8		Unknown, specific for enterobacterial type Ib <i>bcs</i> operons
	BcsP	180	Ic	Q2KWQ2		Unknown, specific for β -proteobacterial type Ic <i>bcs</i> operons
	BcsK	750	III	Q8U4V1		Likely interacts with peptidoglycan
	BcsL	550	IIIa	Q7CS43		Acetyltransferase
	BcsM	380	IIIa	Q7CS44		Zn-dependent amidohydrolase, may deacylate modified glucose residues
	BcsN	330	III	B9K194		Unknown, periplasmic, membrane-anchored (1 TM), specific for α -proteobacteria
	BcsS	250	III	A9W2G1		Unknown, secreted, specific for α -proteobacteria
	BcsT	620	III	A9CEZ9		Membrane protein (8 TM) with a cytoplasmic glycosyltransferase domain
	BcsU	170	III	Q7CS50		Unknown, secreted, specific for Rhizobiales
	BcsV	320	III	Q7CS49		Beta-mannanase, secreted, specific for Rhizobiales
	BcsW	440	IV	Q8YQR4		Unknown, predicted periplasmic
	BcsX	260	–	Q9WX69		Probable cellulose deacylase c
	BcsY	380	–	Q9WX70		Membrane protein (10 TM), probable cellulose acylase

2. Constructs summary of CesA for X-ray crystallography study

Constructs	Organism	Tag	Amino Acid Coverage	Expression	Purification	Crystallization	Resolution
AtCesA1CatD	<i>A. thaliana</i>	His-Tev-	333-807	Y	Y		
AtCesA1CatDLD0	<i>A. thaliana</i>	His-Tev-	333-807(Δ647-691)	Y	Y		
AtCesA3VC001	<i>A. thaliana</i>	His-Tev-	315-843	Y	Insoluble		
AtCesA3VC002	<i>A. thaliana</i>	His-Tev-	337-843	Y	Insoluble		
AtCesA3VC003	<i>A. thaliana</i>	His-Tev-	352-843	Y	Insoluble		
AtCesA3VC004	<i>A. thaliana</i>	His-Tev-	369-843	Y	Insoluble		
AtCesA3VC005	<i>A. thaliana</i>	His-Tev-	315-793	Y	Insoluble		
AtCesA3VC006	<i>A. thaliana</i>	His-Tev-	337-793	Y	Insoluble		
AtCesA3VC007	<i>A. thaliana</i>	His-Tev-	352-793	Y	Insoluble		
AtCesA3VC008	<i>A. thaliana</i>	His-Tev-	369-793	Y	Insoluble		
AtCesA3VC009	<i>A. thaliana</i>	His-Tev-	325-828	Y	Insoluble		
AtCesA3VC010	<i>A. thaliana</i>	-His	325-828	Y	Insoluble		
AtCesA3VC011	<i>A. thaliana</i>	His-Tev-	317-792	Y	Y		
AtCesA3VC012	<i>A. thaliana</i>	His-Tev-	315-792	Y	Y		
AtCesA3VC013	<i>A. thaliana</i>	His-Tev-	337-792	Y	Insoluble		
AtCesA3VC014	<i>A. thaliana</i>	His-Tev-	352-792	Y	Insoluble		
AtCesA3VC015	<i>A. thaliana</i>	His-Tev-	369-792	Y	Insoluble		
AtCesA3VC019	<i>A. thaliana</i>	His-Tev-	317-810	Y	Insoluble		
AtCesA3VC020	<i>A. thaliana</i>	His-Tev-	317-817	Y	Insoluble		
AtCesA3VC021	<i>A. thaliana</i>	His-Tev-	317-823	Y	Insoluble		
AtCesA3VC022	<i>A. thaliana</i>	His-Tev-	317-827	Y	Insoluble		
AtCesA3VC023	<i>A. thaliana</i>	His-Tev-	317-833	Y	Insoluble		
AtCesA3VC024	<i>A. thaliana</i>	His-Tev-	317-843	Y	Insoluble		
AtCesA3CatDLD0	<i>A. thaliana</i>	His-Tev-	317-792(Δ630-673)	Y	Y	Y	2.6Å
AtCesA3(317-792)	<i>A. thaliana</i>	His-Tev-	317-792(Δ630-698::GGGGS)	Y	Y	Y	2.1Å
AtCesA3(317-810)	<i>A. thaliana</i>	-His	317-810(Δ630-698::GGGGS)	Y	Y	Y	2.95Å
AtCesA6CatD	<i>A. thaliana</i>	His-Tev-	334-812	Y	Aggregation		
AtCesA6CatDLD0	<i>A. thaliana</i>	His-Tev-	334-812(Δ637-654)	Y	Precipitate		
AtCesA7CatD	<i>A. thaliana</i>	His-Tev-	296-753	N	N		
AtCesA8CatD	<i>A. thaliana</i>	His-Tev-	242-710	Y	Y		
AtCesA8CatDLD0	<i>A. thaliana</i>	His-Tev-	242-710(Δ555-579)	Y	Y		
AtCesA10CatD	<i>A. thaliana</i>	His-Tev-	320-794	Y	Y		
AtCesA10CatDLD0	<i>A. thaliana</i>	His-Tev-	320-794(Δ628-671)	Y	Y		
HvCesA1CatD	<i>H. Vulgare</i>	His-Tev-	112-597	Y	Y		
HvCesA1CatDLD0	<i>H. Vulgare</i>	His-Tev-	112-597(Δ425-467)	Y	Y		
HvCesA2CatD	<i>H. Vulgare</i>	His-Tev-	335-819	Y	low yield		
HvCesA3CatD	<i>H. Vulgare</i>	His-Tev-	328-790	Y	Y		
HvCesA3CatDLD0	<i>H. Vulgare</i>	His-Tev-	328-790(Δ641-677)	Y	Y		
HvCesA4CatD	<i>H. Vulgare</i>	His-Tev-	261-770	Y	low yield		
HvCesA5CatD	<i>H. Vulgare</i>	His-Tev-	243-712	Y	low yield		
HvCesA6CatD	<i>H. Vulgare</i>	His-Tev-	314-786	Y	Y		
HvCesA6CatDLD0	<i>H. Vulgare</i>	His-Tev-	314-786(Δ627-670)	Y	Y		
HvCesA8CatD	<i>H. Vulgare</i>	His-Tev-	326-780	Y	Y		
HvCesA8CatD-CSR1-LD0	<i>H. Vulgare</i>	His-Tev-	326-780(Δ639-670)	Y	Y		
HvCesA8CatD-CSR1-LD8	<i>H. Vulgare</i>	His-Tev-	326-780(Δ639-670::((GGGGS)G3))	Y	Y		
HvCesA8CatD-CSR1-LD16	<i>H. Vulgare</i>	His-Tev-	326-780(Δ639-670::((GGGGS)3G))	Y	Y		
HvCesA8CatD-CSR1-LD18	<i>H. Vulgare</i>	His-Tev-	326-780(Δ639-670::GG(GGGGS)3G)	Y	Y		
HvCesA8CatD-CSR2-LD4	<i>H. Vulgare</i>	His-Tev-	326-780(Δ627-662::GGGS)	Y	Insoluble		
HvCesA8CatD-CSR2-LD6	<i>H. Vulgare</i>	His-Tev-	326-780(Δ627-662::GGGSGG)	Y	Insoluble		
HvCesA8CatD-CSR2-LD10	<i>H. Vulgare</i>	His-Tev-	326-780(Δ627-662::(GGGGS)2)	Y	Insoluble		
OsCesA4CatD	<i>O. Sativa</i>	His-Tev-	247-715	N	N		
OsCesA8CatD	<i>O. Sativa</i>	His-Tev-	334-808	Y	low yield		
OsCesA8CatDLD0	<i>O. Sativa</i>	His-Tev-	334-808(Δ647-689)	Y	Y	Y	

3. Data collection and refinement statistics of structures in Chapter 1

	AtCesA3 ³¹⁷⁻⁷⁹²	UDP-AtCesA3 ³¹⁷⁻⁷⁹²	UDP-Glc-AtCesA3 ³¹⁷⁻⁷⁹²	AtCesA3 ³¹⁷⁻⁸¹⁰
Data collection				
PDB ID				
Wavelength	0.9537	0.9537	0.9537	
Resolution range	43.23 - 2.43 (2.60 - 2.43)	43.47 - 2.1 (2.17 - 2.1)	43.72 - 2.10 (2.26 - 2.10)	19.8 - 2.98 (3.06 - 2.98)
Space group	P 1 2 ₁ 1	P 1 2 ₁ 1	P 1 2 ₁ 1	P 2 ₁ 2 ₁ 2 ₁
Unit cell				
a, b, c (Å)	54.62, 147.03, 92.51	54.38, 148.29, 93.75	54.46, 148.42, 93.99	49.14, 58.43, 321.89
α, β, γ (°)	90, 102.05, 90	90, 102.97, 90	90, 103.22, 90	90, 90, 90
Total reflections	376895(69396)	592151(95029)	576070(105404)	539158(43739)
Unique reflections	53351 (9633)	82752 (8168)	81610 (14633)	20005(1578)
Multiplicity	7.07(7.15)	7.16(7.29)	7.06(7.20)	26.95(27.71)
Completeness (%)	99.48 (99.41)	98.27 (97.16)	98.78 (97.53)	99.52 (99.95)
Mean I/sigma(I)	9.43(1.15)	12.34(1.12)	10.17(1.17)	6.55(1.04)
Wilson B-factor	59.84	47.69	45.94	66.64
R-meas	0.12(1.58)	0.09(1.74)	0.12(1.75)	0.80(5.29)
CC1/2	99.8 (65.5)	99.9(62.3)	99.8(65.6)	99.3(50.0)
Refinement				
Resolution range	43.23 - 2.43	43.47 - 2.1	43.72 - 2.10	19.8 - 2.98
R-work	0.2275 (0.4777)	0.2084 (0.4659)	0.2088 (0.4579)	0.2526 (0.3229)
R-free	0.2596 (0.4933)	0.2309 (0.4883)	0.2262 (0.4759)	0.3121 (0.3282)
Number of non-hydrogen atoms	5571	5810	5849	5816
macromolecules	5552	5530	5539	5816
ligands		54	76	
solvent	19	226	234	
Protein residues	705	702	703	738
RMS(bonds)	0.01	0.013	0.013	0.009
RMS(angles)	1.43	1.29	1.26	1.54
Ramachandran favored (%)	91.05	94.06	93.2	91.23
Ramachandran allowed (%)	6.2	4.35	4.2	8.08
Ramachandran outliers (%)	2.74	1.59	2.6	0.7
Rotamer outliers (%)	5.19	3.37	3.87	2.41
Clashscore	11.17	10.14	10.09	30.56
Average B-factor	76.93	60.97	60.21	155.99
macromolecules	76.97	61.43	60.12	156.01
ligands		24.7	75.47	
solvent	63.61	58.25	58.51	

Statistics for the highest-resolution shell are shown in parentheses.

4. Secondary prediction of full-length AtCesA3

5. Multiple sequence alignment of AtCesAs

```

1      10      20      30      40      50      60
CESA1 MEASAGLVAGSYRRNELVIRHESDGGTKPLKXNMNGQIQGLGDDVGLAETDVFVACNE
CESA3 .....MESEGE..TAGHPKXIVPQTQSSNWKTVDDRRVACDI
CESA4 .....MEPNTMA.....FDDEHRSSTSAKIKVCCDEVKDDNDQTVACNV
CESA6 MNTGGRLIAGSHNRNEFVLINADENARINSVQELSGQTQLCRDEILTVDGEFVACNE
CESA7 MEASAGLVAGSHNRNELVVIHNNH.EEPKPLKXNLDGQFCEICGQQLGLTVEGDFVACNE
CESA8 .....MESEGE.....MESRSPIQNTGGEELGVKSNGEFVACHE

70      80      90      100     110
CESA1 CAFVVRCPQVEYERKDKTQCPCQKTRFRHRGSPRVEGDEEDDDVDIE.NEFNYAQGA
CESA3 CSFVVRCPQVEYERKDKGNQSCPCQKTRRYKRLKGSFAIPGDKLEDGLADEGTVFNYPQK.
CESA4 CVFVVRCPQVEYERSSGNKCCPCQKNTLYKRHKGSFKIAGDEENNGPISD.DELNICYRQ.
CESA6 CAFVVRCPQVEYERRRGNQACPCQKTRRYKRLKGSFRVEGDEEDDIDLL.NEFYGNNG
CESA7 CGFVVRCPQVEYERRRGTQNCPCQKTRRYKRLKGSFRVEGDEEDDIDIE.YFNIEHEQ
CESA8 CSFVVRCPQVEYERRRGTQNCPCQKTRRYKRLKGSFRVEGDEEDDIDIE.YFNIEHEQ

120     130     140     150     160
CESA1 NKARHQHGEFSSSRH.....ESQPIPLLTHGHTVSGEIRTPDTQSVRT
CESA3 ...EKISERMGLGWHLRGKGEEMGEFYDKVSHNHLRRLTSRQDTSGEFSAASPERLSV
CESA4 DGSSIH.Q.....NPAYGSENGDY.NSK...QQ.....WR.PNGRAFS.
CESA6 IGFDDQVSEGMISRRNSGFPQS...DLDSAPPQSQIFLLTYGDEDV.EISS.DRHALIV
CESA7 DKHKSAEAMLYGKMVYGRGPEDDEN.....GRGPVI.AGHSGEFPVVG.....
CESA8 .....T.Q.....SIVPQTQNTNSQD.....

170     180     190     200     210     220
CESA1 TSGLPGSDRNAISSPYIDPRQPVPRIVDPSPKDLNSYGLGNVDKEAVECHLLKOEKNM
CESA3 SSTI.IAGGK...RLPYSSDVNQSPN....RRIVDPVGLGNVAKERVDGKMKOEKNT
CESA4 STC.....SVLGKDFEAERDGYTDAEAKERVDKMKAEKRG
CESA6 .P.SLGGHGGRVHPVSLSDPTVAAPRPMVPQKDLAVYGYGSVAKERMEBEMKQNEKL
CESA7 .G.YGNGE.....HGLHRRVHPYP.....SSEAGSEGGERRMDDWKLQHG.N.L
CESA8 .SG...IHA.....RHISTVSTIDSELNDEYGNPIKRRVESMKDKKDKS

230     240     250     260     270     280
CESA1 LQMTGKYHEGKGGE...IE..GTGSNGEELQMAADTQLPMSRVVPIPSRRLTPYRVVIL
CESA3 GPVSTQAASERGV...DIDASTDILADALINDEARQPLSKKVSIPSSRINPYRMVIML
CESA4 LVTKGQGTNE.....DREDDDEEYLDASARQPLMKKVPSSSKISPRIVIVL
CESA6 QVVRHEGD...P...DF...EDGDDAFPMDEGQPLSKKIPKSSKINPYRMVIL
CESA7 GPFPDD.....DPMGLIDEARQPLSKKVPKSSKINPYRMVIA
CESA8 KKKKDKPKATKAEQHEAQIPTQQHMEDTPNTESGATDLSVVVIPHRTKITSYRIVIM

290     300     310     320     330     340
CESA1 RLHLCLFFLVYRTHPVKNAYPLWLWSVCEIWFASWLLDQFPKWPINRETYMDRLAT
CESA3 RLVLCLFLHVRITNPVKNAYPLWLWSVCEIWFASWLLDQFPKWPINRETYMDRLAL
CESA4 RLVLVFFFRIRILTPAKDAYPLWLWSVCEIWFASWLLDQFPKWPINRETYMDRLSM
CESA6 RLVLCLFFHRLHVKDAHALWLSVCEIWFASWLLDQFPKWPINRETYMDRLSL
CESA7 RLVLVFFFRIRILTPVDSAYPLWLWSVCEIWFASWLLDQFPKWPINRETYMDRLSL
CESA8 RLHLCLFFLVYRTHPVDSAYPLWLWSVCEIWFASWLLDQFPKWPINRETYMDRLSA

350     360     370     380     390     400
CESA1 VYDRDCEPSQLVVPVDFVSTVDPLKEPPLVLTANTVLSITLVDPYVDVACVYSDDGSAIML
CESA3 RYDERCEPSQAAVDFVSTVDPLKEPPLVLTANTVLSITLVDPYVDKVS CVYSDDGSAIML
CESA4 RPERDCEPSQAAVDFVSTVDPLKEPPLITANTVLSITLVDPYVKNVSCVYSDDGSAIML
CESA6 RYKEKCEPSQAAVDFVSTVDPLKEPPLITANTVLSITLVDPYVDKVS CVYSDDGSAIML
CESA7 RYERECEPSQAAVDFVSTVDPLKEPPLVLTANTVLSITLVDPYVKEIS CVYSDDGSAIML
CESA8 RPERDCEPSQAAVDFVSTVDPLKEPPLITANTVLSITLVDPYVDKVS CVYSDDGSAIML

410     420     430     440     450     460
CESA1 TFESLSEAEFAKKWVPFCKKYSIEPRAPETVFAQRLDYLKDKTQPSFVKERRAMKRREYE
CESA3 SFESLSEAEFAKKWVPFCKKYSIEPRAPETVFAAQLDYLKDKVOTPFVKERRAMKRREYE
CESA4 LFDTLSEAEFAKKWVPFCKKYNVEPRAPETVSEKLDYLKDKVOTPFVKERRAMKRREYE
CESA6 TFESLSEAEFAKKWVPFCKKYSIEPRAPETVFCHELDYLKDKVHPAFVKERRAMKRREYE
CESA7 TFESLSEAEFAKKWVPFCKKYSIEPRAPETVFTLKVLYLQDKVHPAFVKERRAMKRREYE
CESA8 SFESLSEAEFAKKWVPFCKKYSIEPRAPETVFTLKVLYLQDKVQPSFVKERRAMKRREYE

470     480     490     500     510     520
CESA1 SFKVRINAVAKAKKPEEGWTMQDGTBPWPGNTRDHPGMIQVFLGSGGLTDGNELPR
CESA3 SFKVRINAVSKAKKPEEGWTMQDGTBPWPGNTRDHPGMIQVFLGNGGLDAGNELPR
CESA4 SFKVRINAVAKAKKPEEGWTMQDGTBPWPGNTRDHPGMIQVFLGKEGAFIDGNELPR
CESA6 SFKVRINAVATAKPEEGWTMQDGTBPWPGNSVRDHPGMIQVFLGSDGVRVNNELPR
CESA7 SFKVRINAVAKAKKPEEGWTMQDGTBPWPGNTRDHPGMIQVFLGSGGDFVEGHELPR
CESA8 SFKVRINAVAKAKKPEEGWTMQDGTBPWPGNTRDHPGMIQVFLGSGCARIDGNELPR

530     540     550     560     570     580
CESA1 LLYVSREKRPGEQHKKKAGAMNALIRVSAVLNCGAYILNVDCDHYFNNSKALREANMCFPM
CESA3 LLYVSREKRPGEQHKKKAGAMNALIRVSAVLNCGFFILNDCDHYFNNSKALREANMCFPM
CESA4 LLYVSREKRPGEQHKKKAGAMNAMVRVSAVLNCGAPPFLNDCDHYFNNSKALREANMCFPM
CESA6 LLYVSREKRPGEQHKKKAGAMNSLIRVSGVLNCGAPPYLLNVDCDHYFNNSKALREANMCFPM
CESA7 LLYVSREKRPGEQHKKKAGAMNALIRVSAVLNCGAPPFLNDCDHYFNNSKALREANMCFPM
CESA8 LLYVSREKRPGEQHKKKAGAMNALIRVSAVLNCGAPPFLNDCDHYFNNSKALREANMCFPM

590     600     610     620     630     640
CESA1 DPAIGKQCVVQFFQRFDGIDLRDYNRRNVVFFDINMKGLDGGGPFVYVGTGCVFRROA
CESA3 DPAIGKQCVVQFFQRFDGIDLRDYNRRNVVFFDINMKGLDGGGPFVYVGTGCVFRROA
CESA4 DPAIGKQCVVQFFQRFDGIDLRDYNRRNVVFFDINMKGLDGGGPFVYVGTGCVFRROA
CESA6 DPAIGKQCVVQFFQRFDGIDLRDYNRRNVVFFDINMKGLDGGGPFVYVGTGCVFRROA
CESA7 DPAIGKQCVVQFFQRFDGIDLRDYNRRNVVFFDINMKGLDGGGPFVYVGTGCVFRROA
CESA8 DPAIGKQCVVQFFQRFDGIDLRDYNRRNVVFFDINMKGLDGGGPFVYVGTGCVFRROA

```

```

        650          660          670
CESA1 LYCYDFVLTEDLEPNI....IVKSCGSRKKKXSSKK..YN.....
CESA3 LYCYEPPPKVKHKKKPL....LSKLC...GGSRKKNSSAKKE.....
CESA4 LYCYEPPPVSEKRRKMTCDWPSWICCCGGGNNRKSDSSKKKSGIKSLFSKLKKTKKK
CESA6 LYCFDAPKKKKGPRKTCNCWPKWCLLCFSGRKNRKAKTVAADK.....
CESA7 LYCYEPPKGGPR..PKM.....ISCCPCFGRNNKKKFSK.....
CESA8 LYCYSPPKSKPILPQSS.....SSSCCLTKKKKQPDSE.....

        680          690          700          710          720
CESA1 YE...KRRGINRSDSNAPTFNMPDIDEG..FEGYD.DERSILMSQSVDRPCQSDVF
CESA3 SD...KKKSGRHTDSTVPVFNLDIDEEG.VEGAGPJ.DEXALMSQSVLEKRFQSDAVVF
CESA4 SDDKTMSSYSRKRSTEAIFDLEDIEEGL...EGYDELEKSSLSQKKNFERFCMSDFVI
CESA6 .....KKNREASKQTHALENIEEGRVTKGSN..VEQSTEAMOMKLEKRFQSDPVVF
CESA7 .....IKYDARREELDAAIFNLDGL.....DNYDEYDRSMLSQTSFERFQGLSTVF
CESA8 .....IYKDAARREELDAAIFNLDGL.....DNYDEYDRSMLSQTSFERFQGLSTVF

        730          740          750          760          770          780
CESA1 AATFMEGGGIPPTNFATLKEATHVISCYEDKTEWGRKEIGWIYGSVFEDILTGFRMMA
CESA3 ASITLMENGGVFPFATFENMLKEATHVISCYEDKSDWGRKEIGWIYGSVFEDILTGFRMMA
CESA4 ASITLMENGGVFPFATFENMLKEATHVISCYEDKTEWGRKEIGWIYGSVFEDILTGFRMMA
CESA6 ASARMENGGMARNAFACLKEATHVISCYEDKTEWGRKEIGWIYGSVFEDILTGFRMMA
CESA7 TSTLMENGGVFPFSSHAVLKEATHVISCYEDKTEWGRKEIGWIYGSVFEDILTGFRMMA
CESA8 ESTLMENGGVFPDSVNFSTLKEATHVISCYEDKTEWGRKEIGWIYGSVFEDILTGFRMMA

        790          800          810          820          830          840
CESA1 RGVISIVCNPPRPAFKGSAPINLSDRLNQVLRWALGSITLLSRHGFNIYGYHG.RRLRL
CESA3 RGVRSIYCMPLKPAFKGSAPINLSDRLNQVLRWALGSVITLFSRHGFNIYGYNG.RLKKL
CESA4 RGVRSIYCMPLKPAFKGSAPINLSDRLNQVLRWALGSVEIFFSRHGFNIYAWGG.RLKKL
CESA6 RGVRSIYCMPLKPAFKGSAPINLSDRLNQVLRWALGSVEIFFSRHGFNIYGYNG.RLKKL
CESA7 RGVRSIYCMPLKPAFKGSAPINLSDRLNQVLRWALGSVEIFFSRHGFNIYGYNG.RLKKL
CESA8 RGVRSIYCMPLKPAFKGSAPINLSDRLNQVLRWALGSVEIFFSRHGFNIYGYNG.RLKKL

        850          860          870          880          890          900
CESA1 ERLAVNTIVPFTSPLIAYCTIPAVCLDGRFIPETISNYASIMFTLFTSTAVGCHL
CESA3 ERLAVNTIVPFTSPLIAYCTIPAVCLDGRFIPETISNYASIMFTLFTSTAVGCHL
CESA4 ERLAVNTIVPFTSPLIAYCTIPAVCLDGRFIPETISNYASIMFTLFTSTAVGCHL
CESA6 ERLAVNTIVPFTSPLIAYCTIPAVCLDGRFIPETISNYASIMFTLFTSTAVGCHL
CESA7 ERLAVNTIVPFTSPLIAYCTIPAVCLDGRFIPETISNYASIMFTLFTSTAVGCHL
CESA8 ERLAVNTIVPFTSPLIAYCTIPAVCLDGRFIPETISNYASIMFTLFTSTAVGCHL

        910          920          930          940          950          960
CESA1 ELRWSGQVSIEDLWRNEQFWVIGGSAHLFAVFOGLKVLACIDTNFTVTSRA..TDEDGDF
CESA3 ELRWSGQVSIEDLWRNEQFWVIGGSAHLFAVFOGLKVLACIDTNFTVTSRA..TDEDGDF
CESA4 ELRWSGQVSIEDLWRNEQFWVIGGSAHLFAVFOGLKVLACIDTNFTVTSRA..TDEDGDF
CESA6 ELRWSGQVSIEDLWRNEQFWVIGGSAHLFAVFOGLKVLACIDTNFTVTSRA..TDEDGDF
CESA7 ELRWSGQVSIEDLWRNEQFWVIGGSAHLFAVFOGLKVLACIDTNFTVTSRA..TDEDGDF
CESA8 ELRWSGQVSIEDLWRNEQFWVIGGSAHLFAVFOGLKVLACIDTNFTVTSRA..TDEDGDF

        970          980          990          1000          1010          1020
CESA1 AELYLKWTALLIPPFTVLLVNLIGVAGVSYAVNSGYQSWGPLFGRLFFALWVIAHLYP
CESA3 AELYLKWTALLIPPFTVLLVNLIGVAGVSYAVNSGYQSWGPLFGRLFFALWVIAHLYP
CESA4 AELYLKWTALLIPPFTVLLVNLIGVAGVSYAVNSGYQSWGPLFGRLFFALWVIAHLYP
CESA6 AELYLKWTALLIPPFTVLLVNLIGVAGVSYAVNSGYQSWGPLFGRLFFALWVIAHLYP
CESA7 AELYLKWTALLIPPFTVLLVNLIGVAGVSYAVNSGYQSWGPLFGRLFFALWVIAHLYP
CESA8 AELYLKWTALLIPPFTVLLVNLIGVAGVSYAVNSGYQSWGPLFGRLFFALWVIAHLYP

        1030          1040          1050          1060          1070          1080
CESA1 FLKGLMGQRNRPTIVVWSVLLASIFSLLWVRNPFFVDAHPNANNFNKGKGVF....
CESA3 FLKGLMGQRNRPTIVVWSVLLASIFSLLWVRNPFFVDAHPNANNFNKGKGVF....
CESA4 FLKGLMGQRNRPTIVVWSVLLASIFSLLWVRNPFFVDAHPNANNFNKGKGVF....
CESA6 FLKGLMGQRNRPTIVVWSVLLASIFSLLWVRNPFFVDAHPNANNFNKGKGVF....
CESA7 FLKGLMGQRNRPTIVVWSVLLASIFSLLWVRNPFFVDAHPNANNFNKGKGVF....
CESA8 FLKGLMGQRNRPTIVVWSVLLASIFSLLWVRNPFFVDAHPNANNFNKGKGVF....

```

6. Secondary structure prediction of AtPRF3

7. Data collection and refinement statistics of structures in Chapter 2

Structure	AtPRF2	AtPRF2-Poly-P	AtPRF3Δ37	AtPRF3Δ22
PDB code	6IQ1	6IQJ	6IQF	6IQK
Wavelength	1.5406	1.5406	0.9537	0.9764
Resolution range	27.7 - 2.4 (2.48 - 2.40)	36.77 - 1.92 (1.991 - 1.92)	37.96 - 1.46 (1.509 - 1.46)	49.24 - 3.6 (3.729 - 3.6)
Space group	P 3 2 1	P 2 ₁ 2 ₁ 2 ₁	C 1 2 1	P 2 ₁ 2 ₁ 2 ₁
Unit cell	58.68 58.68 83.94 90 90 120	51.52 65.16 73.53 90 90 90	80.98 52.86 32.41 90 110.38 90	135.9 153.37 196.68 90 90 90
Total reflections	54868	65543	143911	131516
Unique reflections	12582 (1274)	18928 (1596)	22378 (2174)	48279 (4769)
Multiplicity	4.2	3.5	6.43	27.2
Completeness (%)	99.12 (100.00)	97.37 (84.44)	96.7(85.3)	99.80 (99.79)
Mean I/sigma(I)	25.8(1.7)	9.3(2.3)	8.77(0.88)	5.99(0.75)
Wilson B-factor	52.38	13.96	25.02	115.08
R-meas	0.058	0.13	0.077	0.58
CC1/2	99.7(73.4)	99.8(79.3)	99.6 (63.4)	99.1(27.5)
Reflections used in refinement	12569 (1274)	18913 (1596)	22357 (2166)	48230 (4763)
Reflections used for R-free	1168(96)	929 (72)	1062 (105)	4822 (477)
R-work	0.2623 (0.3891)	0.183(0.2333)	0.1963(0.4002)	0.3014(0.4123)
R-free	0.2968 (0.4503)	0.2183 (0.3123)	0.2156 (0.4381)	0.3469 (0.4396)
Number of non-hydrogen atoms	1969	2447	1024	11558
macromolecules	1950	2114	975	11558
solvent	19	333	49	0
Protein residues	260	283	130	1647
RMS(bonds)	0.004	0	0.01	0.003
RMS(angles)	0.87	1.03	1.23	0.63
Ramachandran favored (%)	96.09	98.18	96.9	92.34
Ramachandran allowed (%)	3.91	1.82	3.1	7.47
Ramachandran outliers (%)	0	0	0	0.19
Rotamer outliers (%)	0	0	0	0
Clashscore	12.23	2.15	2.58	5.77
Average B-factor	64.83	15.87	30.61	125.61
macromolecules	64.59	14.38	30.27	125.61
solvent	89.28	25.32	37.32	

7. Publication

- He, Q., Toh, J. D., Ero, R., **Qiao, Z.**, Kumar, V., Serra, A., . . . Gao, Y. G. (2019). The unusual di-domain structure of *Dunaliella salina* glycerol-3-phosphate dehydrogenase enables direct conversion of dihydroxyacetone phosphate to glycerol. *Plant J.* doi:10.1111/tpj.14619
- Qiao, Zhu**, Sun, He, Ng, Justin Tze Yang, Ma, Qianqian, Koh, Si Hui, Mu, Yuguang, . . . Gao, Yong-Gui. (2019). Structural and computational examination of the Arabidopsis profilin-Poly-P complex reveals mechanistic details in profilin-regulated actin assembly. *Journal of Biological Chemistry.* doi:10.1074/jbc.RA119.011307
- Sun, H., **Qiao, Z.**, Chua, K. P., Tursic, A., Liu, X., Gao, Y. G., . . . Miao, Y. (2018). Profilin Negatively Regulates Formin-Mediated Actin Assembly to Modulate PAMP-Triggered Plant Immunity. *Curr Biol*, 28(12), 1882-1895 e1887. doi:10.1016/j.cub.2018.04.045