Computational Study of Plant Cell Wall Synthases

by

Alexander Adams

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Engineering)
in the University of Michigan
2023

Doctoral Committee:

Professor Sharon C. Glotzer, Co-Chair
Dr. Heather B. Mayes, National Renewable Energy Laboratory, Co-Chair
Professor Ron G. Larson
Professor Erik Nielsen

Alexander Adams

xadams@umich.edu

ORCID iD: 0000-0002-1383-224X

Table of Contents

List of Tables v

List of Figures vi

Abstract vii

Chapter 1 Introduction 1

1.1 Plant Cell Growth 1

1.2 Objective and Outline 3

Chapter 2 Comparative Modeling & *De Novo* Structure Predictions Reveal Extensive Conservation and Key Differences in *Arabidopsis* CESA and CSLD Proteins 6

2.1 Abstract 6

2.2 Introduction 6

2.3 Methods 7

2.3.1 Coarse-grain simulations 7

2.3.2 Atomistic simulations 9

2.4 Results 9

2.4.1 Resolved Secondary Structure of CESA is conserved across families 9

2.4.2 Putative Stalk Structure Differentiates CESA and CSLD 11

2.4.3 Generation of a putative stalk structure 13

2.4.4 Structural Evaluation of a Putative Stalk Structure 15

2.4.5 Cellulose Synthases Assemble into Trimers or Smaller 16

2.4.6 Refuting of a Dimeric Crystal Structure 17

2.5 Discussion 19

Chapter 3 A Temperature Sensitive FERONIA Mutant Allele that Alters Root Hair Growth 22
3.1 Chapter Introduction 22
3.2 Abstract 22
3.3 Introduction 23
3.4 Results 25
3.4.1 The G41S mutation in fer-ts is a highly conserved glycine residue in CrRLK1L subfamily proteins 25
3.4.2 Pectin is not predicted to bind near the G41S mutation site..... 27
3.4.3 Molecular dynamics studies show no temperature-dependent rearrangement..... 28
3.5 Discussion 28
Chapter 4 Computational Tool Development 33
4.1 Chapter Introduction 33
4.2 InLoFi: A reformed tool for quantification of lobed cells 33
4.2.1 Introduction 33
4.2.2 Software Design 36
4.2.3 Conclusions and Future Work 38
4.3 On-the-Fly Autocorrelation as a HOOMD-blue plug-in 38
4.3.1 Introduction 39
4.3.2 Software Design 41
4.3.3 Testing and Validation 44
4.3.4 Testing and Validation 46
4.3.5 Conclusions and Future Work 48
Chapter 5 Conclusions, Limitations, and Future Work 50
5.1 Research Summary 50
5.2 Limitations and Future Work 51
Appendix 53

A.1 Equations 53
A.1.1 Calculation of Relative Per-Protein Energies 53
A.1.2 Propagation of Uncertainty of Relative Per-Protein Energies 53
A.2 Sequence Alignments 53
A.2.1 RING finger Sequence Alignment 54
A.2.2 P-CR Sequence Alignment 54
A.2.3 CSR Sequence Alignment 55
Bibliography 57

List of Tables

Table 2.1. Per protein energies, each relative to the monomeric system for their system size. ... 16

Table 4.4.1: Run Time Analysis of On the Fly and Post-Processed Data 47

List of Figures

Figure 1.1. Simplified representation of the cell wall synthesis process. 2

Figure 2.1. Structure of representative CESA and CSLD sequences. 11

Figure 2.2. Predicted structural conservation across all Arabidopsis CESA and CSLD proteins. 13

Figure 2.3. The F register is only stable after a rearrangement of tryptophan residues, while the B register requires only minor rearrangement. 14

Figure 2.4. Comparison of membrane-integrated full dimeric crystal structure with PttCESA8 trimeric structure. 18

Figure 3.1. Localization of the highly conserved glycine mutation in fer-ts in A. thaliana FERONIA:RALF23:LLG2 and X. laevis malectin structures. 27

Figure 4.1. Fluorescence microscopy images of cotyledon pavement cells. 35

Figure 4.2. Visualization of LobeFinder output, adapated from[95] 35

Figure 4.3. Schematic of InLoFi work flow. 37

Figure 4.4.4 (a) Autocorrelation of xy component of pressure implemented on the fly and as a post-processing method. 46

Figure 4.4.5 Performance data for post-processed and plug-in data. 47

Figure A.1. RING finger regions within the N-terminal demonstrate sequence variation but conserved structures. 54

Figure A.2. P-CR structure is well conserved across protein families. 54

Figure A.3. CSRs are intrinsically disordered and highly variable within and across protein families. 55

Figure A.4. DSL regions of *At*CSLDs do not disrupt trimeric interfaces. 55

Figure A.5. Schematic of Coiled-coil candidate structure work flow. 56

Figure A.6 Consideration of the full CESA structure for a putative dimeric structure introduces extensive steric clashes (red). The members of the dimer are shown in blue and cyan. 56

Abstract

This work combines a range of techniques in the field of computational biology to address both longstanding and more recent questions of plant biology, and contributes additional tools when the current options are insufficient.

First, comparative modeling and *de novo* protein structure prediction were used to generate full structures of *Arabidopsis* Cellulose Synthase A (CESA) and Cellulose Synthase-Like D (CSLD) proteins. Comparison of these structures showed a novel feature in CSLD proteins that we have named CSLD specific loops (DSLs), and that CESAs are capable of forming N-terminal stalk structures while CSLDs are not. Despite these differences, we show that CSLDs should form trimeric interfaces similar to CESAs given the positioning of these differences and extensive structural conservation otherwise.

Then, this N-terminal difference was fully elucidated by developing a method for building coiled-coil structures from low-resolution structures, even without coordinates. This method produced two candidate structures with vastly different interfaces, which allows for a future experimental study to determine which is more likely.

Next, we provide some evidence that CESAs form into trimers. Using coarse-grain molecular dynamics (MD) simulations of different numbers of CESA proteins in a plant membrane environment, we show that assemblies of greater than three proteins are energetically unfavorable by an order of magnitude. We also show that a crystal structure used as evidence for dimers is only possible because it excludes crucial structures, and that a full dimer in the

captured arrangement would introduce steric clashes both between proteins and between proteins and the membrane.

In a different plant growth system, we present a novel temperature-sensitive mutant protein that inhibits root hair growth only at elevated temperatures. Using comparative modeling, docking analyses, and MD studies we show that the mutation does not lead to loss of characterized binding sites, so some other mechanism must be responsible for the temperature-dependent loss of activity.

Finally, we present two programs to assist in the processing of data. Inspired by LobeFinder (InLoFi) quantifies the lobes of plant cells from confocal microscopy images, performing faster and more consistently than manual analysis. The Autocorrelator plug-in for HOOMD-blue introduces a Logger function that performs on-the-fly autocorrelation, dramatically reducing file storage needs and speeding up performance by eliminating these file-writing operations.

In total, this work adds insight to a richly complex body of experimental evidence, and develops methods and tools for further investigation of these complex systems.

Chapter 1 Introduction

Decades of effort and development of analytical techniques have given us a wealth of biological information: billions of sequenced genes, a hundred thousand solved structures, and volumes of experimental insights awaiting theories to knit them together into a rich narrative. Due to the complex, multi-dimensional nature of many of these processes, open questions in the field still seem disparate, and in need of analysis that considers its full context. Answering any single question requires a litany of experimental and computational techniques, and slowly builds towards a body of knowledge that better enables solving the next question, until such a point that the complete toolset offers a path forward for any problem. Computational modeling, in the form of comparative modeling, *de novo* structure prediction, molecular dynamics, and others, has an essential role to play in this progress. Computational methods have the potential to perform in hours what would take months of benchtop experiments with living organisms, providing valuable screening to pare down complex designs that can only be properly studied *in vivo*, while also offering molecular level insight for those questions that can be narrowed down to single proteins and simple systems. In this work, we use a myriad of different approaches, developing additional tools as necessary, to fully utilize the breadth of information available to assemble, from so many pieces, answers to new and longstanding questions in the field of plant biology.

1.1 Plant Cell Growth

Plants' synthesis of carbohydrates, especially cellulose, is vital for the flow of energy and materials into Earth's ecosystems. The scientific community has only a cursory understanding of

this essential process, and further comprehension would allow for the rational design of plant structures with custom properties for direct use as materials, or for eased utilization as a renewable carbon source, among other possibilities. In plants, cellulose is produced by Cellulose Synthase A (CESA) proteins that assemble first into lobes of unknown composition and then into hexameric Cellulose Synthesis Complexes (CSCs) that each produce a single cellulose microfibril, woven from 18-24 glucan strands produced by individual CESAs. These microfibrils are then coated by xylans and crosslinked by additional hemicelluloses such as lignin and pectin to form the cellular walls[1].

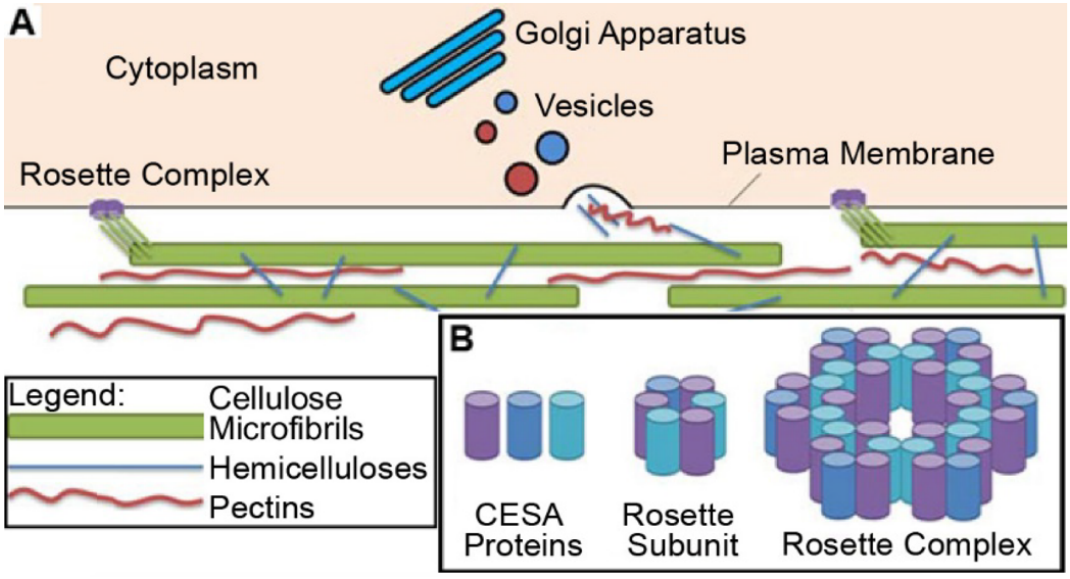

Figure 1.1. Simplified representation of the cell wall synthesis process.

In (A), Rosette Complexes/Cellulose Synthesis Complexes (CSCs) are trafficked to the plasma membrane alongside hemicelluloses and pectins. The CSC's then produce glucan chains that assemble into cellulose microfibrils. (B) shows one possible structure of the CSC: a hexamer of subunits, with each subunit consisting of six CESA proteins.

The ability to study this complex, interwoven biological system is limited, as many CESA mutants or knockouts result in dead organisms[2]–[4]. For this reason, our study of CESA proteins is complemented by study of the structurally similar Cellulose Synthase-Like Family D (CSLD) protein, whose knockout mutants result in non-lethal phenotypes such as the loss of root

hair growth or other tip growth mechanisms. These two protein families are so similar in fact, that a significant portion of chapter 2 is dedicated to characterizing their limited structural differences, and hypothesizing how these distinctions manifest as differing membrane localization and propensity to self-assemble. In chapter 3, we even go so far as to study a receptor protein whose mutation has a cascading effect that eventually disrupts root hair growth, specifically at elevated temperatures. This is the nature of studying these biological environments, and the relative freedom of computational methods allow for the interrogation of specific systems that are otherwise inaccessible.

The computational study of these systems has recently been enabled by a pair of developments. The first is the publication of a CESA protein in a trimeric complex at atomistic resolution. This serves as the basis for comparative models of CESA and related proteins, such as CSLDs, in other organisms, and also serves as an essential data point for machine learning based predictive tools such as AlphaFold[5]. This infusion of data has allowed for long established computational tools, when strung together and interpreted in the context of decades of experimental insight, to provide new progress on these longstanding questions.

1.2 Objective and Outline

The primary aim of this thesis is to address longstanding questions in the field of plant biology by uniting domain knowledge, novel structural information, and an array of computational tools to isolate and interrogate impactful details of prohibitively complex biological systems.

In chapter 2, we address the questions of CSC formation and the difference between the enzymatically interchangeable CESA and CSLD proteins. These insights are made possible by a combination of comparative modeling, *de novo* protein structure prediction, molecular dynamics

simulation, and several specific tools dealing with sequence alignment and coiled-coil prediction and modeling. A particular combination of these tools offer general applicability, particularly utilizing an increasingly available dataset in the form of low resolution cryogenic-electron microscopy (cryo-EM) structures, so is formalized for wider use. Although the included coarse-grain molecular dynamics is limited in its scope and interpretability, it represents an important first step in assessing what further systems can be effectively studied with this method.

In chapter 3, we discover and study a single mutant protein with a profound effect: a temperature-sensitive change in a growth signaling pathway, desensitization towards a signaling ligand, and altered accumulation of reactive oxygen species (ROS). Using a wide combination of techniques, this time experimental and computational in nature, we investigate the full effect of this mutation and potential mechanisms that it can offer a gateway into, as well as potential mechanisms for broadly affecting single amino acid change. Using comparative modeling, docking, and molecular dynamics simulations, we do not achieve a complete understanding of the mechanism, but are able to narrow down the possibilities for future study.

Finally, in chapter 4, we present additional tools developed for specific applications, but that hold the potential to grow to complement the computational arrays used in either of the previous chapter, or to be of use in more disparate fields. Inspired by LobeFinder (InLoFi) is a tool that iterates on an existing one to directly quantify images of lobed plant cells in a way that is faster and more consistent than manual segmentation. The Autocorrelator plug-in for HOOMD-blue provides on-the-fly analysis that dramatically reduces the simulation time and storage requirements for a specific polymer system, but is written in such a way as to easily be adapted for other properties in different systems that may benefit from this feature. Both tools are

developed with resilient testing so they can be extended by us or others as additional uses become clear.

Chapter 2 Comparative Modeling & *De Novo* Structure Predictions Reveal Extensive Conservation and Key Differences in *Arabidopsis* CESA and CSLD Proteins

Disclosure: Reprinted from **Adams, A**, Mayes, H.B., and Nielsen, E. Comparative Modeling & De Novo structure predictions reveal extensive conservation and key differences in *Arabidopsis* CESA and CSLD proteins. *In preparation*

2.1 **Abstract**

In plants, biopolymers are manufactured in a range of contexts by a variety of protein families. It has been shown that cellulose synthase (CESA) proteins and cellulose synthase-like D subfamily (CSLD) proteins are catalytically interchangeable, both encoding glycosyltransferases that synthesize β-1,4-linked glycan chains, the backbone of plant cellulose fibrils[6], [7]. Given this identical chemical activity, an open question remains as to the structural differences driving the differing localization of these proteins. Here, we use comparative modeling, *de novo* structure prediction and computational simulation to reveal key differences between CESA and CSLD proteins, while demonstrating a propensity for both to form trimeric assemblies. To better characterize one such difference, we develop a method for the generation of coiled-coil structures based on low resolution surface models, and used it to produce two candidate structures for this vital region.

2.2 Introduction

Cellulose is the most abundant biopolymer on Earth, but its manufacture is still not fully understood. Crystalline cellulose, which comprises the cell walls of plants, is composed of tightly bundled β-1,4-linked glycan chains produced by protein super-assemblies called

Cellulose Synthesis Complexes (CSCs)[8], [9]. Each CSC is an assembly of six lobes, with each lobe comprised by a number of Cellulose Synthase family A (CESA) proteins. It is still an open question as to the composition of each lobe, both in terms of number of proteins as well as identity of those proteins among the 10 isoforms[10]. Each protein is a catalytically active glycosyltransferase synthesizing β-1,4-linked glycan chains, which subsequently assemble into cell wall cellulose. The publication of a plant CESA8 protein in a homo-trimeric arrangement provided novel structural details as well as additional evidence for trimeric lobes[11]. This model enabled structural comparison of CESA and related proteins through traditional comparative modelling methods, as well as eventually through protein structure prediction tools such as AlphaFold[5], [12]. In addition to CESAs, this superfamily of related proteins includes six Cellulose Synthase-Like (CSL) clades: CSLA, CSLB, CSLC, CSLD, CSLE, and CSLG[13], [14]. Of these, CSLD shows the highest degree of sequence similarity and overall domain organization, and has been shown to be catalytically interchangeable in chimeric studies[6], [7], [15].

In this work, we utilize newly available structural data to determine key differences between and within CESA and CSLD protein families in the model plant *Arabidopsis thaliana*. To fully characterize a difference in the N-terminal, we present a methodology for developing candidate coiled-coil structures based on low-resolution structural data that produces a pair of structures that fit a putative model in the absence of published coordinates. Finally, we provide additional evidence for CESA trimeric lobes through a molecular dynamics (MD) study, as well as examination of a putative dimeric crystal structure.

2.3 Methods

2.3.1 Coarse-grain simulations

All coarse-grain simulations were prepared with CHARMM-GUI and simulations were performed using GROMACS[16]–[20]. All lipid models and parameters used in this study follow the MARTINI v2.0 lipids[21], [22]. The system was energy minimized and subjected to a set of serial equilibration runs in which the protein backbone positions were restrained. A Verlet cut-off scheme with a buffer tolerance of 0.005 kJ mol−1 ns−1 was used. A cut-off of 1.1 nm was used for calculating both the electrostatic and van der Waals interaction terms; both interactions were smoothly shifted beyond the cut- off using the potential-shift-Verlet algorithm. Coulomb interactions were calculated using the reaction-field algorithm implemented in GROMACS. A velocity rescale thermostat with a coupling constant of 1.0 ps was used to maintain the temperature at 303 K. A Parrinello−Rahman barostat with the semi-isotropic scheme was used to maintain 1 bar pressure with 12.0 ps coupling constant. The CG simulations were performed using a 20 fs time step, and the trajectory snapshot was saved at the 100 ps interval for the first 1 μs, then at the 0.5 ns interval for the duration. The simulation times reported here are actual simulation times and are not corrected for the faster dynamics attributed to the smooth potentials of the MARTINI force field[23]. Simulations were run for 18 μs for the largest system and 26 μs for the 2 smaller systems. For every microsecond, the average system-wide electrostatic and non-bonded interactions were calculated using gmx energy, including an estimated standard error of the mean. The energies of the corresponding membrane-only systems were subtracted from the energy of each protein system to yield the protein-specific contributions to the system energy. From there, each energy was divided by the number of proteins to obtain a per-protein energy, and then the monomer Coulombic and Lennard-Jones energies were set to zero to serve as a reference. The operation is summarized in SI Eq A.1.1. Uncertainty was calculated for the total operation, summarized in SI Eq A.1.2.

2.3.2 Atomistic simulations

CHARMM-GUI[16] NAMD and Gromacs inputs[17], [24] and martini maker [18], [19]

All atomistic MD systems were prepared with CHARMM-GUI, and performed using NAMD 2.13[16], [17], [25]. The CHARMM36 force field, including the φ/ψ cross-term map (CMAP) correction for the proteins was used for all the simulations[26]–[28]. Water molecules were described with the TIP3P model[29]. The protein was energy-minimized in solution for 10000 steps using a conjugate gradient algorithm[30]. Simulations were carried out using a 2 fs timestep at 303K constant temperature using Langevin dynamics with a damping coefficient γ of 1 ps−1. NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign.

2.4 Results

2.4.1 Resolved Secondary Structure of CESA is conserved across families

The sequences of multiple CESA isoforms across multiple species have long been available, and allow us to assess sequence conservation within and between CESA families, in various organisms. The recently published cryo-EM structure of isolated PttCESA8[11], allows for the comparative modeling of different CESA isoforms from different organisms. We were curious how PttCESA8 compared to analogous structures in the model plant *Arabidopsis* in the form of CESA proteins, as well as related structures in the form of *Arabidopsis* CSLD proteins, so we generated homology models of AtCESA isoforms 1-9 and AtCSLD isoforms 1-5 using the online SWISS-MODEL workspace[12]. Using PttCESA8 as a basis, we generated models from approximately residue 250 to 1050, which includes all structures except the N and C termini, for which the PttCESA8 model contained insufficient information. These models were aligned based

on the highly conserved transmembrane regions and resulted in the overlap of other secondary structural elements.

Excepting the previously characterized hypervariable region, the class-specific region (CSR), all structures within families were highly conserved, with sequence analysis corroborating complete overlap of the transmembrane, catalytic, and plant-conserved regions of *Arabidopsis* CESAs and CSLDs with the published PttCESA8 (SI Fig. 1). Two representative structures in CESA1 and CSLD3 are shown in Figure 2.1 (c) and (d). The PttCESA8 model, as well as all CESA and CSLD models, include a sequentially non-conserved, intrinsically disordered class-specific region on the periphery of trimeric structure. This region varies both between and within protein families. This high degree of structural conservation makes it likely that these proteins also form trimeric structures since the PCR and TM interfaces are conserved, while also possessing variable regions that may explain the non-redundancy of the CESA isoforms.

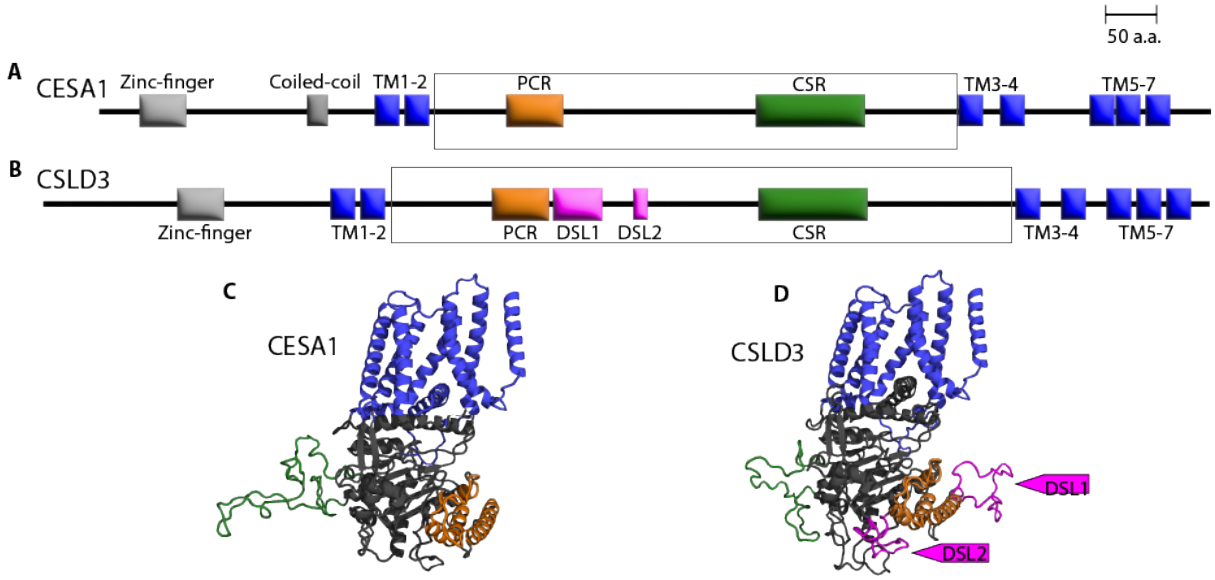

Figure 2.1. Structure of representative CESA and CSLD sequences.

Secondary structure schematics for (A) AtCESA1 (*Arabidopsis thaliana* CEllulose Synthase A1) and (B) AtCSLD3 (Arabidopsis thaliana CEllulose Synthase-Like D3) with TM helices in blue, PCRs in orange, CSR in green, DSLs in pink, and unmodeled N-terminal features in grey. The soluble catalytic region is indicated in black. (C) CESA1 crystal structure colored as in panel A. (D) CSLD3 homology model with DSLs labeled.

Finally, all AtCSLD protein sequences include two short (26 or 36 amino acids long) intrinsically disordered insertions that vary within the CSLD family. These insertions are peripherally located, and are not predicted to occlude the trimeric interfaces (Appendix A.3). Instead, they may present alternative surfaces to accessory proteins or otherwise affect the differential localization of the CESA and CSLD proteins.

2.4.2 Putative Stalk Structure Differentiates CESA and CSLD

Another structural difference between the CESA and CSLD families, and a key feature of the structure published by Purushotham et al., is the putative stalk structure formed by trimerization of the N-termini[11]. The structure was identified as a trimeric coiled-coil, with an unresolved loop stretching towards the atomistically resolved structure and volumes matching

ions coordinated around the base which presumably represents the zinc-fingers. The primary features of the structure have been reproduced in Fig 2.2 Using RaptorX de novo protein structure prediction, we tested whether or not the CESA and CSLD sequences were predicted to form the secondary structural elements necessary for the assembly of this structure[31]–[33]. We determined that all of the CESA proteins, both *Arabidopsis* and PttCESA8, were predicted to form both the RING finger structure and an alpha helical segment of requisite length, while the CSLD proteins were predicted to form only the RING finger. These predictions were corroborated by structures published as part of the AlphaFold database[5], [34]. The inability to form an N-terminal stalk structure, alongside the additional flexible linkers, represent the limited structural differences between the CESA and CSLD families in *Arabidopsis*.

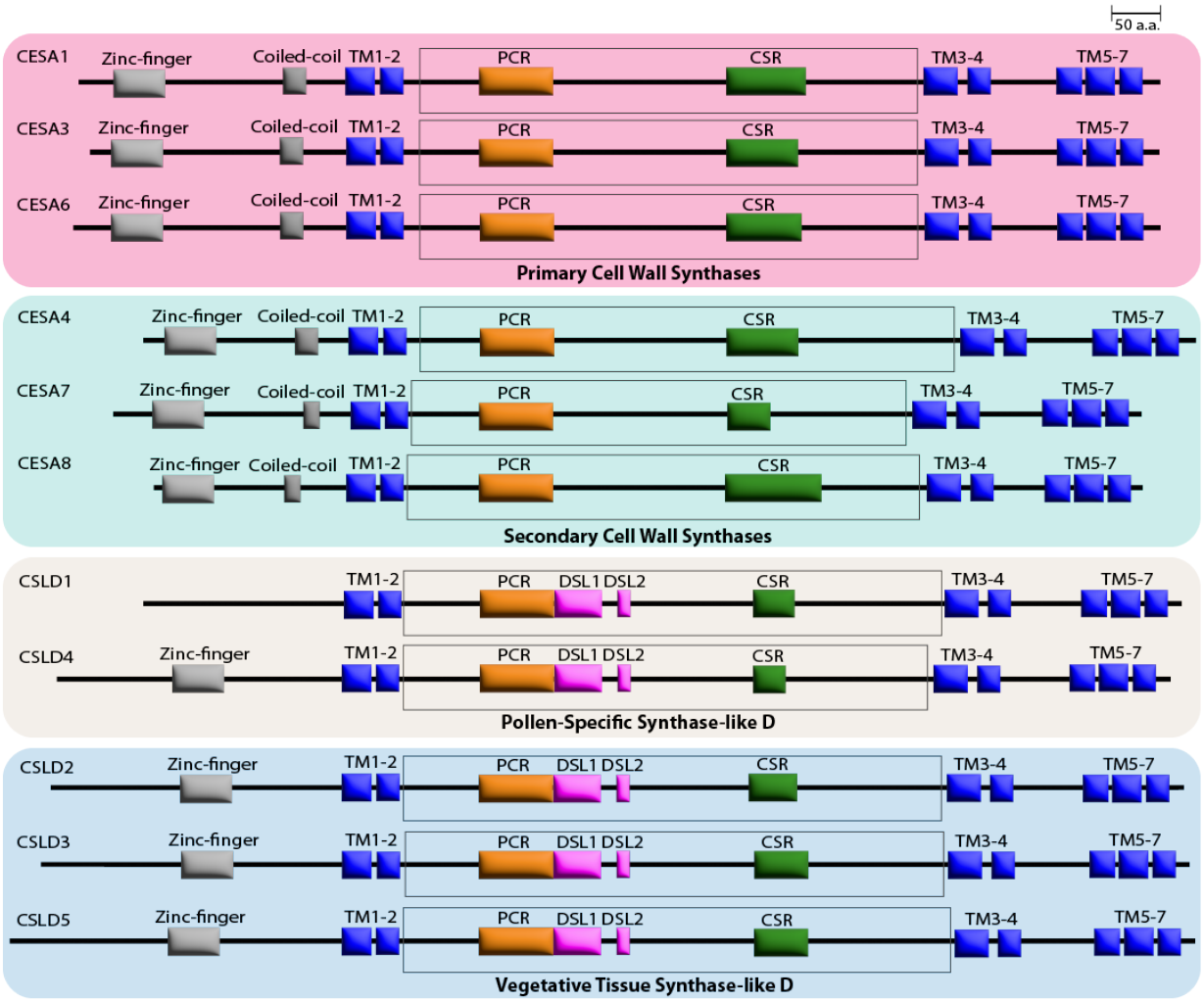

Figure 2.2. Predicted structural conservation across all *Arabidopsis* CESA and CSLD proteins.

All N-terminal (grey) structures, were predicted by RaptorX, and all remaining structures were predicted based on comparative models built using SWISS-Model. All structural publications were later corroborated with the AlphaFold database.

2.4.3 Generation of a putative stalk structure

Due to the novelty of this stalk structure compared with previous models of CESA based on BCSA or other analogs, we attempted to generate an atomistic structure as a starting point for isolation studies. Although a refined structure of the stalk is unavailable, it is clear from the cryo-EM volume and predicted protein structure that the N-termini form a trimeric coiled-coil. Coiled-coils have seven viable conformations, so we constructed all possible arrangements of the predicted alpha helical sequence using CCBuilder2, and used the built-in energy minimization to

optimize each structure[35], [36]. We then ran triplicates of 10 ns simulations for each of these coiled-coils in a 50 Å water box with 0.8 mM of CaCl starting with the optimized structures from CCbuilder[37]. Based on work by Pineiro et al., we expected this timescale to be sufficient to gauge the stability of the coiled-coil structures[38]. A schematic of the procedure is shown in Appendix A.4. The result was the swift dissociation of 5 of the seven registers, with only register B and F remaining as trimeric structures in any replicas. All three simulations of register B maintained a trimeric coiled-coil, while only a single register F simulation showed a stable structure. The stable F register successfully buries a trio of hydrophobic tryptophan residues in the core of the coiled-coil structure compared to the starting structure, shown in Figure 2.3.

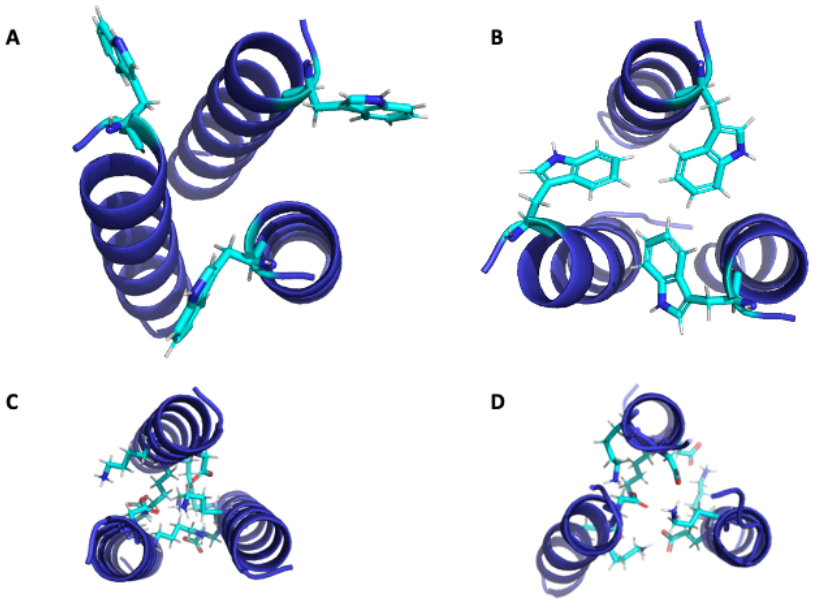

Figure 2.3. The F register is only stable after a rearrangement of tryptophan residues, while the B register requires only minor rearrangement.

(A) shows the initial configuration for the F register predicted by CCbuilder2. (B) shows the final configuration in the only stable replica. This register is stabilized by burying these hydrophobic residues. Conversely, (C) shows the initial configuration of the B register, which only minorly rearranges to the final position shown in (D). This register is stabilized by extensive polar interactions.

For each of these stable registers, we generated a trimeric coiled-coil and aligned the full N-terminal structure predicted by RaptorX with respect to the coiled-coil segment. The RaptorX structure predicts the flexible linker connecting the coiled-coil and RING finger domains as fully relaxed, which places the zinc-finger domains far from the base of the structure where they are seen in the published space-filling volume. To reproduce the cryo-EM surface volume, the RING fingers were manually repositioned at the base of the coiled-coil, introducing non-physical arrangements into the connecting loops. To relax the structure, we performed an MD simulation, restraining the helical structures for the first 10 ns held while the loops were free to rearrange, and the following 30 ns were completely unbiased for the full structure to seek an energy minimized state.. After 30 ns the structure was assessed to be equilibrated based on the interhelical distances of the coiled-coil, and the RMSD values of the RING finger domains.

2.4.4 Structural Evaluation of a Putative Stalk Structure

Coiled-coil structures tend to be stabilized by internal pairs of hydrophobic residues, typically Leucine and Valine[39]. In contrast, the N-termini of the CESA proteins are low in these motifs, which may rationalize why coiled-coil predictors like Marcoil do not predict that these segments will form coiled-coil structures[40], [41]. However, some coils, especially engineered ones, can also be facilitated by lysine interactions, which the CESA8 N-terminal has in abundance, and can be seen interacting with aspartic and glutamic acid residues to stabilize register B (Figure 2.3)[38]. We also see, in the single equilibration run of register F that remains structurally stable, a triad of tryptophan residues are forced together compared to the model predicted by CC builder. Based on our proposed structure, mutating these residues might disrupt stalk formation and provide insight towards its possible function in contributing to either trimerization of CESAs or subsequent assembly into rosette terminal complexes.

2.4.5 Cellulose Synthases Assemble into Trimers or Smaller

To investigate the long time scale stability of PttCESA8 polymers, coarse-grained (CG) models of the CESA8 monomers were constructed on the basis of the MARTINI v2.2p force field using CHARMM-GUI. Systems with identical numbers of lipids, free ions, and water molecules were generated containing 0,1, 2, 3, 4, and 6 copies of PttCESA8. For the initial coordinates, systems with three or fewer proteins were started from the homo-trimeric crystal structure (accession code: 6wlb), selecting the corresponding number of proteins. For systems with four or six proteins, the proteins were arranged with the hydrophobic 7th transmembrane helix oriented towards a neighboring protein as we assessed this was a crucial element of a stable structure. The protein assemblies were then embedded in a CG lipid membrane composed of 45% CHOL (cholesterol), 29% POPE (1-Palmitoyl-2-Oleoyl-sn-Glycero-3-Phosphoethanolamine), and 26% POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine)[42], along with MARTINI water and the appropriate number of counter ions for the 0.8 mM CaCl concentration[37]. For each number of proteins, three systems of different sizes were constructed to monitor the size effects. The systems sizes were: 1304 lipids and 28,125 water molecules, 3104 lipids and 66,948 water molecules, and 5000 lipids and 107,841 water molecules, according to the lipid ratios. The resulting energies are shown in Table 2.1.

Table 2.1. Per protein energies, each relative to the monomeric system for their system size.

Relative energy per protein (kJ/mol)	**Monomer**	**Dimer**	**Trimer**	**Tetramer**	**Hexamer**
Small System	0 ± 119	-38 ± 79	-15 ± 76	8900 ± 72	8820 ± 48
Medium System	0 ± 164	-189 ± 92	-284 ± 90	8705 ± 76	8726 ± 54
Large System	0 ± 206	-202 ± 111	-168 ± 94	8741 ± 62	8674 ± 52

From these values, it is clear that assembling more than three CESA proteins into a single assembly imposes a severe enthalpic penalty, which would eventually drive the complex towards smaller sub-assemblies. Despite this severe difference, the assemblies did not spontaneously decompose over the course of tens of microseconds. There were rearrangements in the assemblies, notably the tetrameric system closed an initial hole within the assembly over the course of the simulation, yet either the timescale was insufficient, or the smoothing of the Martini forcefield removed any local minima that might have been accessed to escape the extensive protein-protein interactions felt in the assembly. Furthermore, we are unable to confidently distinguish an enthalpic preference between monomer, dimer, and trimer assemblies. There is no evidence for monomeric instances of CESA proteins, so this is surely a shortcoming of the method. We attribute this inability to the lack of entropic measurements in our measurements that may be responsible for driving small numbers of proteins together.

2.4.6 Refuting of a Dimeric Crystal Structure

Given the inconclusiveness of our MD study, trimers and dimers remain the primary candidates for CSC lobe composition. Recently, Qiao et al. put forth their crystal structure (accession code:7ck3) as evidence for a potential dimeric intermediate, or final dimeric structure in the CESA assembly process[43]. However, while this data does helpfully update the structure of the CESA cytosolic region to atomistic resolution, it does not present a viable dimeric structure. Only a specific cytoplasmic segment has been expressed, which permits the two halves to interact in a specific way to form this structure. Homology modelling of the full structure using the SWISS-Model suite with PttCESA8 as the target shows this dimeric assembly by necessarily placing the transmembrane region of one protein and the cytoplasmic portion of the other within

the membrane no matter how they are arranged, shown in Figure 2.4. Examining further reveals considerable steric interference when the full structures are included (Appendix A.5).

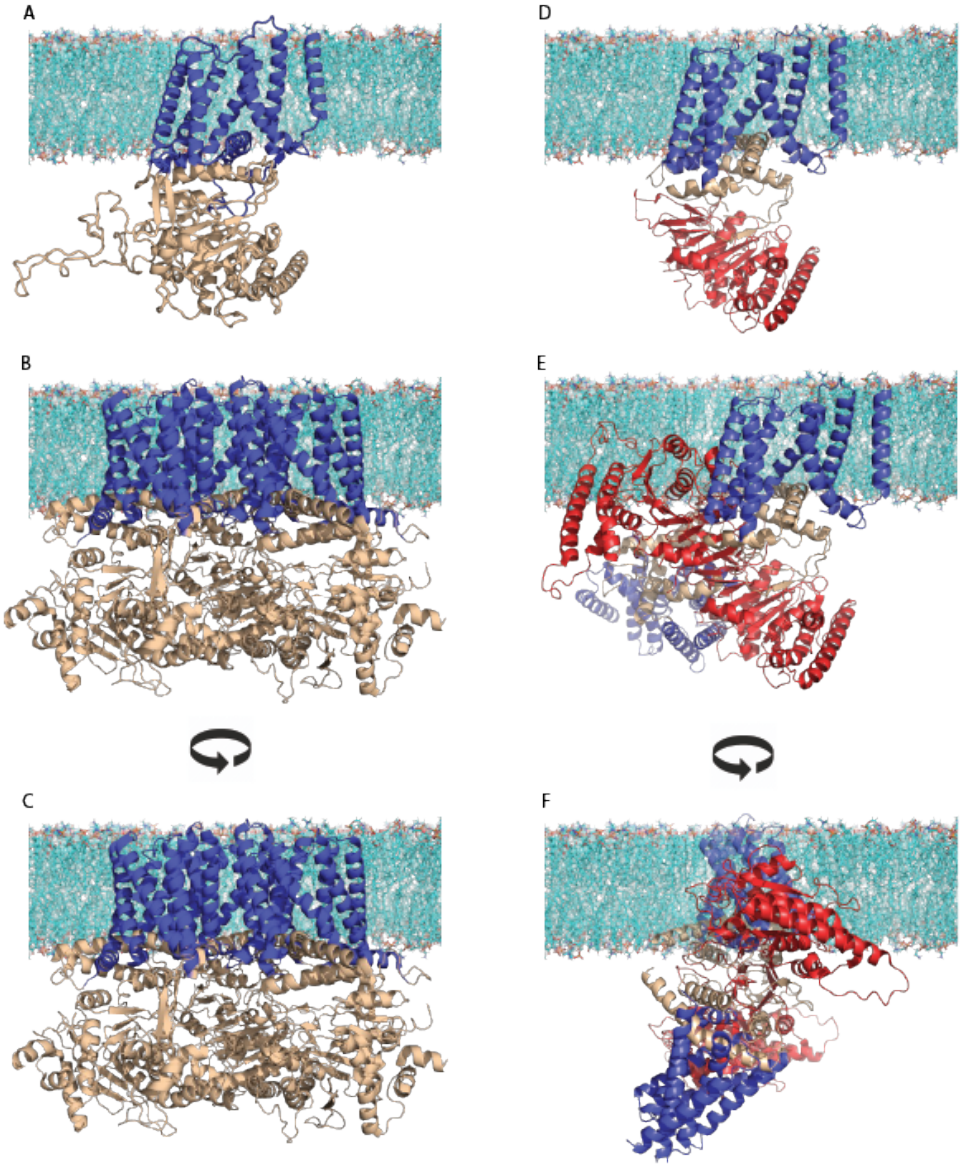

Figure 2.4. Comparison of membrane-integrated full dimeric crystal structure with PttCESA8 trimeric structure.

(A) shows the PttCESA8 monomeric structure with membrane region in blue. (B) Shows the crystallized fraction in red, with the unexpressed transmembrane region in blue and cytosolic region in tan. (C)-(D) show the full CESA trimer integrated into the membrane from the front and the side. (E)-(F) similarly show the dimeric crystal structure from the front and the side. It is clear that the crystal structure interface does not allow for the resulting dimer to successfully integrate into a membrane.

2.5 Discussion

In land plants, the CESA superfamily of proteins is responsible for synthesizing various cell wall polysaccharides. Of these, CESA and CSLD proteins are the most similar, and have been shown to be catalytically interchangeable in chimeric studies [6], [7]. In this study, we used comparative modeling to formalize the structural conservation between these protein families, and to determine key differences that may lead to the difference in localization. The first difference is a pair of insertions that we have called cellulose synthase-like D specific loops: DSL1 and DSL2. These insertions are a conserved feature in a wide range of plant CSLDs, though are variable within each protein family. Within *Arabidopsis*, these loops do not preclude the formation of trimeric lobes based on the *Ptt*CESA8 trimeric structure (Appendix A.3). It is unlikely that the DSL regions prevent the further assembly of lobes into CSC complexes, as a CESA6 chimera with the CSLD3 catalytic domain, including the DSL regions, was able to rescue a *cesa6* mutant[7]. Instead, they present additional surface-accessible area that may be involved in trafficking of the proteins..

The second structural difference between CESA and CSLD proteins is the ability to form an N-terminal trimeric coiled-coil. The *Ptt*CESA8 structure showed this novel feature, and only CESA proteins were predicted to form alpha helical segments capable of assembling this so-called stalk structure. CSLD proteins are not currently believed to form into larger CSC complexes, which may implicate the stalk structure in mediating its assembly. This structure may serve an additional role in localization of each protein, as the CESA6 and CSLD3 chimeras containing their native N-termini still localized to the correct environment. The mechanism for either of these functions is unclear, but of immense interest.

In order to better study potential N-termini structures, we presented a methodology to estimate the structure from low-resolution structures, including a visual representation despite

the lack of complete coordinates. Coiled-coils are one of the simplest and most widespread motifs in nature, consisting of two or more α-helices wound around each other in one of seven viable arrangements [35]. This simplicity allows us to combine a series of tools to predict the sequence responsible for forming the coiled-coil, generate all candidate structures, and evaluate them computationally[5], [25], [36], [40], [41]. The resulting stable structures can quantitatively compared to available coordinates, or qualitatively compared when only a visual is available, as is the case in this study. The relevance of this procedure is likely to increase alongside the rising adoption of cryo-EM and resulting boost in published structures [44]. For this study specifically, the methodology yielded two candidate structures for the *Ptt*CESA8 stalk structure. Without published coordinates, it was difficult to narrow down beyond these two stable structures. Each demonstrates unique interfaces that present clear targets for mutagenic studies to determine the true *in vivo* register.

Neither the difference in stalk structure nor the DSLs prevent the CSLDs from assuming the same polymer configuration as the CESAs, so we performed a coarse-grain MD study to quantify the most energetically favorable arrangement. Increasing evidence indicates trimeric lobes, but it is still an open question lacking a definite answer[11], [45]. Our methods were unable to distinguish between trimeric structures and smaller, likely due to the exclusion of entropy from the calculations, which would have required more rigorous and slow-to-converge sampling methods. We were able to quantify an order of magnitude difference between the enthalpies of trimeric and smaller systems, and tetrameric and hexameric systems. The slow kinetics of these systems, even at a coarse-grain resolution, prevented the energetically favorable dissolution of these assemblies over the course of our simulations. While these specific systems

should not require further study, it inform the potential study of full CSC complexes using similar procedures.

Chapter 3 A Temperature Sensitive FERONIA Mutant Allele that Alters Root Hair Growth

3.1 Chapter Introduction

In this chapter, we present a single-mutant protein with a profound resulting phenotype. My contribution to this work is in modeling and displaying the FERONIA structure (Figure 3.1), which showed the unintuitive finding that although a set of disaccharide coordinating residues are conserved compared to malectin, the overall structure is rearranged in such a way as to separate these conserved residues and bury them in surface inaccessible regions, abolishing the binding site. The manuscript has been revised to highlight these specific contributions, and the published paper contains additional details on experimental results and discussion.

3.2 Abstract

In plants, root hairs undergo a highly-polarized form of cell expansion called tip-growth, in which cell wall deposition is restricted to the root hair apex. In order to identify essential cellular components that might have been missed in earlier genetic screens, we identified conditional temperature sensitive (ts) root hair mutants by ethyl methanesulfonate mutagenesis in *Arabidopsis thaliana*. Here, we describe one of these mutants: feronia-temperature sensitive (*fer-ts*). Mutant *fer-ts* seedlings were unaffected at normal temperatures (20˚C), but failed to form

root hairs at elevated temperatures (30˚C). Map based-cloning and whole-genome sequencing revealed that *fer-ts* resulted from a G41S substitution in the extracellular domain of FERONIA (FER). Molecular modeling and sequence comparison with other Catharanthus roseus receptor-like kinase 1L (CrRLK1L) receptor family members revealed that the mutated glycine in *fer-ts* is highly conserved, but is not located within the recently characterized RALF23 and LORELI-LIKE-GLYCOPROTEIN 2 (LLG2) binding domains, perhaps suggesting that *fer-ts* phenotypes may not be directly due to loss of binding to RALF1 peptides.

3.3 Introduction

In land plants, root hairs are cellular protuberances resulting from the polarized outgrowth of specialized root epidermal cells, known as trichoblasts[46]. Root hair development can be divided into three phases: cell specification, initiation of bulge formation, and polarized tip growth[47]. Recently, a number of receptor-like kinases (RLKs) have been identified that are involved in cellular growth regulation, including the cell elongation associated with root hair tip growth[48]–[50]. In particular, CrRLK1L subfamily proteins, which includes *FERONIA* (*FER*)[51], *ERELUS* (*ERE*)[52], [53], *THESEUS1* (*THE1*)[54], and *ANXUR1/2*[55], have been implicated in cell wall sensing associated with a variety of cellular events such as female fertility, cell elongation, root-hair development, mechano-sensing, and responses to hormones and pathogens[50], [56], [57].

The CrRLK1L subfamily is named after the first member functionally characterized in *Catharanthus roseus* cell cultures[58], and Arabidopsis, contains 17 CrRLK1L subfamily members[59]. The majority of CrRLK1L receptor-like kinase proteins are predicted serine/threonine kinases with a single transmembrane domain between an N-terminal extracellular domain containing two tandem domains with limited homology to the carbohydrate-binding

domain of animal malectin proteins, and a C-terminal cytoplasmic kinase domain[57], [60]. *FER* was initially identified in the regulation of female control of fertility[51], and is highly expressed in the synergid cells of the female gametophyte and in a variety of vegetative tissues, but not in the male gametophyte[61], [62].

In addition to its important roles during fertilization, FER has also been shown to regulate aspects of root hair elongation[63], [64], calcium signaling during mechanical stimulation of roots[65], and cell wall responses to both abiotic and biotic stress[50], [51], [65]–[69]. FER was shown to bind to a ROP guanidine exchange factor 1 (ROPGEF1) that activates ROP GTPase signaling during root hair tip-growth in *Arabidopsis*[63]. *FER*, and other members of the CrRLK1L receptor family, have been proposed to bind to secreted rapid alkalinization factor (RALF) peptide ligands[52], with RALF1 binding the *FER* extracellular domain to suppress cell elongation of the primary root[52]. RALF34 binds to *THE1* during regulation of lateral root initiation[70], and RALF23 binding to *FER* negatively regulates plant immune responses[71]. *FER* was shown to associate with a glycosylphosphatidylinositol (GPI)-anchored co-receptor, *LORELEI* (*LRE*)[72], and a crystal structure of the FER extracellular domain bound to a LRE-like protein (LLG2), and RALF23 shows that LLG2 and RALF23 peptides associate with the second malectin-like domain[73].

Although RALF peptide binding occurs in several CrRLK1L receptor-like kinases, it remains unclear whether this is the only ligand recognized by this class of receptor-like kinases. While RALF peptides appear to only bind the second of two malectin-like domains in the FER extracellular domain, both these domains appear to associate with pectin polysaccharides[74]. We have identified a new temperature-sensitive mutation (*fer-ts*) in a highly conserved glycine residue (G41S) present in the extracellular domain of the FER receptor kinase, as well as other

members of the CrRLK1L receptor-like kinase family and mammalian malectin sequences. The *fer-ts* mutant exhibited rapid cessation of root hair tip-growth within one minute upon transferal to elevated (30^{o}C) growth temperatures. This rapid cessation of root hair tip growth was also observed when *fer-4* null mutants were transformed with a fluorescently-tagged version of the temperature-sensitive FER(G41S)-EYFP fusion protein. FER(G41S)-EYFP was still present in the plasma membrane in root hairs that had ceased tip-growth, indicating that the primary defect of this mutant is due to failure to properly transmit extracellular signals at elevated temperatures. These findings provide a powerful new tool for dissecting how FERONIA, and potentially other CrRLK1L receptor-like kinases, function during plant growth and development.

3.4 Results

3.4.1 The G41S mutation in *fer-ts* is a highly conserved glycine residue in CrRLK1L subfamily proteins

The G41S substitution of the *fer-ts* mutant rapidly inactivated *FER* function during root hair growth at elevated temperatures. In addition, a similar G37D mutation is responsible for inactivation of *THE1*, another member of the CrRLK1L family[54], and multiple sequence alignment analysis with other Arabidopsis CrRLK1L family members as well as animal malectin sequences showed that the G41 residue of FERONIA is absolutely conserved in these malectins and malectin-like 1 (ML1) domains (Figure 3.1a). Interestingly, based on structural studies of animal malectin proteins, five key residues (Y67, Y89, Y116, F117, D186; Figure 3.1b, red residues) were found to form contacts with a bound disaccharide ligand, nigerose, in the active site as determined by structural analysis of the *X. laevis* malectin protein[60], [75]. In this malectin structure these surface exposed residues extend from the malectin fold forming the nigerose binding pocket, with the conserved glycine (G40) located at the bottom of this structural region (Figure 3.1b). While several of the tyrosine and phenylalanine residues shown to be

important for interaction with carbohydrates in animal malectin proteins are maintained in plant malectin-like domains (e.g. FERONIA Y88, Y114, F115, D197) (Figure 3.1a), these are not surface exposed in the ML1 domain of the recently described FER crystal structure (Figure 3.1c, in green) with its co-receptor LLG2 (Figure 3.1c, in blue) and a RALF23 ligand (Figure 3.1c, in magenta)[73]. It is however, notable that in this structure the invariant glycine (G41; Figure 3.1c, red residue) of the FER ML1 domain is not located within or nearby the RALF23 and LLG2 binding surfaces in the ML2 domain[73].

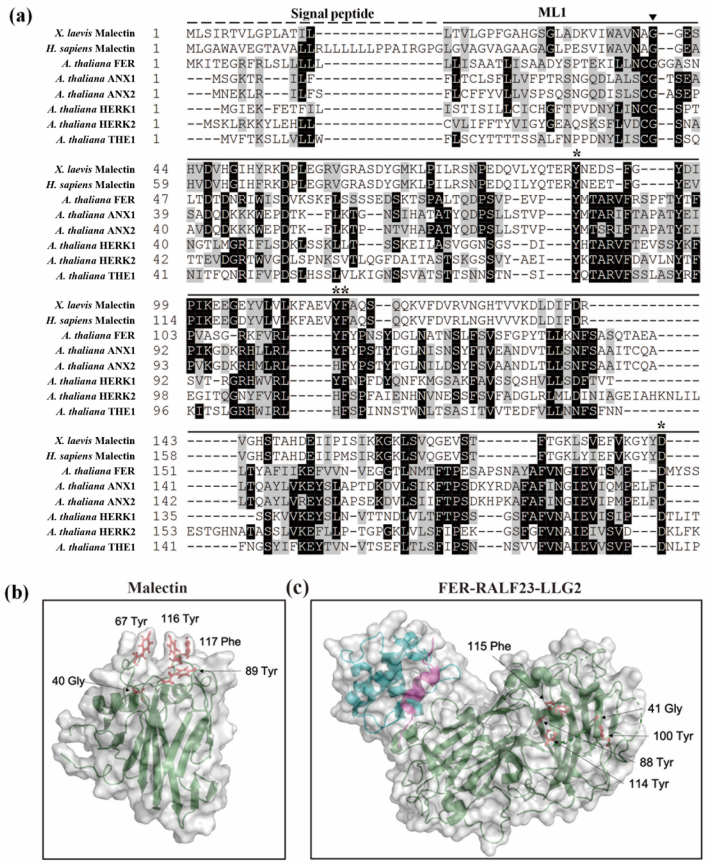

Figure 3.1. Localization of the highly conserved glycine mutation in *fer-ts* in *A. thaliana* FERONIA:RALF23:LLG2 and *X. laevis* malectin structures.

(a) Several residues important for binding carbohydrate ligands are conserved in plant CrRLK1L receptor kinase family members. Sequences analyzed include animal Malectin (X. laevis and H. sapiens) FERONIA and other well-characterized CrRLK1L homologs in A. thaliana (ANXUR1; ANX1, ANXUR1; ANX2, HERCULES1; HERK1 and THESEUS1; THE1). Putative N-terminal signal peptides are indicated as black dashed lines, and malectin and CrRLK1L ML1 domains by solid lines, respectively. The highly conserved G41 of FER is marked by arrowhead. Black boxes indicate fully conserved residues; shaded boxes indicate similar and partially conserved residues. Conserved residues that have been shown to participate in binding nigerose in the X.laevis malectin structure are marked by asterisks. Sequence alignment analysis was performed by CLUSTAL Omega program (http://www.ebi.ac.uk/Tools/msa/clustalo/) and displayed by using BOXSHADE software (www.ch.embnet.org/software/BOX_form.html). (b) Crystal structure of the X. laevis malectin protein (PDB ID: 2K46) with binding pocket aromatic residues and the highly conserved glycine residue based on sequence similarity to FER shown in red. (c) Crystal structure of FER protein (green) in complex with RALF23 ligand (magenta) and GPI-anchored protein LLG2 (blue) (PDB ID: 6A5E). No analogous binding pocket is observed on the ML1 domain, as all conserved aromatic residues (red) are buried within the protein. Both (b and c) were generated using PyMol (DeLano Scientific).

3.4.2 Pectin is not predicted to bind near the G41S mutation site

In addition to RALF23, which is bound at the distant site, *FER* is also known to interact with pectin[74]. To determine if a pectin site was nearby to the mutation site, I performed a ligand docking study with Rosetta[76]–[78]. I assessed pectin-binding proteins in the protein

databank and found three protein surfaces involved in binding: beta sheet surfaces, beta sheet clefts, and alpha helix clefts[79]–[81]. Within the FERONIA structure, the only beta sheet surfaces and clefts are on ML2, far from the mutation site. However, this is an alpha helical cleft directly adjacent to the mutant residue. I simplified pectin as a trisaccharide and generated 100 decoys docking pectin into the putative binding cleft. The result was an energetic score comparable with docking the pectin to any other surface of the protein, suggesting that pectin does not bind to nearby sites so is unlikely to be directly affected by this mutation.

3.4.3 Molecular dynamics studies show no temperature-dependent rearrangement

To determine whether a temperature dependent rearrangement occurred in the mutant *FER*, I used NAMD to simulate triplicate copies of the wild-type (WT) protein at 20ºC and 30ºC, and the mutant protein at 20ºC and 30ºC. Each protein was solvated and run for 30 ns. The final 10 ns of each system was used to calculate a root-mean squared fluctuation (RMSF) for each residue, and the relative change in RMSF with temperature was compared between the WT and mutant proteins. The result was internal variation in each system that exceeded the measured differences between the systems, making it unlikely that a significant rearrangement occurs spontaneously with a rise in temperature.

3.5 Discussion

In eukaryotes, receptor like kinases (RLKs) have been implicated to play an important role in many crucial eukaryotic cellular processes, such as cell cycle progression, cell signaling, embryogenesis, abiotic and biotic stress responses[48], [49], [82]. In this study, we isolated and identified a temperature-sensitive root hair elongation mutant, which we have determined is a new mutant *FER* allele that we have called *fer-ts*. The *fer-ts* mutant displays normal overall growth characteristics at normal temperature (20˚C), but root hair initiation and elongation are specifically

and rapidly inhibited within one minute upon transfer of these plants to elevated temperature (30˚C). We have shown that the *fer-ts* mutant is the result of a substitution mutation in which a highly conserved glycine residue in the FER extracellular domain is changed to serine (G41S). FERONIA is a member of the CrRLK1L subfamily of receptor-like kinases (RLKs) in *Arabidopsis* and the mutated glycine residue (G41S) is highly conserved in multiple members of the CrRLK1L family of receptor proteins as well as in animal malectin proteins.

FER has been implicated in a variety of plant processes, including roles in root hair tip growth as well as crucial plant processes, such as pollen tube reception, hypocotyl elongation, regulation of ABA signaling and controlling seed size[61], [63], [83]–[85]. In many of these processes, FER signaling appears to regulate ROS production. In constitutive *fer* mutants, ROS levels are reduced, and FER overexpression results in increased ROS levels. The observation that the *fer-ts* mutant also displays reduced ROS levels only at elevated temperatures suggests that this mutation affects FER signaling in a similar fashion as other *fer* mutants, perhaps providing a powerful tool for elucidation of downstream signaling events associated with FER function, and indicating that at least one important downstream effect of *FER* signal transduction is regulation of ROS production. This was elegantly explained by the discovery that FER recruits ROPGEFs, which in turn activate ROP GTPases, leading to the stimulation of RHD2 NADPH oxidase dependent ROS production[63]. Therefore, FER mediated regulation of ROS production is likely important and tightly controlled for many cellular functions.

Based on sequence comparison, the extracellular domains of members of the CrRLK1L subfamily of plant RLK proteins were predicted to share some structural similarity to the mammalian malectin protein[60]. Malectin was first identified and characterized in *X. laevis* as a carbohydrate binding protein in the endoplasmic reticulum where it plays an important role in the

early steps of protein N-glycosylation for biogenesis of glycoproteins[60]. Based on NMR structure analysis, there are five key residues in the malectin domain (Y67, Y89, Y116, F117, D186) that are located in a pocket-shaped structure, and these aromatic residues and the aspartate mediate interactions with the glucose residues of maltose and nigerose di-saccharide ligands[60]. In plants, malectin-like domains are mainly found in CrRLK1L subfamily and these display low overall sequence identity with animal malectins[86]. In FER, two malectin-like domains, ML1 and ML2, are found as a tandem-repeat in the extracellular domain. Interestingly, several key residues found in the ligand-binding pocket of the animal malectin structure are maintained in the malectin-like domains of FER and other plant CrRLK1L family members[60]. However, the discovery that members of a family of small secreted peptides, RALFs, rather than cell wall polysaccharides or oligosaccharides, serve as important ligands for FER and other CrRLK1L family receptors[52], [70], [71], [87] might indicate that these extracellular domains may interact with ligands in a manner distinct from their animal counterparts. Indeed the recent structural characterization of ANX1/2 extracellular domains[88] and the FER extracellular domain in complex with RALF23 and the FER co-receptor, LLG2[73] has shown that the RALF23 binding domain and interaction with LLG2 occurs primarily with the ML2 domain, and that conserved tyrosine and phenylalanine residues in CrRLK1L malectin folds in these structures appear to be buried within the ML1 fold, and therefore likely unavailable to interact with cell wall carbohydrates in a manner similar to animal malectins.

On the other hand, analysis of the animal and plant malectin domains, reveals an additional invariant glycine residue, that is present in all animal and plant malectin sequences, and which is also found in close proximity to pocket-shape ligand-binding cleft determined in the structure of the animal malectin protein. This invariant glycine is replaced with a serine (G41S) in the *fer-ts*

mutation described in this paper. The highly conserved nature of this glycine residue, and the rapid elimination of FER signaling at elevated temperatures, suggests a critical role for the FER ML1 domain in ligand binding or transduction of a ligand-binding signal in members of the CrRLK1L family of receptor-like kinases. Indeed, mutation of an analogous glycine residue to aspartic acid (G37D) in the extracellular domain of THESEUS in the *the1-1* mutant also results in a loss of function mutation in this RLK[54]. The *the1-1* mutation also results in its insensitivity to its specific RALF ligand, RALF34 *in vivo*[70]. However, RALF34 was still found to bind the extracellular domain of *the1-1* containing the G37D mutation *in vitro*, suggesting that mutation of this glycine residue did not directly impact RALF34 binding[70]. Similarly, the response of *fer-ts* mutant to treatment with RALF1 peptide was dramatically reduced under elevated temperature conditions. Precisely how the G41S *fer-ts* mutation, which is not found within the RALF23 peptide binding surface in the FER ML2 domain, would directly block RALF peptide perception and signaling is unclear, although it should be noted that the timescales of root growth analysis (>6 hrs) are significantly longer and therefore the increased protein turnover of *fer-ts* receptors may play a role not observed in the rapid cessation of tip-growth in root hairs (<1 min).

Additionally, the rapid cessation of tip-growth in *fer-ts* root hairs occurs even in the absence of added RALF1 peptide, perhaps indicating non-RALF peptide mediated signaling by FER receptor proteins. The recent discovery of links between *FER* signaling and pectin dynamics during salt stress[74], [89] and fertilization events[90] may provide insight into potential non-RALF peptide mediated signaling by FER and related CrRLK1L family proteins. During salt stress, *FER* appears to sense cell wall softening and both FER ML1 and FER ML2 domains were shown to directly interact with pectin *in vitro*[74]. More recently, *FER* function was shown to be required in order to maintain de-esterified pectin levels in the filiform apparatus during pollination

and fertilization events[90]. In addition, other CrRLK1L receptor kinases, such as *ERU*[53] and *THE1*[70] also appear to regulate aspects of pectin methyl esterification and cell wall integrity sensing, respectively. Whether the G41S mutation in *fer-ts*, or other analogous mutations of this invariant glycine residue in other CrRLK1L receptors affect the ability of these receptors to interact with or regulate pectin dynamics in plant cell walls is an intriguing possibility that warrants future investigation.

Chapter 4 Computational Tool Development

4.1 Chapter Introduction

Modern scientific workflows have access to vast quantities of data and computing power. Processing and interpreting this data can require a massive human endeavor, to the point of becoming completely untenable in some contexts. In this work, we will describe the development of two novel tools: a lobe detecting python program called InLoFi, and an On-the-fly Autocorrelation plug-in for the molecular dynamics software HOOMD-blue.

4.2 InLoFi: A reformed tool for quantification of lobed cells

Disclosure: The introduction of this chapter includes excerpts from Yang, J, Sahu, A, Gu, F., **Adams, A**, Mayes, H.B., and Nielsen, E. "Functional Relations of CSLD2, CSLD3, and CSLD5 Proteins during Cell Wall Synthesis." *In preparation*

4.2.1 Introduction

Plant cells are embedded within a load-bearing extracellular matrix, the cell wall. To change the size and shape of plant cells during growth and development, new cell wall material must be properly delivered and incorporated into existing cell walls during cell expansion. Two major mechanisms that control changes in plant cell shape are called diffuse growth and tip growth[91]. During diffuse growth, the primary load-bearing components of cell wall, cellulose microfibrils, are synthesized directly at the plasma membranes by large integral membrane protein complexes called cellulose synthase complexes (CSCs), comprised of multiple catalytic subunits encoded by CESA (Cellulose Synthase) proteins[8], [9]. The new cellulose microfibrils are deposited and integrated in ordered arrays along the entire expanding faces of the cells,

which are often transversely oriented to the major axis of cell expansion[92]. On the other hand, during tip growth, new cell wall material is selectively delivered and deposited by polarized secretion only at restricted plasma membrane domains within the cell, resulting in a highly polarized cellular expansion[93].

Interestingly, in *csld3* mutants, where the CSLD3 protein is not expressed, tip growth root hair is abolished, with root hair precursor cells undergoing cell rupture upon transition to tip-restricted elongation, resulting in a hairless phenotype in *csld3* mutants[94]. To further study the extent of this mutation, we examined the morphology of cotyledon pavement cells, which are characterized by interlocking lobed cells. The lobes of these cells form and extend through tip growth mechanisms, so may be altered by the loss of associated proteins. We used propidium iodide to stain the cotyledon cell walls, and imaged them with fluorescence microscopy (Figure 4.1). The cell boundary coordinates were manually recorded as a region of interest (ROI) file for quantification of the cells.

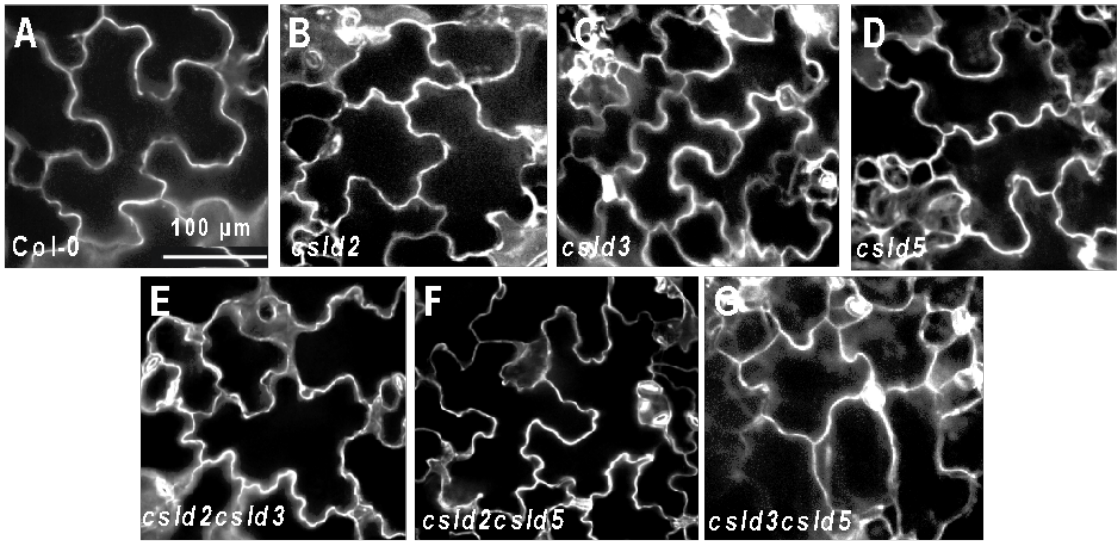

Figure 4.1. Fluorescence microscopy images of cotyledon pavement cells.

(A) shows wild-type cells, (B)-(D) show *csld* single mutants and (E)-(G) show *csld* double mutants.

Manually segmenting the cell coordinates into a 'core' and lobes is a time-intensive process involving numerous subjective judgements. The current automated solution, LobeFinder, is an open-source and freely available MatLab program that determines global properties of a cell, and the distance between cell points and a refined convex hull (Figure 4.2)[95]. The result is a non-unique, inverted form of the cell that markedly improves upon previous automated methods of identifying the number of lobes.

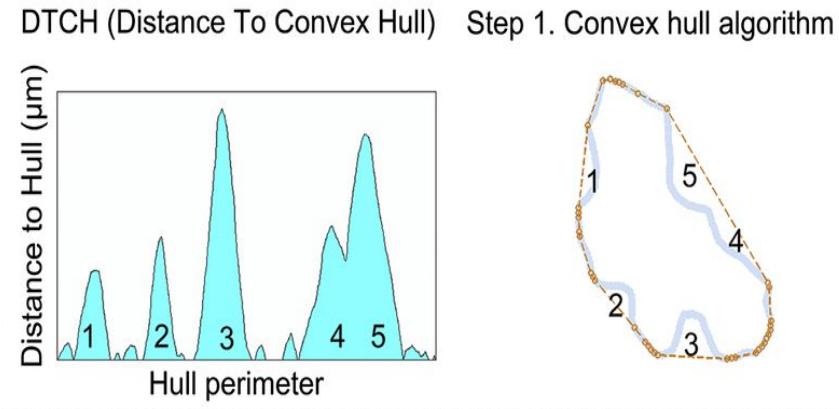

Figure 4.2. Visualization of LobeFinder output, adapated from*[95]*

However, individual lobe parameters, such as the area and height, cannot be determined from this construction, yet are of interest for comparing the change in directional growth resulting from knockout studies, such as the csld mutations. Thus, we present Inspired by LobeFinder (InLoFi), a program based on LobeFinder that directly quantifies lobe dimensions, freely available at https://github.com/xadams/Inspired-by-LobeFinder-InLoFi-.

4.2.2 Software Design

The InLoFi algorithm was developed to reproduce a base set of images with the lobes manually identified. The primary challenge is correctly determining the neck of each lobe: the pair of points along the cell perimeter that bound the lobe. Following this, various lobe dimensions can be trivially calculated. To determine these lobe necks, InLoFi takes cell boundary coordinates, manually identified from microscope images, and performs the following steps:

1. Use Shapely to determine the convex hull of the cell, and calculate the distance between each cell perimeter point and the hull[96].
2. Identify local maxima exceeding a set value. These will serve as the starting points for the lobe necks.
3. For each side lobe, check that the neck point can connect to more than 50% of the points within the lobe without crossing cell boundaries. If it cannot, move the to the next point and repeat.
4. With all of the lobes finalized, measure each area and discount any below a specific threshold. This has been shown to eliminate systematic errors between manual segmentation and automated scripts[97].

5. Quantify the lobe dimensions, currently including neck width, lobe height, lobe width, and lobe to cell area.

A graphical representation of the steps is shown in Figure 4.3.

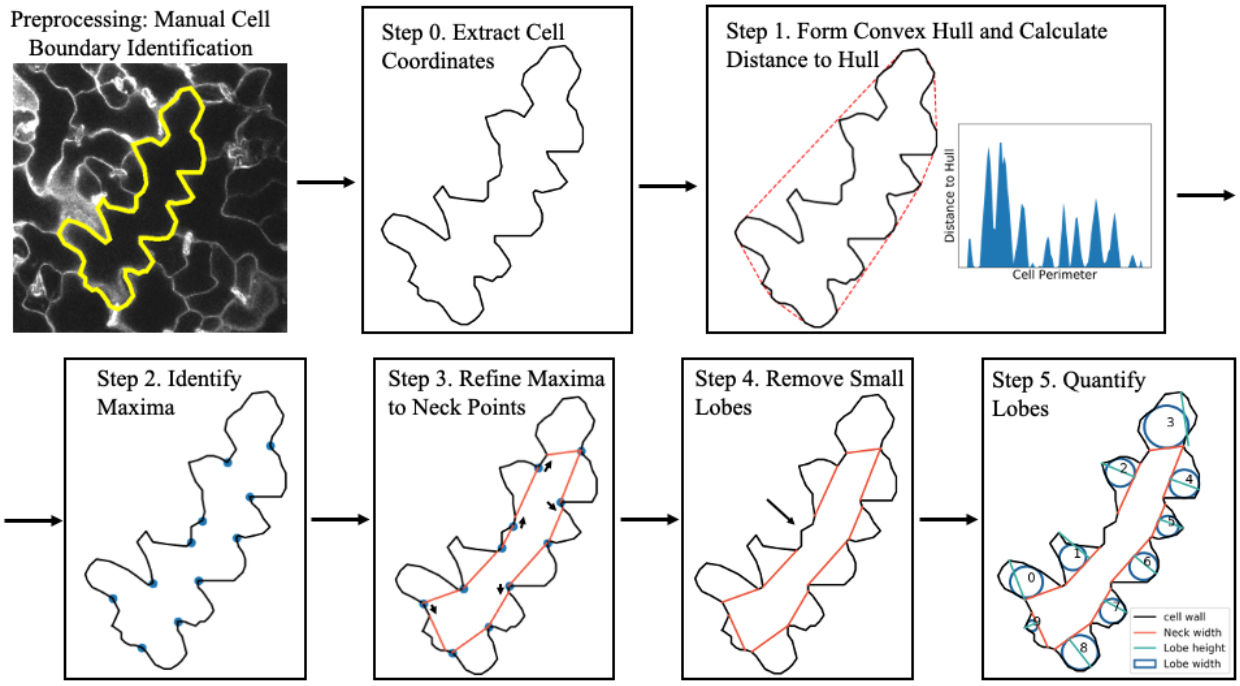

Figure 4.3. Schematic of InLoFi work flow.

In addition to the base functionality, InLoFi includes usability features including the option to append to existing files with redundant files automatically removed, as well as resiliency towards failures within individual files of the full submission, with helpful error messages. We developed a collection of unit tests that test basic functionality, as well as proper output based on a range of different valid inputs. No quantitative test of the output is currently in place, as it will not perfectly match any manually measured lobe quantities. However, manual inspection of the output files identify close agreement with manual segmentation in excess of 90% of cases.

4.2.3 Conclusions and Future Work

In this work, we present a novel program inspired by an existing tool, whose output could not be achieved by simply reprocessing the existing tool. It allows for high-throughput processing of manually identified lobed cell boundary coordinates with consistency and a high degree of accuracy compared to manual segmentation. Performance could be improved by adding additional features present in the original program, specifically resampling of coordinates and refinement of the convex hull. Introduction of these features should be straightforward, as they are based on the existing algorithm, and should increase accuracy by removing artifacts from manual identification of the cell boundary and improving the starting points of the lobe necks.

Following these implementations, more of the internal values can be made available to the user. Currently the user only specifies the input and output files, but parameters such as the lobe cutoff size and percentage of the lobe that must connect within the cell boundaries to the neck could be made accessible for a user to optimize to their specific systems. A final, lofty goal, would be to integrate an ImageJ script to automatically select the cell boundaries from an image, making the entire process quick and free of human subjectivity.

4.3 On-the-Fly Autocorrelation as a HOOMD-blue plug-in

Disclosure: This section includes material adapted from a final project report submitted for NERS 590: *Methods and Practice of Scientific Computing*, completed in collaboration with Alyssa Travitz Ph.D.

Alyssa was responsible for the original idea to implement the Likhtman Autocorrelator in HOOMD-blue, the function calling the algorithm that implements it, and the validation testing. I was responsible for unit tests regarding functionality, the general form of the unit tests, and the developer guide. We both were equally responsible for design of the **Correlator** class, its implementation in Python and C++, and the user documentation.

4.3.1 Introduction

Molecular dynamics (MD) is an N-body simulation method for studying systems of particles, often atoms or molecules. Pairwise interactions are defined between all particles in the system, and after forces between all neighboring particles are solved that that be integrated into Newton's equations of motions to dynamically evolve particle positions and velocities at each time step[98]. There are many Molecular Dynamics software packages, each with their respective strengths and shortcomings. In this project, we will be working with the software package HOOMD-blue (HOOMD), developed by the Glotzer Group here at the University of Michigan[99]. Our primary motivation for contributing to HOOMD is that we are able to work closely and in person with the lead developer, though HOOMD offers other advantages including powerful GPU acceleration and a python wrapper that makes scripting simulations flexible, capable, and accessible to much of the scientific community. HOOMD-blue is a general purpose particle simulation software, in that it also can perform hard particle Monte Carlo and dissipative particle dynamics, in addition to classical molecular dynamics. HOOMD is modular by design which facilitates extending its capabilities.

MD software can readily access system or per-particle data, such as volume, temperature, pressure (as a tensor), or particle velocities at each time step during a simulation. To access dynamic properties of the simulation, such as the complex modulus or diffusion coefficients, we can process the system data with time correlation functions (here on referred to as correlation functions)[100]. Correlation functions rely on the fluctuation-dissipation theorem, which states that a system's response to spontaneous fluctuations within a system is equivalent to its response to small applied force. We take advantage of the fluctuation dissipation theorem and use correlation functions to compute relaxation and transport properties of physical systems. For

example, we can autocorrelate the off-axis stress values to calculate the stress relaxation modulus of a system, or autocorrelate particle velocities to determine the related diffusion coefficients.

The general form of an autocorrelation function, or a correlation of a property with itself over a delay, is:

$$\mathrm{R(t)} = \frac{\sum_{i=1}^{N-\tau}(Y - \bar{Y})(Y_{i+\tau} - Y)}{\sum_{i=1}^{N}(Y_i - \bar{Y})^2} \tag{4.1}$$

where t is the time point for which the autocorrelation is being calculated, Y is the value of a given property at time i, τ is the lag time between measurements, and N is the total number of timesteps.

Calculating correlation functions can require significant compute time and memory, especially if per-particle calculations are necessary. A large part of this overhead is the time and memory required to write all data points to an output file, reading the data into a correlator, then writing an equally large data file containing the correlated values. In 2010, the Likhtman group proposed an algorithm that uses block averaging and a multiple-tau correlation method to produce highly efficient and accurate calculations[100]. As data is correlated, the correlated values at short time scales accumulate redundant data which can be represented accurately as the average of a given time block. Ramirez and coworkers demonstrate that their algorithm maintains accuracy through the block averaging, and that the final amount of raw data produced is greatly reduced. They also use a multiple-tau method, which controls the ratio between the averaging time and the lag time of the correlator. Parameters m and p are user defined as a way of controlling relative error, the smaller the ratio m/p is, the smaller the relative error. The recommended values (and default values in our implementation) are m = 2 and p = 16. This is crucial for calculations spanning very long time scales to reduce the memory and minimize error.

The general structure of the multiple-tau algorithm is as follows, but more detail is available in their paper.

- User-defined parameters: m, p , and S
- D_{ij}: array to store data (S+1 by p)
- C_{ij}: array to store correlation results (S+1 by p)
- N_{ij}: counter array for calculating averages (S+1 by p)
- A_i: accumulator (S+1)
- M_i: counter (S+1)

At each simulation time step, a new data value ω is sent to correlator level I, and the following algorithm takes place:

1. ω is stored at the first position of D, and all other data values are pushed up one value
2. The correlation array is updated as $C_{ij} = C_{ij} + D_{i0}D_{ij}$ and the correlation counter is incremented $N_{ij} = N_{ij} + 1$ (for maximum efficiency, this calculation iterates of j=0…p-1 at level 0 and j=p/m…p-1 for all other values)
3. ω is added to the accumulator as $A_i = A_i+\omega$, and the counter is incremented as $M_i = M_i+1$

When $M_i = m$, A_i/m is sent to the next level, (i+1), and A_i and M_i are reset to 0.

The Likhtman group provides their code as a standalone C++ package (hereon referred to as the Likhtman correlator) that can be used as a post-processing method. However, the algorithm is intended to be used during run time as an on the fly method to avoid the large memory and requirements of writing each data point into an external file, then reading the data into the correlator. Our ultimate goal is to bring this efficiency to the HOOMD-blue software package.

4.3.2 Software Design

Our design decisions while developing this plug-in were motivated by our conversations with HOOMD's lead developer, Joshua Anderson, Ph.D., and our own experiences as HOOMD users. The ultimate goal is to develop a plug-in correlator that is simple and customizable. For simplicity, the only required arguments are the quantities to be correlated and the period at which to send values to the correlator; all other arguments are optional and set to default values if no argument is passed. A powerful tool within HOOMD is the ability to use a callback function, which allows the user to define a lamba function that is evaluated at every time step. We defined the "quantities" argument as a HOOMD vector string, which allows the user to pass callback functions to the correlator in the same way they would pass to a standard HOOMD logger. To implement this, the C++ **Correlator** class inherits from the **Logger** class. However, the **Correlator** directly inherits from the **Analyzer** class in Python, since Logger is not an HOOMD-defined Python class. This gives us all necessary functionalities because **Logger** inherits directly from **Analyzer**, so all inherited methods are preserved.

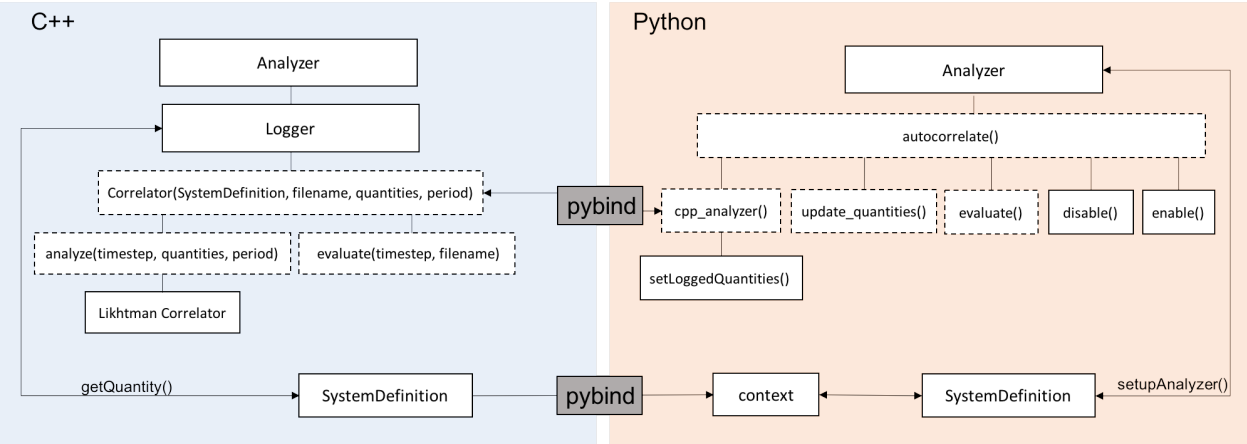

Figure 4.4 Schematic of HOOMD-blue plug-in architecture.

Solid boxes indicate external code (HOOMD main code or the Likhtman Correlator), dashed boxes represent our plug-in. Lines represent inheritance, from top down. Arrows represent communication between components.

We integrated the Likhtman correlator into the architecture of HOOMD-blue, evaluated its accuracy and computational performance, and developed both user documentation and a developer guide. The plug-in is open-source and freely available on GitHub (https://github.com/atravitz/correlator-hoomd-plugin), along with instructions for installing and running the plug-in. Figure 4.4 is a simplified diagram of HOOMD-blue's architecture and depicts how the plug-in is integrated into each level of the software. HOOMD is considerably larger and more complex than we were able to diagram, so we only show the relevant pieces of code that our plug-in directly inherits from or communicates with. HOOMD is a multi-language software package. The user interfaces with HOOMD entirely through Python calls within a script, but the back-end is written entirely in C++. HOOMD uses pybind to communicate between the Python and C++ levels. For this level of implementation, we provided the user with high-level options (file name, quantities to correlate, correlation and logging frequencies) but hard-coded the recommended algorithm parameters (m, p, S) into the C++ code.

Part of the inherent advantage of the Likhtman correlator is its separate functions for correlating and evaluating. During the correlating stage, no values are being actively written to a file, which saves overhead. The evaulate() call then iterates over the current data and writes it to a file. This means that the current state of the correlator can be written to a file at any given time step. This flexible output method is useful but must be used carefully. Dumping the data at every time step would defeat the purpose of the on-the-fly method, but writing the data exclusively at the end of the run can be risky for very long simulations that may experience instabilities or exceed walltime. Therefore, by default we set the correlator to write the data at the end of the simulation, but also created the "eval_period" parameter. The "eval_period" allows the user to specify how frequently the correlator data is written to a file. This is recommended to be only a

few times throughout the simulation, as a sort of "safety," or to allow the user to check in on the progress of their simulation in real time.

We also realize that it may be of interest to evaluate the correlator at specific points in a simulation, such as after volume or temperature are changed in connection with equilibration or other processes. To address this, users can call correlate.evaluate() at a specific line in their HOOMD code. The evaluate method is entirely independent from the primary correlator, but both methods access the same instance of LikhtmanCorrelator. This highlights the importance of our decision to define the instance of LikhtmanCorrelator as a class attribute, so that both methods can access the correlator.

4.3.3 Testing and Validation

We designed regression and unit tests to verify the functionality and efficacy of the plug-in as we continued to develop it. We test basic functionality such as proper initialization and output as simple unit tests. **TestValues** performs a short simulation and compares Correlator plug-in data to a verified post-processing workflow with the same data. Best-practice frameworks such as removing output from routine testing and use of variables are employed to assist with altering or expanding unit-testing as further functionality is added.

The performance of the Likhtman algorithm is well studied in[100], which demonstrates that it is more accurate and efficient when compared with a standard single-tau correlator. Therefore, the purpose of our performance testing is not to demonstrate the performance of the algorithm, but to demonstrate accurate implementation and quantitatively compare the efficiency of the correlator plug-in with the correlator used as a post-processing method. The motivation for conducting this performance test is to ensure that our implementation is efficient in its method of

communicating the simulation values to the correlator. It is possible, if the plug-in is poorly designed, that the on the fly method would be slower than the post-process method.

We chose a bead-spring network as our testing system because it is computationally inexpensive and we are familiar with its stress relaxation behavior from previous research. The network is composed of spheres connected by dumbbell springs, as seen in Figure 4.5b. The xy component of the pressure tensor matrix is the correlated property. All units are left as reduced units, since there is no physical relevance to the system and it is used simply for validation purposes.

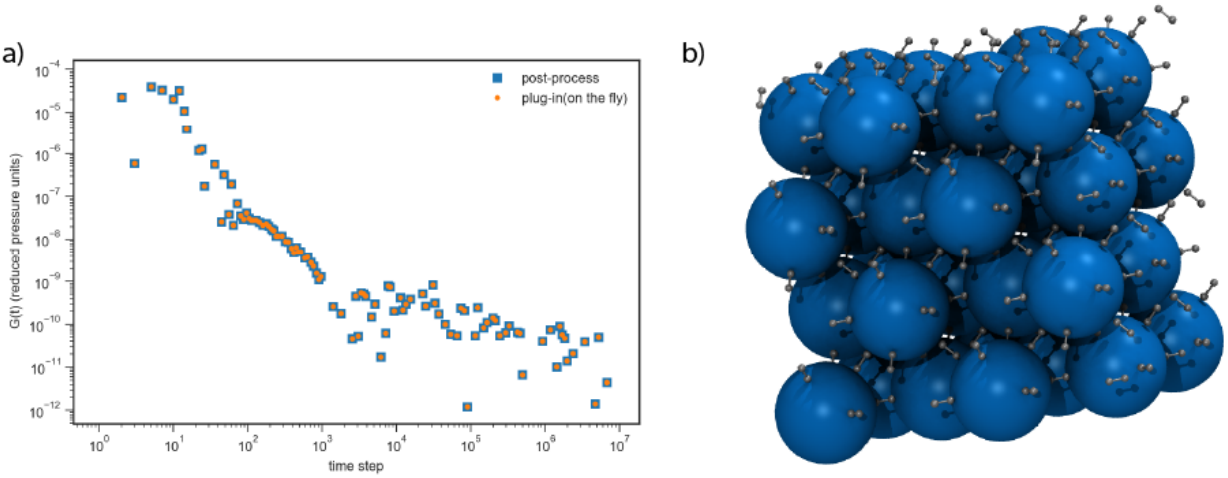

Figure 4.4.4 (a) Autocorrelation of xy component of pressure implemented on the fly and as a post-processing method.

All units are reduced units. Both simulations were equilibrated for $1x10^5$ time steps, then data was sent to the correlator every 10 time steps for 1 x 10^7 time steps. (b) Visualization of the simulation initial configuration. Large spheres are initialized on a face centered cubic lattice and connected with nearest neighbors by harmonic spring dumbbells[101], [102].

4.3.4 Testing and Validation

The same bead-spring network system used to validate the plug-in was also used for performance testing. The plug-in was tested by equilibrating the system, initializing the correlator, running for the given number of time steps, then evaluating the correlator. The post-processed correlation was tested by equilibrating the system, initializing a logger (a HOOMD method), then running for the given number of time steps. The raw data from the logger was then formatted as a Likhtman correlator input script and passed to the Likhtman correlator. The results of this study, Figure 4.6, show that the plug-in is consistently more efficient than the post-processed implementation. Note that in unit testing validation both cases were run in a single simulation, but to assess the performance of the methods we ran the cases as entirely separate simulations.

Table 4.4.1: Run Time Analysis of On the Fly and Post-Processed Data

# time steps	on the fly (sec)	post-process (sec)	seconds per time-step
1.00E+06	151	161	1.50E-05
5.00E+06	669	772	206E-05
1.00E+07	1330	1572	2.42E-05
5.00E+07	6497	8933	4.87E-05
1.00E+08	13446	15785	2.34E-05
		average	**2.64E-05**

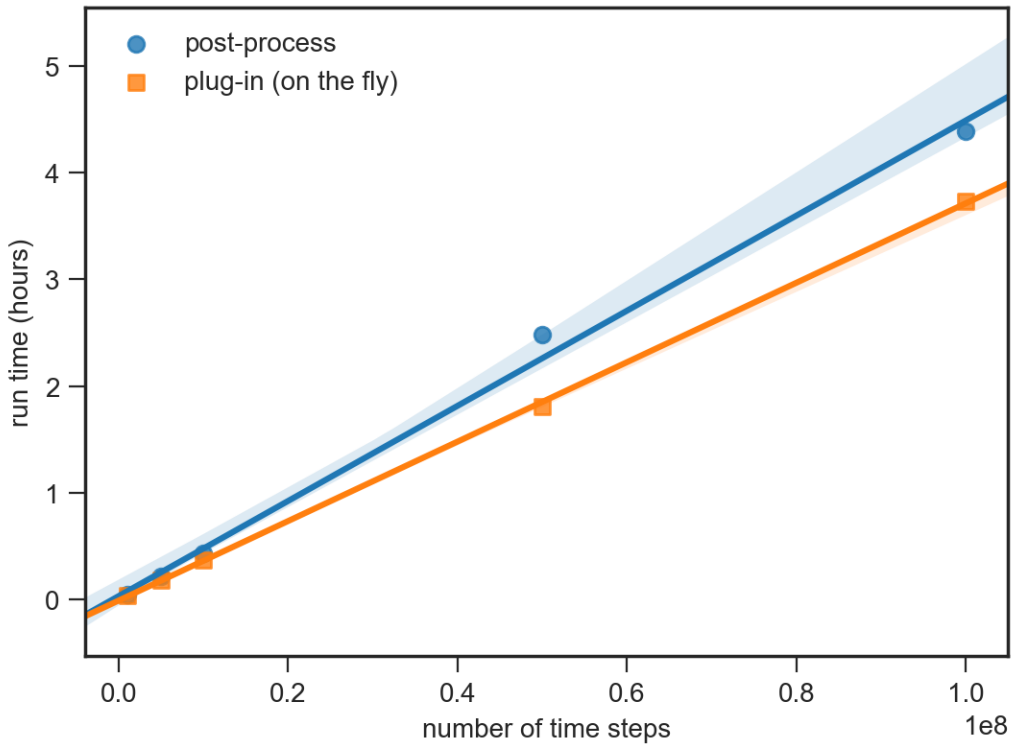

Figure 4.4.5 Performance data for post-processed and plug-in data.

Number of time steps refers to the number of time steps after $1x10^5$ equilibration run. Shaded regions represent 95% confidence interval.

It is important to note that quantifying this improved efficiency as a percent speed increase would be arbitrary, as the reduced time is only a function time steps at which data is written. Instead, we present the decrease in time per time step as a metric for understanding the relative speeds. This is a rough calculation, but gives the order of magnitude of time that using the correlator as a plug-in offers as a speed up. It is common for simulations to run for $1x10^8$ to $1x10^9$ time steps, occasionally more. This translates to a time savings of 45 minutes to 7.5 hours

for a single property to be calculated. For simulations where per-particle properties must be calculated, this time savings scales with the number of particles which is often on the order of $1x10^8$.

The improved speed is only part of the advantage of the on the fly implementation. To use the correlator as a post-processing method, we must write every time step (or every period) to a file, then parse that file into input file for the correlator. This can result in data files that are several gigabytes for very long time scales. Just as with the run time, if we must calculate per-particle values, this amount of memory can difficult to store and manage. The autocorrelation output files are only on the order of kilobytes, essentially eliminating the memory storage burden of the correlator.

4.3.5 Conclusions and Future Work

In this work, we produced a functioning autocorrelation plug-in. It includes additional generalizable features such as the **evaluate()** function and period logging for backup. Currently, we have only tested on individual CPUs, and have not attempted running on GPUs or in parallel. Incorporating GPU compatibility should be a relatively straightforward process of adding in CUDA capabilities through flags in the plug-in. Achieving a correlator that works with a simulation running in parallel should also be straightforward as the entire plugin is lightweight and could be performed on a singular node without performance losses. The last high priority feature is restartable jobs. The framework is there in the periodic logging capabilities, but the process of reading in the multi-tau arrays to initialize the Liktman correlator is non-trivial and was not a priority to develop.

Another area of improvement is adding more user-defined parameters. For example, the Likhtman correlator allows for m, p, and S parameters to be defined as a means of tuning error

and efficiency. These would be simple to add into our plug-in, but would require additional unit testing. Other helpful plug-in arguments might be file parameters such as headers and delimiters to make plotting and processing the correlated data simpler for the user. It would also be beneficial to the user to include helpful error messages and warnings, which will be developed based on experiences with using the plug-in. Although Likhtman's mutiple-tau correlator is favorable for our research, some HOOMD users might prefer a linear correlator. In the future, we can add a linear correlation algorithm into the same plug-in.

Chapter 5 Conclusions, Limitations, and Future Work

5.1 Research Summary

In this work, we present discoveries at the intersection of decades of experimental insight, novel information and a breadth of methods.

In Chapter 2, use a variety of methods to elucidate a novel structure. The publication of a single structure provides a foothold for sequence analysis, *de novo* structure prediction, comparative modelling, and molecular dynamics studies to work in concert towards insights regarding the structure itself, as well as homologous proteins in a different organism. We are able to address questions regarding structure, assembly, and key differences between largely interchangeable protein families.

In Chapter 3, we once again employ a swathe of experimental and computational techniques to discover and interrogate the mechanism of a novel mutant: Feronia temperature-sensitive. This single mutant demonstrates several interesting effects: a loss of root-hair growth at elevated temperatures, an insensitivity to the RALF1 ligand, and accumulation of reactive oxygen species. While the direct mechanism of these effects could not be explained, several hypotheses were investigated, narrowing the scope of future studies. More broadly, this mutant serves as a valuable tool for investigation of any of the pathways that it is involved with.

In Chapter 4, we present multiple computational tools developed to better leverage known information. InLoFi allows for the processing of microscopy images of lobed cells in a way that is tenable and consistent compared to manual segmentation, and more descriptive of individual lobes than the alternative automated method. The Autocorrelator plug-in in HOOMD-blue implements a known algorithm that dramatically reduces the computation time and storage requirements vs comparable post-processing methods, while allowing access to the scalability and powerful scripting capabilities of HOOMD-blue. Both of these tools are narrow in scope, but

were deliberately developed to be flexible and resilient so they can tuned as workflows change, or adapted for entirely new purposes.

5.2 Limitations and Future Work

While the above work demonstrates novel insights and approaches for these biological problems, many obstacles remain for their generalizability and conclusiveness.

Chiefly, biological systems are extremely heterogeneous and complex. Nearly a dozen accessory proteins have been implicated in assembling, trafficking, or regulating the CSC during cellulose production, miring the ability of computational studies to make tight, conclusive arguments[103]. More frequently, dynamics studies are targeted at specific hypotheses, as was done in Chapter 3. However, given the disparity in time, effort, and expense between computation and the study of live organisms, any narrowing of the possibility space is still of tremendous value.

The multi-dimensional nature of these systems made it difficult to establish satisfying controls of known or suspected facts within the plant biology community. Where we strived to quantitatively confirm that trimeric assemblies were the most energetically favorable and then utilize the lens of molecular dynamics to elucidate why, we instead struggled to differentiate between trimeric and monomeric systems when there is no belief that monomeric systems are viable. This further endangers our future efforts to computationally distinguish between homomeric and heteromeric complexes of CESA and CSLD proteins, as the differences in these systems are more subtle than between systems with entirely different numbers of proteins. Most likely, contributions due to entropy account for the differences, but these are computationally expensive to reliably determine, particularly given the coarse-grained and heterogenous nature of the membrane and larger system.

Despite this, computational studies can still inform and drive the experimental work that will more decisively drive the field forward. Specifically, molecular scale insights into the architecture of the CESA lobes and the putative stalk structure allow for the design of cross-linking experiments that can verify the interfaces predicted by these models. In the case of the stalk structure, determination of the primary stabilizing interfaces shows where to disrupt these bonds to confirm the proposed structure.

The methodology to narrow down candidate coiled-coil structures is a smaller system that equilibrates on a much shorter timescale[38]. This means that it would be much more capable of addressing a question such as the preference for homomeric or heteromeric trimers by rapidly testing the relative stabilities of various trimeric coiled-coil candidate stalk structures. Similarly, another candidate system for study is the dynamic assembly of glucan chains into a cellulose microfibril. This system will be on the scale of the coiled-coil systems, and will provide crucial information towards the grand question of whether CSCs are natively stable, or are driven together by the forces of crystallizing cellulose.

Appendix

Supplementary Information for: Comparative Modeling & *De Novo* structure predictions reveal extensive conservation and key differences in *Arabidopsis* CESA and CSLD proteins

A.1 Equations

A.1.1 Calculation of Relative Per-Protein Energies

R_{memb} = reference energy of membrane and solvent
E_n = raw energy of the system with number of proteins N
$R_{monomer}$ = reference energy of monomer with counterions = $E_1 - R_{memb}$
ΔE_n = energy relative to the monomer

$$\Delta E_N = \frac{E_N - R_{memb}}{N} - R_{monomer}$$

A.1.2 Propagation of Uncertainty of Relative Per-Protein Energies

$$\Delta\sigma_N = \sqrt{\sigma_N^2 + \sigma_{memb}^2 + N\sigma_1^2 + N\sigma_{memb}^2}$$

A.2 Sequence Alignments

All sequences were aligned with ClustalW (https://www.genome.jp/tools-bin/clustalw) and visualized with ESPript (https://espript.ibcp.fr)[104].

A.2.1 RING finger Sequence Alignment

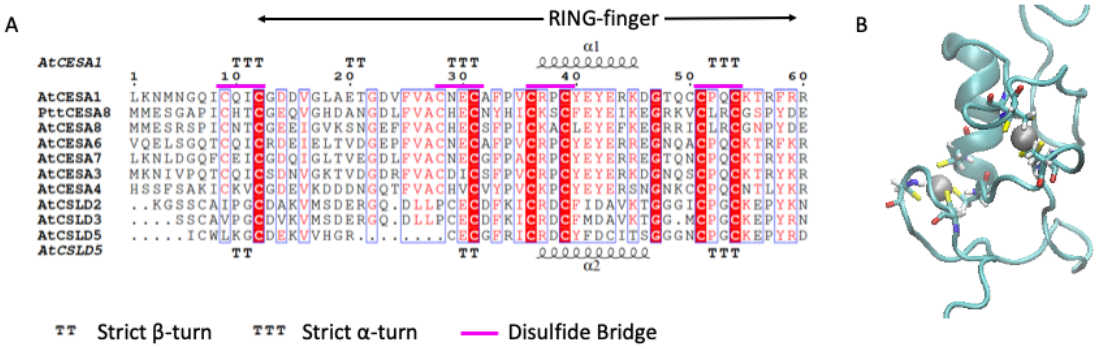

Figure A.1. RING finger regions within the N-terminal demonstrate sequence variation but conserved structures.

(A) shows the alignment of *At*CESA and *At*CSLD proteins in the vicinity of the predicted RING finger domain. (B) shows a representative RING finger domain with cysteine residues shown in licorice representation and Zn^{2+} ions as spheres.

A.2.2 P-CR Sequence Alignment

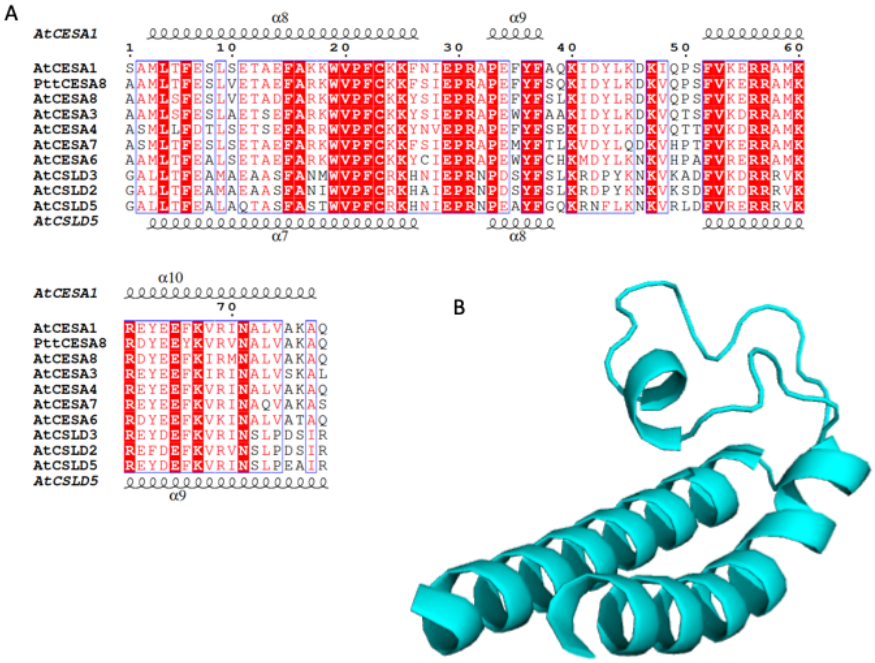

Figure A.2. P-CR structure is well conserved across protein families.

(A) shows the alignment of *At*CESA and *At*CSLD proteins in the vicinity of the P-CR domain. (B) shows a representative P-CR domain in cartoon representation.

A.2.3 CSR Sequence Alignment

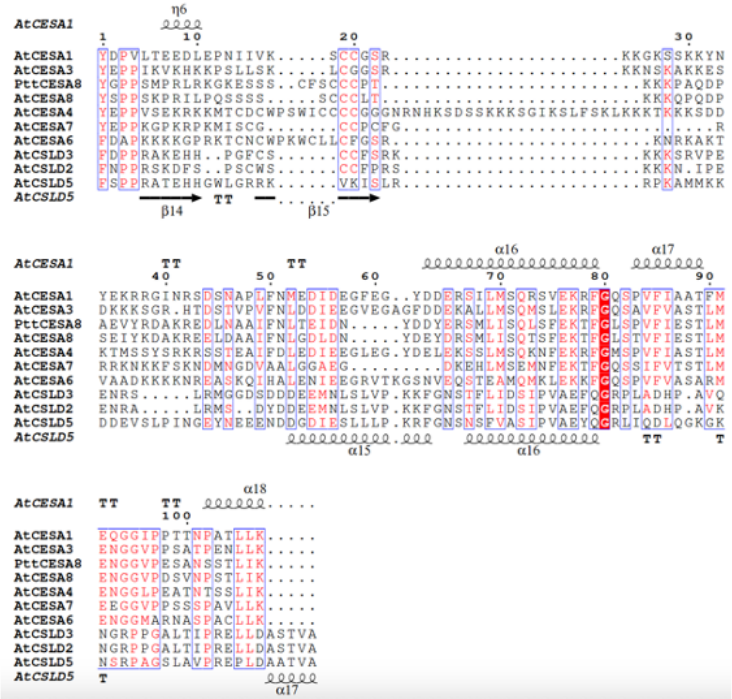

Figure A.3. CSRs are intrinsically disordered and highly variable within and across protein families.

A.3 CESA/CSLD Trimer Comparison

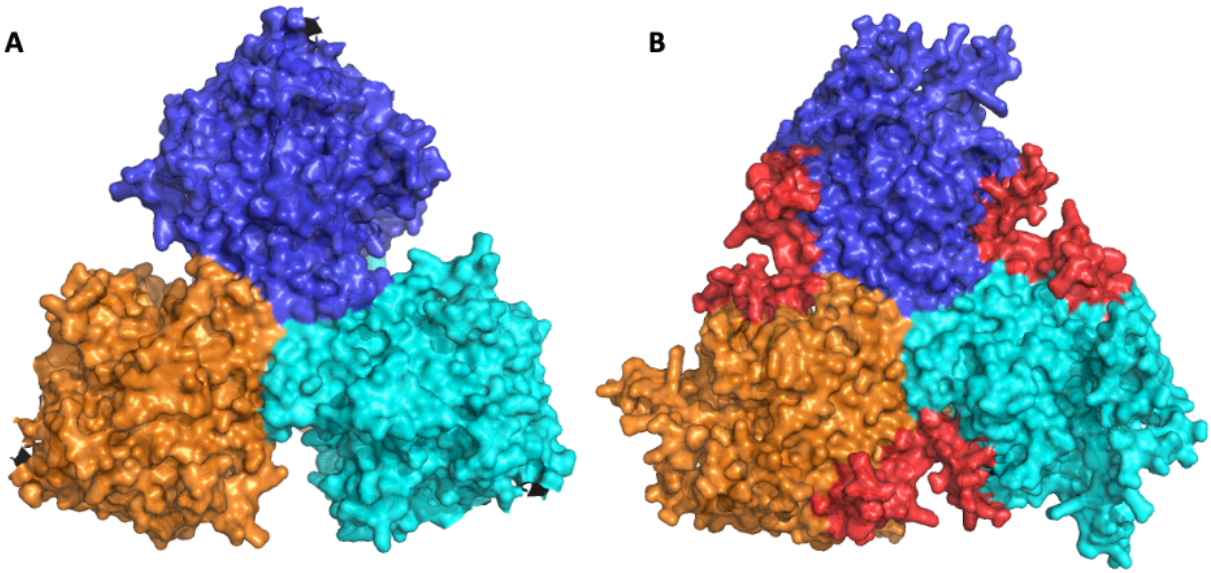

Figure A.4. DSL regions of *At*CSLDs do not disrupt trimeric interfaces.

(A) shows a surface representation of a representative CESA trimer: CESA6. (B) shows an analogous CSLD trimer: CSLD3, with the DSL regions shown in red.

A.4 Schematic for Generating Candidate Structures from Low Resolution Structures

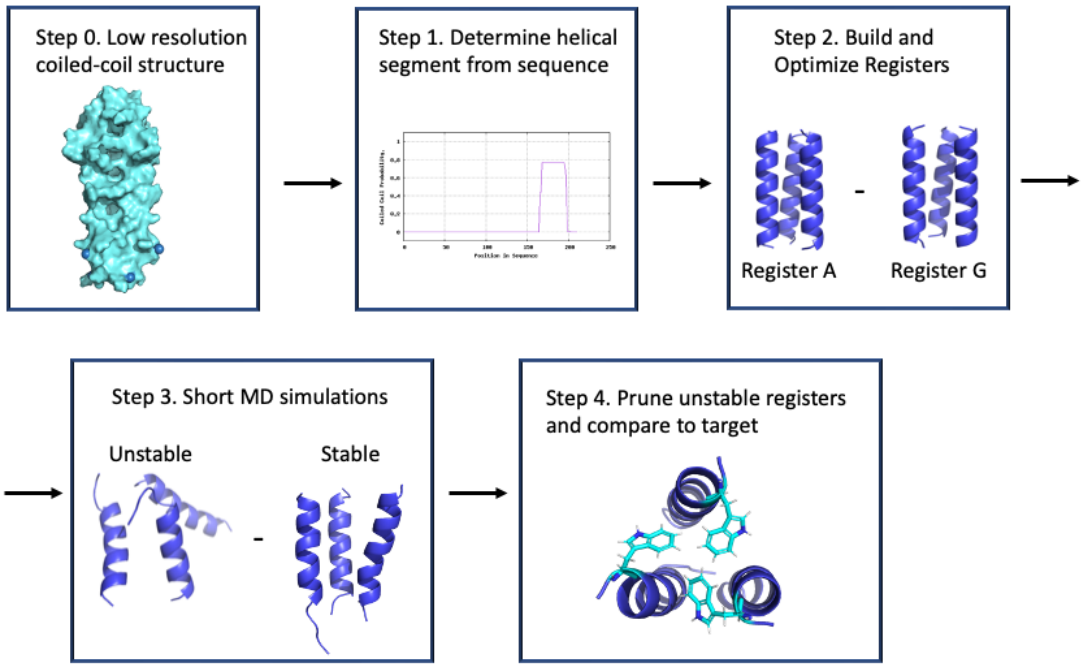

Figure A.5. Schematic of Coiled-coil candidate structure work flow.

A.5 Steric Clashes in a Full Putative Dimeric Structure

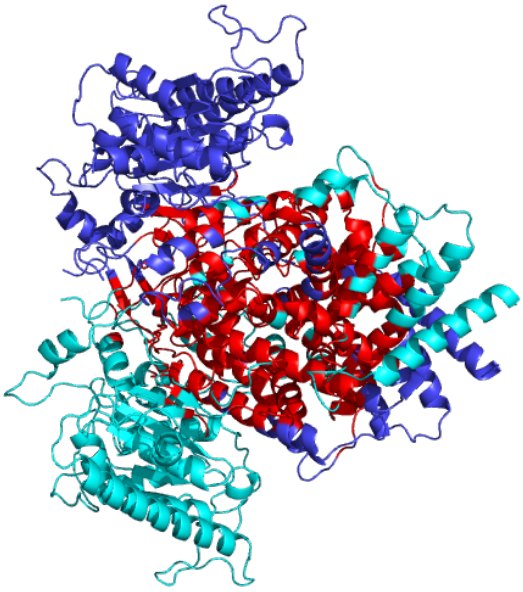

Figure A.6 Consideration of the full CESA structure for a putative dimeric structure introduces extensive steric clashes (red). The members of the dimer are shown in blue and cyan.

Bibliography

[1] M. J. Meents, Y. Watanabe, and A. L. Samuels, "The cell biology of secondary cell wall biosynthesis," *Ann. Bot.*, vol. 121, no. 6, pp. 1107–1125, May 2018, doi: 10.1093/aob/mcy005.

[2] T. Arioli *et al.*, "Molecular Analysis of Cellulose Biosynthesis in *Arabidopsis*," *Science*, vol. 279, no. 5351, pp. 717–720, Jan. 1998, doi: 10.1126/science.279.5351.717.

[3] M. Fagard *et al.*, "PROCUSTE1 Encodes a Cellulose Synthase Required for Normal Cell Elongation Specifically in Roots and Dark-Grown Hypocotyls of Arabidopsis".

[4] S. Persson *et al.*, "Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in *Arabidopsis*," *Proc. Natl. Acad. Sci.*, vol. 104, no. 39, pp. 15566–15571, Sep. 2007, doi: 10.1073/pnas.0706592104.

[5] J. Jumper *et al.*, "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, no. 7873, pp. 583–589, Aug. 2021, doi: 10.1038/s41586-021-03819-2.

[6] S. Park, A. L. Szumlanski, F. Gu, F. Guo, and E. Nielsen, "A role for CSLD3 during cell-wall synthesis in apical plasma membranes of tip-growing root-hair cells," *Nat. Cell Biol.*, vol. 13, no. 8, pp. 973–980, Aug. 2011, doi: 10.1038/ncb2294.

[7] J. Yang *et al.*, "Biochemical and Genetic Analysis Identify CSLD3 as a beta-1,4-Glucan Synthase That Functions during Plant Cell Wall Synthesis," *Plant Cell*, vol. 32, no. 5, pp. 1749–1767, May 2020, doi: 10.1105/tpc.19.00637.

[8] S. C. Mueller and R. M. Brown, "Evidence for an intramembrane component associated with a cellulose microfibril-synthesizing complex in higher plants.," *J. Cell Biol.*, vol. 84, no. 2, pp. 315–326, Feb. 1980, doi: 10.1083/jcb.84.2.315.

[9] T. H. Giddings, D. L. Brower, and L. A. Staehelin, "Visualization of particle complexes in the plasma membrane of Micrasterias denticulata associated with the formation of cellulose fibrils in primary and secondary cell walls.," *J. Cell Biol.*, vol. 84, no. 2, pp. 327–339, Feb. 1980, doi: 10.1083/jcb.84.2.327.

[10] W.-R. Scheible and M. Pauly, "Glycosyltransferases and cell wall biosynthesis: novel players and insights," *Curr. Opin. Plant Biol.*, vol. 7, no. 3, pp. 285–295, Jun. 2004, doi: 10.1016/j.pbi.2004.03.006.

[11] P. Purushotham, R. Ho, and J. Zimmer, "Architecture of a catalytically active homotrimeric plant cellulose synthase complex," *Science*, vol. 369, no. 6507, pp. 1089–1094, Aug. 2020, doi: 10.1126/science.abb2978.

[12] A. Waterhouse *et al.*, "SWISS-MODEL: homology modelling of protein structures and complexes," *Nucleic Acids Res.*, vol. 46, no. W1, pp. W296–W303, Jul. 2018, doi: 10.1093/nar/gky427.

[13] T. Richmond and C. R. Somerville, "The Cellulose Synthase Superfamily," *Plant Physiol.*, vol. 124, pp. 495–498, 2000.

[14] A. Little *et al.*, "Revised Phylogeny of the *Cellulose Synthase* Gene Superfamily: Insights into Cell Wall Evolution," *Plant Physiol.*, vol. 177, no. 3, pp. 1124–1141, Jul. 2018, doi: 10.1104/pp.17.01718.

[15] M. S. Doblin, L. D. Melis, E. Newbigin, A. Bacic, and S. M. Read, “Pollen Tubes of Nicotiana alata Express Two Genes from Different ␣-Glucan Synthase Families,” vol. 125, 2001.
[16] S. Jo, T. Kim, V. G. Iyer, and W. Im, “CHARMM-GUI: A web-based graphical user interface for CHARMM,” *J. Comput. Chem.*, vol. 29, no. 11, pp. 1859–1865, Aug. 2008, doi: 10.1002/jcc.20945.
[17] J. Lee *et al.*, “CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field,” *J. Chem. Theory Comput.*, vol. 12, no. 1, pp. 405–413, Jan. 2016, doi: 10.1021/acs.jctc.5b00935.
[18] Y. Qi, H. I. Ingólfsson, X. Cheng, J. Lee, S. J. Marrink, and W. Im, “CHARMM-GUI Martini Maker for Coarse-Grained Simulations with the Martini Force Field,” *J. Chem. Theory Comput.*, vol. 11, no. 9, pp. 4486–4494, Sep. 2015, doi: 10.1021/acs.jctc.5b00513.
[19] P. Hsu *et al.*, “CHARMM-GUI Martini Maker for modeling and simulation of complex bacterial membranes with lipopolysaccharides,” *J. Comput. Chem.*, vol. 38, no. 27, pp. 2354–2363, Oct. 2017, doi: 10.1002/jcc.24895.
[20] M. J. Abraham *et al.*, “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers,” *SoftwareX*, vol. 1–2, pp. 19–25, Sep. 2015, doi: 10.1016/j.softx.2015.06.001.
[21] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, “The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations,” *J. Phys. Chem. B*, vol. 111, no. 27, pp. 7812–7824, Jul. 2007, doi: 10.1021/jp071097f.
[22] D. H. de Jong *et al.*, “Improved Parameters for the Martini Coarse-Grained Protein Force Field,” *J. Chem. Theory Comput.*, vol. 9, no. 1, pp. 687–697, Jan. 2013, doi: 10.1021/ct300646g.
[23] H. I. Ingólfsson *et al.*, “The power of coarse graining in biomolecular simulations: The power of coarse graining in biomolecular simulations,” *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 4, no. 3, pp. 225–248, May 2014, doi: 10.1002/wcms.1169.
[24] B. R. Brooks *et al.*, “CHARMM: The biomolecular simulation program,” *J. Comput. Chem.*, vol. 30, no. 10, pp. 1545–1614, Jul. 2009, doi: 10.1002/jcc.21287.
[25] J. C. Phillips *et al.*, “Scalable molecular dynamics on CPU and GPU architectures with NAMD,” *J. Chem. Phys.*, vol. 153, no. 4, p. 044130, Jul. 2020, doi: 10.1063/5.0014475.
[26] A. D. Mackerell, M. Feig, and C. L. Brooks, “Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations,” *J. Comput. Chem.*, vol. 25, no. 11, pp. 1400–1415, Aug. 2004, doi: 10.1002/jcc.20065.
[27] S. E. Feller and A. D. MacKerell, “An Improved Empirical Potential Energy Function for Molecular Simulations of Phospholipids,” *J. Phys. Chem. B*, vol. 104, no. 31, pp. 7510–7515, Aug. 2000, doi: 10.1021/jp0007843.
[28] N. Foloppe and A. D. MacKerell, Jr., “All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data,” *J. Comput. Chem.*, vol. 21, no. 2, pp. 86–104, Jan. 2000, doi: 10.1002/(SICI)1096-987X(20000130)21:2<86::AID-JCC2>3.0.CO;2-G.
[29] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, “Comparison of simple potential functions for simulating liquid water,” *J. Chem. Phys.*, vol. 79, no. 2, pp. 926–935, Jul. 1983, doi: 10.1063/1.445869.

[30] O. Axelsson, “On conjugate gradient methods for large sparse systems of linear equations,” in *Numerical Integration of Differential Equations and Large Linear Systems*, vol. 968, 1982. [Online]. Available: https://doi.org/10.1007/BFb0064898
[31] S. Wang, S. Sun, Z. Li, R. Zhang, and J. Xu, “Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model,” *PLOS Comput. Biol.*, vol. 13, no. 1, p. e1005324, Jan. 2017, doi: 10.1371/journal.pcbi.1005324.
[32] S. Wang, W. Li, R. Zhang, S. Liu, and J. Xu, “CoinFold: a web server for protein contact prediction and contact-assisted protein folding,” *Nucleic Acids Res.*, vol. 44, no. W1, pp. W361–W366, Jul. 2016, doi: 10.1093/nar/gkw307.
[33] J. Ma, S. Wang, Z. Wang, and J. Xu, “Protein contact prediction by integrating joint evolutionary coupling analysis and supervised learning,” *Bioinformatics*, vol. 31, no. 21, pp. 3506–3513, Nov. 2015, doi: 10.1093/bioinformatics/btv472.
[34] M. Varadi *et al.*, “AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models,” *Nucleic Acids Res.*, vol. 50, no. D1, pp. D439–D444, Jan. 2022, doi: 10.1093/nar/gkab1061.
[35] M. Gruber, “Historical review: Another 50th anniversary – new periodicities in coiled coils,” *Trends Biochem. Sci.*, vol. 28, no. 12, pp. 679–685, Dec. 2003, doi: 10.1016/j.tibs.2003.10.008.
[36] C. W. Wood and D. N. Woolfson, “CCBuilder 2.0: Powerful and accessible coiled-coil modeling: Powerful and Accessible Coiled-Coil Modeling,” *Protein Sci.*, vol. 27, no. 1, pp. 103–111, Jan. 2018, doi: 10.1002/pro.3279.
[37] B. Sattelmacher, “The apoplast and its significance for plant mineral nutrition,” *New Phytol.*, vol. 149, no. 2, pp. 167–192, Feb. 2001, doi: 10.1046/j.1469-8137.2001.00034.x.
[38] Á. Piñeiro, A. Villa, T. Vagt, B. Koksch, and A. E. Mark, “A Molecular Dynamics Study of the Formation, Stability, and Oligomerization State of Two Designed Coiled Coils: Possibilities and Limitations,” *Biophys. J.*, vol. 89, no. 6, pp. 3701–3713, Dec. 2005, doi: 10.1529/biophysj.104.055590.
[39] L. Zhang and J. Hermans, “Molecular dynamics study of structure and stability of a model coiled coil,” *Proteins Struct. Funct. Genet.*, vol. 16, no. 4, pp. 384–392, Aug. 1993, doi: 10.1002/prot.340160407.
[40] L. Zimmermann *et al.*, “A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core,” *J. Mol. Biol.*, vol. 430, no. 15, pp. 2237–2243, Jul. 2018, doi: 10.1016/j.jmb.2017.12.007.
[41] F. Gabler *et al.*, “Protein Sequence Analysis Using the MPI Bioinformatics Toolkit,” *Curr. Protoc. Bioinforma.*, vol. 72, no. 1, Dec. 2020, doi: 10.1002/cpbi.108.
[42] S. J. Gilmour, R. K. Hajela, and M. F. Thomashow, “Cold Acclimation in *Arabidopsis thaliana*,” *Plant Physiol.*, vol. 87, no. 3, pp. 745–750, Jul. 1988, doi: 10.1104/pp.87.3.745.
[43] Z. Qiao *et al.*, “Structure of *Arabidopsis* CESA3 catalytic domain with its substrate UDP-glucose provides insight into the mechanism of cellulose synthesis,” *Proc. Natl. Acad. Sci.*, vol. 118, no. 11, p. e2024015118, Mar. 2021, doi: 10.1073/pnas.2024015118.
[44] S. K. Burley *et al.*, “RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences,” *Nucleic Acids Res.*, vol. 49, no. D1, pp. D437–D451, Jan. 2021, doi: 10.1093/nar/gkaa1038.

[45] B. T. Nixon *et al.*, “Comparative Structural and Computational Analysis Supports Eighteen Cellulose Synthases in the Plant Cellulose Synthesis Complex,” *Sci. Rep.*, vol. 6, no. 1, p. 28696, Jun. 2016, doi: 10.1038/srep28696.
[46] S. Gilroy and D. L. Jones, “Through form to function: root hair development and nutrient uptake,” *Trends Plant Sci.*, vol. 5, no. 2, pp. 56–60, Feb. 2000, doi: 10.1016/S1360-1385(99)01551-4.
[47] H.-T. Cho and D. J. Cosgrove, “Regulation of Root Hair Initiation and Expansin Gene Expression in Arabidopsis[W],” *Plant Cell*, vol. 14, no. 12, pp. 3237–3253, Dec. 2002, doi: 10.1105/tpc.006437.
[48] S.-H. Shiu and A. B. Bleecker, “Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases,” *Proc. Natl. Acad. Sci.*, vol. 98, no. 19, pp. 10763–10768, Sep. 2001, doi: 10.1073/pnas.181141598.
[49] M. D. Lehti-Shiu, C. Zou, K. Hanada, and S.-H. Shiu, “Evolutionary History and Stress Regulation of Plant Receptor-Like Kinase/Pelle Genes,” *Plant Physiol.*, vol. 150, no. 1, pp. 12–26, May 2009, doi: 10.1104/pp.108.134353.
[50] H. Lindner, L. M. Müller, A. Boisson-Dernier, and U. Grossniklaus, “CrRLK1L receptor-like kinases: not just another brick in the wall,” *Curr. Opin. Plant Biol.*, vol. 15, no. 6, pp. 659–669, Dec. 2012, doi: 10.1016/j.pbi.2012.07.003.
[51] N. Huck, J. M. Moore, M. Federer, and U. Grossniklaus, “The *Arabidopsis* mutant *feronia* disrupts the female gametophytic control of pollen tube reception,” *Development*, vol. 130, no. 10, pp. 2149–2159, May 2003, doi: 10.1242/dev.00458.
[52] M. Haruta, G. Sabat, K. Stecker, B. B. Minkoff, and M. R. Sussman, “A Peptide Hormone and Its Receptor Protein Kinase Regulate Plant Cell Expansion,” *Science*, vol. 343, no. 6169, pp. 408–411, Jan. 2014, doi: 10.1126/science.1244454.
[53] S. Schoenaers *et al.*, “The Auxin-Regulated CrRLK1L Kinase ERULUS Controls Cell Wall Composition during Root Hair Tip Growth,” *Curr. Biol.*, vol. 28, no. 5, pp. 722-732.e6, Mar. 2018, doi: 10.1016/j.cub.2018.01.050.
[54] K. Hématy *et al.*, “A Receptor-like Kinase Mediates the Response of Arabidopsis Cells to the Inhibition of Cellulose Synthesis,” *Curr. Biol.*, vol. 17, no. 11, pp. 922–931, Jun. 2007, doi: 10.1016/j.cub.2007.05.018.
[55] S. Miyazaki *et al.*, “ANXUR1 and 2, Sister Genes to FERONIA/SIRENE, Are Male Factors for Coordinated Fertilization,” *Curr. Biol.*, vol. 19, no. 15, pp. 1327–1331, Aug. 2009, doi: 10.1016/j.cub.2009.06.064.
[56] A. Boisson-Dernier *et al.*, “Disruption of the pollen-expressed *FERONIA* homologs *ANXUR1* and *ANXUR2* triggers pollen tube discharge,” *Development*, vol. 136, no. 19, pp. 3279–3288, Oct. 2009, doi: 10.1242/dev.040071.
[57] A. Y. Cheung and H.-M. Wu, “THESEUS 1, FERONIA and relatives: a family of cell wall-sensing receptor kinases?,” *Curr. Opin. Plant Biol.*, vol. 14, no. 6, pp. 632–641, Dec. 2011, doi: 10.1016/j.pbi.2011.09.001.
[58] P. Schulze-Muth, S. Irmler, G. Schröder, and J. Schröder, “Novel Type of Receptor-like Protein Kinase from a Higher Plant (Catharanthus roseus),” *J. Biol. Chem.*, vol. 271, no. 43, pp. 26684–26689, Oct. 1996, doi: 10.1074/jbc.271.43.26684.
[59] K. Hematy and H. Hofte, “Novel receptor kinases involved in growth regulation,” *Curr. Opin. Plant Biol.*, vol. 11, no. 3, pp. 321–328, Jun. 2008, doi: 10.1016/j.pbi.2008.02.008.

[60] T. Schallus *et al.*, "Malectin: A Novel Carbohydrate-binding Protein of the Endoplasmic Reticulum and a Candidate Player in the Early Steps of Protein *N* -Glycosylation," *Mol. Biol. Cell*, vol. 19, no. 8, pp. 3404–3414, Aug. 2008, doi: 10.1091/mbc.e08-04-0354.
[61] J.-M. Escobar-Restrepo *et al.*, "The FERONIA Receptor-like Kinase Mediates Male-Female Interactions During Pollen Tube Reception," *Science*, vol. 317, no. 5838, pp. 656–660, Aug. 2007, doi: 10.1126/science.1143562.
[62] H. Guo, L. Li, H. Ye, X. Yu, A. Algreen, and Y. Yin, "Three related receptor-like kinases are required for optimal cell elongation in *Arabidopsis thaliana*," *Proc. Natl. Acad. Sci.*, vol. 106, no. 18, pp. 7648–7653, May 2009, doi: 10.1073/pnas.0812346106.
[63] Q. Duan, D. Kita, C. Li, A. Y. Cheung, and H.-M. Wu, "FERONIA receptor-like kinase regulates RHO GTPase signaling of root hair development," *Proc. Natl. Acad. Sci.*, vol. 107, no. 41, pp. 17821–17826, Oct. 2010, doi: 10.1073/pnas.1005366107.
[64] G. Huang *et al.*, "Arabidopsis Rop GEF 4 and Rop GEF 10 are important for FERONIA -mediated developmental but not environmental regulation of root hair growth," *New Phytol.*, vol. 200, no. 4, pp. 1089–1101, Dec. 2013, doi: 10.1111/nph.12432.
[65] H.-W. Shih, N. D. Miller, C. Dai, E. P. Spalding, and G. B. Monshausen, "The Receptor-like Kinase FERONIA Is Required for Mechanical Signal Transduction in Arabidopsis Seedlings," *Curr. Biol.*, vol. 24, no. 16, pp. 1887–1892, Aug. 2014, doi: 10.1016/j.cub.2014.06.064.
[66] N. Rotman, F. Rozier, L. Boavida, C. Dumas, F. Berger, and J.-E. Faure, "Female Control of Male Gamete Delivery during Fertilization in Arabidopsis thaliana," *Curr. Biol.*, vol. 13, no. 5, pp. 432–436, Mar. 2003, doi: 10.1016/S0960-9822(03)00093-9.
[67] Q. Duan *et al.*, "Reactive oxygen species mediate pollen tube rupture to release sperm for fertilization in Arabidopsis," *Nat. Commun.*, vol. 5, no. 1, p. 3129, May 2014, doi: 10.1038/ncomms4129.
[68] Q. A. Ngo, H. Vogler, D. S. Lituiev, A. Nestorova, and U. Grossniklaus, "A Calcium Dialog Mediated by the FERONIA Signal Transduction Pathway Controls Plant Sperm Delivery," *Dev. Cell*, vol. 29, no. 4, pp. 491–500, May 2014, doi: 10.1016/j.devcel.2014.04.008.
[69] C. Li, H.-M. Wu, and A. Y. Cheung, "FERONIA and Her Pals: Functions and Mechanisms," *Plant Physiol.*, vol. 171, no. 4, pp. 2379–2392, Aug. 2016, doi: 10.1104/pp.16.00667.
[70] M. Gonneau *et al.*, "Receptor Kinase THESEUS1 Is a Rapid Alkalinization Factor 34 Receptor in Arabidopsis," *Curr. Biol.*, vol. 28, no. 15, pp. 2452-2458.e4, Aug. 2018, doi: 10.1016/j.cub.2018.05.075.
[71] M. Stegmann *et al.*, "The receptor kinase FER is a RALF-regulated scaffold controlling plant immune signaling," *Science*, vol. 355, no. 6322, pp. 287–289, Jan. 2017, doi: 10.1126/science.aal2541.
[72] X. Liu *et al.*, "The Role of LORELEI in Pollen Tube Reception at the Interface of the Synergid Cell and Pollen Tube Requires the Modified Eight-Cysteine Motif and the Receptor-Like Kinase FERONIA," *Plant Cell*, vol. 28, no. 5, pp. 1035–1052, May 2016, doi: 10.1105/tpc.15.00703.
[73] Y. Xiao *et al.*, "Mechanisms of RALF peptide perception by a heterotypic receptor complex," *Nature*, vol. 572, no. 7768, pp. 270–274, Aug. 2019, doi: 10.1038/s41586-019-1409-7.

[74] W. Feng *et al.*, “The FERONIA Receptor Kinase Maintains Cell-Wall Integrity during Salt Stress through Ca2+ Signaling,” *Curr. Biol.*, vol. 28, no. 5, pp. 666-675.e5, Mar. 2018, doi: 10.1016/j.cub.2018.01.023.
[75] L. N. Müller, C. Muhle-Goll, and M. B. Biskup, “The Glc2Man2-fragment of the N-glycan precursor – a novel ligand for the glycan-binding protein malectin?,” *Org. Biomol. Chem.*, vol. 8, no. 14, p. 3294, 2010, doi: 10.1039/c004502k.
[76] J. Meiler and D. Baker, “ROSETTALIGAND: Protein-small molecule docking with full side-chain flexibility,” *Proteins Struct. Funct. Bioinforma.*, vol. 65, no. 3, pp. 538–548, Nov. 2006, doi: 10.1002/prot.21086.
[77] I. W. Davis and D. Baker, “RosettaLigand Docking with Full Ligand and Receptor Flexibility,” *J. Mol. Biol.*, vol. 385, no. 2, pp. 381–392, Jan. 2009, doi: 10.1016/j.jmb.2008.11.010.
[78] G. Lemmon and J. Meiler, “Chapter 10: Rosetta Ligand Docking with Flexible XML Protocols,” in *Computational Drug Discovery and Design*, 2012.
[79] D. W. Abbott and A. B. Boraston, “A Family 2 Pectate Lyase Displays a Rare Fold and Transition Metal-assisted β-Elimination,” *J. Biol. Chem.*, vol. 282, no. 48, pp. 35328–35336, Nov. 2007, doi: 10.1074/jbc.M705511200.
[80] M. Alahuhta *et al.*, “The catalytic mechanism and unique low pH optimum of Caldicellulosiruptor bescii family 3 pectate lyase,” *Acta Cryst*, pp. 1946–1954, 2015.
[81] A. S. Luis *et al.*, “Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides,” *Nat. Microbiol.*, vol. 3, no. 2, pp. 210–219, Dec. 2017, doi: 10.1038/s41564-017-0079-1.
[82] S. A. Morillo and F. E. Tax, “Functional analysis of receptor-like kinases in monocots and dicots,” *Curr. Opin. Plant Biol.*, vol. 9, no. 5, pp. 460–469, Oct. 2006, doi: 10.1016/j.pbi.2006.07.009.
[83] S. D. Deslauriers and P. B. Larsen, “FERONIA Is a Key Modulator of Brassinosteroid and Ethylene Responsiveness in Arabidopsis Hypocotyls,” *Mol. Plant*, vol. 3, no. 3, pp. 626–640, May 2010, doi: 10.1093/mp/ssq015.
[84] F. Yu *et al.*, “FERONIA Receptor Kinase Controls Seed Size in Arabidopsis thaliana,” *Mol. Plant*, vol. 7, no. 5, pp. 920–922, May 2014, doi: 10.1093/mp/ssu010.
[85] F. Yu *et al.*, “FERONIA receptor kinase pathway suppresses abscisic acid signaling in *Arabidopsis* by activating ABI2 phosphatase,” *Proc. Natl. Acad. Sci.*, vol. 109, no. 36, pp. 14693–14698, Sep. 2012, doi: 10.1073/pnas.1212547109.
[86] S.-H. Shiu and A. B. Bleecker, “Expansion of the Receptor-Like Kinase/Pelle Gene Family and Receptor-Like Proteins in Arabidopsis,” *Plant Physiol.*, vol. 132, no. 2, pp. 530–543, Jun. 2003, doi: 10.1104/pp.103.021964.
[87] Z. Ge *et al.*, “*Arabidopsis* pollen tube integrity and sperm release are regulated by RALF-mediated signaling,” *Science*, vol. 358, no. 6370, pp. 1596–1600, Dec. 2017, doi: 10.1126/science.aao3642.
[88] S. Du, L.-J. Qu, and J. Xiao, “Crystal structures of the extracellular domains of the CrRLK1L receptor-like kinases ANXUR1 and ANXUR2: Crystal Structures of CrRLK1L Receptor,” *Protein Sci.*, vol. 27, no. 4, pp. 886–892, Apr. 2018, doi: 10.1002/pro.3381.
[89] J. Chen *et al.*, “FERONIA interacts with ABI2-type phosphatases to facilitate signaling cross-talk between abscisic acid and RALF peptide in *Arabidopsis*,” *Proc. Natl. Acad. Sci.*, vol. 113, no. 37, Sep. 2016, doi: 10.1073/pnas.1608449113.

[90] Q. Duan *et al.*, "FERONIA controls pectin- and nitric oxide-mediated male–female interaction," *Nature*, vol. 579, no. 7800, pp. 561–566, Mar. 2020, doi: 10.1038/s41586-020-2106-2.
[91] A. Geitmann and J. K. E. Ortega, "Mechanics and modeling of plant cell growth," *Trends Plant Sci.*, vol. 14, no. 9, pp. 467–478, Sep. 2009, doi: 10.1016/j.tplants.2009.07.006.
[92] C. Grierson, E. Nielsen, T. Ketelaar, and J. Schiefelbein, "Root Hairs," in *The Arabidopsis Book*, American Society of Plant Biologists, 2014.
[93] G. Guerriero, J.-F. Hausman, and G. Cai, "No Stress! Relax! Mechanisms Governing Growth and Shape in Plant Cells," *Int. J. Mol. Sci.*, vol. 15, no. 3, pp. 5094–5114, Mar. 2014, doi: 10.3390/ijms15035094.
[94] B. Favery *et al.*, "*KOJAK* encodes a cellulose synthase-like protein required for root hair cell morphogenesis in *Arabidopsis*," *Genes Dev.*, vol. 15, no. 1, pp. 79–89, Jan. 2001, doi: 10.1101/gad.188801.
[95] T.-C. Wu, S. A. Belteton, J. Pack, D. B. Szymanski, and D. M. Umulis, "LobeFinder: A Convex Hull-Based Method for Quantitative Boundary Analyses of Lobed Plant Cells," *Plant Physiol.*, vol. 171, no. 4, pp. 2331–2342, Aug. 2016, doi: 10.1104/pp.15.00972.
[96] S. Gillies, "Shapely: manipulation and analysis of geometric objects." toblerity.org, 2007. [Online]. Available: https://github.com/Toblerity/Shapely
[97] G. N. Thyssen *et al.*, "A Gly65Val substitution in an actin, GhACT_LI1, disrupts cell polarity and F-actin organization resulting in dwarf, lintless cotton plants," *Plant J.*, vol. 90, no. 1, pp. 111–121, Apr. 2017, doi: 10.1111/tpj.13477.
[98] D. and B. Frenkel and Smit, "Chapter 4 - Molecular Dynamics Simulations," in *Understanding Molecular Simulation*, Second., Academic Press, 2002, pp. 63–107. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780122673511500067
[99] J. A. Anderson, J. Glaser, and S. C. Glotzer, "HOOMD-blue: A Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations," *Comput. Mater. Sci.*, vol. 173, p. 109363, Feb. 2020, doi: 10.1016/j.commatsci.2019.109363.
[100] J. Ramírez, S. K. Sukumaran, B. Vorselaars, and A. E. Likhtman, "Efficient on the fly calculation of time correlation functions in computer simulations," *J. Chem. Phys.*, vol. 133, no. 15, p. 154103, Oct. 2010, doi: 10.1063/1.3491098.
[101] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *J. Mol. Graph.*, vol. 14, no. 1, pp. 33–38, Feb. 1996, doi: 10.1016/0263-7855(96)00018-5.
[102] J. E. Stone, "An efficient library for parallel ray tracing and animation".
[103] J. S. Griffiths and H. M. North, "Sticking to cellulose: exploiting Arabidopsis seed coat mucilage to understand cellulose biosynthesis and cell wall polysaccharide interactions," *New Phytol.*, vol. 214, no. 3, pp. 959–966, May 2017, doi: 10.1111/nph.14468.
[104] X. Robert and P. Gouet, "Deciphering key features in protein structures with the new ENDscript server," *Nucleic Acids Res.*, vol. 42, no. W1, pp. W320–W324, Jul. 2014, doi: 10.1093/nar/gku316.