

ABSTRACT

KIRSCH, MICHAEL PAUL. In Silico Insights into Cellulose Synthase Homotrimer Stability. (Under the direction of Dr. Yaroslava Yingling).

Cellulose synthases, glycosyltransferase family 2 integral membrane proteins, produce cellulose, the main structural component of plant cell walls. In plants, functionalization of these enzymes requires their assembly into trimers, which further assemble into six-trimer cellulose synthesis complexes. Recent investigations into soybean cellulose synthases suggest a structural preference for homotrimeric over heterotrimeric assembly. However, the structural and energetic interactions responsible for this behavior remain unknown. To gain insight into the structural basis of cellulose synthase trimerization, homotrimeric and heterotrimeric models comprised of *Physcomitrium patens* cellulose synthase isoforms 5, 7, and 8 were constructed and investigated using all-atom molecular dynamics simulations in a lipid bilayer environment. The results indicate that the interfaces of homotrimers display greater stability than those of heterotrimers formed by the replacement of one monomer, consistent with their experimentally observed assembly preferences. These findings support a biological preference for homotrimer formation during cellulose synthase trimerization and contribute to improved understanding of the composition of higher-order cellulose synthesis complexes.

© Copyright 2026 by Michael Paul Kirsch

All Rights Reserved

In Silico Insights into Cellulose Synthase Homotrimer Stability

by
Michael Paul Kirsch

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Chemistry

Raleigh, North Carolina

2026

APPROVED BY:

Dr. Yaroslava Yingling
Committee Chair

Dr. Alexej Smirnov

Dr. Melissa Pasquinelli

DEDICATION

This degree is dedicated to my family:

Mom, Dad, and Zach.

Who have been with me every step of my graduate journey.

Your love and support mean more than I could ever say.

And to my grandfather, Dr. Paul Kirsch.

Who has always encouraged me to “follow my North Star” wherever it leads.

BIOGRAPHY

Michael Kirsch was born to Stephen and Linda Kirsch in Dryden, NY. He completed high school at Dryden High School in 2011. After working for several years as a paraprofessional, he completed a Liberal Arts and Sciences - Math/Science Associate Degree at Tompkins Cortland Community College (TC3) in 2018. This was followed by completing a Biochemistry Bachelor's Degree with a Mathematics Minor at SUNY Oswego, where he graduated Summa Cum Laude in 2020. Since joining the Chemistry program at North Carolina State University in 2021 he has worked in the Yingling research group, using simulations to investigate protein structure and function. In addition to his research in the Yingling group, Michael has gotten the opportunity to be a Teaching Assistant (TA) for 22 Organic Chemistry lab sections, including those restricted to majors-only Chemistry students. He hopes to combine the knowledge gained from research and TAing into a future career of teaching in higher education.

ACKNOWLEDGMENTS

I would like to thank Dr. Yingling for both the opportunity to work in her research group as well as for her invaluable feedback as I've progressed through my master's program. I've learned much under her mentorship.

I would also like to thank Dr. Albert Kwansa, whose knowledge has greatly assisted the development of my own computational skills including the setup, running and analysis of MD simulations.

I would like to thank Dr. Alison Roberts, professor at The University of Rhode Island, for her role as a collaborator on this project and her shared insights into cellulose synthase function.

I would like to thank my groupmates and friends in the program, whose support and camaraderie has been a great joy over the course of the program.

Finally, I would like to thank the North Carolina State University Department of Chemistry at large, especially Dr. Reza Ghiladi, for their support as I've pursued this degree.

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	vii
Chapter 1: Current Insights into Cellulose Synthase Structure, Assembly, and Function ..	1
Chapter 2: Modeling Trimeric Assemblies of <i>Physcomitrium patens</i> Cellulose Synthases ...	5
Introduction.....	5
Materials and Methods.....	11
Results.....	13
Discussion.....	26
Chapter 3: Assessment of <i>Physcomitrium patens</i> Cellulose Synthase Trimers using Molecular Dynamics Simulations	28
Introduction.....	28
Materials and Methods.....	33
Results.....	35
Discussion.....	42
References	44
Appendix	52

LIST OF TABLES

Table 2.1: Numbers of generated PpCESA models.....	13
Table 2.2: PpCESA structural domains	16
Table 2.3: Residue counts for each alignment of the PpCESA sequences (5, 7, and 8) to PttCESA8.....	16
Table 2.4: Model quality scores for conserved PpCESA domains.....	17
Table 2.5: Quality scores for conserved PpCESA domains of selected PpCESA models	20
Table 2.6: Number of identical residue pairs for the alignments of the PpCESA sequences (5, 7, and 8) to PttCESA8	23
Table 2.7: RMSD values for each PpCESA monomer within each trimer, measured relative to its corresponding PttCESA8 monomer.....	25
Table 2.8: RMSD values from the structural superposition of each PpCESA trimer onto the template PttCESA8 homotrimer	25
Table 3.1: Trimer interface residue count.....	37
Table S1: Secondary structure composition percentages for all PpCESA trimers	52

LIST OF FIGURES

Figure 1.1: Structural representations of <i>P. patens</i> cellulose synthases highlighting key functional regions and the trimeric assembly state	3
Figure 2.1: Annotated sequences of PpCESAs 5, 7, and 8	15
Figure 2.2: I-TASSER modeling of the N-terminal domain of PpCESA5	19
Figure 2.3: ChimeraX alignments of the PpCESA sequences (5, 7, and 8) to PttCESA8	21
Figure 2.4: Generated PpCESA trimers	24
Figure 3.1: Overall trimer secondary structure	36
Figure 3.2: PpCESA 555 homotrimer interface residues, represented using VMD's "NewCartoon" drawing method	37
Figure 3.3: Post-convergence trimer interface RMSD values	38
Figure 3.4: Trimer monomer-monomer and monomer-"rest of the trimer" nonbonding energy, arranged by trimer	39
Figure 3.5: Trimer monomer-monomer and monomer-"rest of the trimer" nonbonding energy, arranged by decreasing energy	40
Figure 3.6: Trimer interfacial total nonbonding energy, arranged by decreasing energy	41

AUTHORSHIP STATEMENT

Michael Paul Kirsch is the sole author of this thesis.

AI (Perplexity) was utilized to find papers relevant to the research described, though the content of papers identified by Perplexity was assessed without the use of AI.

AI (ChatGPT) was used to refine word choice and tone within this document, though all writing was originally done without the use of AI.

CHAPTER 1: Current Insights into Cellulose Synthase Structure, Assembly, and Function

Cellulose, the most abundant biopolymer in nature, is a linear polysaccharide comprised of β -(1 \rightarrow 4)-linked glucose residues and is the main structural component of plant cell walls.¹ It is a highly promising biomaterial, particularly as a sustainable biofuel, owing to its abundance and renewable nature.²⁻⁵ Understanding the structural and functional properties of cellulose-producing proteins can provide a rational basis for future efforts to modify cellulose biosynthesis.

Cellulose in plants is synthesized by cellulose synthases (CESAs), glycosyltransferase family 2 integral membrane proteins.⁶ Many plant species, including poplar, cotton, rice, and moss, express multiple CESA isoforms, though the functional significance of this diversification is not well understood.⁷ Previous investigations into cellulose synthases have supported that each CESA monomer catalyzes the polymerization of UDP-glucose into a single β -(1 \rightarrow 4)-linked glucan chain, which is then extruded through the cell membrane and into the apoplast via the CESA's transmembrane channel.^{1,6,8,9} These enzymes contain a Rossmann nucleotide binding fold and a conserved D, D, D, QxxRW active site motif within their catalytic domain (CatD).^{10,11} Plant CESAs share several conserved structural features, including a transmembrane channel consisting of seven alpha helices, a plant-conserved region (PCR) consisting of two antiparallel alpha helices, and a class specific region (CSR) (Figure 1.1).¹²⁻¹⁶ Both the CSR and the N-terminus (N-term) of CESAs are predicted to contain intrinsically disordered regions (IDRs) that may contribute to the formation and regulation of higher-order CESA assemblies.^{17,18} An IDR does not fold into a single, well-defined secondary structure motif, but instead exists as an ensemble of interconverting conformations, where any one structural state is insufficient to fully

describe the conformational heterogeneity of the region.¹⁹ Though the mobile nature of these regions hinders their structural understanding via traditional resolution techniques, in silico modeling and simulations have made progress towards understanding the conformational ensemble these IDRs sample within the overall CESA structure.^{16,20}

CESA monomers self-associate to form stable oligomeric assemblies. Cryo-EM visualization of *Populus tremula x tremuloides* CESA isoform 8, *Gossypium hirsutum* CESA isoform 7, and *Glycine max* CESAs 1, 3, and 6 have shown that CESA monomers can form stable homotrimers (Figure 1.1).^{1,17,21} Dimerization of *Arabidopsis thaliana* CESA isoform 3 has also been observed, but the residues involved differ from those implicated in the trimeric interface and its functional role, including whether it precedes trimer formation, remains unresolved.^{5,22} FF-TEM imaging of *Physcomitrium patens* CESAs show that CESA trimers will further assemble into six-trimer cellulose synthesis complexes (CSCs) called “rosettes”.²³ This arrangement is also consistent with current models of cellulose microfibril structure, which propose that each microfibril contains 18 glucan chains, matching the number of CESA monomers in a rosette CSC.²⁴

Certain conserved structural regions of CESAs are proposed to facilitate their higher-order assembly into CSCs. PCR interactions are believed to form the main stabilized interface between each CESA monomer in a trimeric assembly.^{1,5,21} Recently, the CSR has been implicated in CSC formation, with inter-trimer contacts in *Glycine max* CSCs theorized to be the result of fuzzy CSR-CSR interactions.¹⁷ Ho et al. also noted a distinct lack of heterotrimer formation between co-expressed *Glycine max* CESA isoforms, additionally suggesting that the PCR of individual isoforms may be specialized enough to only permit the stable formation of homotrimeric CESAs.¹⁷

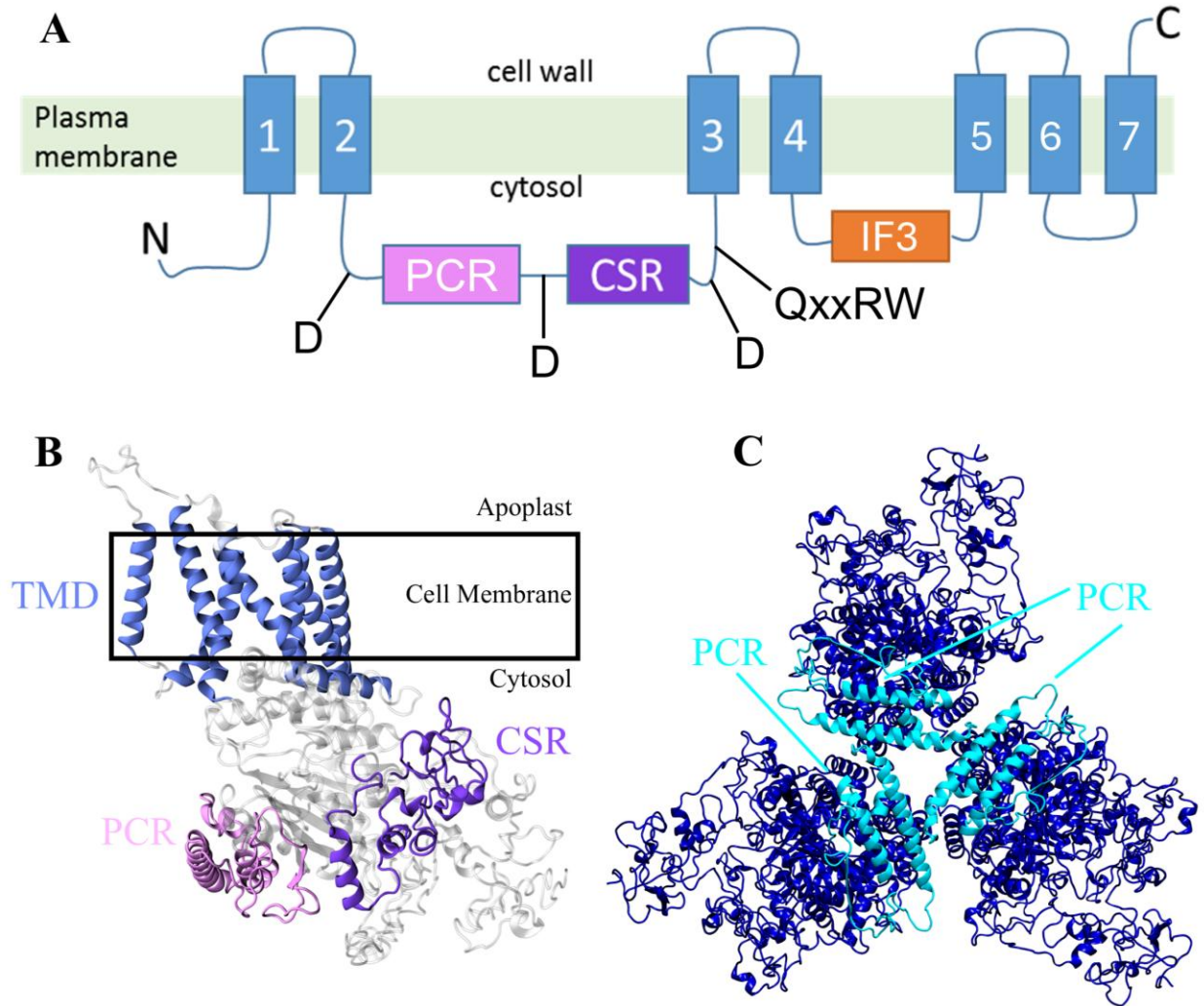


Figure 1.1: Structural representations of *P. patens* cellulose synthases highlighting key functional regions and the trimeric assembly state. (A) Domain architecture of *P. patens* cellulose synthases highlighting the locations of the transmembrane domain (TMD), consisting of transmembrane helices 1-7 and interfacial helix 3 (1-7, IF3), class specific region (CSR), plant-conserved region (PCR), and active site residues (D, D, D, QxxRW). (B) Side view of a PpCESA5 monomer highlighting the TMD (blue), CSR (purple), and PCR (lavender), represented using VMD's "NewCartoon" drawing method. (C) Cytosolic view of a PpCESA8 homotrimer (blue), represented using VMD's "NewCartoon" drawing method. The PCRs at the trimeric interface are lighter blue for enhanced visibility.

Investigations of CESA activity in the moss *Physcomitrium patens* have shown that the removal of certain CESA isoforms by genetic knockout can severely impair overall cellulose production, even when other isoforms remain present.²⁵ *P. patens* contains seven cellulose-producing CESA isoforms (named PpCESAs 3, 4, 5, 6, 7, 8, and 10), which are subdivided into two clades. Clade A contains PpCESAs 3, 5, and 8 and Clade B contains PpCESAs 4, 6, 7, and 10. Of these CESAs, only PpCESA5 can produce cellulose in the absence of any other isoform. For every other isoform, the presence of both Clade A and Clade B isoforms is necessary for cellulose production. Thus, for example, co-expressed PpCESAs 7 and 8, 7 being from Clade B and 8 being from Clade A, can produce cellulose, while co-expressed PpCESAs 3 and 8, both from Clade A, cannot.²⁵ Consequently, with the exception of PpCESA5, functional *P. patens* CSCs must be assembled from either heterotrimers or a hetero-oligomeric combination of homotrimers. Despite the recent evidence supporting preferential homotrimeric CESA formation and hetero-oligomeric CSC composition, the structural interactions that influence CSC stability and enable cellulose synthesis remain unclear, providing the rationale for investigating homotrimeric and heterotrimeric assemblies.

Chapter 2 describes the use of structural prediction, evaluation, and alignment tools to construct computational trimer models of PpCESA isoforms 5, 7, and 8.

Chapter 3 employs all-atom explicit solvent MD simulations to elucidate structural and energetic differences between homotrimeric and heterotrimeric PpCESA interfaces.

CHAPTER 2: Modeling Trimeric Assemblies of *Physcomitrium patens* Cellulose Synthases

Introduction

Protein structure prediction has progressed from the initial identification of secondary structure motifs by Pauling and Corey in 1951, and the early physical homology modeling developed by Browne et al. in 1969, to the computational methods widely used today.^{26–28} Modern protein structure prediction strategies include homology modeling, threading, and deep learning-based approaches for determining protein structures from their residue sequences.²⁹ In the homology modeling approach, programs like BLAST and HHblits are used to find homologs of the input protein sequence.³⁰ These homologs are used as a template to provide atomic coordinates to the input sequence based on their alignment. While this method can lead to high-quality protein structures, the quality of the output model is largely dependent on the available template structures, with low-homology template models leading to poor quality models.^{31,32} Threading takes a different template-based approach to protein modeling. Based on the idea that protein folding motifs can be conserved across a diverse space of protein sequences, threading instead uses common protein folds as templates, threading regions of the sequence to these templates to find the best-fit templates, which are then used to assign atomic coordinates.³³ Like homology modeling, threading quality can be limited by the folds available.³¹ Deep learning, a subfield of artificial intelligence, uses neural networks to predict protein structures directly from amino acid sequences without the need for structural templates.²⁹ By analyzing multiple sequence alignments, these neural networks capture co-evolutionary relationships

between residues that inform predictions of spatial proximity and structural constraints.^{34,35} The neural networks then iteratively refine their initial predictions to generate the final 3D structure.

Homology modeling, threading, and deep learning all exhibit reduced predictive accuracy for CESAs due to the intrinsically disordered N-term and CatD. X-ray crystallography has been unable to resolve IDRs, resulting in a shortage of experimentally derived templates for these regions in the Protein Data Bank, limiting homology-based modeling approaches.³⁶ Although nuclear magnetic resonance (NMR) spectroscopy has been employed to characterize IDRs, it faces limitations in resolving long disordered regions and proteins containing both ordered and disordered domains, as well as in achieving comprehensive sampling of the IDR conformational ensemble.³⁷⁻³⁹ Template-free prediction methods, including deep learning-based structure prediction algorithms, exhibit limited predictive accuracy for IDRs, tending to over-compact disordered segments due to their training on datasets dominated by stable, well-structured proteins.⁴⁰ Threading techniques also exhibit limited predictive accuracy for IDRs. Their template folds, obtained from experimentally solved structures in the PDB, fail to encompass the range of conformations that can be natively adopted by an IDR.⁴¹⁻⁴³ Even when an IDR is modeled as part of a protein structure, the resulting single conformation cannot capture the dynamic conformational ensemble characteristic of these regions.⁴² Moreover, some IDRs can adopt partially ordered conformations upon interaction with binding partners, such as forming fuzzy complexes, which cannot be accurately modeled in the absence of their interaction partners.⁴⁴ Fuzzy complex formation has been predicted for the CSRs of CESAs.¹⁷

One way to overcome the inherent limitations of single-conformation protein models is to explore their conformational landscapes through MD simulations, which are described in more detail in Chapter 3. Briefly, for CESA trimers in particular, MD simulations can provide insight

into both the intrinsic flexibility of their IDRs and the monomer-monomer interface interactions that occur over the time scale of the simulation.^{5,45} MD simulations can also capture conformational changes IDRs undergo in response to interactions with nearby molecules.⁴⁶ Additionally, a single conformation cannot account for the effects of lipid-protein interactions on a protein, making it necessary to embed CESAs within membranes to capture lipid-mediated effects on protein conformation.⁴⁷ The versatility of MD simulation environments even permits the incorporation of tailored mixed-composition lipid bilayers, enabling the approximate modeling of CESA native plant cell membrane conditions.⁴⁸ Due to their capacity to model atomic motion and interactions under specified environmental parameters over time, MD simulations provide deeper and more accurate molecular insights than single-conformation protein structures, which cannot account for intrinsic flexibility or dynamic interactions.

To generate starting structures for these simulations, the protein modeling programs SWISS-MODEL, I-TASSER, I-TASSER-MTD, Phyre2, and AlphaFold were initially considered. SWISS-MODEL uses homology modeling, I-TASSER, I-TASSER-MTD, and Phyre2 use threading, and AlphaFold uses deep learning.^{34,43,49-51} Phyre2, known to exclude modeling regions without homologous folds, which includes IDRs, was purposefully excluded.⁵¹ At that time, AlphaFold had known difficulty predicting initial IDR structures with confidence and thus was also excluded.⁵²⁻⁵⁴ Thus, SWISS-MODEL, I-TASSER, and I-TASSER-MTD were selected for protein structure prediction. A complete hybrid model of GhCESA1 with 974 resolved residues developed by Kwansa et al. was used as a structural template to guide modeling.²⁰ This model incorporates the structurally resolved regions of a cryo-EM determined PttCESA8 structure (720 out of 978 residues), including the ordered transmembrane domain and

most of the catalytic domain.¹ These ordered domains were supplemented with in silico-determined structure of less structured regions, including the N-term and CSR.^{16,20}

Following model generation, validation was performed using ERRAT, ProSA-web, and QMEANDisCo to evaluate stereochemical quality and confirm biological plausibility prior to MD simulations.⁵⁵⁻⁵⁸ Each tool provides complementary insights into model quality, and the following sections describe their methodologies, beginning with ERRAT. In proteins, noncovalent atomic interactions occur in specific spatial patterns that are dictated by the protein's amino acid sequence, folding principles, and physicochemical environment.⁵⁶ The frequent occurrence of nitrogen-oxygen nonbonded interactions is attributed to their strong tendency to participate in hydrogen-bonding interactions, which play a central role in stabilizing protein structure. ERRAT leverages this, evaluating structural error based on the relative fractions of C-C, C-N, C-O, N-N, N-O, and O-O nonbonded interactions in submitted structures.⁵⁶ These interactions must be within a 3.5 Å cutoff and occur between atoms in separate residues. ERRAT first calculates the relative fraction of each nonbonded interaction type within a 9-residue sliding window (Equation 2.1). These fractions are then used to generate a six-dimensional vector for the window (Equation 2.2). The deviation of that vector from the ERRAT reference vector, comprised of nonbonded interaction fractions from 96 high-resolution crystal structures from the Brookhaven PDB, is then calculated (Equation 2.3) and incorporated into a quadratic form error function (Equation 2.4).^{56,59}

$$f(XX) = \frac{n_{xx}}{n_{CC}+n_{CN}+n_{CO}+n_{NN}+n_{NO}+n_{OO}} \quad (2.1)$$

Where $f()$ denotes the fraction of a nonbonded interaction type, n is the number of interactions of a given type, and XX is representative of any type of nonbonded interaction.

$$y_i = (f(CC), f(CN), f(CO), f(NN), f(NO), f(OO))_{i-4}^{i+4} \quad (2.2)$$

Where i represents the central residue of the window.

$$x_i = y_i - \bar{y} \quad (2.3)$$

Where \bar{y} is the reference vector comprised of the mean fraction composition of the reference structures.

$$\text{Error function} = x_i^T B x_i \quad (2.4)$$

Where x_i^T is the transposed matrix of x_i and B is a symmetric positive-definite matrix calculated from the reference vectors.

The 95% confidence limit for the nonbonded interactions of a given window corresponds to an error function of ~ 6 . The overall percentage of residues that fall within the 95% confidence limit for a model is considered to be the model's Overall Quality Factor, with a cutoff of 95% for high-quality structures.

ProSA-web uses knowledge-based mean force potentials to give insight into the overall energy of a model based on its $C\alpha$ (alpha carbon) atoms.^{57,60} This in turn gives insight into the accuracy of its folds. Initially, ProSA-web incorporates the model of interest into a polyprotein comprised of 230 proteins of known structure ($\sim 50,000$ residues).⁶¹ Conformations are then generated starting at each residue that can give a full-length structure when threading the model sequence along the polyprotein structure, giving $\sim 50,000 - (L - 1)$ total conformations, where L is the length of the model. For each conformation, the energy of each pair of $C\alpha$ atoms is calculated (Equation 2.5) and the pair energies summed to give the total energy for each conformation, including the model of interest (Equation 2.6).^{60,62} The energy of the model of interest is then compared to the average energy of all conformations and divided by the standard deviation of the energy of all conformations (Equation 2.7), giving a z-score of the model's normalized energy.

$$E^{ij} = -k_B T \ln \left[\frac{P_{ij}(r_{ij})}{P_{ref}(r_{ij})} \right] \quad (2.5)$$

Where k_B is the Boltzmann Constant, T is temperature (293 K), r_{ij} is the distance between Ca atoms i and j, $P_{ij}(r_{ij})$ is the observed probability of two Ca atoms being r distance apart, and $P_{ref}(r_{ij})$ is the random probability of two Ca atoms being r distance apart.⁶⁰

$$E_{conformation} = \sum_{i < j} E^{ij} \quad (2.6)$$

$$z - score = \frac{E_{model} - \bar{E}_{conformations}}{\sigma_{conformations}} \quad (2.7)$$

Where $\bar{E}_{conformations}$ is the mean energy of the set of conformations and $\sigma_{conformations}$ is the standard deviation of the set of conformations.

A model is considered high-quality if its z-score falls within the range of z-scores typical for experimentally determined proteins of the same size in the PDB, which is included in the ProSA-web output. However, due to the high prevalence of z-scores of soluble globular proteins within the database, these z-score ranges may not be applicable for transmembrane or similar proteins.⁵⁷

Rather than the single-metric approaches of ERRAT and Pro-SA web, QMEANDisCo uses a composite scoring function to evaluate protein structure.⁵⁸ It does this by combining individual model scores derived from the structure of the submitted model with a distance constraint (DisCo) score that compares the pairwise residue-residue distances of the model to the pairwise residue-residue distances of an ensemble of homologous protein structures. This weighted linear combination of terms is shown in Equation 2.8. A feed-forward neural network is used to adaptively weigh the contributions of these scoring metrics.

$$QMEANDisCo = w_1 S_{AApot} + w_2 S_{C\beta pot} + w_3 S_{pack_pot} + w_4 S_{torsion} + w_5 S_{SA} \quad (2.8)$$

$$+ w_6 S_{SAagree} + w_7 S_{SSagree} + w_8 S_{C\beta pack} + w_9 S_{red_pot} + w_{10} S_{clash} + w_{11} S_N + w_{12} S_{DisCo}$$

Where S_{AApot} is the all-atom interaction potential, $S_{C\beta pot}$ is the C β (beta carbon) interaction potential, S_{pack_pot} is the packing potential, $S_{torsion}$ is the torsion potential, S_{SA} is the solvent accessibility, $S_{SAagree}$ is the solvent accessibility agreement, $S_{SSagree}$ is the secondary structure agreement, $S_{C\beta pack}$ is the C β packing potential, S_{red_pot} is the reduced potential, S_{clash} is the clash score, S_N is the number of residues within 15 Å, S_{DisCo} is the distance constraint, and w_i is the weight assigned to each term.

QMEANDisCo scores can range from 0 to 1. A model with a QMEANDisCo global score of at least 0.60 is considered to be a high-quality structure. Additionally, the per-residue scores, which also have the same range and quality cutoff, can help identify problematic parts of a model.

While these three structural evaluation tools can provide a foundation for model selection, their scores alone cannot be deemed sufficient for final selection. ERRAT and ProSA-web base their methodologies on crystal structures from the PDB, in which IDRs and transmembrane regions are underrepresented.^{57,63} Thus, in the case of CESAs, closeness of fit of the TMDs of the models to the TMD of the Cryo-EM PttCESA8 model and observation of the modeled IDRs for an initial placement within the expected conformational ensemble should be considered in addition to these scores when selecting candidate models for subsequent CESA trimer simulations.

Materials and Methods

Initially, FASTA sequences of PpCESAs 5 and 7 (1081 and 1096 residues, respectively) were submitted to modeling software SWISS-MODEL, I-TASSER, and I-TASSER-MTD. The PpCESA8 FASTA sequence (1091 residues) was only submitted to SWISS-MODEL and

I-TASSER due to the I-TASSER-MTD being down when its structure was submitted. Each FASTA sequence was submitted with an accompanying GhCESA1 hybrid structure template (974 residues), described by Kwansa et al.²⁰

To find the domains of each model, ChimeraX 1.5 was used to perform structure-based sequence alignment between each model and the cryo-EM derived PttCESA8 monomer described by Purushotham et al.¹ Alignment was completed using the Needleman-Wunsch algorithm and BLOSUM-62 matrix.⁶⁴

ERRAT, ProSA-web, and QMEANDisCo were used to evaluate the quality of each model's transmembrane and catalytic domains, excluding the intrinsically disordered CSR from the catalytic domain.⁵⁶⁻⁵⁸ This was followed by visual inspection of the IDRs of promising models to screen for favorable initial configurations. Finally, the evaluation scores of the most promising models were compared to the standards of high-quality models set forth by each evaluation software to ensure general model quality.

Following the selection of suitable initial models for each of PpCESA5, 7, and 8, an in-house script was used to identify pairs of identical residues in the alignments between each monomer and the PttCESA8 monomer. These identical residue pairs were used as the basis for the alignments of the monomers into PpCESA trimers. Importantly, only identical residue pairs with resolved atomic coordinates in the PttCESA8 cryo-EM structure were used (residues 157 to 549, 610 to 846, and 869 to 958), as several regions in this model remain structurally unresolved.¹ Using this filtered set of identical residue pairs with PttCESA8 spatial coordinates, seven PpCESA trimers were constructed for MD simulations within Visual Molecular Dynamics (VMD).⁶⁵ Trimer construction was completed using a custom Tcl script, which fit the PpCESA monomers to the PttCESA8 homotrimer by minimizing the RMSD (root mean square deviation)

of the C α atoms of the spatially resolved, identical residue pairs. This alignment strategy was chosen to maintain isoform-specific structural differences while ensuring consistency with the resolved template structure. The resulting trimer set included three homotrimers (PpCESAs 555, 777, and 888) and four heterotrimers, comprising combinations of PpCESA5 and 7 (557, 577) and PpCESA7 and 8 (887, 877).

RMSD values were calculated using VMD to assess the fit of each PpCESA monomer to its corresponding PttCESA8 monomer. These calculations were based on the C α atoms of the aligned spatially resolved identical residue pairs. Additionally, RMSD values were calculated for each complete trimer against the PttCESA8 homotrimer, using both C α atoms of the aligned spatially-resolved identical residue pairs as well as the C α atoms of all aligned spatially-resolved residues.

Results

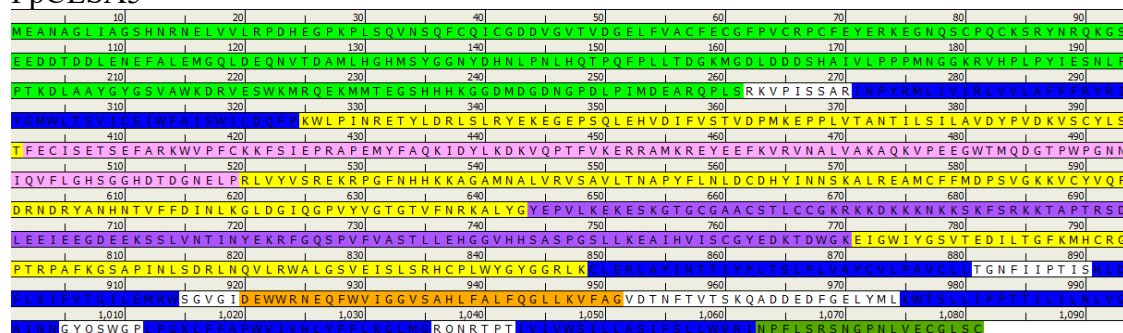
Structural modeling of the PpCESA isoforms resulted in 12 models for PpCESA5, 11 models for PpCESA7, and 6 models for PpCESA8. The corresponding breakdown of models by modeling program is shown in Table 2.1.

Table 2.1: Numbers of generated PpCESA models. Initial models of PpCESAs 5, 7, and 8 were generated using SWISS-MODEL, I-TASSER, and I-TASSER-MTD.

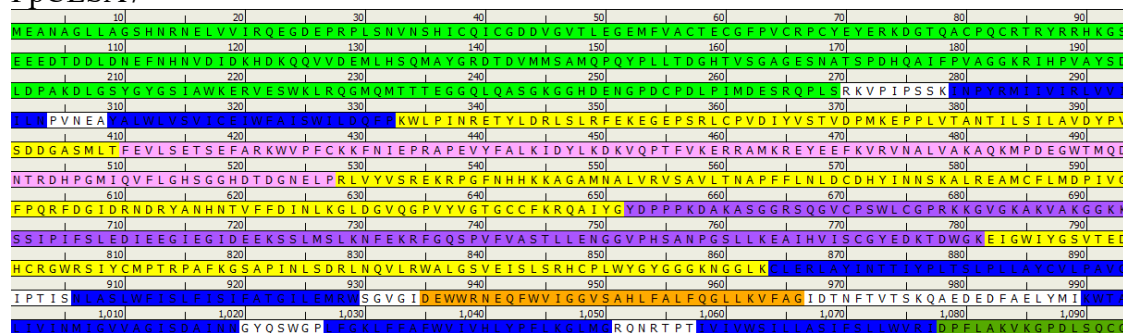
	PpCESA5	PpCESA7	PpCESA8
SWISS-MODEL	2	1	1
I-TASSER	5	5	5
I-TASSER-MTD	5	5	-
Total	12	11	6

Structural domains for each isoform were identified based on alignment to PttCESA8. These domains included the N-term, transmembrane domain (TMD), consisting of all seven transmembrane helices and interfacial helix 3 (TMH1-7 and IF3), CatD, which contains both the Plant-Conserved Region (PCR) and the Class Specific Region (CSR), and the C-terminus (C-term) (Figure 2.1, Table 2.2). Residue counts for each alignment of a PpCESA sequence (5, 7, and 8) to PttCESA8 are shown in Table 2.3, including counts for residues that are within spatially resolved regions of PttCESA8.

PpCESA5



PpCESA7



PpCESA8

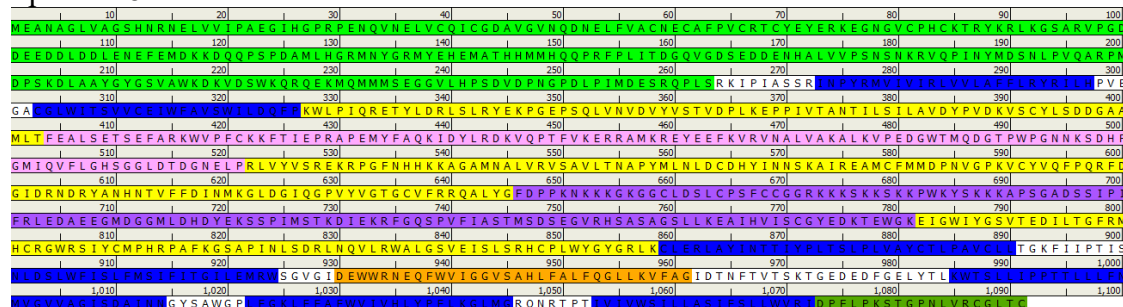


Figure 2.1: Annotated sequences of PpCESAs 5, 7, and 8. Domains of interest for PpCESAs

5, 7, and 8 were determined through alignment to PttCESA8. This includes the N-terminus (N-term) in light green, transmembrane domain (TMD), consisting of all seven transmembrane helices (TMH1-7) in blue and interfacial helix 3 (IF3) in orange, the catalytic domain (CatD) in yellow, which contains both the Plant-Conserved Region (PCR) in pink and the Class Specific Region (CSR) in purple, and the C-terminus (C-term) in green.

Table 2.2: PpCESA structural domains. Domains of interest for PpCESAs 5, 7, and 8 were determined through alignment to PttCESA8. This includes the N-terminus (N-term), transmembrane domain (TMD), consisting of all seven transmembrane helices and interfacial helix 3 (TMH1-7 and IF3), the catalytic domain (CatD), which contains both the Plant-Conserved Region (PCR) and the Class Specific Region (CSR), and the C-terminus (C-term). PttCESA8 domains are included for completeness.

CESA	Length	N-term	TMH1	TMH2	CatD	PCR	CSR
PttCESA8	1-978	1-160	170-194	200-223	224-743	301-418	543-664
PpCESA5	1-1081	1-261	271-295	301-324	325-848	402-519	644-770
PpCESA7	1-1096	1-269	279-303	309-332	333-863	410-527	652-781
PpCESA8	1-1091	1-263	273-297	303-326	327-858	404-521	646-781

CESA	TMH3	TMH4	IF3	TMH5	TMH6	TMH7	C-term
PttCESA8	744-775	786-809	815-846	870-899	907-930	938-957	958-978
PpCESA5	849-880	891-914	920-951	975-1004	1012-1035	1043-1062	1063-1081
PpCESA7	864-895	906-929	935-966	990-1019	1027-1050	1058-1077	1078-1096
PpCESA8	859-890	901-924	930-961	985-1014	1022-1045	1053-1072	1073-1091

Table 2.3: Residue counts for each alignment of the PpCESA sequences (5, 7, and 8) to PttCESA8. Residue counts are reported for both all aligned residues and for all aligned residues that are within spatially resolved regions of PttCESA8.

Alignment to PttCESA8	PpCESA5	PpCESA7	PpCESA8
Total number of aligned residues	973	964	970
Number of aligned residues with spatial coordinates	719	720	718

The quality of the catalytic domain (excluding the intrinsically disordered CSR) and transmembrane domain of each model for each isoform was assessed using ERRAT, ProSA-web, and QMEANDisCo (Table 2.4).

Table 2.4: Model quality scores for conserved PpCESA domains. The catalytic domain (CatD, excluding the disordered CSR) and the transmembrane domain (TMD) were evaluated using ERRAT, ProSA-web (ProSA in table), and QMEANDisCo (QMEAN in table). Results were compiled for each model for each isoform. The top three scores in each category for each isoform are highlighted in green, with the total of top scores per model included in the right column. PttCESA8 scores for these regions were included for comparison.

CESA Type	Model Source and Number	CatD (excluding CSR)			TMD			# of Top 3 Scores
		ERRAT	ProSA	QMEAN	ERRAT	ProSA	QMEAN	
PttCESA8	PDB: 6WLB	78.17	-7.51	0.77	100	1.97	0.67	-
PpCESA5	SWISS-MODEL 1	87.97	-7.26	0.68	98.97	2.57	0.52	1
	SWISS-MODEL 2	88.30	-7.31	0.68	98.97	2.56	0.52	1
	I-TASSER 1	97.15	-7.63	0.78	95.07	2.45	0.61	5
	I-TASSER 2	93.98	-7.26	0.68	96.06	2.84	0.54	0
	I-TASSER 3	92.73	-7.52	0.76	96.55	2.52	0.62	3
	I-TASSER 4	92.37	-7.52	0.77	100	1.72	0.61	5
	I-TASSER 5	89.56	-7.41	0.70	92.12	3.11	0.51	0
	I-TASSER-MTD 1	97.66	-7.25	0.75	96.06	3.18	0.55	1
	I-TASSER-MTD 2	89.61	-7.61	0.75	92.12	2.63	0.55	1
	I-TASSER-MTD 3	83.94	-4.62	0.69	67.49	6.06	0.47	0
	I-TASSER-MTD 4	90.65	-7.48	0.74	84.24	2.74	0.54	0
	I-TASSER-MTD 5	95.32	-7.42	0.75	88.67	2.41	0.54	2
PpCESA7	SWISS-MODEL	87.27	-7.47	0.68	98.47	3.16	0.59	2
	I-TASSER 1	89.35	-7.26	0.76	98.52	2.22	0.66	4
	I-TASSER 2	91.60	-7.38	0.67	96.06	3.07	0.60	1
	I-TASSER 3	91.03	-7.11	0.78	97.54	2.66	0.66	3
	I-TASSER 4	88.02	-7.39	0.76	96.55	2.04	0.66	4
	I-TASSER 5	95.57	-7.28	0.75	99.51	2.58	0.65	3
	I-TASSER-MTD 1	88.27	-7.47	0.74	88.18	3.19	0.61	1
	I-TASSER-MTD 2	77.95	-5.89	0.71	97.04	3.53	0.62	0
	I-TASSER-MTD 3	76.53	-5.62	0.70	97.04	3.52	0.62	0
	I-TASSER-MTD 4	87.10	-6.69	0.67	94.09	3.53	0.62	0
	I-TASSER-MTD 5	87.72	-7.17	0.72	67.98	3.40	0.50	0

Table 2.4: Model quality scores for conserved PpCESA domains (continued).

PpCESA8	SWISS-MODEL	97.63	-7.54	0.79	96.55	2.31	0.58	3
	I-TASSER 1	94.70	-7.48	0.76	98.52	2.27	0.60	6
	I-TASSER 2	90.65	-6.92	0.68	95.57	2.37	0.58	0
	I-TASSER 3	94.15	-7.38	0.76	100	2.29	0.61	5
	I-TASSER 4	89.82	-7.81	0.76	93.60	1.04	0.57	3
	I-TASSER 5	93.12	-7.41	0.77	98.03	2.44	0.60	3

The transmembrane and catalytic domains of the I-TASSER models exhibited more favorable quality scores than those produced by I-TASSER-MTD or SWISS-MODEL. Consequently, subsequent analyses focused exclusively on the I-TASSER models. The IDRs of these models were examined to identify structural features that appeared atypical even to the conformational ranges of IDRs. Two examples of atypical IDR configurations predicted by I-TASSER for models of PpCESA5 are presented in Figure 2.2. Following IDR evaluation, Model 1 of the generated I-TASSER models for each isoform was selected for final benchmarking against the established quality standards of ERRAT, ProSA-web, and QMEANDisCo. Table 2.5 shows the results of this comparison, with high-quality model scores shown in green. With their structural quality confirmed by their scores, I-TASSER Model 1 was selected as the monomer model that would be used for each isoform.

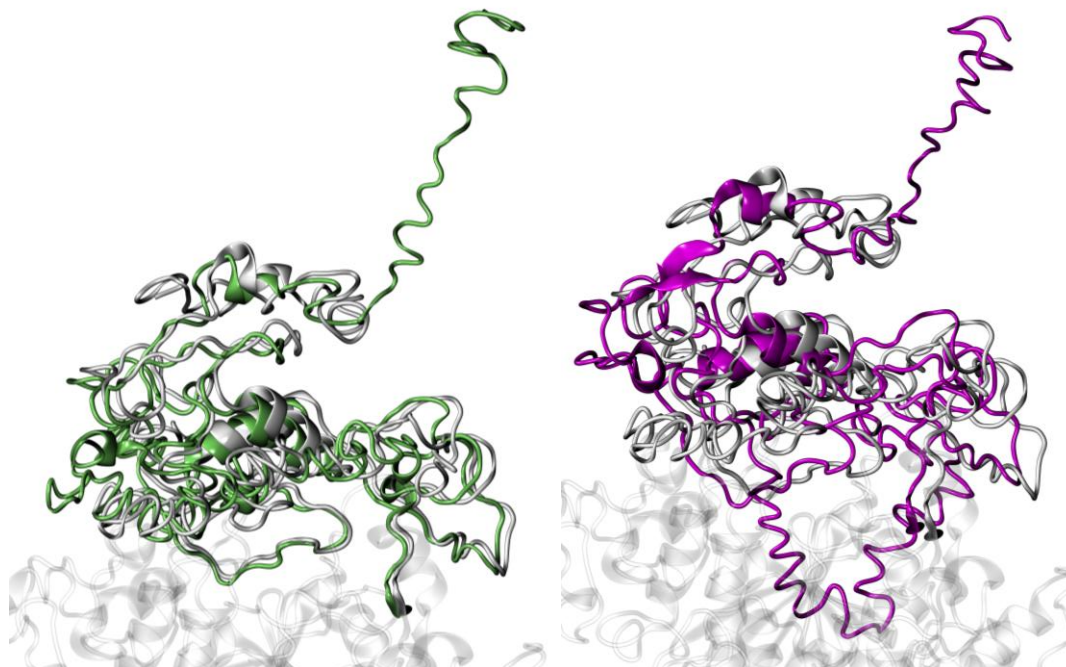


Figure 2.2: I-TASSER modeling of the N-terminal domain of PpCESA5. (A) Comparison of the modeled N-terminal domain of I-TASSER Model 1 (white, New Cartoon) to I-TASSER Model 4 (lime, New Cartoon). Part of the catalytic domain of Model 1 (transparent white, New Cartoon) is included for structural reference. The first 30 residues of Model 4 diverge significantly from their predicted locations in Model 1. (B) Comparison of the modeled N-terminal domain of I-TASSER Model 1 to I-TASSER Model 5 (magenta, New Cartoon). The first 30 residues of Model 5 diverge significantly from their predicted locations in Model 1. The I-TASSER Model 1 representation is the same as in (A).

Table 2.5: Quality scores for conserved PpCESA domains of selected PpCESA models. The catalytic domain (CatD, excluding the disordered CSR) and the transmembrane domain (TMD) were evaluated using ERRAT, ProSA-web (ProSA in table), and QMEANDisCo (QMEAN in table). Scores that fall within that range of “high-quality” for each evaluation method are highlighted in green. PttCESA8 scores for these regions were included for comparison.

CESA Type	Model Source and Number	CatD (excluding CSR)			TMD		
		ERRAT	ProSA	QMEAN	ERRAT	ProSA	QMEAN
PttCESA8	PDB: 6WLB	78.17	-7.51	0.77	100	1.97	0.67
PpCESA5	I-TASSER 1	97.15	-7.63	0.78	95.07	2.45	0.61
PpCESA7	I-TASSER 1	89.35	-7.26	0.76	98.52	2.22	0.66
PpCESA8	I-TASSER 1	94.70	-7.48	0.76	98.52	2.27	0.60

Figure 2.3 displays the alignments of the PpCESA sequences (5, 7, and 8) to PttCESA8. Counts of identical residue pairs for the full alignments, as well as those within the spatially resolved regions of the PttCESA8 monomer, are reported in Table 2.6.

Figure 2.3: ChimeraX alignments of the PpCESA sequences (5, 7, and 8) to PttCESA8.

Identical residue pairs in the alignments are highlighted in green, with the total number of these pairs indicated for each sequence. The ruler corresponds to the full alignment length.

PpCESA5-PttCESA8 alignment - 653 pairs of identical residues

```

-----10-----20-----30-----40-----50-----60-----70-----80-----90-----100
-----MMESGAPICHTCGEQVGH DANGDLFVAGHECNYHIKCKSCFEYEIKERKRVLRGGSRY-----GSP
MEANAGLAGSHNRNRLVLRPDHEGPKPLSQVNSQFCQICGDDYGVTVDGELEFVAGFECGFPVCRPCFEYERKEGNQSLRQCKSRYNRQKGSRPVPGDE
110-----120-----130-----140-----150-----160-----170-----180-----190-----200
DENLDDVEKKGSLNEFALEMGLDEQNVTDAMLHGHSYGGNQGSTMASHLNNSQ-----DVGTHA-----RHISVSTVDS-EMNDE-----
EEDTDDLNEFALEMGLDEQNVTDAMLHGHSYGGNQGSTMASHLNNSQ-----DVGTHA-----RHISVSTVDS-EMNDE-----
210-----220-----230-----240-----250-----260-----270-----280-----290-----300
-----YGNPIWKNRVESWKKDRNKKKSSNTKPEPEPAQVPEEQGMENKPSAESEPLSIVYPIPRNKLTPLYRAVIMRLIILGLFFHYRITNPVDS
PTKDLAAYGVGSVAWKDRVESWKKMRQKMMTEGSH-HKGGMDGDGNDPLIMDEARQLSRKVPIS SARINPYRMLVI RLVVLAFFRIRLNPVWEG
310-----320-----330-----340-----350-----360-----370-----380-----390-----400
AFBLWLSVYCEIWFASFWDQFPKWLIRNRETFIERLSARYEREGEPSQLAAVDFVYSTVDPLKEPPLITANTVLSILAVDYPVKYSCYVSDDGAM
AYGMWLSVYCEIWFASFWDQFPKWLIRNRETFIERLSARYEREGEPSQLAAVDFVYSTVDPMKEPPLVITANTVLSILAVDYPVKYSCYVSDDGAM
410-----420-----430-----440-----450-----460-----470-----480-----490-----500
LTFESLVEIAEFARKWVPPCKKFSIEPRAPEFYFSQKIDYLDKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPDEGWTMQDGTWPWPGNTRDHPG
LTFECISESEFARKWVPPCKKFSIEPRAPEMYFAQKIDYLDKDKVQPTFVKERRAMKREYEEFKVRVNALVAKAQKVPDEGWTMQDGTWPWPGNTRDHPG
510-----520-----530-----540-----550-----560-----570-----580-----590-----600
MIQVFLQNTGARDIEGNELPRLVYVSREKRPQYQHKKKAGAEALNRVSAVLTNAPYILNLDGDHYVNNSKAVREAMCILMDPQVGRDVCYVQFPQRFDG
MIQVFLQNTGARDIEGNELPRLVYVSREKRPQYQHKKKAGAEALNRVSAVLTNAPYILNLDGDHYVNNSKALREAMCFMDPQVGRDVCYVQFPQRFDG
610-----620-----630-----640-----650-----660-----670-----680-----690-----700
IDRSRYANRNIYVFDVNMKGLDGLGGPVMYVGTGCVFNRAQALYGYGPPSMPLRLKKGK-SSSCFSCCPTKKKPAQDPAEYVYRDAK-----REDLNAAIIF
IDRNDRYANRNIYVFDVNMKGLDGLGGPVMYVGTGCVFNRAQALYGYGPPSMPLRLKKGK-SSSCFSCCPTKKKPAQDPAEYVYRDAK-----REDLNAAIIF
710-----720-----730-----740-----750-----760-----770-----780-----790-----800
NLTEDINYDYRSMISQLSFEKTFGLSPVPIESTLMENGGVPEASNSTIKEAITHVIGCGFEETKTEWKEIGWIYGSVTEILSGFKMHCGRGWSIY
SLEIEEGDEKSSVNTINYEKRFQSPVFASTLLEHGGVHSSAPGSLKEAITHVIGCGFEETKTEWKEIGWIYGSVTEILSGFKMHCGRGWSIY
810-----820-----830-----840-----850-----860-----870-----880-----890-----900
CMPVLRPAFKGSAPINLSDRLNQVLRWALGSVEISLSRHCFWYGYGGRLKWKERLAYINTIYVYFSLPIIAYCTIPAVCLLTGKFIIPITLSNLASMLF
CMPTRPAFKGSAPINLSDRLNQVLRWALGSVEISLSRHCFWYGYGGRLKWKERLAYINTIYVYFSLPIIAYCTIPAVCLLTGKFIIPITLSNLASMLF
910-----920-----930-----940-----950-----960-----970-----980-----990-----1,000
LGLFISIVTAVLELRWSGVSEDLWRNEQFVWVIGGVS AHLFAVFOGLKMLAGIDTNTFTVAKAADDTEFGELYMKWITLLIPPTLLIINIYGVVAG
ISLFLSIVTAVLELRWSGVSEDLWRNEQFVWVIGGVS AHLFAVFOGLKMLAGIDTNTFTVAKAADDTEFGELYMKWITLLIPPTLLIINIYGVVAG
1,010-----1,020-----1,030-----1,040-----1,050-----1,060-----1,070-----1,080-----1,090-----1,100
FSDALNKGVEAWGPLFGKFFAFWVILHLYPFLKGLMGRQNRTPTIIVLWSVLLTSVLSLWVWKINPFLSKYVNDNTLAGETCISIDC
ISDAINNGYSWGPLFGKFFAFWVILHLYPFLKGLMGRQNRTPTIIVLWSVLLTSVLSLWVWKINPFLSKYVNDNTLAGETCISIDC

```

PpCESA7-PttCESA8 alignment - 661 pairs of identical residues

```

-----10-----20-----30-----40-----50-----60-----70-----80-----90-----100
-----MMESGAPICHTCGEQVGH DANGDLFVAGHECNYHIKCKSCFEYEIKERKRVLRGGSRY-----GSP
MEANAGLAGSHNRNRLVLRPDHEGPKPLSQVNSQFCQICGDDYGVTVDGELEFVAGFECGFPVCRPCFEYERKEGNQSLRQCKSRYNRQKGSRPVPGDE
110-----120-----130-----140-----150-----160-----170-----180-----190-----200
DENLDDVEKKGSLNEFALEMGLDEQNVTDAMLHGHSYGGNQGSTMASHLNNSQ-----DVGTHA-----RHISVSTVDS-EMNDE-----
EEDTDDLNEFALEMGLDEQNVTDAMLHGHSYGGNQGSTMASHLNNSQ-----DVGTHA-----RHISVSTVDS-EMNDE-----
210-----220-----230-----240-----250-----260-----270-----280-----290-----300
-----YGNPIWKNRVESWKKDRNKKKSSNTKPEPEPAQVPEEQGMENKPSAESEPLSIVYPIPRNKLTPLYRAVIMRLIILGLFFHYRITNPVDS
PTKDLAAYGVGSVAWKDRVESWKKMRQKMMTEGSH-HKGGMDGDGNDPLIMDEARQLSRKVPIS SARINPYRMLVI RLVVLAFFRIRLNPVWEG
310-----320-----330-----340-----350-----360-----370-----380-----390-----400
AFBLWLSVYCEIWFASFWDQFPKWLIRNRETFIERLSARYEREGEPSQLAAVDFVYSTVDPLKEPPLITANTVLSILAVDYPVKYSCYVSDDGAM
AYGMWLSVYCEIWFASFWDQFPKWLIRNRETFIERLSARYEREGEPSQLAAVDFVYSTVDPMKEPPLVITANTVLSILAVDYPVKYSCYVSDDGAM
410-----420-----430-----440-----450-----460-----470-----480-----490-----500
LTFESLVEIAEFARKWVPPCKKFSIEPRAPEFYFSQKIDYLDKDKVQPSFVKERRAMKRDYEEYKVRVNALVAKAQKTPDEGWTMQDGTWPWPGNTRDHPG
LTFECISESEFARKWVPPCKKFSIEPRAPEMYFAQKIDYLDKDKVQPTFVKERRAMKREYEEFKVRVNALVAKAQKVPDEGWTMQDGTWPWPGNTRDHPG
510-----520-----530-----540-----550-----560-----570-----580-----590-----600
DGTWPWPGNTRDHPGMIQVFLGNTGARDIEGNELPRLVYVSREKRPQYQHKKKAGAEALNRVSAVLTNAPYILNLDGDHYVNNSKAVREAMCILMDPQV
DGTWPWPGNTRDHPGMIQVFLGNTGARDIEGNELPRLVYVSREKRPQYQHKKKAGAEALNRVSAVLTNAPYILNLDGDHYVNNSKALREAMCFMDPQV
610-----620-----630-----640-----650-----660-----670-----680-----690-----700
GRDVCYVQFPQRFDGIDRSRYANRNIYVFDVNMKGLDGLGGPVMYVGTGCVFNRAQALYGYGPPSMPLRLKKGK-SSSCFSCCPTKKKPAQDPAEYVYRDAK
GKRVCYVQFPQRFDGIDRSRYANRNIYVFDVNMKGLDGLGGPVMYVGTGCVFNRAQALYGYGPPSMPLRLKKGK-SSSCFSCCPTKKKPAQDPAEYVYRDAK
710-----720-----730-----740-----750-----760-----770-----780-----790-----800
K-----REDLNAAIIFNLETID--NYDDYERSMISQLSFEKTFGLSPVPIESTLMENGGVPEASNSTIKEAITHVIGCGFEETKTEWKEIGWIYGSVTE
KKKPPSRSSSIPISLEIEEGIEGEEKSSMSLKNFERKFGQSPVFASTLLEHGGVHSSAPGSLKEAITHVIGCGFEETKTEWKEIGWIYGSVTE
810-----820-----830-----840-----850-----860-----870-----880-----890-----900
EDILSGFKMHCGRGWSIYCMPTVLRPAFKGSAPINLSDRLNQVLRWALGSVEISLSRHCFWYGYGGRLKWKERLAYINTIYVYFSLPIIAYCTIPAVCLLTGKFIIPITLSNLASMLF
EDILSGFKMHCGRGWSIYCMPTVLRPAFKGSAPINLSDRLNQVLRWALGSVEISLSRHCFWYGYGGRLKWKERLAYINTIYVYFSLPIIAYCTIPAVCLLTGKFIIPITLSNLASMLF
910-----920-----930-----940-----950-----960-----970-----980-----990-----1,000
VCLLTGKFIIPITLSNLASMLFVCLLTGKFIIPITLSNLASMLFVCLLTGKFIIPITLSNLASMLFVCLLTGKFIIPITLSNLASMLFVCLLTGKFIIPITLSNLASMLF
VCLLTGKFIIPITLSNLASMLFVCLLTGKFIIPITLSNLASMLFVCLLTGKFIIPITLSNLASMLFVCLLTGKFIIPITLSNLASMLFVCLLTGKFIIPITLSNLASMLF
1,010-----1,020-----1,030-----1,040-----1,050-----1,060-----1,070-----1,080-----1,090-----1,100
TLLIPPTLLIINIYGVVAGFSDALNKGVEAWGPLFGKFFAFWVILHLYPFLKGLMGRQNRTPTIIVLWSVLLTSVLSLWVWKINPFLSKYVNDNTLAGETCISIDC
TLLIPPTLLIINIYGVVAGFSDALNKGVEAWGPLFGKFFAFWVILHLYPFLKGLMGRQNRTPTIIVLWSVLLTSVLSLWVWKINPFLSKYVNDNTLAGETCISIDC
1,110-----1,120-----1,130-----1,140-----1,150-----1,160-----1,170-----1,180-----1,190-----1,200
AETCISIDC
CITNG-----

```

PpCESA8-PttCESA8 alignment - 648 pairs of identical residues

Table 2.6: Number of identical residue pairs for the alignments of the PpCESA sequences (5, 7, and 8) to PttCESA8. Residue counts are reported both for all aligned residues and for aligned residues located within spatially resolved regions of PttCESA8.

Alignment to PttCESA8	PpCESA5	PpCESA7	PpCESA8
Total number of identical pair aligned residues	653	661	648
Number of identical pair aligned residues with spatial coordinates	562	564	554

All seven PpCESA trimeric combinations were generated as described (Figure 2.4). To assess the quality of the monomer-level alignments, RMSD values were calculated for each PpCESA isoform in each trimer relative to its corresponding PttCESA8 monomer. These RMSDs were calculated using the C α atoms of all aligned, spatially resolved, sequence-identical residue pairs. The resulting values are listed in Table 2.7.

In addition to monomer-level RMSD values, RMSD values were also calculated for each complete PpCESA trimer relative to the PttCESA8 homotrimer. Two sets of calculations were performed for each trimer: one using only the C α atoms of spatially resolved, sequence-identical

residue pairs, and another using all spatially resolved C α atoms from aligned residues, regardless of sequence identity (Table 2.8).

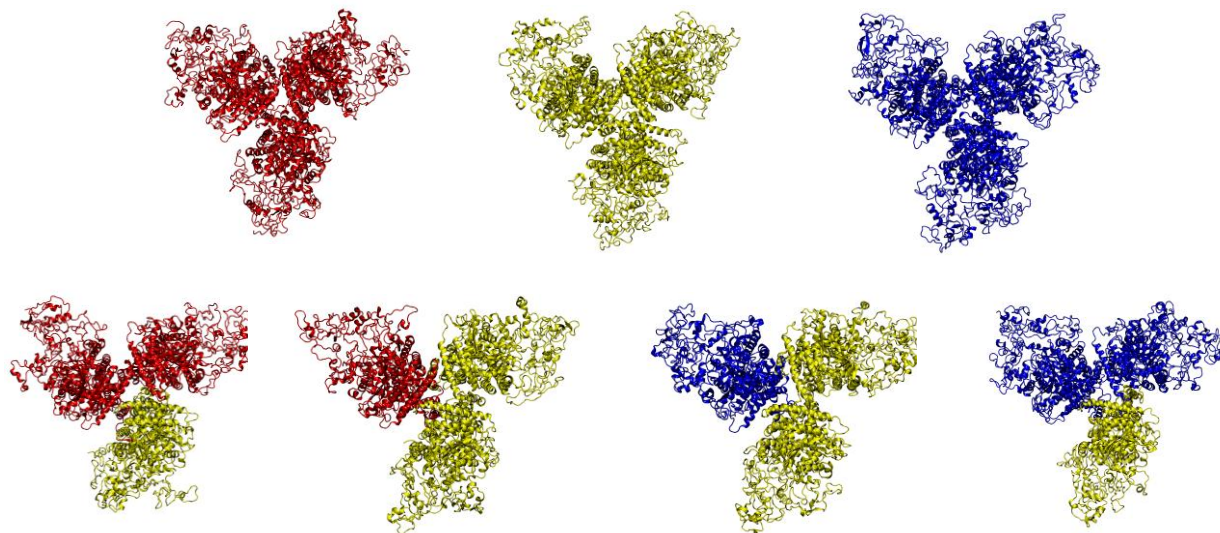


Figure 2.4: Generated PpCESA trimers. All seven generated PpCESA trimers, viewed from the cytosol. Homotrimers are on the top row (from left to right: 555, 777, 888) and heterotrimers are on the bottom row (from left to right: 557, 577, 877, 887). Monomers within the trimer are colored by PpCESA isoform (5 is red, 7 is yellow, and 8 is blue).

Table 2.7: RMSD values for each PpCESA monomer within each trimer, measured relative to its corresponding PttCESA8 monomer. RMSD values were calculated using the C α atoms of all aligned, spatially resolved, sequence-identical residue pairs. The corresponding PpCESA isoform type is included in parentheses below the RMSD value.

	555	557	577	777	877	887	888
PttCESA8 Chain A RMSD (Å)	0.76 (5)	0.76 (5)	0.76 (5)	1.15 (7)	1.10 (8)	1.10 (8)	1.10 (8)
PttCESA8 Chain B RMSD (Å)	0.76 (5)	0.76 (5)	1.16 (7)	1.16 (7)	1.16 (7)	1.10 (8)	1.10 (8)
PttCESA8 Chain C RMSD (Å)	0.76 (5)	1.16 (7)	1.16 (7)	1.16 (7)	1.16 (7)	1.16 (7)	1.10 (8)

Table 2.8: RMSD values from the structural superposition of each PpCESA trimer onto the template PttCESA8 homotrimer. Both RMSD fits were calculated based on C α atom pairs that were spatially resolved in the PttCESA8 homotrimer. RMSD values are reported for both the subset of identical residue pair C α atoms used during alignment and all aligned C α atoms within the resolved regions.

RMSD of all C α atoms from identical pair aligned residues with spatial coordinates

	555	557	577	777	877	887	888
RMSD (Å)	0.76	0.91	1.04	1.16	1.14	1.12	1.10

RMSD of all C α atoms from aligned residue pairs with spatial coordinates

	555	557	577	777	877	887	888
RMSD (Å)	0.98	1.13	1.26	1.38	1.35	1.31	1.28

Discussion

SWISS-MODEL, I-TASSER, AND I-TASSER-MTD each successfully generated complete structures for PpCESA isoforms 5, 7, and 8. However, structural analysis of the models using ERRAT, ProSA-web, and QMEANDisCo revealed overall quality differences in the modeling of the conserved transmembrane and catalytic domains by each modeling software. Specifically, the analysis revealed that the I-TASSER models were overall of a higher quality than those of either I-TASSER-MTD or SWISS-MODEL, based on the overall number of top 3 scores I-TASSER models received from the evaluation software. This, combined with the visual examination of the IDRs of the I-TASSER models, led to the conclusion that the most promising PpCESA model for each isoform was the I-TASSER Model 1.

The evaluations scores for each isoform's I-TASSER Model 1, when compared to the high-quality model thresholds for each evaluation technique, agreed that the I-TASSER Model 1s were promising structural candidates for each PpCESA isoform. ERRAT scores above 95% correspond with higher quality models, which was met by the TMD and CatD of the PpCESA5 model, the TMD of the PpCESA7 model, and the TMD of the PpCESA8 model. The CatD of the PpCESA8 model scored slightly outside the range (94.7%). While the CatD of the PpCESA7 model scored 89.35%, this still exceeds the 78.17% ERRAT score for the CatD of the cryo-EM PttCESA8 monomer, supporting the plausibility of the PpCESA7 model. QMEANDisCo scores for each model met or surpassed the 0.60 threshold for quality models. The ProSA-web z-scores for the CatD of all three models fell within the expected range of z-scores for proteins of their size, while the TMD z-scores did not. The positive z-scores observed for the TMDs, typically indicative of structural errors, instead likely reflect the

under-representation of membrane proteins in the ProSA-web reference set, which is predominantly derived from soluble globular proteins.⁵⁷

The high conformational variability observed at the beginning of the N-terminal region of the generated models likely reflects limitations in modeling this part of the IDR due to the absence of a structural template, as the GhCESA1 hybrid model did not extend to those residues. I-TASSER Model 1 for each isoform generated this region in a compact conformation, while other I-TASSER models generated linear N-terminal conformations. In agreement with previous investigations of intrinsic disorder, which have demonstrated that IDRs tend to adopt some compactness, rather than fully random coils, the compact conformation was selected as a more plausible starting state within the IDR conformational ensemble.^{66,67}

The low RMSD values between the PpCESA monomers and the template PttCESA8 trimer during trimer assembly indicate close structural agreement between the C α atoms of the spatially resolved, identical residue pairs of the trimers and the experimentally derived model. This is further confirmed by the low RMSD values of the completed trimers compared to the experimental PttCESA8 trimer model. While the RMSD values of the C α atoms of the spatially resolved, identical residue pairs for the entire trimers were expected to be low, as these were the residues used to perform the structural alignment, the RMSD values of all C α atoms residue pairs with spatial coordinates being similarly low (less than 1.5 Å for each trimer) indicates close structural agreement of the trimers to the PttCESA8 homotrimer model over all resolved regions.

CHAPTER 3: Assessment of *Physcomitrium patens* Cellulose Synthase Trimers using Molecular Dynamics Simulations

Introduction

Molecular dynamics (MD) simulations were first developed in the 1950s, but the simulation of bovine pancreatic trypsin inhibitor in 1977 marked the beginning of widespread use of MD methods in structural biology.^{68,69} Since then, MD simulations have become an essential tool in computational biophysics, enabling the exploration of molecular motion and energetics with atomic-level precision beyond what static experimental techniques can provide. MD simulations can be used to model complex and diverse biological interactions, including ligand binding, membrane transport, and protein folding, increasing understanding of these processes at an atomic level.^{70,71} MD simulations also enable characterization of IDR structure and dynamics, regions whose conformational heterogeneity complicates resolution by X-ray crystallography.⁷² All-atom explicit solvent MD simulations have previously been employed to investigate CESA systems, including CESA monomer active site interactions with UDP-glucose for 500 ns, CESA trimer CatD interactions for 1 μ s, and CESA monomer interactions with glycan chains for 500 ns.^{5,48,73} This last simulation is of particular note, as the CESAs were modeled in trimeric form and embedded within a heterogeneous lipid bilayer.

Initial parameters of MD simulations include atomic positions, force fields, and environmental conditions such as temperature and pressure. The atomic configuration is propagated over time by numerically integrating Newton's equations of motion, which update positions and velocities at each time step.⁷⁴ From the MD simulations, system information such as energy contributions, pressure, volume, temperature, and density can be obtained over a

defined time period.⁷⁵ These results offer molecular-level insights that can complement and assist the interpretation of wet lab experiments. While there are many simulation software packages that may be utilized for MD simulations, this work will describe the use of Amber 2020 (Amber20 and AmberTools20), chosen for its computational efficiency.⁷⁵

MD simulations require force fields for different types of molecules in the system (water, proteins, ions, lipids, sugars). The energy parameters in these files can be broken down into two categories. Bonded energy (Equation 3.1) is the sum of the energies of bond lengths, atom angles, and dihedral angles. Nonbonded energy (Equation 3.2) is the sum of the 6-12 Lennard Jones (LJ) potential and Coulombic interactions. Together, when calculated over the entire system, these energies sum to the total potential energy of the system (Equation 3.3).⁷⁵

$$U_{bonded} = \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \quad (3.1)$$

$$U_{nonbonded} = \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right] + \sum_{i < j} \left[\frac{q_i q_j}{\epsilon R_{ij}} \right] \quad (3.2)$$

$$U_{total} = U_{bonded} + U_{nonbonded} \quad (3.3)$$

In the bonded energy equation k_r is the bond force constant (kcal/mol/Å²), r is the bond length (Å), r_{eq} is the equilibrium bond length (Å), k_θ is the angle force constant (kcal/mol/rad²), θ is the angle (deg), θ_{eq} is the equilibrium angle (deg), V_n is the torsion force constant (kcal/mol), n is the torsion periodicity (unitless), ϕ is the torsion angle (deg), and γ is the phase shift (deg).

In the nonbonded energy equation A_{ij} is the repulsive coefficient that represents the LJ energy-well depth multiplied by the twelfth power of the equilibrium interatomic distance between two atoms, i and j , (kcal/mol*Å¹²), B_{ij} is the attractive LJ coefficient that represents twice the LJ energy-well depth multiplied by the sixth power of the equilibrium interatomic distance between two atoms, i and j , (kcal/mol*Å⁶), R_{ij} is the interatomic distance between two

atoms, i and j , (\AA), q_i and q_j are the partial charges of atoms i and j (C), and ϵ is the effective permittivity ($\text{C}^2/(\text{kcal/mol}\cdot\text{\AA})$).⁷⁵

The total potential energy of the system can then be used to calculate the force on an atom. Force is defined as the negative gradient of the potential energy with respect to the atom's, i , position and is calculated by taking partial derivatives of the potential energy with respect to the x, y, and z components of the atom's position vector (Equation 3.4).

$$\vec{F}_i = -\nabla_i U = -\left(\frac{\partial U}{\partial x} + \frac{\partial U}{\partial y} + \frac{\partial U}{\partial z}\right) \quad (3.4)$$

Once the force vector has been calculated, the acceleration of the atom can be calculated using Newton's Second Law (Equation 3.5).

$$a_i = \frac{F_i}{m_i} \quad (3.5)$$

Using the acceleration of the atom and its velocity, derived from the acceleration, the atom's position, r , following the simulation timestep, $t + \delta t$, can be calculated (Equation 3.6).

$$r_i(t + \delta t) = r_i(t) + v_i(t)\delta t + \frac{1}{2}a_i(t)\delta t^2 \quad (3.6)$$

Amber20 calculates the velocity of an atom using the Velocity-Verlet algorithm.

Velocity-Verlet calculates the updated velocity by first deriving the velocity at the half-time step, $v_i(t + \frac{1}{2}\delta t)$ (Equation 3.7), and using that half-step velocity and the acceleration of the new time step (calculated from the potential energy of the new atomic positions) to calculate velocity at the new time step, $v_i(t + \delta t)$ (Equation 3.8).

$$v_i\left(t + \frac{1}{2}\delta t\right) = v_i(t) + \frac{1}{2}a_i(t)\delta t \quad (3.7)$$

$$v_i(t + \delta t) = v_i\left(t + \frac{1}{2}\delta t\right) + \frac{1}{2}a_i(t + \delta t)\delta t \quad (3.8)$$

From the calculated atomic coordinates and velocities, simulation trajectories are generated and can be subsequently analyzed to extract structural and energetic descriptors, including RMSD,

linear interaction energy (LIE), and secondary structure, which provide insight into the system's conformational variability and stability.

While atomic motion in MD simulations is fully determined by Newtonian mechanics, the resulting trajectory depends on configurable aspects of the system including the simulation environment and force field selection. PpCESA trimers, being integral membrane proteins in plant cells, require a native-like lipid-bilayer environment for structural fidelity. To create an experimentally relevant *P. patens* membrane environment, essential to ensure native PpCESA folding patterns, this work draws on the findings of Resemann et al., who have characterized the membrane lipid composition of *P. patens*.⁷⁶ They determined that the most prevalent lipid head groups (that were available within the lipid17 force field) were phosphatidylcholine (54%), phosphatidylglycerol (22%), phosphatidylethanolamine (13%), and phosphatidic acid (4%), totaling 93% of the lipid head membrane composition. Similarly, the most prevalent fatty acid tail groups (that were available within the lipid17 force field) were palmitic acid (33.1%) and arachidonic acid (16.1%). While this accounts for only 49.2% of the fatty acid tail composition, the next two most prevalent fatty acid tail types were not available within lipid17. These four fatty acid tails together would have accounted for 74.8% of the fatty acid membrane composition. This work utilized the relative composition of the available lipids to model the bilayer.

An explicit solvent system was used in the simulation. This was required to accurately model the membrane environment, as implicit water can fail to capture water-lipid interactions accurately due to the loss of specific interactions between the water and the polar head groups.⁷⁷ Potassium and chloride ions were added within the solvent in a 0.15 M concentration to mimic

the natural concentration of potassium in plant cells, in agreement with previous PpCESA MD simulation work.^{78,79}

Another consideration in simulation setup is how the selected force fields model IDRs. Many traditional protein force fields produce overly compact structural ensembles.⁸⁰ Protein force fields parameterized specifically for IDPs, such as a99SB-disp, have demonstrated substantial improvements in capturing conformational ensembles while retaining the accuracy of more structured regions.⁸¹ The selection of the water model itself also has a large effect on IDP structure.^{80,82} TIP3P, a popular water model, demonstrates weaker-than-experimental solute-solvent interactions, indicating that the overly compact regions are partially due to understabilization of the protein by the surrounding water. For this same reason TIP3P is also known to overstabilize secondary structure when compared to the OPC water model.⁸⁰ Though the current protein force field the developers of Amber recommend, ff19SB, has shown improvement in modeling flexible peptides when paired with the OPC water model, including modeling less compact IDPs, it can still struggle accurately modeling these structures.^{80,83} The developers of ff19SB have indicated their intent to use this force field pairing in future IDP studies.

This work uses the combination of ff14SB and TIP3P to study these trimers. As CESAs are largely structured proteins and this research focuses specifically on monomer-monomer interface interactions within the PpCESA trimers, where IDRs are fully absent, any limitations of the protein and water force fields in modeling IDRs is expected to have minimal impact on the studied interactions. However, the consideration of force fields will need to be revisited should CSCs be modeled, as the disordered CSR is directly implicated in their trimer-trimer contacts.¹⁷

In total, the force field package for the following simulations is ff14SB for proteins, lipid17 for lipids, TIP3P for water with Joung/Cheatham ion parameters, and GLYCAM_06j-1 for carbohydrates.^{84–88} In addition to their determined suitability for this work, this force field ensemble allows direct comparison with previous studies on CESAs, including a recent investigation of PpCESA5-glucan chain interactions.^{20,48} Had this work been entirely independent from prior work, ff19SB and OPC would have been strongly considered for use as the protein and water force fields.

The all-atom explicit solvent MD simulations of *P. patens* trimers described herein give novel insight into differences between homotrimeric and heterotrimeric PpCESA interfaces. These simulations are in agreement with the findings of Ho et al., as described in Chapter 1.¹⁷ Through the use of these simulations, it was found that PpCESA homotrimeric interfaces exhibit an increased stability when compared to those of heterotrimers made by swapping one monomer in a homotrimer to a different isoform, supporting the hypothesis that stable CSCs will be comprised of homotrimers.

Materials and Methods

These methods describe the pre-simulation preparation and MD simulation of the seven *P. patens* trimers described in Chapter 2. For redundancy, the PpCESA isoform compositions of the trimers are 555, 557, 577, 777, 877, 887, and 888.

Initially, a 22-mer glucan chain was placed into the transmembrane channel of each monomer in each trimer. These chains are an extension of the cellopentaoses originally resolved in the initial PttCESA8 cryo-EM structure determined by Purushotham et al. and have been extended to mimic nascent glucan chains.^{1,20} Then, PACKMOL-Memgen was used to generate

the cellular environment, which included a phospholipid bilayer, 0.15 M K⁺ and Cl⁻ ions, and explicit TIP3P water.^{89,90} The phospholipid bilayer was comprised 39% DPPC, 16% DPPG 9% DPPE, 3% DPPA, 19% DAPC, 8% DAPG, 5% DAPE, and 1% DAPA, where DP and DA (dipalmitoyl and diarachidonyl, respectively) were the fatty acid tail groups and PC, PG, PE, and PA (phosphatidylcholine, phosphatidylglycerol, phosphatidylethanolamine, and phosphatidic acid, respectively) were the lipid head groups.

The Amber 2020 software package was then used to perform the MD simulation, which followed a 10-stage protocol.⁷⁵ CPU-only PMEMD (Particle-Mesh Ewald Molecular Dynamics) was used for minimization and GPU-accelerated single precision-fixed point PMEMD was used for all subsequent stages.^{91,92} Each simulation employed periodic boundary conditions and PME electrostatics with a 10 Å cutoff.⁹³ A Langevin thermostat ($\gamma_{ln}=1$) and a Berendsen barostat (set to 1 atm, 1 ps relaxation time, 44.6×10^{-6} bar⁻¹ compressibility, and anisotropic scaling) were applied in relevant stages.^{94,95} Each simulation stage was conducted under either NVT (constant number of particles, fixed volume, and regulated temperature) or NPT (constant number of particles, regulated pressure, and regulated temperature) ensembles. The ff14SB, Lipid17, GLYCAM_06j-1, and TIP3P water force fields were used along with Joung/Cheatham ion parameters.⁸⁴⁻⁸⁸

Simulations began with an up-to-10,000-step NVT minimization stage, consisting of 5,000 steps of the steepest descent method followed by 5,000 steps of the conjugate gradient method. During this stage, restraints of 10 kcal/mol/Å² were applied to all protein atoms and 2.5 kcal/mol/Å² to all phosphorus atoms. Following minimization, for each stage, bonds involving hydrogen atoms were constrained using the SHAKE algorithm.⁹⁶ The system was then gradually heated from 0 K to 300 K over a 100 ps NVT heating stage with a 1 fs timestep. The

system was kept at 300 K for the remaining equilibration and final production stages. Two 100 ps NVT equilibration stages with a 1 fs timestep followed heating, the first maintaining the same restraints as the minimization stage and the second reducing the protein restraints to 5 kcal/mol/Å². Subsequently, equilibration continued under NPT conditions in five stages: 100 ps with restraints of 2.5 kcal/mol/Å² on protein and 1.0 kcal/mol/Å² on lipids and a 1 fs timestep, 100 ps with 1.0 and 0.5 kcal/mol/Å² restraints and a 2 fs timestep, 100 ps with 0.5 and 0.1 kcal/mol/Å² restraints and a 2 fs timestep, 100 ps with only 0.1 kcal/mol/Å² protein restraints and a 2 fs timestep, and finally 400 ps without any restraints and a 2 fs timestep. Following equilibration, a 500 ns NPT final production stage with a 2 fs timestep was performed without any restraints.

Analyses of the trimers (RMSD, LIE, and secondary structure) were performed using options within the AmberTools20 CPPTRAJ program.⁷⁵

Results

500 ns simulations were completed for each trimer. Convergence of each trimer was determined to have occurred by 200 ns, so the final 300 ns of each simulation was used for analysis.

Values for the average percent secondary structure composition of each trimer following convergence were determined using CPPTRAJ, which utilizes the DSSP algorithm. It was determined that, between all seven trimers, the percent variance in secondary structure was at maximum less than 2 percentage points (Figure 3.1, Table S1).

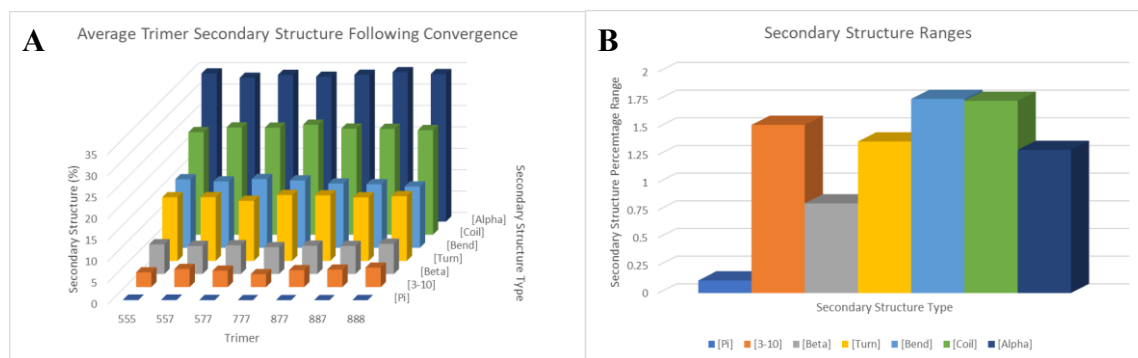


Figure 3.1: Overall trimer secondary structure. (A) Average secondary structure percent composition for each trimer over the last 300 ns of each simulation. (B) Range of trimer secondary structure composition variance was at maximum less than 2 percentage points.

For each monomer in each trimer, interface residues were determined by selecting atoms within a 5 Å cutoff from its neighboring monomers (Figure 3.2). Any residue containing at least one of these atoms was selected as an interface residue and the totals were counted (Table 3.1). They were then compared to each trimer's total residue count to get a % interface residue value for each trimer.

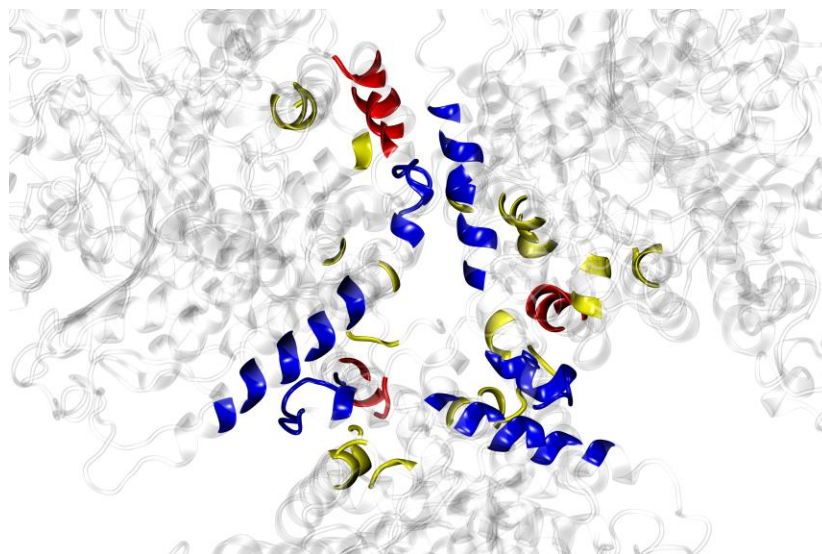


Figure 3.2: PpCESA 555 homotrimer interface residues, represented using VMD’s “NewCartoon” drawing method. Residues were selected if they contained an atom within 5 Å of a neighboring monomer during the last frame of the 500 ns simulation. Residues from different monomer atoms are colored blue, yellow, and red respectively.

Table 3.1: Trimer interface residue count. Number of interface residues for each trimer. Each residue contains at least one atom within 5 Å of a neighboring monomer. The total number of residues and the resultant calculated % interface residues for each trimer are also reported.

Trimer:	555	557	577	777	877	887	888
# Interface Residues:	242	278	263	300	293	288	321
# Total Residues:	3243	3258	3273	3288	3283	3278	3273
% Interface Residues:	7.46	8.53	8.04	9.12	8.92	8.79	9.81

Following the determination of interface residues, CPPTRAJ was used to calculate the RMSD for each trimer’s interface compared to its post-convergence average structure. These RMSD values were examined both for individual trimers as well as for the groupings of homotrimers and heterotrimers (Figure 3.3).

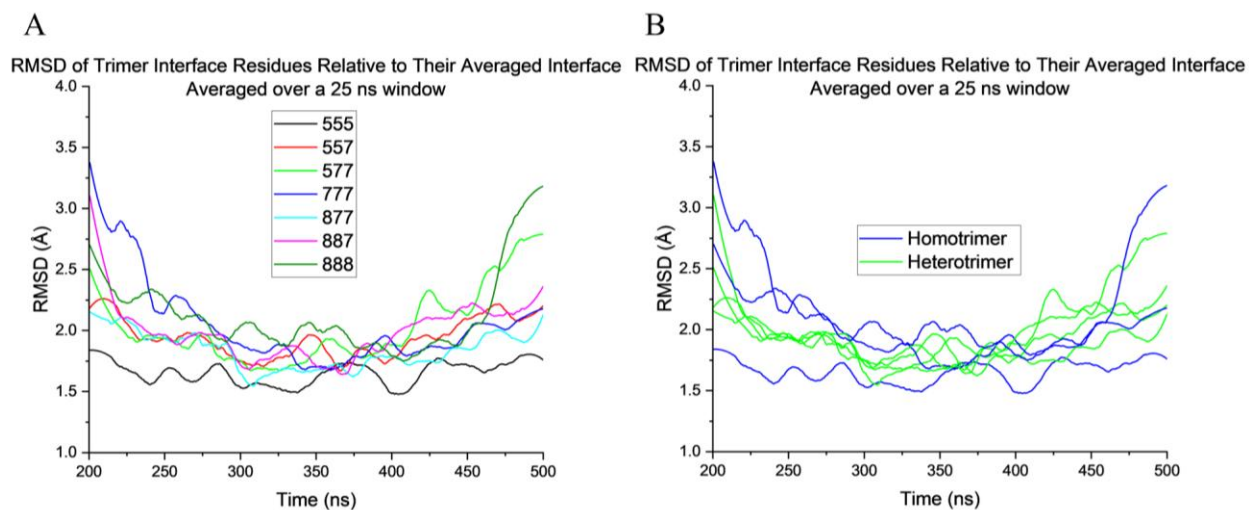


Figure 3.3: Post-convergence trimer interface RMSD values. (A) RMSD of the interfaces of all seven trimers, colored by individual trimer and smoothed over a 25 ns window. (B) RMSD of the interfaces of all seven trimers, colored based on trimer composition (homotrimers are blue and heterotrimers are light green) and smoothed over a 25 ns window. RMSD was calculated based on the average structure for each trimer over the last 300 ns of each simulation.

The potential energy values of the interface interactions were calculated using CPPTRAJ's LIE functionality, which calculates both nonbonded van der Waals and nonbonded Coulombic contributions. The last 300 ns of each simulation was used and the energy values averaged. This was done on a monomer-monomer interaction level as well as on a monomer-“rest of the trimer” interaction level to capture both individual interaction contributions as well as the ability of the rest of a trimer to stabilize each participating monomer. Notation-wise, 555_AB_55 is to be read as the interaction between the first (A) and second (B) monomer of the 555 trimer, with the 55 highlighting that this is an interaction between two PpCESA5 isoforms. Similarly, 887_BC_87 is the interaction of the second (B) and third (C) monomers of 887, which is an interaction between PpCESA isoforms 7 and 8. A, B, and C were

initially arbitrarily designated based on monomers in the initial PttCESA8 structure, though they are consistently assigned for each trimer. Additionally, 555_A_BC is the interaction between the first monomer (A) and both of its neighboring monomers (B and C) (calculated by summing their individual interaction contributions). These interface values were initially investigated based on their trimer of origin, colored by interaction energy contribution types (Figure 3.4). They were additionally investigated based on their total potential energy, colored by the nature of the monomer-monomer interaction (Figure 3.5).

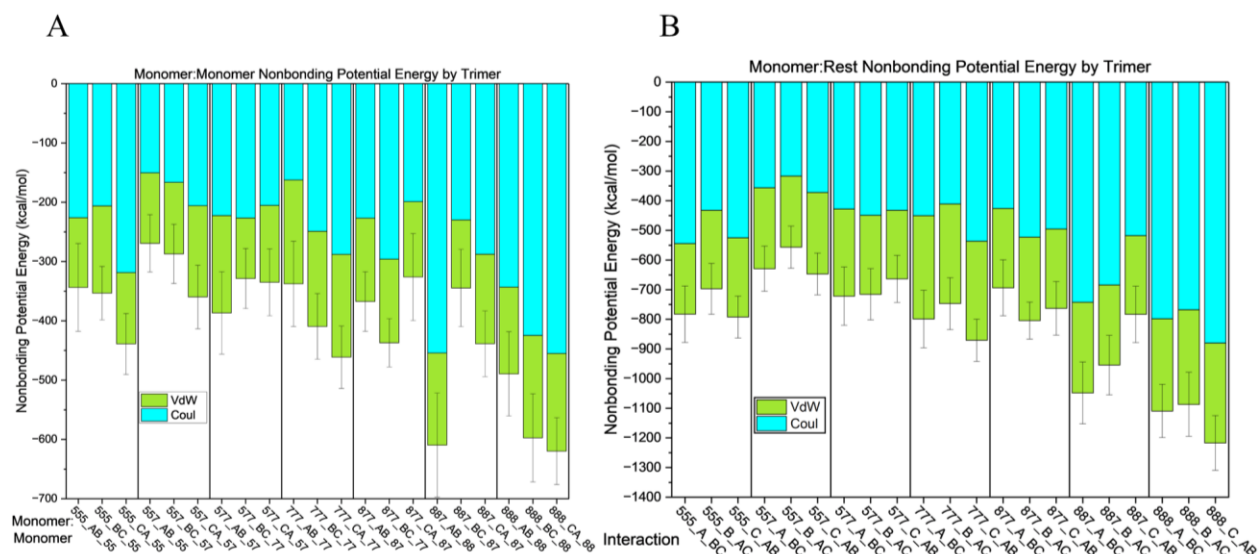


Figure 3.4: Trimer monomer-monomer and monomer-“rest of the trimer” nonbonding energy, arranged by trimer. (A) Nonbonding energy for each monomer-monomer interaction in each trimer. Energy was averaged over the last 300 ns. Van der Waals energy is light green and Coulombic energy is in cyan. Error bars are one standard deviation of the total. (B) Nonbonding energy for the interactions of each monomer and its two neighboring monomers. Color scheme and error bars are the same as (A), as is the energy average.

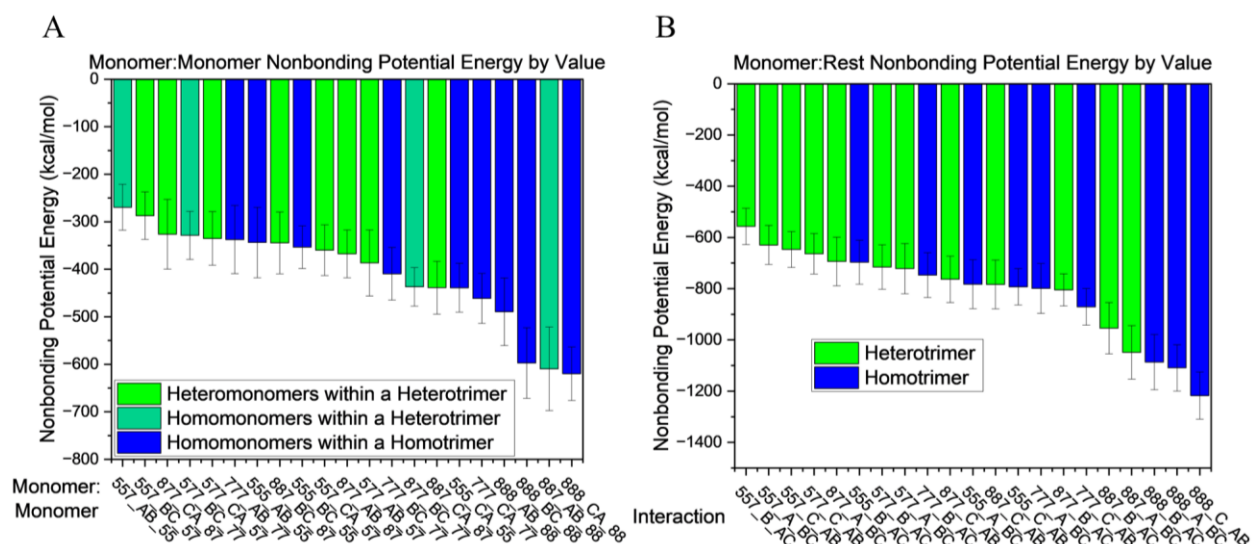


Figure 3.5: Trimer monomer-monomer and monomer-“rest of the trimer” nonbonding energy, arranged by decreasing energy. (A) Nonbonding energy for each monomer-monomer interaction in each trimer, arranged by decreasing energy (and thus increasing stability). Energy was averaged over the last 300 ns. A hetero-interaction in a heterotrimer is light green, a homo-interaction in a heterotrimer is aqua, and a homo-interaction in a homotrimer is blue. Error bars are one standard deviation of the total. (B) Nonbonding energy for the interactions of each monomer and its two neighboring monomers. Color scheme, energy averaging, and error bars are the same as (A), except there are no purely homo-interactions in the heterotrimers.

Then the total potential interface energy of each trimer was calculated by summing the individual monomer-monomer contributions, to get an overall stability order (highest energy to lowest energy) for the trimers (Figure 3.6). The increasing stability order is 557, 577, 877, 555, 777, 887, and 888. Notably, though this will be discussed in the Discussion section, each heterotrimer is less stable than the homotrimer that contains two of its monomers (557 is less stable than 555, 577 and 877 are less stable than 777, and 887 is less stable than 888).

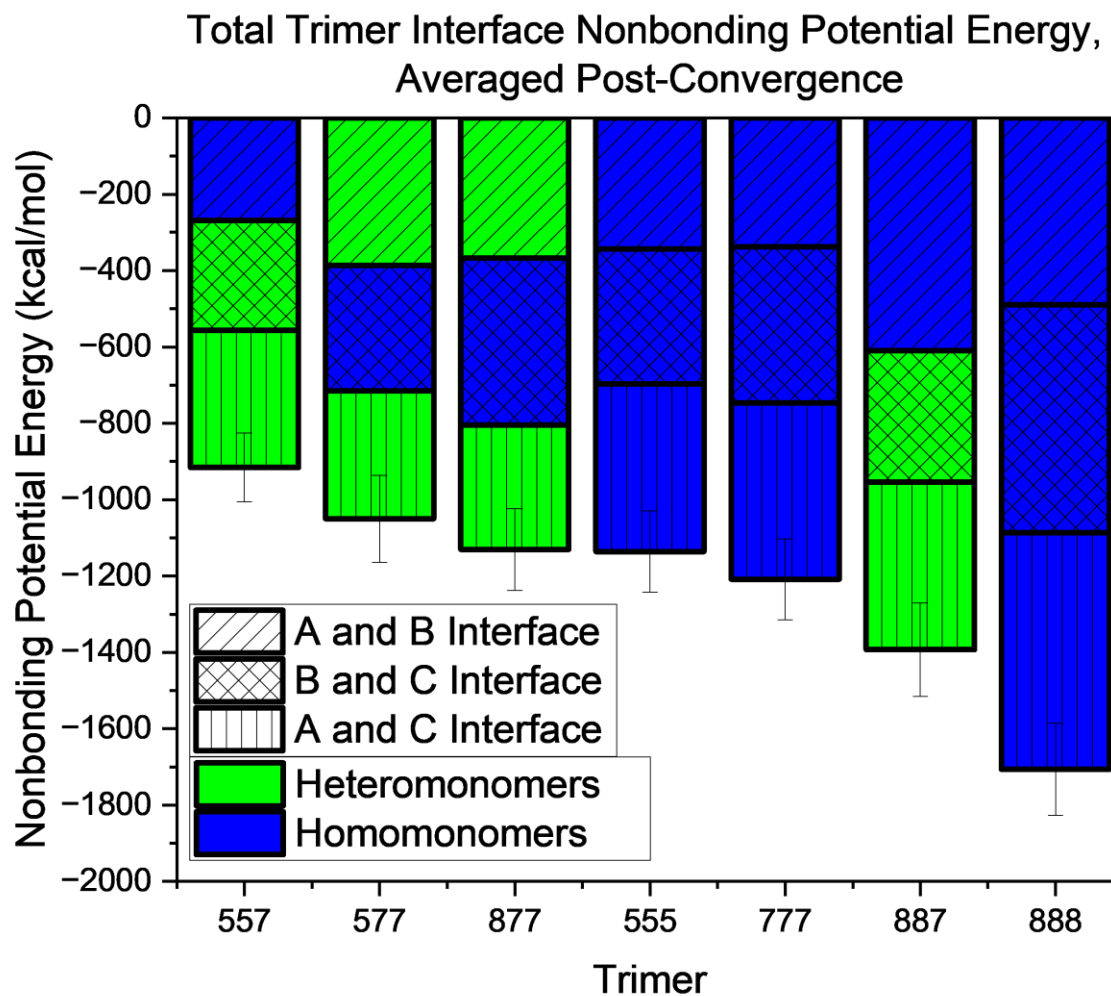


Figure 3.6: Trimer interfacial total nonbonding energy, arranged by decreasing energy.

Total nonbonding potential energy of each trimer's interfaces, averaged over the last 300 ns of the simulation, arranged by decreasing total energy (increasing stability). Each box within a column represents a different interface within the trimer (AB interfaces have diagonal lines, BC have crosshatched patterns, and AC have vertical lines). Additionally, monomer-monomer interactions are colored by monomer interaction type (homo=blue, hetero=green). The error bars are one standard deviation.

Discussion

The results presented indicate that homotrimeric CESA trimer assemblies are more stabilized relative to their heterotrimeric counterparts formed by replacing one monomer of the homotrimer with a different PpCESA isoform. This is evidenced in Figure 3.6, where the 555 homotrimer is more stable than the 557 trimer, the 777 trimer is more stable than either the 577 or 877 trimers, and the 888 homotrimer is more stable than the 887 trimer. This increased stability could indicate a biological preference for CESA homotrimer formation, a computationally-derived conclusion that agrees with the experimental findings of Ho et al., who found that their expressed cellulose-producing CSC rosettes were formed from homotrimers and who did not find tangible evidence of heterotrimer formation in their experiments.¹⁷

This investigation has not yet identified a probable cause for why this destabilization occurs. The overall secondary structure of the trimers does not show much deviation between trimer types, with the greatest range in trimer secondary structure types being less than 2%. Trimer secondary structure includes highly conserved α -helices in the transmembrane domain and PCR and highly conserved β -sheets that form a Rossman fold that comprises part of the catalytic region.⁹⁷ Additionally, the number of residues within 5 Å of the interface do not correlate to the homotrimeric or heterotrimeric nature of the trimer or the length of the monomers in each trimer, with 555 having the least number of residues near the interface and 888 having the most. Finally, the RMSD of the trimeric interfaces does not correspond to the nature of the trimer and seems broadly similar across all CESA trimer types, indicating that the gross movement of the interfaces over the course of the simulations is quite similar.

This could indicate that trimer stability may be mediated by specific inter-monomer contacts, rather than large-scale inter-monomer interactions. Future work will investigate this

possibility by looking into the types and the duration of the contacts at the trimeric interfaces to gain a clearer understanding of the stabilizing forces present.

Additionally, to further investigate the PpCESA functionality work of Li et al., future simulations will investigate different combinations of homotrimeric PpCESA CSCs.²⁵ To better understand CESA functionality at the CSC level, these combinations will involve arrangements with both known or suspected functionality (a CSC comprised only of PpCESA5 homotrimers as well as one comprised of alternating PpCESA 7 and 8 homotrimers) as well as those with known lack-of-function (a CSC comprised of only PpCESA7 homotrimers and one of only PpCESA8 homotrimers). In pursuing this line of inquiry, it is hoped that the resulting insights may help resolve long-standing questions regarding the effects of higher-order assembly on CESA function, ultimately informing future strategies to modulate cellulose synthase activity.

REFERENCES

- (1) Purushotham, P.; Ho, R.; Zimmer, J. Architecture of a Catalytically Active Homotrimeric Plant Cellulose Synthase Complex. *Science* **2020**, *369* (6507), 1089–1094. <https://doi.org/10.1126/science.abb2978>.
- (2) Pauly, M.; Keegstra, K. Plant Cell Wall Polymers as Precursors for Biofuels. *Curr. Opin. Plant Biol.* **2010**, *13* (3), 304–311. <https://doi.org/10.1016/j.pbi.2009.12.009>.
- (3) Bukhari, I.; Haq, F.; Kiran, M.; Aziz, T.; Mehmood, S.; Haroon, M. Lignocellulosic Biomass as a Renewable Resource: Driving Second-Generation Biofuel Innovation from Agricultural Waste. *Biomass Bioenergy* **2025**, *201*, 108133. <https://doi.org/10.1016/j.biombioe.2025.108133>.
- (4) Sulis, D. B.; Lavoine, N.; Sederoff, H.; Jiang, X.; Marques, B. M.; Lan, K.; Cofre-Vega, C.; Barrangou, R.; Wang, J. P. Advances in Lignocellulosic Feedstocks for Bioenergy and Bioproducts. *Nat. Commun.* **2025**, *16* (1), 1244. <https://doi.org/10.1038/s41467-025-56472-y>.
- (5) Du, J.; Vandavasi, V. G.; Molloy, K. R.; Yang, H.; Massenbunrg, L. N.; Singh, A.; Kwansa, A. L.; Yingling, Y. G.; O'Neill, H.; Chait, B. T.; Kumar, M.; Nixon, B. T. Evidence for Plant-Conserved Region Mediated Trimeric CESAs in Plant Cellulose Synthase Complexes. *Biomacromolecules* **2022**, *23* (9), 3663–3677. <https://doi.org/10.1021/acs.biomac.2c00550>.
- (6) Pedersen, G. B.; Blaschek, L.; Frandsen, K. E. H.; Noack, L. C.; Persson, S. Cellulose Synthesis in Land Plants. *Mol. Plant* **2023**, *16* (1), 206–231. <https://doi.org/10.1016/j.molp.2022.12.015>.
- (7) Carroll, A.; Specht, C. D. Understanding Plant Cellulose Synthases through a Comprehensive Investigation of the Cellulose Synthase Family Sequences. *Front. Plant Sci.* **2011**, *2*. <https://doi.org/10.3389/fpls.2011.00005>.
- (8) Herth, W. Arrays of Plasma-Membrane ?Rosettes? Involved in Cellulose Microfibril Formation of Spirogyra. *Planta* **1983**, *159* (4), 347–356. <https://doi.org/10.1007/BF00393174>.
- (9) Morgan, J. L. W.; Strumillo, J.; Zimmer, J. Crystallographic Snapshot of Cellulose Synthesis and Membrane Translocation. *Nature* **2013**, *493* (7431), 181–186. <https://doi.org/10.1038/nature11744>.
- (10) Turner, S.; Kumar, M. Cellulose Synthase Complex Organization and Cellulose Microfibril Structure. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2018**, *376* (2112), 20170048. <https://doi.org/10.1098/rsta.2017.0048>.
- (11) Daras, G.; Templalexis, D.; Avgeri, F.; Tsitsekan, D.; Karamanou, K.; Rigas, S. Updating Insights into the Catalytic Domain Properties of Plant Cellulose Synthase (CesA) and Cellulose Synthase-like (Csl) Proteins. *Molecules* **2021**, *26* (14), 4335. <https://doi.org/10.3390/molecules26144335>.
- (12) Roberts, E.; Roberts, A. W. A CELLULOSE SYNTHASE (*CESA*) GENE FROM THE RED ALGA *PORPHYRA YEZOENSIS* (RHODOPHYTA)¹. *J. Phycol.* **2009**, *45* (1), 203–212. <https://doi.org/10.1111/j.1529-8817.2008.00626.x>.
- (13) Slabaugh, E.; Scavuzzo-Duggan, T.; Chaves, A.; Wilson, L.; Wilson, C.; Davis, J. K.; Cosgrove, D. J.; Anderson, C. T.; Roberts, A. W.; Haigler, C. H. The Valine and Lysine Residues in the Conserved FxVTxK Motif Are Important for the Function of Phylogenetically Distant Plant Cellulose Synthases. *Glycobiology* **2016**, *26* (5), 509–519. <https://doi.org/10.1093/glycob/cwv118>.

- (14) Saxena, I. M.; Brown, R. M. Cellulose Biosynthesis: Current Views and Evolving Concepts. *Ann. Bot.* **2005**, *96* (1), 9–21. <https://doi.org/10.1093/aob/mci155>.
- (15) Sethaphong, L.; Haigler, C. H.; Kubicki, J. D.; Zimmer, J.; Bonetta, D.; DeBolt, S.; Yingling, Y. G. Tertiary Model of a Plant Cellulose Synthase. *Proc. Natl. Acad. Sci.* **2013**, *110* (18), 7512–7517. <https://doi.org/10.1073/pnas.1301027110>.
- (16) Singh, A.; Kwansa, A. L.; Kim, H. S.; Williams, J. T.; Yang, H.; Li, N. K.; Kubicki, J. D.; Roberts, A. W.; Haigler, C. H.; Yingling, Y. G. In Silico Structure Prediction of Full-Length Cotton Cellulose Synthase Protein (GhCESA1) and Its Hierarchical Complexes. *Cellulose* **2020**, *27* (10), 5597–5616. <https://doi.org/10.1007/s10570-020-03194-7>.
- (17) Ho, R.; Purushotham, P.; Wilson, L. F.; Wan, Y.; Zimmer, J. Structure, Function and Assembly of Soybean Primary Cell Wall Cellulose Synthases. *eLife* **2025**, *13*, RP96704. <https://doi.org/10.7554/eLife.96704>.
- (18) Wilson, T. H.; Kumar, M.; Turner, S. R. The Molecular Basis of Plant Cellulose Synthase Complex Organisation and Assembly. *Biochem. Soc. Trans.* **2021**, *49* (1), 379–391. <https://doi.org/10.1042/BST20200697>.
- (19) Dunker, A. K.; Babu, M. M.; Barbar, E.; Blackledge, M.; Bondos, S. E.; Dosztányi, Z.; Dyson, H. J.; Forman-Kay, J.; Fuxreiter, M.; Gsponer, J.; Han, K.-H.; Jones, D. T.; Longhi, S.; Metallo, S. J.; Nishikawa, K.; Nussinov, R.; Obradovic, Z.; Pappu, R. V.; Rost, B.; Selenko, P.; Subramaniam, V.; Sussman, J. L.; Tompa, P.; Uversky, V. N. What's in a Name? Why These Proteins Are Intrinsically Disordered: Why These Proteins Are Intrinsically Disordered. *Intrinsically Disord. Proteins* **2013**, *1* (1), e24157. <https://doi.org/10.4161/idp.24157>.
- (20) Kwansa, A. L.; Singh, A.; Williams, J. T.; Haigler, C. H.; Roberts, A. W.; Yingling, Y. G. Structural Determination of a Full-Length Plant Cellulose Synthase Informed by Experimental and in Silico Methods. *Cellulose* **2024**, *31* (3), 1429–1447. <https://doi.org/10.1007/s10570-023-05691-x>.
- (21) Zhang, X.; Xue, Y.; Guan, Z.; Zhou, C.; Nie, Y.; Men, S.; Wang, Q.; Shen, C.; Zhang, D.; Jin, S.; Tu, L.; Yin, P.; Zhang, X. Structural Insights into Homotrimeric Assembly of Cellulose Synthase Cesa7 from *Gossypium Hirsutum*. *Plant Biotechnol. J.* **2021**, *19* (8), 1579–1587. <https://doi.org/10.1111/pbi.13571>.
- (22) Qiao, Z.; Lampugnani, E. R.; Yan, X.-F.; Khan, G. A.; Saw, W. G.; Hannah, P.; Qian, F.; Calabria, J.; Miao, Y.; Grüber, G.; Persson, S.; Gao, Y.-G. Structure of *Arabidopsis* CESA3 Catalytic Domain with Its Substrate UDP-Glucose Provides Insight into the Mechanism of Cellulose Synthesis. *Proc. Natl. Acad. Sci.* **2021**, *118* (11), e2024015118. <https://doi.org/10.1073/pnas.2024015118>.
- (23) Nixon, B. T.; Mansouri, K.; Singh, A.; Du, J.; Davis, J. K.; Lee, J.-G.; Slabaugh, E.; Vandavasi, V. G.; O'Neill, H.; Roberts, E. M.; Roberts, A. W.; Yingling, Y. G.; Haigler, C. H. Comparative Structural and Computational Analysis Supports Eighteen Cellulose Synthases in the Plant Cellulose Synthesis Complex. *Sci. Rep.* **2016**, *6* (1), 28696. <https://doi.org/10.1038/srep28696>.
- (24) Kubicki, J. D.; Yang, H.; Sawada, D.; O'Neill, H.; Oehme, D.; Cosgrove, D. The Shape of Native Plant Cellulose Microfibrils. *Sci. Rep.* **2018**, *8* (1), 13983. <https://doi.org/10.1038/s41598-018-32211-w>.
- (25) Li, X.; Chaves, A. M.; Dees, D. C. T.; Mansoori, N.; Yuan, K.; Speicher, T. L.; Norris, J. H.; Wallace, I. S.; Trindade, L. M.; Roberts, A. W. Cellulose Synthesis Complexes Are

- Homo-Oligomeric and Hetero-Oligomeric in *Physcomitrium Patens*. *Plant Physiol.* **2022**, *188* (4), 2115–2130. <https://doi.org/10.1093/plphys/kiac003>.
- (26) Pauling, L.; Corey, R. B.; Branson, H. R. The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *Proc. Natl. Acad. Sci.* **1951**, *37* (4), 205–211. <https://doi.org/10.1073/pnas.37.4.205>.
- (27) Pauling, L.; Corey, R. B. The Pleated Sheet, A New Layer Configuration of Polypeptide Chains. *Proc. Natl. Acad. Sci.* **1951**, *37* (5), 251–256. <https://doi.org/10.1073/pnas.37.5.251>.
- (28) Browne, W. J.; North, A. C. T.; Phillips, D. C.; Brew, K.; Vanaman, T. C.; Hill, R. L. A Possible Three-Dimensional Structure of Bovine α -Lactalbumin Based on That of Hen's Egg-White Lysozyme. *J. Mol. Biol.* **1969**, *42* (1), 65–86. [https://doi.org/10.1016/0022-2836\(69\)90487-2](https://doi.org/10.1016/0022-2836(69)90487-2).
- (29) Meng, Y.; Zhang, Z.; Zhou, C.; Tang, X.; Hu, X.; Tian, G.; Yang, J.; Yao, Y. Protein Structure Prediction via Deep Learning: An in-Depth Review. *Front. Pharmacol.* **2025**, *16*, 1498662. <https://doi.org/10.3389/fphar.2025.1498662>.
- (30) Hameduh, T.; Haddad, Y.; Adam, V.; Heger, Z. Homology Modeling in the Time of Collective and Artificial Intelligence. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 3494–3506. <https://doi.org/10.1016/j.csbj.2020.11.007>.
- (31) Bertoline, L. M. F.; Lima, A. N.; Krieger, J. E.; Teixeira, S. K. Before and after AlphaFold2: An Overview of Protein Structure Prediction. *Front. Bioinforma.* **2023**, *3*, 1120370. <https://doi.org/10.3389/fbinf.2023.1120370>.
- (32) Maia, R. T. Protein Structure Prediction by Computational Homology Modeling: A Brief Explanation. *Int. J. Mol. Biol. Open Access* **2024**, *7* (1), 118–120. <https://doi.org/10.15406/ijmboa.2024.07.00180>.
- (33) Bhattacharya, S.; Roche, R.; Shuvo, M. H.; Moussad, B.; Bhattacharya, D. Contact-Assisted Threading in Low-Homology Protein Modeling. In *Homology Modeling*; Filipek, S., Ed.; Methods in Molecular Biology; Springer US: New York, NY, 2023; Vol. 2627, pp 41–59. https://doi.org/10.1007/978-1-0716-2974-1_3.
- (34) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
- (35) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millán, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; Van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* **2021**, *373* (6557), 871–876. <https://doi.org/10.1126/science.abj8754>.
- (36) Balasubramaniam, D.; Komives, E. A. Hydrogen-Exchange Mass Spectrometry for the Study of Intrinsic Disorder in Proteins. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2013**, *1834* (6), 1202–1209. <https://doi.org/10.1016/j.bbapap.2012.10.009>.

- (37) Dyson, H. J.; Wright, P. E. NMR Illuminates Intrinsic Disorder. *Curr. Opin. Struct. Biol.* **2021**, *70*, 44–52. <https://doi.org/10.1016/j.sbi.2021.03.015>.
- (38) Camacho-Zarco, A. R.; Schnapka, V.; Guseva, S.; Abyzov, A.; Adamski, W.; Milles, S.; Jensen, M. R.; Zidek, L.; Salvi, N.; Blackledge, M. NMR Provides Unique Insight into the Functional Dynamics and Interactions of Intrinsically Disordered Proteins. *Chem. Rev.* **2022**, *122* (10), 9331–9356. <https://doi.org/10.1021/acs.chemrev.1c01023>.
- (39) Schiavina, M.; Salladini, E.; Murralli, M. G.; Tria, G.; Felli, I. C.; Pierattelli, R.; Longhi, S. Ensemble Description of the Intrinsically Disordered N-Terminal Domain of the Nipah Virus P/V Protein from Combined NMR and SAXS. *Sci. Rep.* **2020**, *10* (1), 19574. <https://doi.org/10.1038/s41598-020-76522-3>.
- (40) Janson, G.; Feig, M. Transferable Deep Generative Modeling of Intrinsically Disordered Protein Conformations. *PLOS Comput. Biol.* **2024**, *20* (5), e1012144. <https://doi.org/10.1371/journal.pcbi.1012144>.
- (41) Yang, J.; Zhang, Y. Protein Structure and Function Prediction Using I-TASSER. *Curr. Protoc. Bioinforma.* **2015**, *52* (1). <https://doi.org/10.1002/0471250953.bi0508s52>.
- (42) Zheng, W.; Wuyun, Q.; Li, Y.; Liu, Q.; Zhou, X.; Peng, C.; Zhu, Y.; Freddolino, L.; Zhang, Y. Deep-Learning-Based Single-Domain and Multidomain Protein Structure Prediction with D-I-TASSER. *Nat. Biotechnol.* **2025**. <https://doi.org/10.1038/s41587-025-02654-4>.
- (43) Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A Unified Platform for Automated Protein Structure and Function Prediction. *Nat. Protoc.* **2010**, *5* (4), 725–738. <https://doi.org/10.1038/nprot.2010.5>.
- (44) Fuxreiter, M.; Tompa, P. Fuzzy Complexes: A More Stochastic View of Protein Function. In *Fuzziness*; Fuxreiter, M., Tompa, P., Eds.; Advances in Experimental Medicine and Biology; Springer US: New York, NY, 2012; Vol. 725, pp 1–14. https://doi.org/10.1007/978-1-4614-0659-4_1.
- (45) Coppa, C.; Bazzoli, A.; Barkhordari, M.; Contini, A. Accelerated Molecular Dynamics for Peptide Folding: Benchmarking Different Combinations of Force Fields and Explicit Solvent Models. *J. Chem. Inf. Model.* **2023**, *63* (10), 3030–3042. <https://doi.org/10.1021/acs.jcim.3c00138>.
- (46) Uto, T.; Ikeda, Y.; Sunagawa, N.; Tajima, K.; Yao, M.; Yui, T. Molecular Dynamics Simulation of Cellulose Synthase Subunit D Octamer with Cellulose Chains from Acetic Acid Bacteria: Insight into Dynamic Behaviors and Thermodynamics on Substrate Recognition. *J. Chem. Theory Comput.* **2021**, *17* (1), 488–496. <https://doi.org/10.1021/acs.jctc.0c01027>.
- (47) Kuang, G.; Bulone, V.; Tu, Y. Computational Studies of the Binding Profile of Phosphoinositide PtdIns (3,4,5) P3 with the Pleckstrin Homology Domain of an Oomycete Cellulose Synthase. *Sci. Rep.* **2016**, *6* (1), 20555. <https://doi.org/10.1038/srep20555>.
- (48) Kwansa, A. L.; Chaves, A. M.; Del Mundo, J. T.; Pierce, E. T.; Gomez, E. W.; Gomez, E. D.; Haigler, C. H.; Yingling, Y. G.; Roberts, A. W. Empirical Evidence That Glucan-Interacting Amino Acid Side Chains within the Transmembrane Channel Collectively Facilitate Cellulose Synthase Function. *Plant Mol. Biol.* **2025**, *115* (4), 85. <https://doi.org/10.1007/s11103-025-01615-4>.
- (49) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL:

- Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296–W303. <https://doi.org/10.1093/nar/gky427>.
- (50) Zhou, X.; Zheng, W.; Li, Y.; Pearce, R.; Zhang, C.; Bell, E. W.; Zhang, G.; Zhang, Y. I-TASSER-MTD: A Deep-Learning-Based Platform for Multi-Domain Protein Structure and Function Prediction. *Nat. Protoc.* **2022**, *17* (10), 2326–2353. <https://doi.org/10.1038/s41596-022-00728-0>.
- (51) Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nat. Protoc.* **2015**, *10* (6), 845–858. <https://doi.org/10.1038/nprot.2015.053>.
- (52) David, A.; Islam, S.; Tankhilevich, E.; Sternberg, M. J. E. The AlphaFold Database of Protein Structures: A Biologist’s Guide. *J. Mol. Biol.* **2022**, *434* (2), 167336. <https://doi.org/10.1016/j.jmb.2021.167336>.
- (53) Ruff, K. M.; Pappu, R. V. AlphaFold and Implications for Intrinsically Disordered Proteins. *J. Mol. Biol.* **2021**, *433* (20), 167208. <https://doi.org/10.1016/j.jmb.2021.167208>.
- (54) Piovesan, D.; Monzon, A. M.; Tosatto, S. C. E. Intrinsic Protein Disorder and Conditional Folding in ALPHAFOLDDB. *Protein Sci.* **2022**, *31* (11), e4466. <https://doi.org/10.1002/pro.4466>.
- (55) Schreiner, E.; Trabuco, L. G.; Freddolino, P. L.; Schulten, K. Stereochemical Errors and Their Implications for Molecular Dynamics Simulations. *BMC Bioinformatics* **2011**, *12* (1), 190. <https://doi.org/10.1186/1471-2105-12-190>.
- (56) Colovos, C.; Yeates, T. O. Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions. *Protein Sci.* **1993**, *2* (9), 1511–1519. <https://doi.org/10.1002/pro.5560020916>.
- (57) Wiederstein, M.; Sippl, M. J. ProSA-Web: Interactive Web Service for the Recognition of Errors in Three-Dimensional Structures of Proteins. *Nucleic Acids Res.* **2007**, *35* (Web Server), W407–W410. <https://doi.org/10.1093/nar/gkm290>.
- (58) Studer, G.; Rempfer, C.; Waterhouse, A. M.; Gumienny, R.; Haas, J.; Schwede, T. QMEANDisCo—Distance Constraints Applied on Model Quality Estimation. *Bioinformatics* **2020**, *36* (6), 1765–1771. <https://doi.org/10.1093/bioinformatics/btz828>.
- (59) Lay, D. C.; Lay, S. R.; McDonald, J. *Linear Algebra and Its Applications*, Fifth edition.; Pearson: Boston, 2016.
- (60) Sippl, M. J. Calculation of Conformational Ensembles from Potentials of Mean Force. *J. Mol. Biol.* **1990**, *213* (4), 859–883. [https://doi.org/10.1016/S0022-2836\(05\)80269-4](https://doi.org/10.1016/S0022-2836(05)80269-4).
- (61) Sippl, M. J. Recognition of Errors in Three-dimensional Structures of Proteins. *Proteins Struct. Funct. Bioinforma.* **1993**, *17* (4), 355–362. <https://doi.org/10.1002/prot.340170404>.
- (62) Sippl, M. J. Knowledge-Based Potentials for Proteins. *Curr. Opin. Struct. Biol.* **1995**, *5* (2), 229–235. [https://doi.org/10.1016/0959-440X\(95\)80081-6](https://doi.org/10.1016/0959-440X(95)80081-6).
- (63) Bernhofer, M.; Rost, B. TMbed: Transmembrane Proteins Predicted through Language Model Embeddings. *BMC Bioinformatics* **2022**, *23* (1), 326. <https://doi.org/10.1186/s12859-022-04873-x>.
- (64) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. UCSF CHIMERAX : Structure Visualization for Researchers, Educators, and Developers. *Protein Sci.* **2021**, *30* (1), 70–82. <https://doi.org/10.1002/pro.3943>.
- (65) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).

- (66) Van Der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; Kim, P. M.; Kriwacki, R. W.; Oldfield, C. J.; Pappu, R. V.; Tompa, P.; Uversky, V. N.; Wright, P. E.; Babu, M. M. Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **2014**, *114* (13), 6589–6631. <https://doi.org/10.1021/cr400525m>.
- (67) Marsh, J. A.; Forman-Kay, J. D. Structure and Disorder in an Unfolded State under Nondenaturing Conditions from Ensemble Models Consistent with a Large Number of Experimental Restraints. *J. Mol. Biol.* **2009**, *391* (2), 359–374. <https://doi.org/10.1016/j.jmb.2009.06.001>.
- (68) Alder, B. J.; Wainwright, T. E. Phase Transition for a Hard Sphere System. *J. Chem. Phys.* **1957**, *27* (5), 1208–1209. <https://doi.org/10.1063/1.1743957>.
- (69) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267* (5612), 585–590. <https://doi.org/10.1038/267585a0>.
- (70) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99* (6), 1129–1143. <https://doi.org/10.1016/j.neuron.2018.08.011>.
- (71) Simmerling, C.; Strockbine, B.; Roitberg, A. E. All-Atom Structure Prediction and Folding Simulations of a Stable Protein. *J. Am. Chem. Soc.* **2002**, *124* (38), 11258–11259. <https://doi.org/10.1021/ja0273851>.
- (72) Nwanochie, E.; Uversky, V. N. Structure Determination by Single-Particle Cryo-Electron Microscopy: Only the Sky (and Intrinsic Disorder) Is the Limit. *Int. J. Mol. Sci.* **2019**, *20* (17), 4186. <https://doi.org/10.3390/ijms20174186>.
- (73) Verma, P.; Kwansa, A. L.; Ho, R.; Yingling, Y. G.; Zimmer, J. Insights into Substrate Coordination and Glycosyl Transfer of Poplar Cellulose Synthase-8. *Structure* **2023**, *31* (10), 1166–1173.e6. <https://doi.org/10.1016/j.str.2023.07.010>.
- (74) Tuckerman, M. E.; Martyna, G. J. Understanding Modern Molecular Dynamics: Techniques and Applications. *J. Phys. Chem. B* **2000**, *104* (2), 159–178. <https://doi.org/10.1021/jp992433y>.
- (75) D.A. Case; K. Belfon; I.Y. Ben-Shalom; S.R. Brozell; D.S. Cerutti; T.E. Cheatham, III; V.W.D. Cruzeiro; T.A. Darden; R.E. Duke; G. Giambasu; M.K. Gilson; H. Gohlke; A.W. Goetz; R. Harris; S. Izadi; S.A. Izmailov; K. Kasavajhala; A. Kovalenko; R. Krasny; T. Kurtzman; T.S. Lee; S. LeGrand; P. Li; C. Lin; J. Liu; T. Luchko; R. Luo; V. Man; K.M. Merz; Y. Miao; O. Mikhailovskii; G. Monard; H. Nguyen; A. Onufriev; F. Pan; S. Pantano; R. Qi; D.R. Roe; A. Roitberg; C. Sagui; S. Schott-Verdugo; J. Shen; C.L. Simmerling; N.R. Skrynnikov; J. Smith; J. Swails; R.C. Walker; J. Wang; L. Wilson; R.M. Wolf; X. Wu; Y. Xiong; Y. Xue; D.M. York; P.A. Kollman. AMBER 2020, 2020.
- (76) Resemann, H. C.; Lewandowska, M.; G̃ilzmann, J.; Feussner, I. Membrane Lipids, Waxes and Oxylipins in the Moss Model Organism *Physcomitrella Patens*. *Plant Cell Physiol.* **2019**, *60* (6), 1166–1175. <https://doi.org/10.1093/pcp/pcz006>.
- (77) Bransburg-Zabary, S.; Kessel, A.; Gutman, M.; Ben-Tal, N. Stability of an Ion Channel in Lipid Bilayers: Implicit Solvent Model Calculations with Gramicidin. *Biochemistry* **2002**, *41* (22), 6946–6954. <https://doi.org/10.1021/bi0120704>.
- (78) Mérida-Quesada, F.; Vergara-Valladares, F.; Rubio-Meléndez, M. E.; Hernández-Rojas, N.; González-González, A.; Michard, E.; Navarro-Retamal, C.; Dreyer, I. TPC1-Type Channels in *Physcomitrium Patens*: Interaction between EF-Hands and Ca²⁺. *Plants* **2022**, *11* (24), 3527. <https://doi.org/10.3390/plants11243527>.

- (79) Leigh, R. A.; Wyn Jones, R. G. A HYPOTHESIS RELATING CRITICAL POTASSIUM CONCENTRATIONS FOR GROWTH TO THE DISTRIBUTION AND FUNCTIONS OF THIS ION IN THE PLANT CELL. *New Phytol.* **1984**, *97* (1), 1–13. <https://doi.org/10.1111/j.1469-8137.1984.tb04103.x>.
- (80) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Miguez, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16* (1), 528–552. <https://doi.org/10.1021/acs.jctc.9b00591>.
- (81) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States. *Proc. Natl. Acad. Sci.* **2018**, *115* (21). <https://doi.org/10.1073/pnas.1800690115>.
- (82) Shabane, P. S.; Izadi, S.; Onufriev, A. V. General Purpose Water Model Can Improve Atomistic Simulations of Intrinsically Disordered Proteins. *J. Chem. Theory Comput.* **2019**, *15* (4), 2620–2634. <https://doi.org/10.1021/acs.jctc.8b01123>.
- (83) Love, O.; Pacheco Lima, M. C.; Clark, C.; Cornillie, S.; Roalstad, S.; Cheatham Iii, T. E. Evaluating the Accuracy of the AMBER Protein Force Fields in Modeling Dihydrofolate Reductase Structures: Misbalance in the Conformational Arrangements of the Flexible Loop Domains. *J. Biomol. Struct. Dyn.* **2023**, *41* (13), 5946–5960. <https://doi.org/10.1080/07391102.2022.2098823>.
- (84) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>.
- (85) Gould, I.R., Skjevik A.A., Dickson, C.J., Madej, B.D., Walker, R.C. Lipid17: A Comprehensive AMBER Force Field for the Simulation of > Zwitterionic and Anionic Lipids, 2018.
- (86) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935. <https://doi.org/10.1063/1.445869>.
- (87) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112* (30), 9020–9041. <https://doi.org/10.1021/jp8001614>.
- (88) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **2008**, *29* (4), 622–655. <https://doi.org/10.1002/jcc.20820>.
- (89) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. P ACKMOL : A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.* **2009**, *30* (13), 2157–2164. <https://doi.org/10.1002/jcc.21224>.
- (90) Schott-Verdugo, S.; Gohlke, H. PACKMOL-Memgen: A Simple-To-Use, Generalized Workflow for Membrane-Protein-Lipid-Bilayer System Building. *J. Chem. Inf. Model.* **2019**, *59* (6), 2522–2528. <https://doi.org/10.1021/acs.jcim.9b00269>.
- (91) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184* (2), 374–380. <https://doi.org/10.1016/j.cpc.2012.09.022>.

- (92) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9* (9), 3878–3888. <https://doi.org/10.1021/ct400314y>.
- (93) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092. <https://doi.org/10.1063/1.464397>.
- (94) Schneider, T.; Stoll, E. Molecular-Dynamics Study of a Three-Dimensional One-Component Model for Distortive Phase Transitions. *Phys. Rev. B* **1978**, *17* (3), 1302–1322. <https://doi.org/10.1103/PhysRevB.17.1302>.
- (95) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690. <https://doi.org/10.1063/1.448118>.
- (96) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341. [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5).
- (97) Rao, S. T.; Rossmann, M. G. Comparison of Super-Secondary Structures in Proteins. *J. Mol. Biol.* **1973**, *76* (2), 241–256. [https://doi.org/10.1016/0022-2836\(73\)90388-4](https://doi.org/10.1016/0022-2836(73)90388-4).

APPENDIX

Table S1: Secondary structure composition percentages for all PpCESA trimers. Values were determined using CPPTRAJ, which uses the DSSP algorithm.

	Pi	3-10	Beta	Turn	Bend	Coil	Alpha
555	0.13	3.44	6.95	14.88	16.02	23.97	34.62
557	0.12	4.24	6.56	14.90	15.51	25.06	33.61
577	0.05	3.83	6.72	14.03	16.07	25.01	34.28
777	0.02	3.02	6.27	15.40	15.75	25.70	33.84
877	0.02	3.95	6.62	15.32	15.01	24.78	34.30
887	0.02	4.09	6.55	14.86	14.85	24.74	34.90
888	0.03	4.54	7.08	15.13	14.32	24.43	34.47