2014

Structural Mechanisms of Glucan Phosphatase Activity in Starch Metabolism

David A. Meekins
University of Kentucky, dame222@uky.edu

Click here to let us know how access to this document benefits you.

Recommended Citation
https://uknowledge.uky.edu/biochem_etds/14

This Doctoral Dissertation is brought to you for free and open access by the Molecular and Cellular Biochemistry at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Molecular and Cellular Biochemistry by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student’s advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student’s thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

David A. Meekins, Student
Dr. Matthew S. Gentry, Major Professor
Dr. Michael D. Mendenhall, Director of Graduate Studies
STRUCTURAL MECHANISMS OF GLUCAN PHOSPHATASE ACTIVITY IN STARCH METABOLISM

A dissertation submitted in partial fulfillment of the Requirements for the degree of Doctor of Philosophy in the College of Medicine at the University of Kentucky

By

David Adam Meekins

Lexington, Kentucky

Co-Directors: Dr. Matthew S. Gentry, Professor of Mol. and Cell. Biochemistry and Dr. David W. Rodgers, Professor of Mol. and Cell. Biochemistry

Lexington, Kentucky

2014

Copyright © David Adam Meekins 2014
ABSTRACT OF DISSERTATION

STRUCTURAL MECHANISMS OF GLUCAN PHOSPHATASE ACTIVITY IN STARCH METABOLISM

Starch is a water-insoluble glucose biopolymer used as an energy cache in plants and is synthesized and degraded in a diurnal cycle. Reversible phosphorylation of starch granules regulates the solubility and, consequentially, the bioavailability of starch glucans to degradative enzymes. Glucan phosphatases release phosphate from starch glucans and their activity is essential to the proper diurnal metabolism of starch. Previously, the structural basis of glucan phosphatase activity was entirely unknown. The work in this dissertation outlines the structural mechanism of activity of two plant glucan phosphatases called Starch EXcess4 (SEX4) and Like Sex Four2 (LSF2). The crystal structures of SEX4 and LSF2 were determined with and without phospho-glucan ligands bound, revealing the basis of their interaction with an endogenous substrate. The data show that SEX4 and LSF2 interact with starch glucans via distinctive mechanisms. SEX4 binds glucan chains via an aromatic-rich pocket spanning its Carbohydrate Binding Module (CBM) and catalytic Dual Specificity Phosphatase (DSP) domains. Conversely, LSF2 lacks a CBM and, instead, binds glucans at two non-catalytic surface-binding sites that are located distally from its active site. In addition, it was previously reported that SEX4 and LSF2 act upon distinct phospho-glucan substrates: SEX4 preferentially dephosphorylates the C6-position of starch glucans and LSF2 exclusively dephosphorylates the C3-position. The data herein reveal that SEX4 and LSF2 contain differences in their active site topology that serve to position the glucan chain in opposite orientations, therefore accounting for the differences in substrate specificity. Using these insights, SEX4 was engineered with reversed substrate specificity, i.e. preferential C3-specific activity. Previous work has established the interaction between phosphatases and protein, lipid, and nucleic acids; however, the current study represents the first insights into phosphatase interaction with carbohydrate substrates. In addition, the insights gained provide a model that will be used in future studies with the mammalian glucan phosphatase laforin, which is linked to neurodegeneration and the fatal epileptic disorder Lafora’s Disease.
STRUCTURAL MECHANISMS OF GLUCAN PHOSPHATASE ACTIVITY IN STARCH METABOLISM

By

David Adam Meekins

Dr. Matthew S Gentry
Co-Director of Dissertation

Dr. David W. Rodgers
Co-Director of Dissertation

Dr. Michael Mendenhall
Director of Graduate Studies

6/9/14
TABLE OF CONTENTS

List of Tables ........................................................................................................ v

List of Figures ........................................................................................................ vi

Chapter 1. Introduction ......................................................................................... 1
Preface ..................................................................................................................... 1
Starch Structure and Synthesis ................................................................. 2
Starch Breakdown and Phosphorylation ................................................ 4
Glucan Phosphatase SEX4 ........................................................................ 7
Structure of SEX4 .............................................................................................. 10
Glucan Phosphatase LSF2 ...................................................................... 12
Summary and Outline of Research ......................................................... 14

Chapter 2. Materials & Methods ........................................................................ 28
Cloning of Glucan Phosphatase Constructs ......................................... 28
Expression and Purification of Recombinant Proteins ....................... 29
Crystal Structure Determination and Refinement ............................... 30
Phosphatase Assays ...................................................................................... 32
Glucan Binding Assays .............................................................................. 33

Chapter 3. Structural Basis of SEX4-Glucan Interaction ............................... 39
Introduction ....................................................................................................... 39
Results .............................................................................................................. 41
  Structure of SEX4 Bound to Maltoheptaose and Phosphate .......... 41
  Maltoheptaose at the SEX4 CBM .............................................................. 42
  Maltoheptaose and Phosphate at the SEX4 DSP ............................ 43
  SEX4ΔCBM and Activity Against Soluble Phosphoglucans .... 44
Discussion ..................................................................................................... 46

Chapter 4. Structural Basis of LSF2-Glucan Interaction ............................... 61
Introduction ....................................................................................................... 61
Results .............................................................................................................. 63
  Structure of LSF2 Bound to Maltohexaose and Phosphate .......... 63
  Maltohexaose and Phosphate at the LSF2 Active Site ................ 64
  Conformational Changes in the LSF2 Active Site ....................... 65
  Non-Catalytic Glucan Binding Sites ..................................................... 67
Discussion ..................................................................................................... 70
Chapter 5. Glucan Phosphatase Substrate Specificity ........................................ 91
  Introduction ......................................................................................................... 91
  Results .................................................................................................................. 93
    Structural Evidence for Substrate Specificity in SEX4 and LSF2 .................. 93
    C3-specific Bonding Interactions at the LSF2 Active Site ................................. 94
    DSP Domain Directs SEX4 Specificity .............................................................. 95
    Reversal of SEX4 Substrate Specificity .............................................................. 96
    Laforin does not Contain Substrate Specificity ............................................... 98
  Discussion .......................................................................................................... 99

Chapter 6. Comparative Phosphatase Analysis .................................................. 113
  Introduction ......................................................................................................... 113
  Structural Comparison of SEX4 and LSF2 ........................................................ 114
  Comparison with Laforin and LSF1 .................................................................. 119
  Structural Comparison Among Glucan Phosphatase and Other DSPs ............ 124

Chapter 7. Summary and Future Directions .................................................... 154
  Summary ............................................................................................................. 154
  Future Directions .............................................................................................. 156

References ............................................................................................................ 165

Vita ....................................................................................................................... 183
LIST OF TABLES

Table 1.1 – Data collection and refinement statistics for Δ89-SEX4 C198S........16
Table 3.1 – Data collection and refinement statistics for glucan-bound Δ89-SEX4 C198S..................................................................................................................49
Table 4.1 – Data collection and refinement statistics for glucan-bound and citrate-bound Δ78-LSF2, C193S and WT .................................................................74
Table 6.1 – Sequence identity / similarity in glucan phosphatase family........129
LIST OF FIGURES

Figure 1.1 – Composition and structure of starch ........................................ 17
Figure 1.2 – Diurnal metabolism of transitory starch in *Arabidopsis* ........... 18
Figure 1.3 – Reversible phosphorylation in starch degradation .................. 19
Figure 1.4 – Substrate specificity in glucan dkinases and glucan phosphatases 20
Figure 1.5 – Starch excess phenotype in SEX4-deficient *Arabidopsis* plants .... 21
Figure 1.6 – Glucan phosphatase domain outlines ..................................... 22
Figure 1.7 – Phosphatase activity of plant glucan phosphatases ................. 23
Figure 1.8 – Non-glucan bound structure of SEX4 .................................. 24
Figure 1.9 – SEX4 DSP catalytic triad ...................................................... 25
Figure 1.10 – CT motif in SEX4 ................................................................ 26
Figure 1.11 – *In vivo* results of LSF2 and SEX4 silencing in *Arabidopsis* ...... 27
Figure 2.1 – Generic pNPP phosphatase assay ........................................... 36
Figure 2.2 – Malachite green assay for detecting glucan phosphatase activity... 37
Figure 2.3 – Radiolabeled starch assay for determining C6- vs. C3-position dephosphorylation .......................................................... 38
Figure 3.1 – Crystal structure of At-SEX4 bound to a glucan and phosphate .... 50
Figure 3.2 – Interaction of maltoheptaose with the SEX4 CBM ................... 51
Figure 3.3 – Alignment of SEX4 orthologs with domains and secondary structures .............................................................. 52
Figure 3.4 – Structural alignment between the SEX4 CBM and AMPKβ ........... 53
Figure 3.5 – Conformational changes in the SEX4 CBM upon glucan binding.... 54
Figure 3.6 – Levels of dephosphorylation of SEX4 mutants and SEX4ΔCBM against pNPP ................................................................. 55
Figure 3.7 – Interaction of maltoheptaose and phosphate at the SEX4 DSP active site .................................................................................. 56
Figure 3.8 – DSP subdomains in SEX4 ...................................................... 57
Figure 3.9 – Conformational changes in the SEX4 DSP ............................. 58
Figure 3.10 – Purification of SEX4ΔCBM .................................................. 59
Figure 3.11 – Glucan phosphatase ability of SEX4ΔCBM ......................... 60
Figure 4.1 – Structure of LSF2 bound to maltohexaose and phosphate ....... 75
Figure 4.2 – Electron density map of maltohexaose and phosphate at the active site
Figure 4.3 – DSP subdomains in LSF2
Figure 4.4 – The LSF2 aromatic pocket
Figure 4.5 – Conservation in LSF2 orthologs
Figure 4.6 – Specific activity of LSF2 and mutants against para-nitrophenyl phosphate (pNPP)
Figure 4.7 – Conformational rearrangement of the LSF2 catalytic site
Figure 4.8 – Structure of LSF2 bound to citrate
Figure 4.9 – Structural alignment of DSP catalytic triad with LSF2
Figure 4.10 – Activation in pNPP activity upon addition of malto-oligosaccharides
Figure 4.11 – LSF2 glucan binding sites
Figure 4.12 – Electron density map of maltohexaose at LSF2 Site-2
Figure 4.13 – LSF2 glucan binding Site-2
Figure 4.14 – Electron density map of maltohexaose at LSF2 Site-3
Figure 4.15 – LSF2 glucan binding Site-3
Figure 4.16 – Sequence conservation of LSF2 and SEX4
Figure 5.1 – Glucan orientation at the SEX4 catalytic site
Figure 5.2 – Interaction between the glucan ligand and SEX4 catalytic triad
Figure 5.3 – Glucan orientation at the LSF2 catalytic site
Figure 5.4 – Interaction between the glucan ligand and LSF2 catalytic triad
Figure 5.5 – Glucan interacting residues at the LSF2 active site
Figure 5.6 – Orientation of maltoheptaose bound to SEX4
Figure 5.7 – Substrate specificity in SEX4, CBM mutants and SEX4ΔCBM
Figure 5.8 – Glucan interacting residues at the SEX4 active site
Figure 5.9 – Active site topologies of SEX4 and LSF2
Figure 5.10 – Reversal of substrate specificity in SEX4
Figure 5.11 – Substrate specificity in LSF2 active site mutants
Figure 5.12 – Laforin substrate specificity
Figure 6.1 – Structural comparison between SEX4 and LSF2 glucan binding methods
Figure 6.2 – CT motif in LSF2 and SEX4
Figure 6.3 – Continuation of directionality in LSF2 active site and Site-2 glucans ................................................................. 132
Figure 6.4 – Aromatic pocket in the SEX4 and LSF2 DSPs ......................... 133
Figure 6.5 – Aromatic pocket residues and interactions.............................. 134
Figure 6.6 – Directionality of glucan chains at the SEX4 and LSF2 active sites 135
Figure 6.7 – Dihedral torsion angles about the glycosidic bonds in the SEX4-bound glucan .................................................................................................................................................. 136
Figure 6.8 – Dihedral torsion angles about the glycosidic bonds in the LSF2 active site-bound glucan .................................................................................................................................................. 137
Figure 6.9 – Alignment of LSF2, SEX4, and LSF1 CT motifs ......................... 138
Figure 6.10 – Sequence alignment of SEX4, LSF1, and laforin CBMs............. 139
Figure 6.11 – Glucan binding sites in glucoamylase and SEX4 CBMs ............. 140
Figure 6.12 – Sequence alignment of SEX4, LSF2, LSF1, and laforin DSPs .... 141
Figure 6.13 – Sequence alignment of glucan phosphatase DSP subdomains .. 142
Figure 6.14 – Structural alignment of SEX4 with 5 closest structural homologs 143
Figure 6.15 – Structural alignment of LSF2 with 5 closest structural homologs 144
Figure 6.16 – Variable loop (V-loop) in the glucan phosphatases and other DSPs .................................................................................................................................................. 145
Figure 6.17 – D-loop substrate-interacting residues in DSPs ......................... 146
Figure 6.18 (AB) – Alignment of glucan phosphatase subdomains with DSPs . 147
Figure 6.18 (C) – Alignment of glucan phosphatase subdomains with DSPs ... 148
Figure 6.18 (D) – Alignment of glucan phosphatase subdomains with DSPs ... 149
Figure 6.18 (E) – Alignment of glucan phosphatase subdomains with DSPs.... 150
Figure 6.19 – Active site topology of DSPs with different target substrates ...... 151
Figure 6.20 – Overlayed active site topologies of DSPs with different target substrates .................................................................................................................................................. 152
Figure 6.21 – Position of ancillary domains in DSPs with respect to the DSP domain .................................................................................................................................................. 153
Figure 7.1 – Model for SEX4 interaction with starch granules within reversible starch phosphorylation ............................................................... 160
Figure 7.2 – Model for LSF2 interaction with starch granules within reversible starch phosphorylation ............................................................... 161
Figure 7.3 – Potential branch and main chain pathways within the SEX4 structure .................................................................................................................................................. 162
Figure 7.4 – Potential LSF2-glucan complexes formed by multiple binding sites
..............................................................................................................163

Figure 7.5 – Proposed site-directed mutagenesis of SEX4 and LSF1 to define minimum requirements for glucan phosphatase activity...............................................164
1. INTRODUCTION

Preface

It is difficult to overestimate the importance of starch in the biosphere of our planet. Starch is the primary glucose storage molecule in plants and is the most abundant energy reserve carbohydrate. It plays an integral part in the life of humankind as a fundamental component of food, fuel, and industrial products. Starch is the primary substance consumed from crop plants and the localized availability of starch-containing plants formed the basis of the agricultural revolution. Today, humans derive 50-80% of daily caloric intake directly from starch (1, 2). In addition, starch is also consumed indirectly as a major component of feed for livestock. Starch is also the main starting product for first-generation biofuel via its conversion into ethanol (3). Furthermore, starch metabolism plays a key role in the production of hydrogen from algae, which is projected to be an important source of fuel in the future (4). Lastly, starch is processed into a feedstock used in industry to generate a myriad of products that are essential for everyday life; including paper, plastics, adhesives, and building materials (5). Due to its central role in food, fuel, and industrial products, starch has, and will continue to be, one of the most important biomolecules in human life.

The world is currently faced with issues that will require an increase in the production of starch without an equivalent increase in the land available for cultivation. Rising population, loss of arable land, and climate change will undoubtedly create greater pressures to increase starch yield from individual plants. Current estimates indicate that demand will require the production of 15 billion tones of starch in 2015 (6). Meeting this demand will require a thorough understanding of the molecular mechanisms of starch metabolic pathways in order to manipulate them to increase yields. In addition, the use of starch in industrial products requires excessive heating/cooling cycles that are costly as well as harsh chemical modifications during processing with undesirable
environmental impacts (5, 7). Developing methods to manipulate the chemistry of starch in planta is therefore of significant interest to diversify its applicability without negative collateral effects.

The current study investigates the structure and function of glucan phosphatases, which are enzymes that dephosphorylate starch and are critical regulators of its degradation. These enzymes represent a potential avenue to both increase starch yields as well as modify starch chemically with a reduction in economic and environmental impacts. It is the hope of this author that the work contained in this dissertation, regarding the structural mechanisms of glucan phosphatase function, will be used to develop biotechnological methods that will have a positive impact on the health and well being of the earth and its inhabitants.

**Starch Structure and Synthesis**

Starch is a relatively simple biomolecule, yet the overall structure of starch granules is quite complex. Starch is composed almost entirely of the glucose homopolymers amylose and amylopectin. Amylose is a minor component of starch (10-30%) and is formed from linear α-1,4-glycosidic linked chains. Amylopectin is the major component of starch (70-90%) and is also formed from α-1,4-glycosidic linked chains, but with α-1,6-branches every 20-25 glucose units (8, 9). Adjacent amylopectin chains interact to form double helices that organize into crystalline lamellae and account for the water-insoluble property of starch (Figure 1.1) (10). The ability for amylopectin to form secondary helical structures is the driving property towards the ultimate formation of starch granules, which are dense, stable, and therefore effective storage molecules (9).

The formation of starch granules allows plants to partition glucose for both short and long-term storage of energy and raw materials (11). Medium- and long-term storage starch is partitioned into vesicular structures called amyloplasts within non-photosynthetic tissues, such as seeds, tubers, and roots. These starch-containing organs are those harvested from crop plants, i.e. the
endosperm of corn, wheat, and rice and the tubers of potato and cassava. In addition, plants partition starch in the chloroplasts of leaves for short-term storage, which is referred to as transitory starch (12). In contrast to long-term storage starch, the metabolism of transitory starch is directly linked to the diurnal cycle of photosynthesis.

The model organism *Arabidopsis* accumulates transitory starch almost exclusively (13). Roughly half of the glucose produced by *Arabidopsis* is partitioned into its leaf starch pools during the day while photosynthesis is occurring (14). This starch is then degraded and utilized during the night to facilitate growth during the non-photosynthetic period (15). Therefore, transitory starch in leaves undergoes a complete cycle of synthesis and degradation every day (*Figure 1.2*). Furthermore, the plants can alter the rate of synthesis and degradation according to the length of the photosynthetic period without skipping a cycle (16). In general, transitory starch granules have the same overall structure as storage starch. The main differences are granule size and shape, which vary among species regardless. *Arabidopsis* starch granules range in size from 1-2 µm, but starch from potato tubers can reach sizes up to 100 µm (9).

Despite these differences in granular morphology, the genes involved in metabolism of *Arabidopsis* transitory starch are highly conserved in other species (17, 18). Therefore, as with other aspects of plant biology, *Arabidopsis* serves as the primary model organism to study starch metabolism.

Starch is synthesized during the day in chloroplasts via the combined activity of starch synthases (SSs), branching enzymes (BEs), and debranching enzymes (DBEs). SSs catalyze the formation of α-1,4-likages via transfer of ADP-glucose to the non-reducing end of an existing chain, thus elongating it (19). This results in a radially organized starch granule with the non-reducing ends of glucan chains pointed towards the outside. As linear chains reach an adequate length, BEs transfer chain segments of 6 or more glucose units to form α-1,6-branches (20). SSs and BEs therefore catalyze the basic formation of amlopectin. Importantly, DBEs are necessary to form the final structure of
starch. DBEs regulate the distance between branch points, removing branches that are too close together (21). The positioning of branch points is essential for consistent formation of amylopectin double helices. Loss of DBE activity results in the formation of soluble phytoglycogen, which is a less effective molecule for glucose storage (22). Together, SS chain elongation, BE branching, and DBE regulation of branch spacing produce the final starch granule.

Starch Breakdown and Phosphorylation

During the night, transitory starch breakdown occurs at a constant rate, and nearly all of the starch synthesized during the day is degraded by the end of the night phase (Figure 1.2) (11, 12, 23). Starch glucans are degraded by the activity of enzymes that hydrolyze $\alpha$-1,4- and $\alpha$-1,6-glycosidic linkages. Linear glucan chains are hydrolyzed by $\beta$-amylases (BAMs) and branch points are removed by the DBEs isoamylase (ISA) and limit dextrinase (LDA) (24-26). BAMs and DBEs are responsible for the majority of starch degradation, but $\alpha$-amylase, $\alpha$-glucan phosphorylase, and disproportionating enzymes are also present and act on glucan chains released from the granule (26-30). The collective activity of these enzymes ultimately results in the release of maltose and glucose that are exported from the chloroplast for use in the cell (31, 32).

In 1991, the Somerville lab identified an interesting Arabidopsis mutant ($sex1$) that contained a starch excess phenotype (33). The $sex1$ leaves contained excess starch at the end of the night, the starch granules were enlarged, and the plants had a decreased rate of growth. The peculiar feature of this mutant line was that the hydrolytic enzymes that degrade the starch granule were active and present in wild type concentrations. Therefore, it was hypothesized that the $sex1$-encoded protein was a novel protein that participates in an unknown regulatory mechanism in the process of starch degradation. A decade later, it was confirmed that, indeed, the $sex1$ plants cannot transition from synthesis to breakdown and that the $sex1$ gene product initiates this transition (34).
The formation of glucan chain secondary helical structures and the consequential exclusion of water has two clear evolutionary benefits: 1) it permits a large quantity of glucose to be stored in a minimal amount of space, and 2) it precludes access to hydrolytic enzymes that would degrade starch while synthesis is occurring. This perceived benefit results in a molecular dilemma for the plant: starch is synthesized in a manner that inhibits its breakdown, but efficient breakdown must occur each night. In other words, starch granules are constructed as impenetrable structures to stop enzymes from degrading them while being built, but their impenetrability also inhibits enzymes from degrading them when energy needs to be released. To overcome this obstacle, plants utilize enzymes that specifically phosphorylate starch to alter its biophysical properties and transition from synthesis to degradation (35, 36). The sex1 gene encodes a protein called α-Glucan Water Dikinase (GWD) that phosphorylates glucose moieties within starch (37). The introduction of phosphate into amylopectin helices results in a steric hindrance and helical unwinding (38). This phosphorylation event initiates a phase transition of α-glucans in the outer starch granule from a highly ordered and insoluble state to a soluble state that can be accessed by hydrolytic enzymes (Figure 1.3A,B) (39, 40). In effect, the plant diminishes the properties that make starch an effective form of glucose storage to instigate breakdown. The process of starch phosphorylation is conserved among algae, moss, and land plants and is therefore an essential mechanistic component in starch metabolism (35).

GWD is a member of a small group of phosphotransfer enzymes that transfer two phosphate molecules from ATP to acceptor molecules. GWD binds ATP and performs an autophosphorylation of orthophosphate at a conserved histidine residue. It then transfers the γ-phosphate to water and the β-phosphate to the C6-position of a starch glucose moiety (Figure 1.4A) (41). GWD is a large monomeric enzyme with a size of about 155 kDa and contains two N-terminal Carbohydrate Binding Modules (CBMs) in the CBM45 family and a C-terminal catalytic pyruvate phosphate dikinase domain (PPDK) (42). The CBM45 family...
has only been identified in three plant enzymes and has been shown to have a low affinity for glucans (Kd=0.68 ± 0.02 mM for β-cyclodextrin), possibly permitting reiterative GWD activity (43). Recombinant GWD can phosphorylate native starch granules from multiple organisms as well as crystalline maltodextrins prepared in laboratory conditions (39). However, GWD is unable to phosphorylate soluble maltodextrins, preferring instead an insoluble crystalline substrate.

After the characterization of GWD, an additional glucan dikinase was identified, called Phosphoglucon Water Dikinase (PWD), that operates downstream of GWD to further solubilize the outer starch granule and facilitate starch breakdown (Figure 1.3A) (44-46). PWD uses the same dikinase mechanism as GWD, but it phosphorylates starch glucans at the C3-position of starch glucose and only acts upon starch granules that have been pre-phosphorylated at the C6-position by GWD (Figure 1.4A). PWD is also a large monomeric enzyme of approximately 130 kDa. In addition to its C-terminal PPDK catalytic domain, PWD contains a CBM20 family carbohydrate-binding module (44). The CBM20 family of CBMs is more thoroughly characterized than the CBM45 of GWD and has a greater affinity for glucans (0.38 ± 0.07 mM Kd for β-cyclodextrin) (47). The different class of CBM in GWD and PWD suggests that the enzymes act on physically differentiated substrates, i.e. PWD phosphorylates a less ordered, more solvent exposed outer starch granule than GWD (47).

Despite differences in substrate preference, loss of PWD in Arabidopsis also results in a starch excess phenotype with a 2-fold increase in starch content at the end of the night (46). However, the PWD-deficient plants had a less severe starch excess phenotype than the GWD-deficient plants and had relatively normal levels of plant growth. Increasing the length of the non-photosynthetic period resulted in some starch degradation in PWD-deficient plants. These results indicate that PWD has a facilitatory effect on starch phosphorylation, but that both GWD and PWD are essential for normal patterns of starch metabolism. In addition, NMR and modeling studies have demonstrated that the C3-
phosphate has a more drastic effect on amylopectin helix unwinding than C6-phosphate, and suggests that C6-phosphate may be more easily accommodated by the amylopectin helices without disrupting them (38, 48, 49). Cumulatively, these results indicate that both PWD and GWD are essential for normal patterns of starch metabolism and stress the importance of starch phosphorylation, at both the C6- and C3-positions, in the transition from starch synthesis to degradation.

Glucan Phosphatase SEX4

An additional layer of complexity was added to the role of phosphate in starch metabolism by the identification of another starch excess mutant called sex4, identified by the Smith and ap Rees labs (50). The sex4 mutants contained more than double the content of leaf starch at the end of the day, compared to wild type plants, and the rate of growth was less than half (Figure 1.5A,B). Interestingly, the sex4 plants displayed no decrease in the activity of starch-degrading enzymes, similar to results found previously in the GWD-deficient sex1 mutants. The Gatehouse lab then determined that the mutated sex4 locus (At3g52180) encoded an active Dual Specificity Phosphatase (DSP) enzyme that was conserved in multiple plant species (51). Following these results, the protein was also found to also contain a chloroplast Targeting Peptide (cTP) and a CBM (Figure 1.6), bind to starch granules within the chloroplast, and follow an expression and pattern of regulation that closely resembled GWD (52-54). In addition, the starch granules within the sex4 plants were found to be extremely large compared to wild type granules, indicating a breakdown in degradation but not synthesis (Figure 1.5C) (52). All the groups involved in the initial sex4 characterizations hypothesized that the Starch EXcess4 (SEX4) protein, as it was eventually called, dephosphorylated enzymes located at the starch granules and thus regulated starch degradation.

Interestingly, the combination of a DSP domain and a CBM, found in SEX4, was only known to exist in one other protein; a human protein called
laforin (53, 55). Laforin is conserved in all vertebrates and a small number of unicellular protists (56). It contains a DSP and CBM, just like SEX4, but the domains are in the opposite orientation (Figure 1.6). Laforin is notable for its role in the fatal myoclonic epileptic disorder called Lafora Disease (LD), which was first described in 1910 by Dr. Gonzalo Lafora (57, 58). Dr. Lafora identified the presence of starch-like material in the neural cells of patients suffering from a previously uncharacterized disease that would later bear his name (59, 60).

Subsequent studies found that the starch-like deposits were found in cells from most tissues of LD patients and were a form of abnormal glycogen that were hyperphosphorylated, had decreased branching, were composed of longer glucan chains, and were insoluble (61-64). The insoluble form of glycogen found in LD patients was called a Lafora body. Thus, multiple lines of evidence suggested that the activity of laforin plays a role in the regulation of glycogen solubility (65).

Glycogen is the main glucose storage molecule in animals, fungi, and bacteria and is analogous to amylopectin in plants and algae. Glycogen is also formed from $\alpha$-1,4-glycosidic linked chains with $\alpha$-1,6-branches, but glycogen has more branch points and shorter chains compared to amylopectin. This arrangement hinders the formation of secondary structures and therefore glycogen remains water-soluble. The solubility of glycogen allows it to be a consistent source of energy for organisms that rely upon bursts of metabolic requirements, rather than the predictable diurnal pattern found in plants. Intriguingly, Lafora bodies in LD patients have a similar branching pattern compared to amylopectin, and they also contain a significant amount of phosphate (61-63, 66). This link between solubility and phosphate in Lafora Bodies eventually led to the discovery in the Dixon lab that laforin dephosphorylates glycogen directly (65, 67, 68). Thus, laforin became the founding member of the glucan phosphatase family.

Emerging information regarding the role of phosphate in starch metabolism and the establishment of laforin as a glucan phosphatase led to
discovery that SEX4 also dephosphorylates starch glucans directly, rather than proteins involved in starch metabolism (Figure 1.3C, Figure 1.4B, Figure 1.7) (69). Thus, SEX4 was also inaugurated into the family of glucan phosphatases. Further studies indicated that SEX4 dephosphorylates both the C6- and C3-positions of starch glucans, with a distinct preference for the C6-position (Figure 1.4B, Figure 1.7C) (70). These results indicate that SEX4 can dephosphorylate both GWD- and PWD-derived phosphates from the starch granule.

The necessity for SEX4 activity in starch metabolism is clearly evidenced by the inability for sex4 plants to properly degrade starch. However, this begs the question: why is starch dephosphorylation necessary for degradation? It was previously established that phosphorylation induces a phase change in the starch granules that allows hydrolytic enzymes, most importantly β-amylase and isoamylase, to degrade the granule. However, it was reported in 1981 that β-amylase is inhibited by the presence of covalently bound phosphate on glucans (71). This is evidenced by the increase in amylopectin-bound phosphate within the starch granules of sex4 plants (69). In addition, sex4 plants had a peculiar accumulation of soluble phospho-glucans with an average chain length of six to seven glucose moieties (69). The chloroplast-localized debranching enzyme isoamylase3 (ISA3) and the endo-amylase α-amylase3 (AMY3) were predicted to be responsible for the generation of these phospho-glucans. These enzymes can remove entire amylopectin chains from the phosphorylated outer starch granule, and the authors hypothesized that they were unable to be exported from the chloroplast (due to their size) and unable to be further degraded by β-amylase (due to the presence of phosphate). As predicted, the double mutants sex4/isa3 and sex4/amy3 both had significant reductions in the accumulation of soluble phospho-glucans, while retaining the starch excess phenotypic characteristics. It was further determined that SEX4 can dephosphorylate the soluble phospho-glucans and is therefore equipped to dephosphorylate both starch itself and the intermediate products of starch degradation (69).
The necessity of SEX4 activity in starch metabolism highlights an additional molecular obstacle that plants face when transitioning from starch synthesis to degradation. Phosphate must be incorporated into the outer starch granule to induce solubilization and permit access to hydrolytic enzymes, but a primary hydrolytic enzyme, β-amylase, is inhibited by the presence of phosphate. Therefore, SEX4 must clear the starch granule of phosphate before it can be fully degraded (Figure 1.3C,D) (69). SEX4 therefore functions in the final step of the reversible phosphorylation cycle. However, the entire process is likely dynamic, with glucan dkinases, glucan hydrolases, and the glucan phosphate SEX4 working in concert (9). Taken together, these results indicate that SEX4 is an essential player in starch metabolism and represents an important target in controlling the transition from starch synthesis to degradation.

**Structure of SEX4**

DSPs, such as SEX4, are categorized within a subgroup of the heterogeneous Protein Tyrosine Phosphatase (PTP) superfamily (65, 72, 73). All PTPs are characterized by a conserved H\textit{C}xxGxx\textit{RS}/T (CX\textsubscript{5}R) catalytic motif, in which the cysteine functions as a nucleophile and is essential for activity. The PTP name is a historical classification and is, in fact, a misnomer. The classical PTPs only dephosphorylated phospho-Tyr residues on proteins, while the classical DSPs dephosphorylate phospho-Tyr, phospho-Ser, and phospho-Thr residues. However, in addition to proteinaceous substrates, some DSPs act on non-proteinaceous substrates, notably nucleic acids and lipids (74-76). For example, the DSP Phosphatase and TENsin homolog (PTEN) is a tumor suppressor that dephosphorylates PI(3,4,5)P at the plasma membrane and negatively regulates the Akt signaling pathway (77-79). SEX4 exists within this designation of DSPs with non-proteinaceous target substrates.

The crystal structures of numerous DSPs have been determined, providing detailed insights into the interaction between phosphatases and protein substrates, lipid substrates, and nucleic acid substrates (73, 77, 80, 81).
However, the ability of a phosphatase to dephosphorylate a glucan substrate was previously unknown. Therefore, determining the structural basis of SEX4 glucan phosphatase activity was of particular interest. In 2010, our lab, in collaboration with the Vander Kooi lab at the University of Kentucky, provided the first steps towards understanding the structural basis of glucan phosphatase activity by determining the ligand-free crystal structure of Arabidopsis thaliana SEX4 to a resolution of 2.4Å (Figure 1.8, Table 1.1) (82). The crystallized SEX4 construct was a truncated version containing the DSP domain and CBM, and lacking the chloroplast targeting peptide (Δ89-SEX4). This structure provided the first information regarding the fold of SEX4 and the physical relationship between domains.

The DSP domain (residues 90-252) of SEX4 contained a characteristic αβα DSP fold with a central five-stranded β-sheet flanked by eight α-helices (Figure 1.8) (82). The catalytic site Cx₅R sequence of SEX4 is known as the PTP-loop and is located between β5 and α6 at the base of the active site pocket. The catalytic cysteine (mutated to a serine (C198S) in the crystallized construct, rendering it inactive) functions as a nucleophile during catalysis, attacking the phosphorous atom of the substrate, and forms a phospho-enzyme intermediate. In addition, the PTP-loop arginine (R204) is necessary for positioning the phosphate within the catalytic site and a conserved aspartate (D166) in SEX4 lies in the adjacent D-loop region and is positioned as a general acid catalyst to enhance hydrolysis of the phospho-enzyme intermediate (Figure 1.9) (82).

The SEX4 CBM (residues 253-338) is a member of the CBM48 family and contained a characteristic CBM48 fold as defined by the database of Carbohydrate-Active enzymes (CAZy) (Figure 1.8) (82-84). The SEX4 CBM contains six β-strands that fold into a characteristic compact β-sandwich composed of antiparallel sheets. The conserved binding site of the SEX4 CBM domain is positioned between the β-sandwich and an adjacent loop region (82).

An interesting and surprising feature of the SEX4 structure was the intimate and extended interdomain interaction between the DSP and CBM
domains (Figure 1.8) (82). The DSP and CBM interact with 457Å² of surface area, which forms an uninterrupted pocket connecting the putative CBM binding site and the catalytic site in the DSP. Although a glucan ligand was not present in the structure, this arrangement suggested that the CBM and DSP work in concert to bind and position a glucan chain in the active site for dephosphorylation. The structure also identified a C-Terminal (CT) motif (residues 338-379) that was predicted to be unstructured. The CT motif is connected to the CBM but makes extensive interdomain interaction with the DSP (Figure 1.8, Figure 1.10A). The SEX4 CT motif is composed of an extended loop followed by two α-helices that intimately associate with the DSP with 1010Å² of interfacial accessible surface area. To determine the function of the CT motif, we generated a construct lacking the CT motif (SEX4ΔCT) and expressed it in *E. coli* (Figure 1.10B). We found that SEX4ΔCT yielded an insoluble protein that aggregated and pelleted upon low speed centrifugation (Figure 1.10C) (82). These data indicate that SEX4 is dependent upon a close CBM-DSP interdomain interaction, which relies upon structural support from the previously unrecognized CT motif.

In summary, the SEX4 structure demonstrated the essential tertiary architecture necessary for the glucan phosphatase activity of SEX4. The structure revealed multiple features unique to SEX4 compared to other DSPs, most notably the extensive DSP-CBM interactions that align the DSP active site and CBM binding site, and a novel CT motif that is structurally integrated with the two domains. These features serve to structurally differentiate SEX4 from phosphatases that dephosphorylate protein, lipid, and nucleic acid substrates.

**Glucan Phosphatase LSF2**

The identification of the CT motif in SEX4 provided additional insights into the requirements for glucan phosphatase activity. The CT motif is highly conserved in SEX4 orthologs, but a similar motif failed to be identified in laforin or any other characterized DSP. However, distinct sequence similarity to the SEX4 CT motif was found in two additional putative phosphatases in plants that were
subsequently named Like Sex Four1 (LSF1) and Like Sex Four2 (LSF2) (51, 85, 86). It was predicted that LSF1 and LSF2 were related to SEX4 and may be additional glucan phosphatases involved in starch degradation.

LSF1 is a 65 kDa protein that is similar in overall domain organization to SEX4 (Figure 1.6) (51, 86). It contains a cTP, DSP, CBM and CT motif with an additional PDZ domain, which historically functions in protein-protein interaction, and an extended Domain of Unknown Function (DUF). The DSP, CBM, and CT motif of LSF1 have approximately 54%, 53%, and 27% sequence similarity, respectively, to the same domains in SEX4. LSF1 is conserved in land plants and mosses, but no LSF1 genes are found in any known algal species (86). In addition, LSF1 localizes to starch granules within the chloroplast. LSF1-deficient plants (lsf1) display a starch excess phenotype due to a reduced rate of starch degradation, but the increase in starch was muted compared to sex4 plants (86). Thus, LSF1 is indeed involved in starch degradation. Surprisingly, the authors observed that LSF1 did not possess any phosphatase activity and they postulated that LSF1 is not a glucan phosphatase. LSF1 exhibited no dephosphorylation activity against Arabidopsis starch granules, solubilized amylopectin, or even the phospho-tyrosine mimetic para-nitrophenyl phosphate (pNPP), which is a generic substrate used to determine phosphatase activity in vitro (Figure 1.7) (86). In addition, leaf extracts from lsf1 plants contained a normal level of glucan phosphatase activity and lsf1 starch granules contained normal phosphorylation patterns. Thus, although LSF1 contains similarities with SEX4, at this time it is not designated as active glucan phosphatase and its precise role in starch metabolism is not understood.

LSF2 is a much smaller protein than LSF1, with a size of 32 kDa, and contains a cTP, DSP, and CT motif (Figure 1.6) (85). The DSP and CT motif of LSF2 contain approximately 63% and 28% sequence similarity, respectively, with the equivalent domains in SEX4. Interestingly, LSF2 does not contain a CBM. The structure of SEX4 suggested that the CBM is important for its glucan phosphatase function and it was believed that any glucan phosphatase would
require a CBM for activity. Therefore the ability for LSF2 to possess glucan phosphatase activity was dubious. However, in collaboration with the Zeeman lab, we found that LSF2 dephosphorylates pNPP, solubilized amylopectin, and Arabidopsis starch granules (Figure 1.7) (85). Thus, in spite of not containing a CBM, LSF2 is an additional member of the glucan phosphatase family. 

Intriguingly, we also found that LSF2 exclusively dephosphorylates the C3-position of Arabidopsis starch (Figure 1.4B, Figure 1.7C) (85). This C3-specificity is in direct contrast with SEX4, which can dephosphorylate both positions but shows a distinct preference for the C6-position (70). Indeed, Arabidopsis plants lacking LSF2 had normal patterns of starch degradation and normal levels of plant growth, but the starch contained an increase in C3-bound phosphate (Figure 1.11) (85). These results indicate that SEX4 has some compensatory activity that can accommodate the loss of LSF2. In addition, the dual mutation of SEX4 and LSF2 in Arabidopsis (sex4/lsf2) resulted in an extreme starch excess phenotype that was more severe than the sex4 mutation alone (Figure 1.11) (85). These results indicate that, despite not containing a CBM, LSF2 is an active glucan phosphatase that dephosphorylates the C3-position of starch granules and is a critical player in the process of reversible starch phosphorylation (Figure 1.3C).

**Summary and Outline of Research**

Reversible starch phosphorylation is a critical process necessary for the diurnal metabolism of transitory starch in plants (Figure 1.3). The glucan phosphatases SEX4 and LSF2 remove phosphate molecules from the C6- and C3-positions of starch glucans, thereby allowing them to be fully degraded. The loss of glucan phosphatase activity in Arabidopsis results in an increase of starch content, a decrease in plant growth, and/or changes in the phosphorylation pattern of starch. Due to an increasing demand for starch as food, fuel, and a component in a myriad of industrial products, developing a means to increase starch yield and diversify its chemical and biophysical characteristics is of clear
interest to the scientific community. As critical regulators of starch metabolism, glucan phosphatases represent ideal targets for biotechnological applications.

SEX4 and LSF2 were previously characterized enzymatically and the structure of ligand-free SEX4 was previously determined. However, the structural mechanisms of their glucan phosphatase activity are unknown. The SEX4 structure suggested that it binds phosphoglucans via a combined CBM-DSP mechanism, but failed to provide precise molecular details of this interaction. In addition, LSF2 does not contain a CBM; therefore its ability to bind and dephosphorylate glucans is inherently puzzling. Furthermore, the presence of two glucan phosphatases in the reversible starch phosphorylation cycle is likely due to their opposing substrate specificities. However, the structural basis for preferential C6-dephosphorylation by SEX4 and exclusive C3-dephosphorylation by LSF2 is entirely unknown. Gaining an understanding into these fundamental issues of glucan phosphatase enzymology are essential for their application.

The overarching hypothesis being tested in this dissertation is that SEX4 and LSF2 have unique structural properties as phosphatases that allow them to dephosphorylate glucan substrates and that they possess inherent mechanistic differences that account for differences in their domain organization and substrate specificity. This hypothesis will be tested using a combination of x-ray crystallography, structural analysis, and enzymatic assays. The specific approach is to determine the crystal structures of SEX4 and LSF2 with and without a phospho-glucan ligand bound. Once determined, site directed mutagenesis will be used, in combination with enzymatic and binding assays, to decipher their mechanisms of activity. Upon achieving these goals, the structures will then be compared with other related phosphatases to define the glucan phosphatase as a family. The intent of this research is to provide fundamental information that will be both useful and interesting to future researchers studying phosphatases, starch, and metabolism in general.
Table 1.1 - Data collection and refinement statistics for Δ89-SEX4 C198S

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>Δ89-SEX4 (C198S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB Code</td>
<td>3NME</td>
</tr>
<tr>
<td>Beamline</td>
<td>APS 22-ID</td>
</tr>
<tr>
<td>Space group</td>
<td>P2_1_2_1</td>
</tr>
<tr>
<td>Wavelength</td>
<td>0.9792</td>
</tr>
<tr>
<td>Unit-cell parameters</td>
<td>59.408, 73.934, 162.382</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>28453</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.2 (94.7)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>2.4 (2.49-2.40)</td>
</tr>
<tr>
<td>( R_{\text{merge}} ) (%)</td>
<td>10.0 (44.7)</td>
</tr>
<tr>
<td>Redudancy</td>
<td>3.2 (2.5)</td>
</tr>
<tr>
<td>( I/\sigma(I) )</td>
<td>11.4 (2.04)</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
</tr>
<tr>
<td>Resolution limits (Å)</td>
<td>20.0-2.40</td>
</tr>
<tr>
<td># reflections / # to comute</td>
<td>24660 / 2000</td>
</tr>
<tr>
<td>( R (R_{\text{free}}) )</td>
<td>19.8 (24.9)</td>
</tr>
<tr>
<td># protein residues</td>
<td>579</td>
</tr>
<tr>
<td># solvent molecules</td>
<td>197</td>
</tr>
<tr>
<td># phosphate molecules</td>
<td>2</td>
</tr>
<tr>
<td>Ramachandran</td>
<td></td>
</tr>
<tr>
<td>Most favored</td>
<td>93.3</td>
</tr>
<tr>
<td>Add. Allowed</td>
<td>6.3</td>
</tr>
<tr>
<td>Generously</td>
<td>0.4</td>
</tr>
<tr>
<td>Disallowed</td>
<td>0.0</td>
</tr>
<tr>
<td>RMS deviation</td>
<td></td>
</tr>
<tr>
<td>Bond, Å</td>
<td>0.006</td>
</tr>
<tr>
<td>Angle, °</td>
<td>0.997</td>
</tr>
</tbody>
</table>

From (82).
Figure 1.1 - Composition and Structure of Starch. A. Amylopectin, the major component of starch, is formed from α-1,4-glycosidic linked chains with α-1,6-glycosidic linked branches (green circle). Adjacent amylopectin chains interact with each other to form double helices. B. Amylose, the minor component of starch, is formed from α-1,4-glycosidic linked chains with little to no branches. Single amylose chains form helical structures. C. Composite electron micrograph of a starch granule showing internal ring structure. The ring structure is formed from alternating layers of crystalline and amorphous lamellae (insert). The crystalline lamellae are formed from interactions between adjacent amylopectin double helices. Modified from (11).
Figure 1.2 – Diurnal Metabolism of Transitory Starch in Arabidopsis. Starch undergoes a daily cycle of synthesis and degradation in the leaves of Arabidopsis. Starch is continually synthesized during the day (white bar) until Arabidopsis plants contain approximately 7 mg / g fresh weight starch. Starch content in leaves can be visualized as a dark brown color via iodine staining. This starch is almost completely degraded during the night (black bar), until the leaves are virtually devoid of starch. Modified from (87) and (50).
Figure 1.3 – Reversible Phosphorylation in Starch Degradation. Intact amylopectin helices in the starch granule are represented by gray bars, unwound amylopectin chains as black lines, and phosphate molecules as red circles. A. Amylopectin double helices are phosphorylated at the C6-position by GWD, resulting in partial solubilization. PWD phosphorylation at the C3-position results in complete solubilization. B. Amylases degrade the solubilized amylopectin chains. β-amylase degrades amylopectin chains, releasing maltose, until a phosphate group is reached. C. SEX4 releases amylopectin-bound phosphate preferentially from the C6-position and LSF2 releases phosphate exclusively from the C3-position. D. Amylases (primarily β-amylase and isoamylase) release maltose and longer glucans from starch. E. The cycle is continued with phosphorylation by GWD.
Figure 1.4 – Substrate specificity in Glucan Dikinases and Glucan Phosphatases. 

**A.** GWD and PWD phosphorylate two separate positions of starch glucose moieties. GWD phosphorylates the C6-position of starch glucose first. After GWD phosphorylation, PWD phosphorylates the C3-position of starch glucose.

**B.** SEX4 and LSF2 also dephosphorylate two separate positions of starch glucose moieties. SEX4 dephosphorylates both the C6- and C3-positions of starch glucose, but preferentially dephosphorylates the C6-position. LSF2 exclusively dephosphorylates the C3-position of starch glucose.
Figure 1.5 – *Starch Excess Phenotype in SEX4-deficient Arabidopsis* plants. **A.** WT and SEX4-deficient (*sex4*) Arabidopsis plants stained with iodine to visualize starch content at the end of the day and night. Starch is visualized by the dark brown color and is increased in *sex4* plants at both timepoints. Modified from (85). **B.** Graphical representation of starch content in leaves over a diurnal cycle in WT and *sex4* plants. Modified from (50). **C.** Electron micrograph of starch granules from WT and *sex4* plants, showing increased size of *sex4* starch granules. Modified from (86).
Figure 1.6 – Glucan Phosphatase Domain Outlines. Domain outlines of the glucan phosphatase family, showing plant glucan phosphatases SEX4 and LSF2, mammalian glucan phosphatase laforin, and non-catalytic plant glucan phosphatase LSF1. Legend in corner defines domain abbreviations.
Figure 1.7 – Phosphatase Activity of Plant Glucan Phosphatases.

Phosphatase activity of SEX4 C/S (C198S, inactive mutant), and wild-type SEX4, LSF2, and LSF1. **A.** Activity against p-Tyr mimetic para-nitrophenyl phosphate (pNPP) measured in µmol P released / min*mg. **B.** Activity against solubilized amylopectin, detected using a malachite green assay, measured in nmol P released / min*mg. **C.** Activity against *Arabidopsis* starch radiolabeled with $^{33}$P at the C6- (blue) or C3-position (yellow) measured in nmol P released / min*mg.
Figure 1.8 – Non-glucan bound structure of SEX4. Surface and ribbon diagrams of the Δ89-SEX4 crystal structure containing the Dual-Specificity Phosphatase (DSP) domain (blue) with the catalytic site (red), the Carbohydrate Binding Module (CBM, pink), and C-Terminal (CT) motif (tan), determined to a resolution of 2.4Å. α-helices and β-sheets are identified on the ribbon diagram. A single phosphate molecule (orange) is bound in the active site. The CBM and DSP domains show an extended interdomain interaction. Modified from (82).
Figure 1.9 – SEX4 DSP Catalytic Triad. The PTP-loop (red) is located between β5 and α6 in the DSP domain and contains the Cx₅R catalytic motif. C198 (mutated to serine to render it inactive in the crystallized construct) generates a nucleophilic attack against the phospho-ligand, generating a phospho-enzyme intermediate. R204 is necessary for catalysis and functions to orient the phosphate within the catalytic site. D166 is located in the D-loop between β4 and α5 and acts as a general acid/base, functioning in expulsion of the dephosphorylated substrate. The catalytic triad is conserved in all active PTP and DSP family phosphatases. Modified from (82).
Figure 1.10 – CT motif in SEX4. **A.** The SEX4 CT motif (tan) consists of a loop and two $\alpha$-helices that are connected to the CBM (pink) and make extensive contacts with the DSP domain (blue). **B.** A SEX4ΔCT construct was created, based upon the Δ89-SEX4 WT construct, to determine the importance of the CT motif. **C.** Expression profile of SEX4 WT and SEX4ΔCT in *E. coli* cells. SEX4ΔCT is insoluble and cannot be purified to homogeneity. UI – uninduced, I – induced, P – insoluble pellet fraction, S – soluble fraction, E – purified elution fraction. Modified from (82).
Figure 1.11 – *In vivo* results of LSF2 and SEX4 silencing in *Arabidopsis*. A. *Arabidopsis* wild type, *lsf2, sex4*, and *lsf2sex4* plants at the end of the day and end of the night. Iodine staining indicates the presence of starch in leaves via dark brown color. B. Plant growth of wild type and mutant *Arabidopsis* plants. C. Starch content in wild type and mutant *Arabidopsis* plants. D. Ratio of C3 vs. C6-position phosphate in wild type and mutant *Arabidopsis* plants. Modified from (85).
2. MATERIALS & METHODS

Cloning of Glucan Phosphatase Constructs

Based on data from secondary structure prediction, disorder predictions, and Hydrogen/Deuterium Exchange Mass Spectrometry (H/DXMS), we generated a construct of *A. thaliana* SEX4 lacking the first 89 amino acids (Δ89-SEX4) (82). Δ89-SEX4 lacks the cTP (predicted to be residues 1-54) along with residues up to the DSP recognition domain. Δ89-SEX4 was subcloned into a pET28b vector (Novagen) using NdeI and XhoI restriction sites to encode an N-terminal His$_6$ tag, a thrombin cleavage site, and SEX4. Based on data from secondary structure predictions, disorder predictions, sequence similarity to SEX4, and analysis of LSF2 orthologs, we generated an *Arabidopsis thaliana* LSF2 construct lacking the first 78 amino acids (Δ78-LSF2) (85). Δ78-LSF2 does not contain the chloroplast targeting peptide (predicted to be residues 1 to 65) along with residues up to the DSP recognition domain. Δ78-LSF2 was subcloned into pET28b (Novagen) using NdeI and Xhol restriction sites to encode a His$_6$ tag, a thrombin cleavage site, and LSF2. Full-length *Homo sapiens* Laforin was subcloned into a pET21 (Novagen) vector using NdeI and Xhol restriction sites to encode Laforin and a C-terminal His$_6$ tag (68). SEX4ΔCBM was designed using the apo SEX4 structure and was cloned via a three-piece ligation with the DSP amplified with NdeI and KpnI restriction sites and the CT motif amplified with KpnI and XhoI restriction sites. The two amplified DNA segments were then subcloned into a pET28b vector to encode a His$_6$ tag, a thrombin cleavage site, and SEX4ΔCBM. A point mutation of each construct was created whereby the catalytic cysteine residue was mutated to a serine, an established technique that generates a catalytically inactive construct used to trap the substrate for crystallography and as a negative control in enzymatic assays. All point mutants were generated using a site-directed mutagenesis kit (Agilent) or mutagenesis service (GenScript). All DNA sequencing (ACGT inc) was confirmed using...
MacVector. Amino acid sequences of LSF2 orthologs were aligned with ClustalW in MacVector.

**Expression and Purification of Recombinant Proteins**

All protein constructs were expressed and purified via the following protocol except where noted. BL21-CodonPlus *Escherichia coli* cells (Stratagene) were transformed with expression vector. Cells were grown at 37°C in 2xYT to an O.D.\textsubscript{600} of 0.6-0.8, placed in ice for 20 min, induced with 1 mM isopropyl β-D-thiogalactoside (IPTG), grown at 16°C for ~16 hr, and harvested by centrifugation. Cells were lysed in 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, and 2 mM dithiothreitol (DTT), centrifuged, and the proteins were purified using a Profinia IMAC column with Ni\textsuperscript{2+} beads (Bio-Rad) with a Profinia protein purification system (Bio-Rad). Proteins were eluted in lysis buffer containing 300 mM imidazole. Protein was then dialyzed in 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, and 2 mM dithiothreitol (DTT) overnight in the presence of thrombin. Protein was then reverse purified over the Profinia IMAC column. Protein was purified to homogeneity using a HiLoad 26/60 Superdex 200 size exclusion column.

Laforin, VHR, and constructs used for western analysis were purified without cleavage of the His\textsubscript{6} tag due to the lack of a thrombin cleavage site (Laforin) or for use of the His\textsubscript{6} epitope for antibody recognition. For these constructs, expression and purification was performed as described above, but proteins were not dialyzed in the presence of thrombin. Instead, protein eluted from the Profinia purification system (Bio-Rad) were immediately purified to homogeneity using a HiLoad 16/60 Superdex 200 size exclusion column.

Recombinant *Solanum tuberosum* GWD and recombinant *Arabidopsis thaliana* PWD for \textsuperscript{33}P labeling were purified as previously described with the following modifications (37, 44, 45). GWD and PWD were transformed into BL21 cells and expressed similarly to constructs described above. GWD was lysed in buffer (50 mM Tris/HCl, pH 7.5, 2.5 mM EDTA, 2.5 mM DTT, and 0.5
mM PMSF), and proteins were purified using an anion exchange column (Q-sepharose-FF; GE Healthcare) with a salt gradient (50 mM Tris/HCl, pH 7.5, 2.5 mM ETA, 2.5 mM DTT, 0.5 mM PMSF, and 0.5 NaCl) to elute the protein. Fractions were collected and protein was further purified using a HiLoad 26/60 Superdex 200 size exclusion column (GE Healthcare) in new buffer (100 mM MOPS/KOH, pH 7.6, 1 mM EDTA, 2 mM DTT, 0.5 mM PMSF, and 150 mM NaCl). GWD was then put over a desalting column using a Bio-Scale Mini Bio-Gel P6 desalting column (Bio-Rad) using a Profinia protein purification system (Bio-Rad). PWD was lysed in buffer (50 mM HEPES/NaOH, pH 8.0, 200 mM NaCl, 10 mM imidazole, and 0.5 mM PMSF) and centrifuged, and proteins were purified using a Profinia IMAC Ni²⁺ column (Bio-Rad) with a Profinia protein purification system (Bio-Rad). PWD was further purified using a HiLoad 26/60 Superdex 200 size exclusion column (GE Healthcare).

Crystal Structure Determination and Refinement

For glucan-bound Δ89-SEX4 crystals, Arabidopsis thaliana Δ89-SEX4 C198S (inactive mutant) protein used for crystallization was pre-incubated with 25 mM maltoheptaose (Sigma-Aldrich). The crystallization screen trials were set up via hanging drop vapor diffusion using a Mosquito liquid handling robot (TTPLabtech) using a 200 nL drop with a 1:1 ratio of condition and 12 mg/mL protein. Ligand-bound crystals were obtained in 0.2M magnesium chloride and 20% polyethylene glycol (PEG) 3350. Crystals were briefly soaked in mother liquor with 20% glycerol and flash frozen. A single crystal was used for data collection and structural determination. Data was collected on the 22-ID beamline of SER-CAT at the Advanced Photon Source (APS), Argonne National Laboratory at 120K at a wavelength of 1.00Å. There was one molecule in the asymmetric unit and the structure was determined using molecular replacement with the B-chain of the previously determined SEX4 structure as the search model (82).

For citrate-bound Δ78-LSF2 crystals, Arabidopsis thaliana Δ78-LSF2 wild-
type protein was used in crystallization screen trials that were set up via hanging drop vapor diffusion using a Mosquito liquid handling robot (TTPLabtech) using a 200 nL drop with a 1:1 ratio of condition and 4.8 mg/mL protein. Δ78-LSF2 wild-type crystals were obtained in 100 mM tri-sodium citrate, pH 5.8, 16% 2-propanol, 31% polyethylene glycol (PEG) 4000, and 2% glycerol at 18˚C. Crystals were briefly soaked in mother liquor with 20% glycerol and flash frozen. A single crystal was used for data collection and structural determination. Data was collected on the 22-ID beamline of SER-CAT at the Advanced Photon Source (APS), Argonne National Laboratory at 110K at a wavelength of 1.0Å. There was one molecule in the asymmetric unit and the structure was determined using molecular replacement using the B-chain of the ligand-free SEX4 DSP and α10 of the CT motif as search models (82).

For glucan-bound Δ78-LSF2 crystals, Arabidopsis thaliana Δ78-LSF2 C193S (inactive mutant) protein used for crystallization was pre-incubated with 25 mM maltohexaose (Sigma-Aldrich). The crystallization screen trials were set up via hanging drop vapor diffusion using a Mosquito liquid handling robot (TTPLabtech) using a 200 nL drop with a 1:2 ratio of condition and 4.8 mg/mL protein. Ligand-bound crystals were obtained in 100 mM di-ammonium hydrogen phosphate, pH 5.7, 17% 2-propanol, and 31% polyethylene glycol (PEG) 4000. Crystals were briefly soaked in mother liquor with 20% glycerol and flash frozen. A single crystal was used for data collection and structural determination. Data were collected on the 22-ID beamline of SER-CAT at the Advanced Photon Source (APS), Argonne National Laboratory at 120K at a wavelength of 1.00Å. There was one molecule in the asymmetric unit and the structure was determined using molecular replacement with the citrate-bound Δ78-LSF2 structure as the search model.

For all structures in this dissertation structure determination, refinement, and analysis was performed with the following programs. PHENIX was used for molecular replacement (88). The structures were fully built and refined via
iterative model building and refinement using Coot (89) and Refmac5 (90), respectively. Stereochemistry of the models was analyzed using MolProbity (91). Analysis and molecular graphics were prepared using Pymol (92). Density maps were produced using the FFT program in CCP4 (93). Comparative structure analyses were performed using the DaliServer (94) and DaliLite (95). Protein-ligand contact analyses were performed with Areaimol (96).

**Phosphatase Assays**

Phosphatase assays using pNPP have been previously described and were performed with the following modifications (Figure 2.1) (65, 68, 97). Hydrolysis of pNPP was performed in 50 µL reactions, containing 1x phosphatase buffer (0.1M sodium acetate, 0.05 M bis-Tris, 0.05 M Tris-HCl, pH 7.0, and 2 mM DTT), 50 mM pNPP, and 1 µg of enzyme at 37˚C for 15 min. The reaction was terminated by the addition of 200 µL of 0.25 M NaOH, and absorbance was measured at 410 nm. The assay was performed with each protein six times or more to determine specific activity. Activation assays were performed as above with the addition of varying concentrations of maltose, maltotriose, maltotetraose, maltopentaose, maltohexaose, and maltoheptaose (Sigma-Aldrich) (Figure 2.1).

Phosphatase assays against amylopectin, as determined via released free phosphate by malachite green detection, were performed as previously described with the following modifications (Figure 2.2) (65, 68, 97). Reactions were performed in 20 µL reactions, containing 1x phosphatase buffer (0.1M sodium acetate, 0.05 M bis-Tris, 0.05 M Tris-HCl, pH 7.0, and 2 mM DTT), 100-1000 ng protein, and 45 µg amylopectin. Amylopectin was solubilized using the Roach method (55). The reaction was stopped by the addition of 20 µL of 100 mM N-ethylmaleamide and 80 µL of malachite green reagent. Absorbance was measured at 620 m. The assay was performed with each protein six times or more to determine specific activity.
Phosphate release from $^{33}$P-labeled starch granules was performed as previously described with the following variations (Figure 2.3) (70, 85). C6-$^{33}$P-labeled starch was generated by isolating phosphate-free starch granules from the Arabidopsis sex1-3 mutant (34), phosphorylating the starch with $^{33}$P at the C6-position by GWD followed by washing until all unincorporated $^{33}$P had been removed. Phosphorylation with unlabeled ATP at the C3-position by PWD was performed as previously described. C3-$^{33}$P-labeled starch was generated by isolating phosphate-free starch granules from the Arabidopsis sex1-3, phosphorylating the starch with unlabeled ATP at the C6-position by GWD followed by phosphorylation with $^{33}$P at the C3-position by PWD and washing until all unincorporated $^{33}$P had been removed, as previously described. In both the C6-$^{33}$P- and C3-$^{33}$P-labeled cases, the starch granules were phosphorylated at both positions; however, the $^{33}$P-label was located at only one or the other position. [$^{33}$P]ATP was obtained from Hartmann Analytic.

Recombinant proteins (150 ng) were incubated in dephosphorylation buffer (100 mM sodium acetate, 50 mM bis-Tris, 50 mM Tris-HCl, pH 6.5, 0.05% [v/v] Triton X-100, 1 µd/µL [w/v] BSA, and 2 mM DTT) with C6- or C3-prelabeled starch (4 mg/mL) in a final volume of 150 µL on a rotating wheel for 5 min at 25°C. The reaction was terminated by the addition of 50 µL of 10% SDS. The reaction tubes were then centrifuged at 13,000 rpm for 5 min to pellet the starch. $^{33}$P release into 150 µL of supernatant was determined using a 1900 TR liquid scintillation counter (Packard). The assay was performed with each protein six time or more to determine specific activity.

**Glucan Binding Assays**

Glucan Binding assays using free amylopectin were performed as previously described with the following modifications (65, 98). All proteins used in the glucan binding assays were purified as described above without cleavage of the His6 tag to maintain the epitope for immunoblot analysis. Amylopectin, from potato starch (Sigma-Aldrich), was solubilized using the Roach method (99) at a
concentration of 5 mg/mL. Five milligrams of amylopectin was then pre-pelleted via centrifugation at 110,000g for 1.5 h at 4°C to collect only pelletable amylopectin and then resuspended in 0.5mL binding buffer (50mMTris, 150mMNaCl, pH7.5, and 2 mM DTT). One microgram of recombinant protein was incubated in amylopectin solution for an hour at 4°C with rocking. The solution was then centrifuged at 110,000g for 1.5 h. Co-sedimentation with amylopectin was measured by centrifuging the samples at 110,000g for 1.5 h. All supernatant was removed and protein was precipitated with 4 volumes of acetone stored at -20°C. Precipitated protein was then pelleted via centrifugation at 21,130g for 30 min, and excess acetone was removed using a SpeedVac concentrator (Savant) at 65°C for 1.25 h. Both soluble and pellet fractions were then resuspended in 30 mL RIPA buffer before the addition of 30 mL SDS-PAGE buffer (60 mL total volume for both the soluble and pellet fractions). Fifteen microliters of the pellet and soluble fraction were then resolved via SDS-PAGE, and relative concentration of pellet and soluble protein was analyzed by immunoblotting with anti-His6 antibody. Quantification of the signal was determined using ImageJ. The assay was performed three times or more with each protein to determine binding capacity.

Glucan binding assays using amylopectin coupled to Concanavalin A (ConA) beads were performed as previously described with the following modifications (100), and in a similar manner to the glucan binding assay described above. 100μL Concanavalin A (ConA) Sepharose 4B beads (GE Healthcare) were washed in ConA buffer (67mM Na-HEPES pH 7.5, 0.2mM CaCl₂ 10mM MgCl₂) 3 times. 125μL of 5mg/mL Amylopectin solution was added to the ConA solution. Amylopectin was pre-solubilized using the Roach method (55). The mixture was incubated for 1 hr at 4°C on nutator to allow amylopectin binding to ConA beads. The solution was then centrifuged at 10,000 rpm for 30 seconds and washed 3 times with ConA buffer. 10μL 100mM dithiothreitol (DTT), 10μL Protease Inhibitor Cocktail and 10μg of His-tagged enzyme were then added to the ConA solution. The mixture was incubated for 1 hr at 4°C on a
neutator and centrifuged at 10,000 rpm for 30 seconds. 30μL of the supernatant (S) was added to 10μL of 4x SDS PAGE buffer. Beads were then washed 3 additional times with 500μL ConA buffer. 30μL for ConA bead solution (P) was then added to 10μL 4x SDS-PAGE buffer. Both (S) and (P) samples were heated for 10 minutes at 100°C and (P) sample was centrifuged at 15,000 rpm for 1 minute to pellet the ConA beads. 30μL of each sample was then run in electrophoresis gel and protein concentration was visualized via western analysis with an α-His antibody.
Figure 2.1 – Generic pNPP phosphatase assay. Generic phosphatase activity is measured using the phospho-Tyr mimetic para-nitrophenyl phosphate (pNPP). 50 mM pNPP (gray circle) is added to a tube containing phosphatase buffer. The phosphatase (red star) is then added and phosphate is released (P-circle), resulting in para-nitrophenolate and a colorimetric change. The reaction is quenched in the linear phase by the addition of NaOH and enzymatic activity is determined by measuring the optical density (O.D.) at 410 nm. For the activation assay, non-phosphorylated malto-oligosaccharides are added, ranging from 2 to 7 glucose units.
Figure 2.2 – Malachite green assay for detecting glucan phosphatase activity. Dephosphorylation of amylopectin is determined by the malachite green assay. 500 mg/mL Amylopectin (yellow circle) is solubilized and added to the tube. The phosphatase is added and may dephosphorylate the amylopectin. Ammonium molybdate is then added, which forms a complex with free phosphate, resulting in a colorimetric change. Activity is determined by measuring optical density (O.D.) at 630 nm.
Figure 2.3 – Radiolabeled starch assay for determining C6- vs. C3-position dephosphorylation. A. Starch purified from phosphate free sex1, GWD-deficient *Arabidopsis* plants is divided in two separate pools. One pool is incubated with recombinant GWD and $[^{33}P]β$-ATP, then recombinant PWD with unlabeled ATP (right). The other pool is incubated with recombinant GWD and unlabeled ATP, then recombinant PWD and $[^{33}P]β$-ATP. This results in two separate pools, phosphorylated at both positions, but only $^{33}P$-radiolabeled at the C6- or the C3-position. B. Radiolabeled starch pools are incubated separately with a glucan phosphatase. 10% SDS is added to stop the reaction and then the starch is pelleted. The level of radioactive phosphate left in solution is determined as a measure of glucan phosphatase activity at the C6- vs. the C3-position.
3. Structural Basis of SEX4-Glucan Interaction

Introduction

Starch granules are composed of the glucose homopolymers amylose and amylopectin (8, 9). Amylopectin, the major component of starch, is formed from $\alpha$-1,4-glycosidic linked chains with $\alpha$-1,6-branches (10, 101). Adjacent amylopectin chains interact to form double helices that cause starch to be water-insoluble, which is an essential feature for its function as a glucose cache (9, 101, 102). However, the outer starch granular surface must be solubilized during non-photosynthetic periods so that glycolytic enzymes can access and degrade starch granules and meet the metabolic needs of the plant (38, 103). Plants regulate the solubility of the starch granular surface via a cycle of reversible phosphorylation by dikinases and phosphatases (45, 69, 85). Phosphorylation of amylopectin chains, at the C6- and C3-position of glucose moieties, causes helical unwinding and local solubilization of the outer starch granule (48, 104, 105). This phosphorylation event permits catabolism of surface glucans, but with an important molecular obstacle: the major degradative enzyme $\beta$-amylase cannot degrade starch past phosphorylated glucans (38, 40, 71). Therefore glucan phosphatases must release phosphate from starch in order to reset the reversible phosphorylation cycle and accomplish complete degradation (69, 85).

The first glucan phosphatase discovered in plants was Starch EXcess4 (SEX4) (68, 69). SEX4 has been shown to dephosphorylate both the C6- and C3-positions of starch \textit{in vivo} and \textit{in vitro}, with a distinct preference for the C6-position (70). SEX4 activity is essential for starch catabolism and its mutation in

Arabidopsis leads to an excess of leaf starch, an accumulation of soluble phospho-glucans, and a decrease in plant growth (50). Another glucan phosphatase, Like Sex Four 2 (LSF2) dephosphorylates starch glucans exclusively at the C3-position, and therefore also functions in starch catabolism (85). Cumulatively, the process of reversible phosphorylation requires the concerted activity of dikinases and phosphatases with SEX4 activity being essential for normal patterns of starch metabolism and plant growth.

SEX4 is a member of the protein tyrosine phosphatase (PTP) superfamily characterized by a conserved H\textit{C}xx\textit{G}xx\textit{RS/T} (Cx$_5$R) catalytic motif (65, 72, 73). The glucan phosphatases belong to a heterogeneous subset of PTPs called dual specificity phosphatases (DSPs), with some DSPs dephosphorylating p-Tyr and p-Ser/Thr residues of proteinaceous substrates and other DSPs dephosphorylating lipids, nucleic acids, or glucans (74-76). Previous DSP structures have provided extensive information regarding phosphatase interaction with protein, lipid, and nucleic acid substrate, but the interaction between phosphatases and glucan substrates was entirely unknown. Furthermore, due to their critical function in carbohydrate metabolism, understanding the structural basis of glucan phosphatase activity is of particular interest. Towards this goal, we previously determined the ligand-free structure of SEX4 and identified an extensive interdomain interaction between its Dual-Specificity Phosphatase (DSP) domain and Carbohydrate Binding Module (CBM) that is maintained in part by a previously unrecognized C-Terminal (CT) motif (82). These results suggested that SEX4 binds and dephosphorylates glucans using both its DSP and CBM domains in concert. Therefore, we hypothesized that SEX4 contains a glucan-binding platform that forms a connection between the two domains to integrate glucans into the active site.

Herein, we elucidate the basis of SEX4-phosphoglucan interaction by determining the structure of SEX4 bound to the phosphoglucan products maltoheptaose and phosphate. SEX4 engages glucan chains via an extended interface of primarily aromatic residues that spans the CBM and DSP domains.
Moreover, we establish that the SEX4 CBM is primarily responsible for glucan binding, whereas the DSP positions the phospho-glucan substrate into the active site for dephosphorylation. We generated a SEX4 construct lacking the CBM (SEX4ΔCBM) and demonstrated that this construct is still capable of dephosphorylating solubilized glucans. Thus, the SEX4 DSP domain can engage and dephosphorylate solubilized glucan substrates and the CBM is minimally important for this activity. Cumulatively, this study establishes the molecular basis for SEX4 substrate engagement and provides useful insights into the basis for general glucan phosphatase activity.

**Results**

**Structure of SEX4 bound to maltoheptaose and phosphate**

The structure of the glucan phosphatase *Arabidopsis thaliana* Starch EXcess4 (SEX4) (residues 90-378, C198S (inactive mutant)), with maltoheptaose and phosphate bound in the active site, was determined to a resolution of 1.65Å (*Figure 3.1A, Table 3.1*). The crystallized SEX4 construct contains the catalytic Dual-Specificity Phosphatase (DSP) domain, the Carbohydrate Binding Module (CBM), and the C-Terminal (CT) motif. Maltoheptaose is composed of seven glucose moieties with α-1,4-glycosidic linkages, and clear electron density allowed modeling of six glucose units of the maltoheptaose chain (*Figure 3.1B*).

The maltoheptaose chain is located within an extended pocket that spans the DSP and CBM domains, with Glc1 located at the DSP and Glc6 located at the CBM (numbered from non-reducing to reducing end). In addition, a single phosphate molecule was found within the catalytic site (PTP-loop), directly below Glc2, at a distance of 2.5Å from the catalytic residue S(C)198. The DSP-CBM pocket is ~9Å deep and ~33Å long with a total contact area of 610Å². Of this contact area, 40% of the interactions occur via the CBM domain and 60% via the DSP. Together, maltoheptaose and phosphate represent the product of dephosphorylation of the endogenous SEX4 phosphoglucan substrate, thus demonstrating the structural basis for SEX4-phosphoglucan interactions.
Maltoheptaose at the SEX4 CBM

The SEX4 CBM interacts with the maltoheptaose chain moieties Glc4-6. The central platform for this interaction is a dual-tryptophan motif formed from W278 and W314, which combines with H330 to interact with both faces of the glucan chain (Figure 3.2A). In addition, N332 and K307 are positioned behind maltoheptaose and form hydrogen bonding interactions with the O3 groups of Glc5 and Glc6, respectively. These five residues are highly conserved among SEX4 orthologs (Figure 3.3) and form a concerted glucan-interacting interface.

The SEX4 CBM belongs to the CBM48 family and the dual-tryptophan platform (W278/W314) and K307 represent a conserved functional motif in CBM48 and the related CBM20 family (47). The most similar CBM to that of SEX4 is found in the AMP-activated protein kinase β-subunit (AMPK-β) (PDB: 1Z0N), which has a root mean square deviation (RMSD) of 1.4Å compared to the SEX4 CBM (Figure 3.4) (106). AMPK-β contains all the SEX4 CBM glucan-interacting residues except for a threonine where H330 is located in SEX4. A comparison between the maltoheptaose bound and ligand-free SEX4 structures reveals that H330 undergoes a conformational shift upon glucan binding, bringing H330 directly in line with Glc6 and W314 (Figure 3.5) (82). This glucan-bound configuration is ideally structured for positioning a linear glucan chain into the SEX4 CBM binding interface.

To determine the contribution of the CBM to overall SEX4 activity, we mutated the aforementioned CBM residues to alanine and tested the mutants’ ability to dephosphorylate native Arabidopsis starch, the endogenous substrate of SEX4 (Figure 3.2B). Alanine point mutations of W278, K307, W314, H330, and N332 resulted in a decrease of total phosphatase activity ranging from 66-97%. The largest decrease resulted from the mutation of W278, which is located closest to the interface between the CBM and DSP domains of SEX4. All CBM mutants had generic para-nitrophenyl phosphate (pNPP) dephosphorylation levels comparable to wild type, indicating that decreases in glucan phosphatase
activity were not due to aberrant folding (Figure 3.6). Since CBMs typically engage substrates, these results suggest that a loss of starch phosphatase activity may indicate a decrease of starch binding. To test this, we incubated SEX4 proteins with amylopectin coupled to Concanavalin A (ConA) sepharose. ConA is a lectin from *Canavalia ensiformis* that binds a variety of carbohydrate substrates, including α-D-glucosyl groups (107). Therefore, it is used to provide an amylopectin matrix on the bead to investigate protein-glucan interactions. After incubating SEX4 with the ConA-amylopectin, we washed the beads and used Western analysis to determine if the protein was bound to amylopectin in the pellet or remained in the supernatant. We found that mutation of CBM residues resulted in a dramatic decrease in amylopectin binding (Figure 3.2C). Together, these data demonstrate that the SEX4 CBM is essential for glucan binding and, consequentially, dephosphorylation of starch.

**Maltoheptaose and Phosphate at the SEX4 DSP**

The binding site at the SEX4 CBM interface is continuous with a corresponding interface in the DSP that guides Glc1-4 of the maltoheptaose chain directly over the catalytic site (Figure 3.7A). The maltoheptaose chain has a curved configuration at the DSP active site. The concave surface of the maltoheptaose chain interacts with F167. The convex surface interacts with K237, F235, Y90, F140, and Y139. All of these residues are located in DSP subdomains whose variability among DSP family members corresponds with the specific substrate requirements of each particular phosphatase. Y90 is located in the Recognition Domain (residues 90-98), F140 and Y139 are in the Variable (V-) loop (131-157), F235 and K237 are located in the R-motif (230-249), and F167 is located in the D-loop (162-168) (Figure 3.8). A comparison of the glucan-bound DSP with the non-glucan bound SEX4 structure reveals that residues F167, F235, Y139, and K237 undergo a conformational shift upon glucan binding to engage the glucan chain (Figure 3.9). Based on these data, we hypothesized that the DSP active site is also essential for dephosphorylation of starch glucan.
chains.

To test this hypothesis, we generated alanine mutations of the identified DSP residues and determined the mutants’ ability to dephosphorylate *Arabidopsis* starch granules (*Figure 3.7B*). Mutation of Y90, Y139, F140, F167A, F235, and K237 to alanine resulted in a decrease of total starch dephosphorylation ranging from 10-80%. Each mutant had generic pNPP dephosphorylation levels comparable to wild type, indicating that decreases were not due to aberrant folding (*Figure 3.6*). Interestingly, the average decrease in SEX4 starch dephosphorylation activity upon DSP mutation (38%) was lower than the average decrease upon CBM mutation (89%). Furthermore, we found that the DSP mutants maintained near wild type binding to amylpectin (*Figure 3.7C*). These data were surprising, given that the DSP has a larger relative contact area with the phosphoglucon substrate (60%) than the CBM (40%). However, the CBM is in contact with only the glucan while the DSP is in contact with both phosphate and glucan. Cumulatively, we observed a significant decrease in starch binding and dephosphorylation of starch granules upon CBM mutation, but only a minimal decrease in binding and dephosphorylation upon DSP mutation. These results indicate that the CBM is critical for binding and dephosphorylation of starch granules, whereas the DSP functions to integrate phosphoglucans into the catalytic site.

**SEX4ΔCBM and activity against soluble phosphoglucans**

The above data demonstrate that the CBM is critical for binding to starch granules and the DSP functions to integrate phosphoglucans into the active site. However, the structure indicates that the DSP makes extensive contacts with the glucan chain and contains an extended aromatic glucan-interacting platform. To further investigate the ability of the SEX4 DSP to interact with glucans, we generated a construct whereby the DSP was directly fused to the CT motif, effectively removing the CBM (SEX4ΔCBM) (*Figure 3.10A*). The SEX4ΔCBM construct was expressed in *E. coli* cells, was soluble, and was able to be purified
to homogeneity (**Figure 3.10B**). In addition, SEX4ΔCBM exhibited stable DSP folding, as evidenced by pNPP activity nearly identical to SEX4 (**Figure 3.6**).

To determine the ability of the SEX4 DSP to dephosphorylate starch granules without a CBM, we quantified the ability of SEX4ΔCBM to dephosphorylate *Arabidopsis* starch. As expected, SEX4ΔCBM dephosphorylation of starch was decreased by 98% compared to wild type SEX4 (**Figure 3.11A**). However, SEX4ΔCBM was able to release some phosphate from starch granules at an amount measurably greater than the catalytically inactive SEX4 C/S. Therefore, we tested if increased amounts of SEX4ΔCBM could dephosphorylate starch granules over an extended period of time. We found that SEX4ΔCBM dephosphorylated starch when the enzyme concentration was increased and incubation time was extended (**Figure 3.11B**). These results indicate that the SEX4 DSP domain is capable of dephosphorylating starch, although at a severely diminished capacity.

Since SEX4ΔCBM could dephosphorylate starch granules to some degree, we decided to test its activity versus other glucan substrates. Amylopectin is often used as a starch substitute in both binding and enzymatic assays because it can be solubilized via acid/base treatment, whereas starch cannot. To determine if SEX4ΔCBM could dephosphorylate amylopectin, we utilized a malachite green-based assay that measures the amount of inorganic phosphate released from a substrate. We determined that SEX4ΔCBM dephosphorylates solubilized amylopectin at a rate nearly identical to SEX4 (**Figure 3.11C**). These results indicate that the SEX4 DSP is capable of interacting with and dephosphorylating glucan substrates by itself, but requires a CBM to dephosphorylate insoluble starch granules. This result also highlights how different solubilized amylopectin is from starch and that starch is a more recalcitrant macromolecule to dephosphorylate.
Discussion

Reversible starch phosphorylation is the central regulatory event governing the transition from starch synthesis to starch breakdown in plant cells. SEX4 is critical to starch catabolism and these data reveal the structural basis of its interaction with phosphoglucans. We determined that SEX4 binds glucans via a continuous interface spanning both the CBM and DSP domains. In addition, we determined that the CBM is critical for binding starch granules and that the DSP functions to position the phosphoglucon into the active site. Lastly, we determined that the SEX4 DSP is independently capable of glucan phosphatase activity against soluble glucans, but requires a CBM for efficient dephosphorylation of insoluble starch. These results significantly expand our understanding of the basis of glucan phosphatase interaction with glucan substrates.

The glucan phosphatase activity of SEX4 requires both sufficient binding to glucan substrates overall and integration of the substrate into the catalytic site. Our structure indicates that SEX4 binds glucans in a single pocket that spans the DSP and CBM domains, but our binding and enzymatic data indicate that the two domains have an unequal distribution of responsibilities. We have established that the CBM is primarily responsible for overall binding to starch granules. This activity of the ancillary CBM domain is consistent with ancillary non-catalytic domains found in other enzymes that serve to localize catalysis to protein, lipid, carbohydrate, or nucleic acid substrates (75). However, the non-catalytic domains of these enzymes are often physically removed from the site of catalysis by extended linkers. These enzymes only require the non-catalytic domains to localize the protein to a specific site, rather than depending on the close spatial relationship of binding and catalysis found in the SEX4 structure. Although rare, this coordination of CBM binding and catalysis is also found in CBM-containing enzymes such as SpuA, which is a large multi-modular enzyme found in *Streptococcus pneumoniae* that degrades host cell glycogen (108), Cel9G, a cellulase from *Clostridium cellulolyticum* (109), and cellulase E4 from
Thermomonospora fusca (110). Therefore, although the coordination between CBM and DSP in SEX4 is unique, it represents a successful strategy for linking overall binding and catalysis found in other systems.

The coordination between the CBM and DSP in SEX4 requires the presence of an aromatic-based glucan interacting platform found in the DSP. DSP subdomains account for the substrate preferences of the particular class of DSPs, and the amino acids within these subdomains are tailored for optimum substrate binding at the site of catalysis (76). The DSP subdomains in SEX4 are unique in that they contain an unprecedented number of aromatic residues localized around the DSP active site. Aromatic residues historically function in forming stacking interactions with the rings of glucan chains (111). Aromatic-based glucan interaction platforms often surround the active sites of glycosyl hydrolases to guide α-glucans of starch, β-glucans of cellulose, or other xylans that are components of the plant cell over the active site for modification. The identification that these residues operate as a concerted glucan-binding platform in SEX4 suggests that this feature may be a critical component of the glucan phosphatase family, and may be identified in additional glucan phosphatases. Furthermore, the integration of the phospho-glucan into the catalytic site by the DSP, with minimal contribution to overall glucan binding, is optimal for both the catalytic formation of the specific phospho-enzyme intermediate and disengagement of the product. Thus, the SEX4 DSP is ideal for enhancing substrate/product turnover.

Despite being composed of only α-linked glucose, starch is a complex substrate containing two glucose polymers, α-1,6-branch points, variable chain lengths, and helical secondary structure. Thus, SEX4 must access phospho-glucans within this heterogeneous landscape. Although phosphate positioning within glucan chains of amylopectin is currently unknown, our structure suggests that multiple C6 phosphate modifications could be accommodated at Glc1, Glc2, Glc5, and Glc6. Additionally, the structure indicates that an α-1,6-branch could be accommodated at both glucan termini and at Glc5. Glucan chains within
amylopectin also possess an inherent torsion as they wrap into double helical structures. It is of note that the bound glucan chain possesses both curvature and torsion, with the C6 hydroxyls of Glc5-6 in the CBM pointed towards the solvent and away from the protein, while they are pointed towards the protein in Glc1-3 at the DSP. Thus, SEX4 accommodates a helical glucan and may influence the geometry of the chain during binding.

The structure reveals that SEX4 accommodates six glucan monomers within its extended active site, with Glc2 positioned over the active site. Because both ends of the bound glucan chain are directed towards solvent, it is possible that SEX4 can accommodate register shifts that would facilitate processivity. Given the domain-specific contributions to glucan binding and positioning, it is possible that SEX4 is able to diffuse along the glucan chain via the CBM and then integrate a phospho-glucan into the catalytic site. In *Arabidopsis* leaf starch, the frequency of glucosyl phosphorylation is ~1 in 2000. Although phosphorylation is likely higher during starch degradation, the challenges of locating a phosphorylated glucosyl residue may require this more coordinated, and possibly processive, phospho-glucan engagement with the DSP-CBM operating as an integrated unit.

In summary, the structure of SEX4 bound to the phosphoglucan products maltoheptaose and phosphate is the first information regarding the mechanism of the glucan phosphatase family of enzymes. The coordination of a DSP and CBM to bind and integrate glucans into the active site may be conserved in the evolutionarily unrelated glucan phosphatase laforin. However, the discovery of the glucan phosphatase LSF2, which lacks a CBM, indicates that LSF2 must utilize a distinct method for interaction with glucans. Our determination that SEX4ΔCBM does possess the ability to interact with glucan substrates indicates that LSF2 may contain a more extensive DSP glucan-platform than SEX4 that interacts more effectively with phosphoglucan substrates.
Table 3.1 - Data collection and refinement statistics for glucan-bound Δ89-SEX4 C198S

<table>
<thead>
<tr>
<th></th>
<th>Δ89-SEX4, C198S: maltoheptaose, phosphate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection</strong></td>
<td></td>
</tr>
<tr>
<td>Beamline</td>
<td>APS 22-ID</td>
</tr>
<tr>
<td>Space group</td>
<td>P2;2;2;1</td>
</tr>
<tr>
<td>Cell Dimensions</td>
<td>33.48, 77.89, 117.21</td>
</tr>
<tr>
<td>Wavelength</td>
<td>1.00</td>
</tr>
<tr>
<td>Unique Reflections</td>
<td>37941</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>95.2 (91.0)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>20.0-1.65 (1.71-1.65)</td>
</tr>
<tr>
<td>$R_{\text{merge}}$ (%)</td>
<td>8.0 (43.6)</td>
</tr>
<tr>
<td>$I / \sigma(I)$</td>
<td>11.0 (2.3)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
</tr>
<tr>
<td>Resolution Limits (Å)</td>
<td>19.43 – 1.65</td>
</tr>
<tr>
<td># reflections / # to compute $R_{\text{free}}$</td>
<td>34257 / 1807</td>
</tr>
<tr>
<td>$R (R_{\text{free}})$</td>
<td>17.9 (21.8)</td>
</tr>
<tr>
<td># protein residues</td>
<td>293</td>
</tr>
<tr>
<td># solvent molecules</td>
<td>355</td>
</tr>
<tr>
<td># ligands</td>
<td>2</td>
</tr>
<tr>
<td><strong>B-factors</strong></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>17.3</td>
</tr>
<tr>
<td>Ligand / ion</td>
<td>18.9 (maltoheptaose) 9.1 (phosphate)</td>
</tr>
<tr>
<td><strong>RMS Deviation</strong></td>
<td></td>
</tr>
<tr>
<td>Bond, Å</td>
<td>0.009</td>
</tr>
<tr>
<td>Angle, °</td>
<td>1.39</td>
</tr>
<tr>
<td><strong>Ramachandran</strong></td>
<td></td>
</tr>
<tr>
<td>Favored (%)</td>
<td>98.4</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>1.6</td>
</tr>
<tr>
<td>Disallowed (%)</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Figure 3.1 – Crystal structure of At-SEX4 bound to a glucan and phosphate.

A. Surface/ribbon diagram of At-SEX4 (residues 90-379, C198A (catalytically inactive mutant)) bound to maltoheptaose (green) and phosphate (orange) determined to a resolution of 1.65Å. The SEX4 structure contains the DSP domain (blue) with the catalytic site (red), the CBM (pink), and the CT-motif (tan). The maltoheptaose chain (green) is located in an extended pocket spanning the CBM and DSP domains, and a single phosphate molecule is located at the base of the catalytic site directly beneath Glc2.

B. Close-up view rotated 45˚ showing the 2Fo-Fc electron density map (1.5σ) of the maltoheptaose chain (green) and phosphate (orange) bound to SEX4. The density permitted modeling of six glucose moieties in the maltoheptaose chain (numbered from non-reducing to the reducing end) and clear assignment of glucan orientation. Two conformers were modeled for Glc1, differing in the orientation of the O6-group.
Figure 3.2 – Interaction of maltoheptaose with the SEX4 CBM. A. Close-up of Glc4-6 (green) bound to the SEX4 CBM (pink). CBM residues (yellow) interact with Glc5 and Glc6 to form a binding interface. B. Specific activity of CBM mutants against Arabidopsis starch. Phosphate-free starch was purified from gwd-deficient Arabidopsis and the starch was pre-labeled with $^{33}$P at either the C6- or C3-position and then incubated with recombinant protein. Starch dephosphorylation over time was linear and was measured via release of $^{33}$P at both the C6- and C3-position after 5 minutes. Error bars represent the ± SD of six replicates. Inactive mutant SEX4 C198S was used as a negative control. C. Results of co-sedimentation assay of SEX4 and CBM mutants with amylopectin. Recombinant His-tagged proteins were incubated with 5 mg/mL amylopectin that was pre-incubated with ConA beads. The amylopectin-beads were pelleted by centrifugation, and proteins in the pellet (P) and supernatant (S) were separated by SDS-PAGE and visualized by immunoblot analysis. Amylopectin-bound proteins are found in the pellet (P) and unbound proteins are found in the supernatant (S) fraction.
Figure 3.3 – Alignment of SEX4 orthologs with domains and secondary structures. Alignment of the residues included in the crystallized SEX4 construct between SEX4 orthologs. Secondary structure of At-SEX4 is represented above (α-helices = ovals, β-strands = boxes). Domains are colored, as are subdomains within the DSP domain. Relevant CBM and DSP aromatic pocket residues that interact with the maltoheptaose chain, mentioned in the text, are marked with an asterisk.
Figure 3.4 – Structural alignment between the SEX4 CBM and AMPKβ.
Structural alignment between the SEX4 CBM (pink) and the AMPKβ subunit (1ZON, gray) (106). Both are classified as CBM48 family members by the CAZy database (83). Glc3-6 from the SEX4 structure are shown (green), as is the bound β-cyclodextrin in the AMPKβ structure (gray). Location of glucan-interacting residues in SEX4 (yellow) and AMPKβ (gray) are shown in the inset box.
Figure 3.5 – Conformational changes in the SEX4 CBM upon glucan binding. Structural alignment of the maltoheptaose-bound SEX4 CBM (pink) with the B-chain CBM of the non-glucan-bound SEX4 (3NME, gray) (82). Residues in the glucan-bound CBM (yellow) and the non-glucan-bound CBM (gray) show a conformational change upon glucan binding.
Figure 3.6 – Levels of dephosphorylation of SEX4 mutants and SEX4ΔCBM against pNPP. Activity against the generic p-Try mimetic substrate para-nitrophenyl phosphate (pNPP). pNPP dephosphorylation over time was linear and measured via colorimetric change at 410 nm. Reaction time was 5 minutes. Activity is reported in µmol phosphate released per minute per mg protein. Error bars represent the ± SD of six replicates.
Figure 3.7 – Interaction of maltoheptaose and phosphate at the SEX4 DSP active site. A. Close-up of maltoheptaose (green), phosphate (orange), and the DSP (blue) active site (red). Side chains of interacting residues are colored yellow. B. Specific activity of DSP mutants against *Arabidopsis* starch. Phosphate-free starch was purified from *gwd*-deficient *Arabidopsis* and the starch was pre-labeled with $^{33}$P at either the C6- or C3-position and then incubated with recombinant protein. Starch dephosphorylation over time was linear and was measured via release of $^{33}$P at both the C6- and C3-position after 5 minutes. Error bars represent the ± SD of six replicates. C. Results of co-sedimentation assay of SEX4 an DSP mutants with amylopectin. Recombinant His-tagged proteins were incubated with 5 mg/mL amylopectin that was pre-incubated with ConA beads. The amylopectin-beads were pelleted by centrifugation, and proteins in the pellet (P) and supernatant (S) were separated by SDS-PAGE and visualized by immunoblot analysis. Amylopectin-bound proteins are found in the pellet (P) and unbound proteins are found in the soluble (S) fraction.
Figure 3.8 – DSP subdomains in SEX4. Ribbon diagram of SEX4 bound to maltoheptaose (green) and phosphate (orange) with DSP subdomains colored: PTP-loop (red, residues 198-204), recognition domain (purple, residues 90-98), V-loop (orange, residues 131-157), D-loop (WPD-loop, blue, residues 162-168), and R-motif (cyan, residues 230-249).
Figure 3.9 – Conformational changes in the SEX4 DSP. Structural alignment of the maltoheptaose-bound SEX4 DSP (blue) with the B-chain of the non-glucan bound SEX4 (3NME, gray) (82). Residues in the glucan-bound SEX4 DSP (yellow) and the non-glucan bound SEX4 (gray) show a conformational change upon glucan binding.
Figure 3.10 – Purification of SEX4\(\Delta\)CBM. A. Domain outline of \(\Delta89\)-SEX4 and \(\Delta89\)-SEX4\(\Delta\)CBM (DSP = Dual Specificity Phosphatase domain, CBM = Carbohydrate Binding Module, CT = C-Terminal motif). B. Purification of SEX4\(\Delta\)CBM from \textit{E. coli} cells (MW = molecular weight marker, UI = Uninduced, I = Induced, P = insoluble Pellet fraction, S = Soluble fraction, E = Elution after His\(_6\) affinity purification, E2 = Elution after His\(_6\) cleavage with thrombin, S2 = Sample after size-exclusion chromatography.)
Figure 3.11 – Glucan phosphatase ability of SEX4ΔCBM. A. SEX4 C/S (inactive mutant), SEX4 WT, and SEX4ΔCBM activity against *Arabidopsis* starch. Phosphate-free starch was purified from *sex1* *Arabidopsis* plants (34) and was separated radiolabeled with $^{33}$P at the C6- and C3-positions using recombinant GWD and PWD. Starch dephosphorylation was measured by determining the concentration of $^{33}$P released from the starch by the glucan phosphatases. B. SEX4ΔCBM activity against *Arabidopsis* starch over a period of 40 min. C. SEX4 C/S (inactive mutant), SEX4 WT, and SEX4ΔCBM activity against solubilized amylopectin (99) using a malachite green-based assay (97) to determine phosphate released.
4. Structural Basis of LSF2-Glucan Interaction

Introduction

Starch is synthesized and degraded in a diurnal manner in plant leaves via the concerted activity of starch synthases, branching enzymes, isoamylases, and amylases (1, 9). Water insolubility is an essential feature of starch that underlies its ability to function in energy storage by regulating the access of starch hydrolyzing enzymes (e.g. amylases) (34, 38). Reversible starch phosphorylation, via glucan dikinases and phosphatases, is necessary to solubilize glucans in the outer layers of the starch granule and permit their catabolism during non-photosynthetic periods (9, 35, 38, 103). Thus starch degradation is based on a cyclical enzymatic process involving glucan dikinases, amylases, and glucan phosphatases.

In plants, there are two glucan dikinases that phosphorylate starch. \( \alpha \)-Glucan Water Dikinase (GWD) phosphorylates the C6-position of Glc moieties on the starch granule surface, and this event triggers C3-phosphorylation by Phosphoglucan Water Dikinase (PWD) (37, 44-46). These phosphorylation events impose steric effects that result in amylopectin helix unwinding and local solubilization of surface glucans, thus permitting amylolytic degradation by \( \beta \)-amylases (38, 40, 48). However, \( \beta \)-amylase activity is inhibited when a phosphate group is reached; therefore glucan phosphatases must release phosphate from starch to allow processive starch breakdown (69, 71). Plants

contain two glucan phosphatases that dephosphorylate starch and allow further degradation by β-amylase: called Starch EXcess4 (SEX4) and Like Sex Four2 (LSF2) (53, 65, 69, 85).

Glucan phosphatases are members of the protein tyrosine phosphatase (PTP) superfamily characterized by a conserved Cx₅R catalytic motif (65, 72, 73). The PTPs include a heterogeneous group of phosphatases called the dual specificity phosphatases (DSPs) that dephosphorylate phospho-Ser, -Thr, and –Tyr residues of proteinaceous substrates as well as more diverse substrates such as lipids, nucleic acids and glucans (74-76). The basis for DSP interaction between protein, lipid, and nucleic acid substrates has been previously reported, but the interaction between DSPs and glucans is still largely unknown. The preceding chapter outlined the interaction between SEX4 and its endogenous phosphoglucan ligand by describing the structure of SEX4 bound to maltoheptaose and phosphate. SEX4 binds glucans via a coordinated interdomain pocket that spans its CBM and DSP domains. We determined that the CBM is essential for binding and dephosphorylation of starch. LSF2, however, does not contain a CBM. Instead it possesses only a chloroplast Targeting Peptide (cTP), a DSP domain, and a CT motif (85). Therefore, the mechanism of glucan interaction and dephosphorylation must be distinct from that utilized by SEX4. We hypothesized that LSF2 must either contain a DSP domain capable of a higher glucan-binding affinity than the SEX4 DSP domain, or it contains additional binding interfaces not found in SEX4.

Herein, we elucidate the basis of LSF2-phosphoglucan interaction by determining the structure of LSF2 with and without the phosphoglucan products maltohexaose and phosphate. The structures reveal that the DSP active site of LSF2 contains a heavily aromatic pocket that acts as a platform for glucan interaction at the site of catalysis. Comparison of the glucan-bound and unbound LSF2 structures reveals that the catalytic site undergoes and conformational transition upon glucan binding conducive to catalysis. Moreover, we discovered evidence that the LSF2 DSP becomes activated upon glucan interaction, which
appears to be a phenomenon common to the plant glucan phosphatase enzymes. Finally, we determined that LSF2 contains two additional, non-catalytic secondary binding sites (SBSs) located >20Å from the active site that are associated with the CT motif and are essential for LSF2 glucan binding and dephosphorylation. These results provide evidence of a novel glucan-interaction mechanism in glucan phosphatases, unique to LSF2.

**Results**

**Crystal Structure of LSF2 Bound to Maltohexaose and Phosphate**

The structure of the *Arabidopsis* LSF2 glucan phosphatase (residues 79 to 282, C193S (catalytically inactive)) bound to maltohexaose and phosphate was determined to a resolution of 2.30Å using molecular replacement with one molecule in the asymmetric unit (Figure 4.1A, Table 4.1). The LSF2 DSP domain (residues 79 to 244) possesses a characteristic αβα PTP fold consisting of a central five-stranded β-sheet region flanked by eight α-helices. The CT motif (residues 245 to 282) consists of a loop region culminating in an α-helix that integrally folds into the DSP domain. The CT motif was first identified in the glucan phosphatase SEX4 (82). A search for structural homologs of the LSF2 DSP domain (residues 79 to 244) identified the DSP domain of Arabidopsis SEX4 (residues 90 to 250, root mean square deviation (RMSD) of 1.1 Å, PDB code 3NME) (82) and mouse PTPMT1 (residues 105 to 256, RMSD of 2.2 Å, PDB code 3RGQ) (80) as the structures most similar to LSF2 despite the fact that the LSF2 DSP is only 48 and 17% identical at the amino acid level to the DSP domain of SEX4 and PTPMT1, respectively.

Maltohexaose is composed of six Glc moieties with α-1,4-glycosidic linkages; thus, it is similar to the unwound helices on the starch granular surface. In the structure, maltohexaose is bound to the LSF2 active site and two distal sites (Figure 4.1A). The LSF2 active-site region contains a single maltohexaose chain and phosphate molecule within the catalytic pocket (Figure 4.1B, Figure 4.2). Multiple conserved DSP active-site motifs converge to form an extended
active-site binding pocket within LSF2 that is ~19Å long and ~9Å deep with 511 Å² contact area. These motifs include a recognition domain from α1 through β1 (83 to 92), a variable (V-) loop from α3 through α4 (132 to 150), a D-loop between β4 and α5 (158 to 163), a PTP-loop between b5 and a6 (192 to 199) that contains the LSF2 catalytic signature (Cx5R) motif, and an R-motif between α7 and α8 (225 to 244) (Figure 4.3).

Maltohexaose and Phosphate at the LSF2 Active Site

Five highly conserved aromatic residues delineate the boundaries of an extended active-site pocket, forming extensive interactions with the Glc rings of the maltohexaose chain (Figure 4.4A). These five aromatic residues provide the majority of the interface between the LSF2 active site and the substrate. Y83, Y85, Y135, and W136 form one side of the channel and interact with Glc moieties Glc1-5. Y83 and Y85 are located within the recognition domain, directly adjacent to the R-motif and PTP-loop, respectively. Y135 and W136 are both located in the V-loop and form a continuous interaction surface with Y85. F162, located in the D-loop, forms the opposite side of the aromatic channel and interacts with Glc2-6. These Glc moieties form a helical structure around F162 and interact with both faces of the phenylalanine ring. The residues that form the aromatic channel are strictly conserved in all land plants as well as in most single-celled members of Kingdom Plantae, with only one non-conservative substitution (Volvox carteri L83) (Figure 4.5).

To investigate the functionality of these aromatic residues, we generated alanine mutants of each channel residue and tested their ability to dephosphorylate starch granules isolated from Arabidopsis (Figure 4.4B). Single alanine point mutations of Y83, Y85, Y135, W136, and F162 resulted in a decrease of C3-dephosphorylation of 38 to 96%. Mutation of both sides of the channel (W136A/F162A) resulted in a 99% loss of glucan phosphatase activity. Importantly, the observed decreases in specific glucan phosphatase activity were not due to destabilization of the active site or misfolding of the protein as
evidenced by near-wild-type phosphatase activity for all mutant proteins toward the generic substrate para-nitrophenyl phosphate (pNPP) (Figure 4.6A). Thus, our findings indicate that LSF2 possesses an aromatic channel that forms an extended active site uniquely suited to bind to polyglucan substrates and necessary for glucan phosphatase activity.

Conformational Changes in the LSF2 Active Site

Intriguingly, one of these key aromatic residues, F162, is found in the D-loop directly after D161, which serves as the general acid/base in DSP catalysis (Figure 4.7A). F162 makes the most contact of any active-site residue with the maltohexaose chain (136Å², 27% of the total contact area). To compare the catalytic site of glucan-bound and unbound LSF2, we crystallized LSF2 without maltohexaose and determined the structure to a resolution of 1.65Å using molecular replacement (Table 4.1). The glucan-free LSF2 produced a different crystal form and contained one molecule in the asymmetric unit and one bound citrate molecule from the crystallization buffer (Figure 4.8). The RMSD of the glucan and citrate-bound DSP domain (residues 79 to 244) structures was 1.0Å. A comparison of product-bound and citrate-bound LSF2 structures revealed a substrate-dependent rearrangement of the D-loop architecture upon glucan binding (Figure 4.7B). In the product-bound structure, the orientations of D161 and F162 are significantly different. The D-loop aromatic residue F162, important for the specific activity of LSF2, interacts with multiple Glc moieties of the glucan chain and shifts toward the V-loop. This movement is associated with a reorientation of the critical general acid/base, residue D161. Comparison of the two structures reveals that the terminal carboxylate of D161 is 3.1Å closer to the catalytic cysteine and directly in contact with the O3 group of Glc3. To further analyze the position of D161, we compared the LSF2 DSP with structures of the glucan phosphatase SEX4 (PDB 3NME) (82) and the prototypical protein phosphatases VHR (PDB 1VHR) (73) and SSH-2 (PDB 2NT2) (112) (Figure 4.9). Analysis of the catalytic triad from each of these structures reveals that the
substrate-bound orientation of LSF2 D161 is in a catalytically competent orientation only in the product-bound structure. Thus, LSF2 undergoes a substrate-induced conformational change with the LSF2 product-bound form ideally positioned for catalysis of a phosphogluca

Most DSPs possess a short-chain hydrophilic reside, S/T/N/H, at the +1 residue from the general acid/base aspartate. However, F162 is invariant in LSF2, and the corresponding residue is also strictly conserved in SEX4, F167. It was previously demonstrated that mutating SEX4 F167 to a short-chain hydrophilic residue (F167S) resulted in a 50% decrease in glucan/pNPP phosphatase activity (82). Cumulatively, these data demonstrate an important role for D-loop movement in order to correctly position the catalytic triad and maximize glucan phosphatase activity.

In addition, the rearrangement of LSF2 active site aromatic residues may provide insights into a phenomenon we uncovered when investigating LSF2 pNPP activity. As stated above, pNPP is a non-natural p-Tyr mimetic used to determine generic phosphatase activity. We found that the pNPP activity of LSF2 increased up to >20-fold upon the addition of non-phosphorylated oligosaccharides ≥ 3 glucose units long (i.e. maltotriose) (Figure 4.10A). We tested LSF2 pNPP against oligosaccharides ranging from 2 to 7 glucose units (maltose to maltoheptaose) at various concentrations. No increase in activity was observed with maltose, but LSF2 pNPP activity increased with the addition of each of the other oligosaccharides. We observed that the level of LSF2 activation was correlated with increasing chain length and concentrations of the sugars. With the discovery of this activation, we investigated if a similar activation occurred with the other glucan phosphatases. A nearly identical pattern of activation was found in SEX4 (Figure 4.10B) with each oligosaccharide except maltose increasing SEX4 pNPP activity. However, the mammalian glucan phosphatase laforin was not activated by the addition of oligosaccharides and was actually inhibited (Figure 4.10C). We also investigated what effect the oligosaccharides have on the SEX4ΔCBM protein that we previously
characterized. We found that the SEX4ΔCBM construct also demonstrated activation upon the addition of oligosaccharides (Figure 4.10D), consistent with our hypothesis that the unique aromatic active site is responsible for this phenomenon. Although the complete implications of this activation are unknown, it demonstrates that the plant glucan phosphatases SEX4 and LSF2 contain active sites that are primed for catalysis upon glucan interaction, and that this characteristic is not shared with the mammalian glucan phosphatase laforin.

Non-catalytic Glucan Binding Sites

Based upon the initial identification of SEX4 and laforin, glucan phosphatases were originally defined as enzymes that contain both a DSP and CBM domain, and the CBM has been shown to be critical for endogenous substrate binding and biological activity for both of these enzymes (56, 65). Because LSF2 lacks a CBM, the nature of its substrate binding ability has been unclear (85). Therefore, we investigated LSF2 glucan binding to amylopectin (Figure 4.11A). LSF2 was incubated with amylopectin, and the amylopectin was then pelleted by ultracentrifugation. Proteins in the pellet and supernatant were separated by SDS-PAGE and visualized by immunoblot analysis. The prototypical protein phosphatase VHR does not bind amylopectin and was found in the supernatant, whereas the prototypical glucan phosphatase SEX4 possesses robust glucan binding and was largely in the pellet. LSF2 also robustly binds amylopectin and, similar to SEX4, was largely in the pellet. Next, we sought to define how mutations in the aromatic channel affect LSF2 glucan binding. The LSF2 W136A/F162A mutant, which had a 99% decrease in specific glucan phosphatase activity, showed only a moderate (32%) decrease in amylopectin binding. This suggests that, while the active site is necessary for glucan phosphatase activity, other regions primarily determine substrate binding. These data are consistent with LSF2 containing additional glucan binding sites distinct from the active-site aromatic channel.

Indeed, the maltohexaose-bound LSF2 structure revealed two additional
glucan binding sites >20Å from the active site (Figure 4.11B). Thus, we hypothesized that one or both of these additional glucan-binding sites could functionally replace a CBM domain and be critical for the biological activity of LSF2.

One maltohexaose chain is located in a binding pocket (Site-2) formed by residues from the DSP domain and CT motif on the opposite side of the V-loop ~21Å from the active site (Figure 4.12). The maltohexaose chain makes extensive contacts (391Å²) with residues in β4 and α5 of the DSP as well as the C terminus of the CT domain (Figure 4.13A). The maltohexaose chain wraps around the CT-loop, forming hydrogen bonds with R153 and R157 and van der Waals contact with W180 and M155. As with the active-site residues, Site-2 residues are highly conserved in LSF2 orthologs (Figure 4.5). To determine the effect of Site-2 glucan binding on LSF2 activity, we tested the ability of alanine point mutants as well as a C-terminal truncation to dephosphorylate starch granules (Figure 4.13B). Alanine mutations of W180, M155, R153, and R157 resulted in decreases of specific glucan phosphatase activity of 24 to 50%. Truncation of the three C-terminal residues (R280/G281/T282, ΔRGT) decreased activity by 46%. All mutant proteins maintained near wild-type pNPP activity, indicating that the observed effects are specific (Figure 4.6B). While mutation of Site-2 resulted in a substantial decrease in glucan phosphatase activity, we also investigated the effect of Site-2 mutants on substrate binding. We tested the ability of LSF2 R157A, which showed the greatest reduction in specific activity, to bind amylopectin and found that it displayed a substantial (64%) decrease in amylopectin binding (Figure 4.13C). This decrease was markedly greater than that observed for the W136A/F162A active-site mutant. These data demonstrate that Site-2 functions as a glucan binding interface and that this binding site is important for the biological activity of starch dephosphorylation by LSF2.

Two additional maltohexaose chains were found in a binding pocket (Site-3) formed by the CT-loop region ~23Å from the active site (Figure 4.14). Five Glc moieties from two maltohexaose chains (Hex-1 and Hex-2) could be resolved.
The two chains form a helical structure, reminiscent of an amylopectin helix, with a contact area of $338\,\text{Å}^2$. LSF2 primarily interacts with Hex-1, forming hydrogen bonding interactions with K245 and E268 and van der Waal’s interactions with F261 (Figure 4.15A). As with the active site and Site-2 residues, Site-3 residues are highly conserved in LSF2 orthologs (Figure 4.5). To determine the effect of Site-3 glucan binding on LSF2 activity, we tested the ability of alanine point mutants to dephosphorylate starch granules (Figure 4.15B). Alanine mutations of E268, K245, and F261 resulted in decreases of specific glucan phosphatase activity of 35 to 87%. All mutant proteins maintained near wild-type pNPP activity, again indicating that the observed effects are specific. As with Site-2, we also investigated the effect of Site-3 on substrate binding (Figure 4.15C). We found that the Site-3 mutant F261A, which showed the greatest reduction in specific activity, showed the most dramatic (73%) decrease in amylopectin binding. In comparison to Site-2, Site-3 mutants showed greater effects on both substrate binding and specific glucan phosphatase activity.

To determine the contribution of the individual glucan binding sites to the overall binding of LSF2, we generated combination mutants. The combination of Site-2 and Site-3 mutations (R157A/F261A) led to an even more dramatic (87%) decrease in amylopectin binding, but slight binding still remained (Figure 4.15C). Only a mutant combining all three sites (F162A/W136A/R157A/F261A) resulted in a protein that possessed no detectable glucan binding ability. Thus, each of the three glucan-binding sites contributes to substrate binding, and none of the sites are individually sufficient for wild-type levels of glucan binding. In addition, all of the LSF2 residues directly involved in non-catalytic glucan binding are not conserved in the CBM-containing SEX4 (Figure 4.16). Taken together, these data demonstrate that rather than requiring a scaffold protein or CBM, LSF2 possesses three glucan binding sites that are each critical for its ability to bind glucans and function as a specific glucan phosphatase.
Discussion

Phosphatases dephosphorylate each of the four organic macromolecules: proteins, lipids, nucleic acids, and carbohydrates. While previous studies have defined the structural bases of phosphatase activity with proteins, lipids, and nucleic acids, the basis for phosphatase-glucan interaction was previously unclear. The structure of SEX4 bound to a phosphoglucon ligand outlined its CBM-based method for glucan interaction. However, we examined the structural and biochemical basis of LSF2 glucan phosphatase activity, demonstrating a method for glucan interaction distinct from SEX4. Therefore, the current study provides a novel model of glucan interaction in LSF2 that, combined with the previous study of SEX4, illustrates the basic method of glucan interaction in the plant glucan phosphatase family.

At the LSF2 active site, an integrated network of aromatic residues forms an extended binding pocket that allows specific glucan interaction and dephosphorylation. While these aromatic residues are within the core DSP domain, in each case they are uniquely suited to promote phospho-glucan binding. Conservation of aromatic residues in the active site in SEX4 and LSF2 suggests that this is a key general characteristic differentiating glucan phosphatases from other phosphatases. It is also particularly notable that both LSF2 and the previously determined SEX4 structure possess an \( \alpha \)-helical V-loop containing conserved aromatic residues: Y135 and W136 in the LSF2 V-loop and Y139 and F140 in the SEX4 V-loop (82). These residues are integral for LSF2 glucan binding and dephosphorylation and form the basis of the conserved theme of an aromatic pocket within the active site of glucan phosphatases.

Aromatic-Glc stacking interactions are a central structural element for CBMs (42, 111) and are found in the catalytic channel of glycosyl hydrolases (113, 114). However, a similar glucan-aromatic channel interface had not been previously observed in any phosphatase, except SEX4. Our crystallography data revealed that the LSF2 active site maintains interactions with all six Glc moieties on the maltohexaose chain, thus indicating that the active site of LSF2 combines
both a DSP active site and glucan-binding platform.

Indeed, we identified a link between LSF2 substrate binding and catalysis mediated by residue F162. In addition to being a part of the active-site aromatic channel, F162 is located at the +1 position from the catalytic triad residue D161. Rotation of these residues upon glucan binding is required for the correct catalytic positioning of D161. This connection suggests an inherent mechanism for phosphate recognition that is tied directly to the architecture of the glucan phosphatase aromatic channel. Our data demonstrating catalytic activation of pNPP activity in the plant glucan phosphatase active site upon glucan-binding also indicates that the aromatic platform is a defining feature of SEX4 and LSF2. It is important to note that aromaticity may be a general feature of this enzyme class since this residue is conserved as an aromatic/long-chain hydrophobic residue in all known glucan phosphatases, whereas other phosphatases typically possess short-chain hydrophilic residues at this position (82).

Despite similarities with other glucan phosphatases, LSF2 is in fact unique among known glucan phosphatases in that the enzyme functions independent of a CBM. There has been debate as to whether or not other functionality or bridging proteins are required for the glucan phosphatase activity of LSF2 (85, 86). However, our data clearly establish that LSF2 uses three glucan binding sites located in the phosphatase domain for carbohydrate binding. The central function attributed to CBMs is substrate localization, and our data demonstrate that the non-catalytic glucan binding sites identified in the LSF2 structural data adopt this functionality. Mutations of Site-2 and Site-3, the two binding sites located away from the active site, result in dramatic decreases in LSF2 glucan binding and dephosphorylation, similar to decreases observed for CBM mutants of SEX4 and laforin (55, 65, 115). It should be noted that Site-2 and Site-3 both use residues from the glucan phosphatase-specific CT motif (82, 85). Thus, this unique elaboration on the core LSF2 phosphatase domain provides novel functionality.

Carbohydrate active enzymes (CAZymes) as defined by the CAZy
database (http://www.cazy.org) are a diverse collection of enzymes that synthesize and degrade an extremely heterogeneous group of complex carbohydrates and glycoconjugates (83). These enzymes cover >250 protein families, including glycoside hydrolases, glycosyltransferases, polysaccharide lyases, carbohydrate esterases, and non-enzymatic proteins that contain a CBM (83). A CBM is a contiguous amino acid sequence with a conserved tertiary fold that possesses carbohydrate-binding ability and is contained within a carbohydrate-modifying enzyme (42, 83, 111). Many of the enzymes that synthesize and degrade carbohydrates utilize a CBM to bind their carbohydrate substrate and then enzymatically act on the carbohydrate via a distinct catalytic module. This model of a binding domain and enzymatic domain is true for the other identified glucan phosphatases (65, 68, 69, 82). Indeed, the glucan phosphatases were originally defined as any protein containing a phosphatase domain and a CBM (65). While LSF2 is a carbohydrate-modifying enzyme that binds carbohydrates, it does not contain a classical CBM and is not classified under the CAZy classification.

Alternatively, LSF2 uses a glucan binding architecture referred to as secondary binding sites (SBSs) (113, 116, 117). SBSs are an emerging theme found in some glycoside hydrolases (117). Many glycoside hydrolases possess one or more CBMs, but recent structural studies have identified a subset of glycoside hydrolases that contain both a CBM and SBSs, such as SusG (114), or that only possess SBSs, such as barley (Hordeum vulgare) α-amylase (113, 118), human salivary and pancreatic α-amylase (119, 120), and yeast glucoamylase (121). Indeed, the two SBSs of barley α-amylase, which are remote from its glucan binding active site, are directly involved in substrate binding and hydrolysis and act synergistically (122).

Our results establish that LSF2 independently binds and dephosphorylates starch. As starch is a complex, insoluble substrate, the presence of multiple glucan binding interfaces may permit LSF2 to uniquely engage the complex multivalent glucan surface of its endogenous substrate. Indeed, the combined-
site mutants (R157A/F261A and F162A/W136A/R157A/F261A) show additive effects, implying that the sites function together. This suggests a model whereby starch binding involves the engagement of longer or multiple glucan chains by the three glucan-binding sites on LSF2. Moreover, the helical glucan chains at Site-3 are reminiscent of an amylopectin double helix, suggesting LSF2 may interact with complex starch granules with distinct helical characteristics. The functional significance of SBSs in glycosyl hydrolases has been extensively reviewed and various additional functions have been postulated, including substrate disruption, allosteric regulation, enhancing processivity, and relaying of reaction products (117). Due to the limitations of using the short-chain glucan maltohexaose as a ligand, ongoing studies using more diverse glucan chains will be required to examine possible cooperativity or connectivity between the additional LSF2 glucan binding sites. This may also provide insights into the position of LSF2 binding and C3-phosphorylation relative to amylopectin branch points.

In conclusion, we have now defined the structural basis of glucan interaction in both of the glucan phosphatases SEX4 and LSF2. These enzymes contain a similar aromatic pocket at the active site, showing that the method of glucan integration is conserved. However, we also revealed that SEX4 and LSF2 contain distinct mechanisms to interact with the heterogeneous and insoluble substrate starch: SEX4 uses a CBM and LSF2 uses SBSs. The difference in substrate binding mechanism may have implications in their different substrate specificities, i.e. C6- vs. C3-position dephosphorylation. Moreover this difference in substrate interaction/specificity may justify the presence of multiple glucan phosphatases within the plant.
Table 4.1 - Data collection and refinement statistics for glucan-bound and citrate-bound Δ78-LSF2, C193S and WT

<table>
<thead>
<tr>
<th></th>
<th>Δ78-LSF2 C193S: maltohexaose, phosphate</th>
<th>Δ78 LSF2 wild-type: citrate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PDB code</strong></td>
<td>4KYR</td>
<td>4KYQ</td>
</tr>
<tr>
<td><strong>Data Collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beamline</td>
<td>APS 22-ID</td>
<td>APS 22-ID</td>
</tr>
<tr>
<td>Spacegroup</td>
<td>P6$_2$22</td>
<td>P2$_1$2$_1$2</td>
</tr>
<tr>
<td><strong>Cell Dimensions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$, $b$, $c$ (Å)</td>
<td>92.78, 92.78, 144.94</td>
<td>51.82, 99.58, 37.76</td>
</tr>
<tr>
<td>$\alpha$, $\beta$, $\gamma$ (°)</td>
<td>90, 90, 120</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>20.0-2.30 (2.38-2.30)</td>
<td>20.0-1.65 (1.71-1.65)</td>
</tr>
<tr>
<td>$R_{merge}$ (%)</td>
<td>10.8 (39.1)</td>
<td>6.6 (59.3)</td>
</tr>
<tr>
<td>$I/\sigma(I)$</td>
<td>9.0 (2.2)</td>
<td>20.5 (2.3)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>94.8 (87.3)</td>
<td>90.1 (86.3)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.7 (2.5)</td>
<td>4.7 (4.7)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>2.30</td>
<td>1.65</td>
</tr>
<tr>
<td># reflections</td>
<td>15288</td>
<td>20833</td>
</tr>
<tr>
<td>$R (R_{free})$</td>
<td>17.4 / 22.9</td>
<td>15.7 / 19.9</td>
</tr>
<tr>
<td><strong># atoms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>1685</td>
<td>1711</td>
</tr>
<tr>
<td>Ligand / ion</td>
<td>251</td>
<td>13</td>
</tr>
<tr>
<td>Water</td>
<td>88</td>
<td>161</td>
</tr>
<tr>
<td><strong>B-factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>45.4</td>
<td>19.8</td>
</tr>
<tr>
<td>Ligand / ion</td>
<td>37.6 (phosphate)</td>
<td>14.6 citrate</td>
</tr>
<tr>
<td></td>
<td>70.1 (maltohexaose, active site)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>59.1 (maltohexaose, Site-2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>60.8 (maltohexaose, Site-3, Hex-1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>86.1 (maltohexaose, Site-3, Hex-2)</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>49.7</td>
<td>31.5</td>
</tr>
<tr>
<td><strong>RMS deviation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond, Å</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>Angle, °</td>
<td>2.1</td>
<td>1.7</td>
</tr>
<tr>
<td><strong>Ramachandran</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored (%)</td>
<td>97.6</td>
<td>97.7</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>2.4</td>
<td>2.3</td>
</tr>
<tr>
<td>Disallowed (%)</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Figure 4.1 – Structure of LSF2 bound to maltohexaose and phosphate. A. Ribbon diagram of LSF2 (blue, residues 79 to 282, C198S) determined to a resolution of 2.30Å. Maltohexaose chains (green, cyan, orange, and pink) and phosphate (teal) are shown. Elements of secondary structure are numbered consecutively from N- to C-termini. B. Maltohexaose chain (green) and phosphate (teal) at the active site (red) pocket. Image correlates with the red box in A. Glc moieties are numbered from non-reducing to reducing end. The total contact area of the active-site pocket with phosphate and maltohexaose is 511Å².
Figure 4.2 – Electron density map of maltohexaose and phosphate at the active site. **A.** 2Fo-Fc electron density map ($1.1\sigma$) and **B.** omit Fo-Fc electron density map ($2.5\sigma$) of maltohexaose (green) and phosphate (teal) at the active site.
Figure 4.3 – DSP subdomains in LSF2. Multiple DSP subdomains converge to interact with the maltohexaose chain (green) at the LSF2 active site. DSP subdomains are colored as follows: recognition domain (residues 83-92, brown), V-loop (variable loop, residues 132-150, pink), D-loop (WPD-loop, residues 158-163, orange), PTP-loop (residues 192-199, red), and R-motif (residues 225-249, blue).
Figure 4.4 – The LSF2 aromatic pocket. A. A maltohexaose chain (green) interacts with aromatic pocket residues (yellow) at the LSF2 active site. Y83 and Y85 are located in the recognition domain N-terminal from α1. Y135 and W136 are located in the V-loop at α3. F162 is located in the D-loop between β4 and α5. The total contact area of the aromatic pocket residues with maltohexaose is 266Å². B. Specific activity of aromatic pocket mutants against the C3-position of Arabidopsis starch. Phosphate-free starch was purified from gwd-deficient Arabidopsis and the starch was pre-labeled with ³³P at the C3-position and then incubated with recombinant protein. Starch dephosphorylation over time was linear and was measured via release of ³³P at the C3-position after 5 minutes. Each bar is the ± SD of six replicates.
Figure 4.5 – Conservation in LSF2 orthologs. Sequence conservation from 21 genomes containing LSF2 from Kingdom Plantae/Archaeplastida. Primary sequence alignment was produced using ClustalW and *Arabidopsis thaliana* LSF2 residues 79-282. The chloroplast Targeting Peptide (cTP) was highly variable between organisms, thus it was excluded for clarity. Secondary structure from the LSF2 structure is shown above. DSP subdomains are depicted and colored. Residues involved in glucan binding in the active site (asterisk), Site-2 (circle), and Site-3 (triangle) are highlighted.
Figure 4.6 – Specific activity of LSF2 and mutants against para-nitrophenyl phosphate (pNPP). Levels of pNPP dephosphorylation of LSF2 and A. active site mutants, B. Site-2 mutants, and C. Site-3 mutants including combinatorial Site-2, Site-3, and active site mutations. pNPP dephosphorylation over time was linear and measured via colorimetric change at 410 nm. Reaction time was 5 minutes. Each bar is the ± SD of six replicates.
Figure 4.7 – Conformational rearrangement of the LSF2 catalytic site. A. LSF2 active site catalytic triad (yellow) residues interact with maltohexaose (green) and phosphate (teal). S(C)193 (catalytically inactive mutant, C193S) and R199 are located within the PTP-loop between β5 and α6. D161 is located within the D-loop between β4 and α5. B. Substrate-dependent positional rearrangement of the D-loop. D-loop residues F162 (aromatic pocket) and D161 (catalytic residue) lie between β4 and α5 and undergo significant movement in the maltohexaose/phosphate bound (yellow) versus the unbound (gray) LSF2 structures.
Figure 4.8 – Structure of LSF2 bound to citrate. The structure of LSF2 (residues 79 to 282, catalytic site in red) bound to citrate was determined to a resolution of 1.65Å. Inset shows the presence of citrate (green) at the LSF2 active site with omit Fo-Fc electron density (3.0σ).
Figure 4.9 Structural alignment of DSP catalytic triad with LSF2. Structural alignment of the catalytic triad of the DSPs SEX4 (3NME) (82), VHR (1VHR) (73), and Slingshot-2 (SSH2, 2NT2) (112). LSF2 residues S(C)193 (catalytically inactive mutant C198S), D161 and R199 are labeled. **A.** Structural alignment of LSF2 (blue) with SEX4 (green). **B.** Structural alignment of LSF2 (blue) with VHR (cyan). **C.** Structural alignment of LSF2 (blue) with SSH2 (dark blue).
Figure 4.10 – Activation in pNPP activity upon addition of malto-oligosaccharides. Activity against para-nitrophenyl phosphate (pNPP) was determined upon the addition of non-phosphorylated malto-oligosaccarides from 0-50mM. Different malto-oligosaccharides were used, ranging from 2 to 7 glucose units (maltose = 2 glucose units, maltotriose = 3, maltotetraose = 4, maltopentaose = 5, maltohexaose = 6, maltoheptaose = 7). Changes in activity were measured for A. LSF2, B. SEX4, C. laforin, and D. SEX4ΔCBM. pNPP dephosphorylation over time was linear and measured via colorimetric change at 410 nm.
Figure 4.11 – LSF2 glucan binding sites. A. Results from co-sedimentation assay of protein and amylopectin (amylopectin binding assay). Recombinant His-tagged proteins were incubated with 5 mg/mL amylopectin, amylopectin was pelleted by ultracentrifugation, and proteins in the pellet (P) and supernatant (S) were separated by SDS-PAGE and visualized by immunoblot analysis. Amylopectin-bound proteins are found in the pellet and unbound proteins are found in the supernatant. B. Surface model of LSF2 showing DSP domain (blue) and CT motif (green). Maltohexaose chains at the active site (green), Site-2 (cyan), and Site-3 (orange and pink) are shown.
Figure 4.12 – Electron density map of maltohexaose at LSF2 Site-2. A. $2F_o - F_c$ electron density map ($1.1\sigma$) and B. omit $F_o - F_c$ electron density map ($2.5\sigma$) of maltohexaose (cyan) at Site-2.
Figure 4.13 – LSF2 glucan binding Site-2. A. Transparent surface model of LSF2 Site-2 showing DSP domain (blue) and CT motif (green) interaction with the maltohexaose chain (cyan). R153 and M155 are located on \( \beta_4 \). R157 is located between \( \beta_4 \) and \( \alpha_5 \). W180 is located on \( \alpha_5 \). T282 is located on the C-terminus after \( \alpha_{11} \). Glc moieties are numbered from non-reducing to reducing end. The total contact area of Site-2 mutants with maltohexaose is 391 Å\(^2\). B. Specific activity of Site-2 mutants against the C3-position of Arabidopsis starch granules. Phosphate-free starch was purified from gwd-deficient Arabidopsis and the starch was pre-labeled with \(^{33}\)P at the C3-position and then incubated with recombinant protein. Starch dephosphorylation over time was linear and was measured via release of \(^{33}\)P at the C3-position after 5 minutes. Each bar is the ± SD of six replicates. C. Amylopectin binding assay of LSF2 R157A mutant, showing amylopectin-bound protein in the pellet (P) and unbound proteins in the supernatant (S).
Figure 4.14 – Electron density map of maltohexaose at LSF2 Site-3. A. 2Fo-Fc electron density map (1.1σ) and B. omit Fo-Fc electron density map (2.5σ) of maltohexaose chains (orange and pink) at Site-3.
A transparent surface model of LSF2 at the CT motif loop (green) showing interactions with maltohexaose chains Hex-1 (orange) and Hex-2 (pink). K245 is located between $\alpha$8 and $\alpha$9. F261 is located between $\alpha$10 and $\alpha$11. E268 is located in $\alpha$11. The total contact area of Site-3 with the maltohexaose chain is 338Å².

B. Specific activity of Site-3 mutants against the C3-position of *Arabidopsis* starch granules. Phosphate-free starch was purified from *gwd*-deficient *Arabidopsis* and the starch was pre-labeled with $^{33}$P at the C3-position and then incubated with recombinant protein. Starch dephosphorylation over time was linear and was measured via release of $^{33}$P at the C3-position after 5 minutes. Each bar is the ± SD of six replicates.

C. Amylopectin binding assay of LSF2 Site-3 mutant F261A, Site-2/Site-3 mutant (R157A/F261A), and quadruple mutation of the active site, Site-2 and Site-3 (F162A/W136A/R157A/F261A) showing amylopectin-bound protein in the pellet (P) and unbound proteins in the supernatant (S).
Figure 4.16 – Sequence conservation of LSF2 and SEX4. *Arabidopsis thaliana* Δ78-LSF2 sequence alignment with *Arabidopsis thaliana* Δ81-SEX4. Secondary structure of LSF2 and SEX4 is depicted above the primary sequences (α-helices = ellipses, β-sheets = boxes). DSP domain (blue), SEX4 CBM domain (orange) and CT motif (green) are labeled. Residues involved in LSF2 glucan binding are highlighted within the aromatic pocket (asterisk), Site-2 (circle), and Site-3 (triangle).
5. GLUCAN PHOSPHATASE SUBSTRATE SPECIFICITY

Introduction

Amylopectin is the major constituent of starch and is composed of α-1,4-glycosidic linked chains with α-1,6-glycosidic linked branches every 20-25 units (8, 9). Interaction between adjacent amylopectin chains results in the formation of double helices that interact to form crystalline lamellae, significantly contributing to the water-insolubility of starch (9, 10). Water-insolubility is an essential feature for the effectiveness of starch as a glucose cache, as it excludes glycolytic enzymes (e.g. amylases) that would degrade the starch during periods of synthesis (38). However, transient starch must be degraded each night to provide energy during the non-photosynthetic period (9, 35, 38, 103). To overcome this paradox, amylopectin glucose moieties are phosphorylated, at the C6- and C3-position, which causes steric effects and unwinding of the amylopectin double helices (40, 45). This results in local solubilization of the outer starch granules, permitting access to glycolytic enzymes and catabolism of starch.


Starch phosphorylation is initiated by phosphorylation at the C6-position of glucose moieties by α-Glucan Water Dikinase (GWD) (34, 123). Subsequent to C6-position phosphorylation, Phosphoglucan Water Dikinase (PWD) phosphorylates starch at the C3-position of glucose moieties (44-46). Although biophysical studies of starch phosphorylation are in their infancy, the data indicate that C6- and C3-position phosphorylation have distinct effects on helical unwinding and solubilization of the outer starch granule (38). NMR and modeling studies indicate that C6-position phosphate can be tolerated by amylopectin helix geometry, but C3-position phosphate causes a steric clash that directly leads to helical unwinding (48, 49). This model indicates that C6-phosphates act as tracker molecules that initiate C3-phosphorylation, which ultimately breaks the amylopectin helices (38). However, in vivo studies of PWD-deficient Arabidopsis plants indicate that starch is degraded, although inefficiently, without C3-position phosphorylation (46). Future studies are necessary to fully understand the molecular implications of starch phosphorylation, but it is clear that both C6- and C3-position phosphorylation cooperatively influence the biophysical properties of starch and are necessary for catabolism.

Similar to the glucan dikinases, the glucan phosphatases Starch EXcess4 (SEX4) and Like Sex Four2 (LSF2) also demonstrate substrate specificity (124). SEX4 dephosphorylates both the C6- and C3-position of Arabidopsis starch glucose but has been postulated to have a preference for the C6-position (70). Conversely, LSF2 exclusively dephosphorylates the C3-position of Arabidopsis starch glucose (85). In fact, the contrasting substrate specificities exhibited by SEX4 and LSF2 likely justifies the necessity for multiple glucan phosphatases within the plant (9). However, the structural basis of C6- or C3-position specificity in both the glucan dikinases and glucan phosphatases is currently unknown.

Due to the variance in biophysical effects generated by C6- and C3-position phosphates on the starch superstructure, understanding the basis of phosphoglucan substrate specificity is of particular interest. Starch is used in a myriad of industrial products that often require physical and chemical modification
prior to their application (5, 85). Often these modifications require economically and environmentally costly methods to generate the desired starch-based product (5, 125, 126). Phosphorylation is the only known natural modification of starch and therefore represents a potentially lucrative method to process starch (7, 11). Understanding the structural basis of glucan dkinase and glucan phosphatase specificity could potentially lead to biotechnological methods for producing designer starches with biophysical properties tailored for specific purposes.

Herein, we provide insights into the C6- and C3-substrate specificities of SEX4 and LSF2, respectively. Electron density of the maltoheptaose ligand in our SEX4 crystal structure shows that it is oriented with the C6-group towards the catalytic site. Conversely, the maltohexaose ligand in the LSF2 structure is oriented with the C3-group towards the catalytic site. Using structural analysis, we determined that conserved residues within the V-loop and R-motif of the DSP domain influence substrate specificity in the glucan phosphatases. Mutation of these residues in SEX4 to the corresponding residues found in LSF2 reversed the substrate preference of SEX4 from the C6- to the C3-position. Furthermore, evidence suggests that residues outside of the active site may also influence C3-specificity in LSF2. Lastly, we found that laforin does not display substrate specificity and dephosphorylates both the C6- and C3-positions equally. Cumulatively, these data provide extensive insights into the basis of C6- and C3-position substrate specificity in the glucan phosphatases.

**Results**

**Structural evidence for substrate specificity in SEX4 and LSF2**

It was previously suggested that SEX4 preferentially dephosphorylates the C6-position of starch glucose (70). Consistent with this result, our SEX4 structure shows that the maltoheptaose chain is clearly positioned in a C6-specific orientation at the catalytic site (Figure 5.1). The O6 group of Glc2 interacts with the phosphate in the catalytic site at a distance of 2.6Å, compared with 7.1Å for
the O3 group. In addition, the glucose moieties upstream (Glc1) and downstream (Glc3) of Glc2 are also oriented with the O6 group pointed towards the catalytic site. Furthermore, the orientation of the PTP-catalytic triad (S(C)198, R204, and D166) is proximal to the Glc2 O6 and phosphate and poised for catalysis (Figure 5.2). The C6-oriented glucan-ligand within the SEX4 structure reaffirms the substrate specificity indicated by our enzymatic data.

Similar, yet reversed, findings are made by analysis of the LSF2 structure. It was previously established that LSF2 exclusively dephosphorylates the C3-position of starch glucose (85). Consistent with this result, our LSF2 structure shows that the maltohexaose chain is clearly positioned in a C3-specific orientation at the catalytic site (Figure 5.3). The O3 group of Glc3 interacts with the phosphate in the active site at a distance of 2.4Å, compared with 7.0Å for the O6 group. Once again, the glucose moieties upstream (Glc4) and downstream (Glc2) are also oriented with the O3 group towards the catalytic site. We found that the glucan-bound LSF2 PTP-catalytic triad is also proximal to the Glc3 O3 group and phosphate and is therefore poised for catalysis (Figure 5.4). Together, the C6-oriented glucan ligand within the SEX4 structure and the C3-oriented glucan ligand within the LSF2 structure reaffirm the substrate specificities indicated by our enzymatic data and provide an opportunity for structural analysis to determine the underlying basis for the differences in substrate specificity in the glucan phosphatases.

C3-specific bonding interactions at the LSF2 active site

To investigate the basis for maltohexaose orientation at the LSF2 active site we analyzed residues that make close contact with Glc2-4 (Figure 5.5). Glc2 is located directly over the LSF2 R-motif. The amino group of R-motif residue G230 makes hydrogen-bonding interactions with the O6 group of Glc2 at a distance of 2.7Å, positioning it in an orientation consistent with C3-specific dephosphorylation. Glc3 is located directly over the catalytic PTP-loop. The D-loop residue F162 forms a cap over Glc3 at a distance of 3.8Å, forming van der
Waals stacking interactions with the ring of the sugar. Lastly, Glc4 is located directly adjacent to the $\alpha$-helical LSF2 V-loop. The indole group of W136 makes hydrogen-bonding interactions with the O3 group of Glc4 at a distance of 3.2Å, positioning it in an orientation consistent with LSF2 C3-specific activity. We have previously established that both the F162 and W136 aromatic pocket residues are important for LSF2 glucan phosphatase activity. Single or double (F162A/W136A) alanine mutations resulted in significant decreases in activity ranging from 60-98% compared to wild type (Figure 4.4). These results provide a structural basis for the C3-specific orientation of the glucan chain at the LSF2 active site.

**DSP domain directs SEX4 specificity**

The multi-domain mechanism of SEX4 glucan integration into the catalytic site complicates the structural basis of its maltoheptaose glucan orientation. The connectivity between the DSP and CBM domains in SEX4 suggests that the CBM may also have a role in directing the orientation of maltoheptaose. Interestingly, we find that the glucan chain actually reverses its orientation between the CBM and DSP (Figure 5.6). At the CBM the O3-group is pointed towards the protein and the O6-group is pointed towards the solvent, but, conversely, at the DSP the O6 group is pointed towards the protein and the O3 group is pointed towards the solvent. These results indicate that the CBM may influence SEX4 C6-position specificity.

However, biochemical analysis of SEX4 indicates that the CBM is not responsible for directing substrate specificity. Previously, we reported that the starch phosphatase activity of CBM mutants H330A and K307A was decreased by 66% and 89%, respectively (Figure 3.2). Interestingly, despite the decrease in overall activity, both CBM mutants maintained a C6-position site-specificity that is nearly identical to wild type SEX4 (Figure 5.7). To further examine this result, we tested the substrate specificity of SEX4ΔCBM, thereby removing all influence from the CBM. We found that SEX4ΔCBM also retains a level of C6-position specificity.
specificity nearly identical to wild type SEX4 (Figure 5.7). These results indicate that the DSP domain functions to control glucan orientation and is therefore responsible for dictating the C6-position substrate specificity of SEX4. In addition, we previously found that the CBM is essential for glucan binding, but the DSP is only minimally important. Therefore, it appears that the CBM is critical for glucan binding and that the DSP functions in orienting the glucan chain within the catalytic site.

Reversal of SEX4 substrate specificity

To determine the specific interactions that drive glucan orientation at the SEX4 active site, we analyzed the residues that interact with Glc1-3 in the same manner as above with LSF2 (Figure 5.8). Glc3 is located directly adjacent to the SEX4 R-motif. The phenylalanine ring of F235 forms van der Waals stacking interactions with the sugar ring of Glc3 at a distance of 3.8Å. Glc2 is located directly above the PTP-loop catalytic site. As with LSF2, a D-loop phenylalanine, F167, caps Glc2 over the SEX4 active site at a distance of 3.8Å. Glc1 is located directly adjacent to the α-helical SEX4 V-loop. The side of the phenylalanine ring of F140 forms van der Waals interactions with the sugar ring of Glc1 at a distance of 3.6Å. Our SEX4 structure indicates that these interactions orient the maltoheptaose chain in a C6-specific orientation at the catalytic site.

These interactions at the SEX4 active site are based entirely on aromatic van der Waals interactions with the glucan chain. This is in contrast with LSF2, where hydrogen-bonding interactions are formed at both the R-motif and V-loop. A comparison between SEX4 and LSF2 indicates that the SEX4 interacting residues (F245 and F140) are replaced with hydrogen-bond contributing residues at the same positions in LSF2 (G230 and W136) (Figure 5.9A,B). In addition, these residue differences create distinct active site surface topologies, i.e. shapes, between SEX4 and LSF2. The active site pockets of DSPs are formed by the connected surface of the V-loop, PTP-loop, and R-motif (77). The surface topology of DSP enzymes reflects the substrate targeting by the enzyme (77).
The active site topologies of SEX4 and LSF2 are both shallow and wide to accommodate three moieties of an extended glucan chain (Figure 5.9A,B). However, at the R-motif, SEX4 contains a distinct ridge formed by the β-carbon of F235 and LSF2 contains a groove formed by the absence of a side chain at G230 at the same position (Figure 5.9C). Therefore, we hypothesized that the structural basis of C6- and C3-position substrate in SEX4 and LSF2, respectively, is derived from differences in the bonding and topology of their respective active sites.

To test this hypothesis in SEX4, we mutated F235 and F140 in SEX4 to the corresponding residues found in LSF2. If these residues influence substrate specificity in SEX4, then their mutation should result in a shift from preferential C6-position dephosphorylation to preferential C3-position dephosphorylation. We generated single point mutants (SEX4-F140W and SEX4-F235G) and a double mutant (SEX4-F140W/F235G) and determined the relative ability of each mutant to dephosphorylate the C6- and C3-position of Arabidopsis starch granules (Figure 5.10A,B). Wild-type SEX4 dephosphorylates the C3-position of starch glucose at a rate of 29% of total dephosphorylation. Strikingly, both the F140W and F235G mutations increased the rate of C3-dephosphorylation to 51% of total dephosphorylation, effectively removing SEX4 C6-specificity. Even more remarkably, the F140W/F235G double mutant increased the ratio of C3-dephosphorylation to 77%, fully reversing the substrate preference of SEX4 from the C6- to the C3-position. These results clearly support our hypothesis and indicate that F235 and F140 in the SEX4 DSP active site influence preferential C6-dephosphorylation of starch granules.

We also investigated whether the substrate specificity of LSF2 could be reversed similar to SEX4. To test this, we generated the reverse mutations of G230 and W136 to the corresponding residues found in SEX4. If these residues are responsible for C3-position substrate specificity in LSF2, then their mutation should result in C6-position dephosphorylation. We generated single point mutants (LSF2-W136F and LSF2-G230F) and a double mutant (LSF2-
W136F/G230F) and determined the relative ability of each mutant to dephosphorylate the C6- and C3-position of *Arabidopsis* starch (*Figure 5.11*). However, the LSF2 mutants were unable to dephosphorylate the C6-position of *Arabidopsis* starch at a detectable level. These mutations simply made LSF2 less efficient at dephosphorylating the C3-position. These results indicate that LSF2 C3-position substrate specificity may be influenced by the secondary binding sites (SBSs) outside of the active site. Cumulatively, these results indicate that the active site of both SEX4 and LSF2 are structured to influence both C6- and C3-position substrate specificity and that their specificities are achieved via different mechanisms.

**Laforin does not contain substrate specificity**

The data presented here indicates that SEX4 and LSF2 have developed distinct substrate specificities to target both GWD- and PWD-derived starch phosphates in the plant system. In the mammalian system, laforin is the only known glucan phosphatase (56). Laforin dephosphorylates glycogen and regulates its solubility within the cell (127). No data regarding the substrate specificity of laforin has been published. The position of phosphate within glycogen glucose moieties is controversial, but it is possible at the C2-, C3-, and/or the C6-position. To determine if laforin contains inherent substrate specificity, we measured the activity of laforin against both the C6- and C3-positions of *Arabidopsis* starch glucose (*Figure 5.12*). Interestingly, we found that laforin does not display any substrate specificity in this system and dephosphorylates both the C6- and C3-position at the same rate. These results indicate that substrate specificity is a feature common to the plant glucan phosphatases and laforin has evolved as a more promiscuous enzyme that targets both positions with the same frequency.
**Discussion**

The reversible phosphorylation of starch glucose, at the C6- and C3-positions, is the central signaling event orchestrating starch breakdown, but the structural basis of substrate specificity in the glucan dikinase and glucan phosphatases was previously unknown. Our LSF2 and SEX4 structures reveal that their respective DSP active sites are structured to influence the orientation of the glucan chain. Furthermore, we demonstrated that the substrate specificity of SEX4 could be reversed via site-directed mutagenesis of residues within the V-loop and R-motif. These results provide useful information that may be applied towards engineering glucan phosphatases for agricultural and industrial purposes.

The data clearly demonstrate that the bonding interactions and topology of the DSP active sites influence the orientation of the glucan chain at the active site. However, the data reveal that additional factors may also be involved in directing SEX4 substrate specificity. Incredibly, the substrate specificity of SEX4 could be completely reversed upon mutation of two residues. This clearly demonstrates the critical role played by the DSP domain in directing glucan orientation and the inherent malleability of SEX4 substrate specificity. However, the mutation did not grant complete C3-specificity to SEX4, indicating that the CBM could potentially still influence specificity. Moreover, the structure does not reveal the basis for basal C3-dephosphorylation by SEX4. Additional structural information may shed light on the mechanism for integration of the C3-position phosphate into the catalytic site.

In addition, mutagenesis of LSF2 did not bestow any detectable C6-dephosphorylation to LSF2. This likely indicates that LSF2 has complexities in its C3-specific activity that were not uncovered in the current study. We showed that SEX4ΔCBM maintained C6-specificity without a CBM, but used high concentrations of proteins at extended timepoints. Longer incubations of the LSF2 mutants discussed above may produce C6-dephosphorylation as well. Furthermore, our LSF2 structure used short six-unit glucose chains to uncover
LSF2 glucan-binding sites. However, there may be connectivity between the sites that may influence C3-specific glucan orientation at the active site. Further studies, using longer glucan chains and addition point mutations will be necessary to decipher possible connectivity between the LSF2 binding sites and determine their influence on specificity.

In addition, we have determined that laforin possesses no substrate specificity against the C6- or C3-position of starch glucose. Laforin is the only known glucan phosphatase in vertebrates and several protist species; therefore it must be capable of removing all phosphates introduced (56). The means by which phosphate is introduced into glycogen is controversial and currently under investigation (128-130). Furthermore, the position of glycogen phosphate has been proposed at the C6-, C3, and C2-positions, but a clear answer to this question has not been provided (129, 130). Regardless, it is clear that laforin has unselective activity against phosphoglucan substrates and therefore is capable of functioning autonomously within the mammalian system.

Phosphorylation is the only known natural modification of starch, and has direct influences on starch hydration, crystallinity, freeze-thaw stability, viscosity, and transparency that are central to its various commercial applications (7, 131, 132). Furthermore, it is clear that C6- and C3- phosphorylation of starch influence its properties differently, as studies show that C3- phosphate has a more direct effect on starch granule solubilization (48, 49). The current study provides insights into the structural basis of SEX4 and LSF2 specificity and illustrates that SEX4 specificity can be altered via engineering of a discrete number of DSP residues. Engineered SEX4 could provide a means to generate designer starches with tailored patterns of phosphorylation and physical characteristics useful in industrial settings. Elucidation of the structural basis of SEX4 activity provides valuable insights necessary for its application in biotechnology.
Figure 5.1 – Glucan orientation at the SEX4 catalytic site. Omit Fo-Fc electron density map (2.5σ) of maltoheptaose (green) and phosphate (orange) shows clear glucan orientation of the O6 group, pointing towards the catalytic site (red). O6 and O3 groups of the Glc2 moiety are labeled.
Figure 5.2 – Interaction between the glucan ligand and SEX4 catalytic triad. The maltoheptaose chain (green) and phosphate (orange) are located in the SEX4 PTP-loop catalytic site (red). The catalytic triad residues S(C)198 (catalytically inactive mutant C198S), R204, and D166 are depicted in yellow. D166 is oriented directly towards the O6 group of Glc2.
Figure 5.3 – Glucan orientation at the LSF2 catalytic site. 2Fo-Fc electron density map (1.3σ) of maltohexaose (green) and phosphate (orange) shows clear glucan orientation of the O3 group, pointing towards the catalytic site (red). O6 and O3 groups of the Glc2 moiety are labeled.
Figure 5.4 – Interaction between the glucan ligand and LSF2 catalytic triad.
The maltohexaose chain (green) and phosphate (orange) are located in the LSF2 PTP-loop catalytic site (red). The catalytic triad residues S(C)193 (catalytically inactive mutant C193S), R199, and D161 are depicted in yellow. D161 is oriented directly towards the O3 and O2 groups of Glc3.
Figure 5.5 – Glucan interacting residues at the LSF2 active site.
Maltohexaose (green) and phosphate (orange) located at the LSF2 active site. The active site is formed by the R-motif, PTP-loop (red, catalytic site), and V-loop. The amino group of R-motif residue G230 makes hydrogen-bonding interactions with the O6 group of Glc3. D-loop residue F162 makes van der Waals stacking interactions with Glc3 over the catalytic site, capping the moiety. The indole group of V-loop residue W136 makes hydrogen-bonding interactions with the O3 group of Glc4. These interactions serve to orient the maltohexaose chain in a C3-specific orientation.
Figure 5.6 – Orientation of maltoheptaose bound to SEX4. Ribbon diagram showing the SEX4 DSP (blue) with the catalytic site (red) and the CBM (pink). The maltoheptaose chain (green) and phosphate (orange) are depicted. The orientation of the C3 and C6 hydroxyl groups, with respect to the protein and solvent, are reversed at the CBM and DSP.
Figure 5.7 – Substrate specificity in SEX4, CBM mutants and SEX4ΔCBM.

Relative specific activity of SEX4 at the C6- (blue) and C3-position (yellow) of *Arabidopsis* starch represented as percentage of C6- and C3-position dephosphorylation relative to total dephosphorylation. Phosphate-free starch was purified from *gwd*-deficient *Arabidopsis* and the starch was pre-labeled with $^{33}$P at either the C6- or C3-position and then incubated with recombinant protein. Starch dephosphorylation over time was linear and was measured via release of $^{33}$P at the C6- or C3-position after 5 minutes. Each bar is the ± SD of six replicates. Error bars represent the ± SD of six replicates.
Figure 5.8 – Glucan interacting residues at the SEX4 active site.
Maltoheptaose (green) and phosphate (orange) located at the SEX4 active site. The active site is formed by the R-motif, PTP-loop (red, catalytic site), and V-loop. The phenylalanine ring of R-motif residue F235 makes van der Waals stacking interactions with Glc3. D-loop residue F167 makes van der Waals stacking interactions with Glc2 over the catalytic site, capping the moiety. The edge of the phenylalanine ring of V-loop residue F140 makes van der Waals interactions with Glc1. These interactions serve to orient the maltohexaose chain in a C6-specific orientation.
Figure 5.9 – Active site topologies of SEX4 and LSF2. A. Active site of SEX4. A cross section of the active site surface (red lines) is overlayed onto the equivalent positions in the model. B. Active site of LSF2. A cross section of the active site surface (purple lines) is overlayed onto the equivalent positions in the model. C. Superimposed cross sections of the active site surface from SEX4 (red) and LSF2 (purple). Positions of F235 and G230 (F, G) and F140 and W136 (F, W) are denoted.
Figure 5.10 – Reversal of substrate specificity in SEX4. A. Relative specific activity of SEX4 active site mutants at the C6- (blue) and C3-position (yellow) of Arabidopsis starch represented as the percentage of total position dephosphorylation per minute per mg protein. Phosphate-free starch was purified from gwd-deficient Arabidopsis and the starch was pre-labeled with $^{33}$P at either the C6- or C3-position and then incubated with recombinant protein. Starch dephosphorylation over time was linear and was measured via release of $^{33}$P at the C6- or C3-position after 5 minutes. Error bars represent the ± SD of six replicates. B. Representation of data from A. as a percentage of C3-dephosphorylation relative to total dephosphorylation.
Figure 5.11 – Substrate specificity in LSF2 active site mutants. Relative specific activity of LSF2 active site mutants at the C6- (blue) and C3-position (yellow) of Arabidopsis starch represented as the percentage of total position dephosphorylation per minute per mg protein. Phosphate-free starch was purified from gwd-deficient Arabidopsis and the starch was pre-labeled with $^{33}$P at either the C6- or C3-position and then incubated with recombinant protein. Starch dephosphorylation over time was linear and was measured via release of $^{33}$P at the C6- or C3-position after 5 minutes. Error bars represent the ± SD of three replicates.
**Figure 5.12 – Laforin substrate specificity.** Relative specific activity of laforin C/S (inactive mutant) and wild type at the C6- (blue) and C3-position (yellow) of *Arabidopsis* starch represented as the percentage of total position dephosphorylation per minute per mg protein. Phosphate-free starch was purified from *gwd*-deficient *Arabidopsis* and the starch was pre-labeled with $^{33}$P at either the C6- or C3-position and then incubated with recombinant protein. Starch dephosphorylation over time was linear and was measured via release of $^{33}$P at the C6- or C3-position after 5 minutes. Error bars represent the ± SD of six replicates.
6. COMPARATIVE PHOSPHATASE ANALYSIS

Introduction

The determination of the mechanism of both SEX4 and LSF2 activity provides an opportunity for comparative structural analyses that will shed further light on glucan phosphatase activity. Plant glucan phosphatases must bind glucan chains in the starch granule, integrate the chain into the catalytic site, and orient the C6- or C3-position of the glucan chain into the catalytic pocket of the DSP domain. This chapter compares and contrasts the differences and similarities regarding how SEX4 and LSF2 interact with phosphoglucan chains to achieve dephosphorylation.

In addition, our structures provide a basis for comparison of the only other two phosphatases that contain a CBM: laforin and LSF1. Laforin is a proven glucan phosphatase and it dephosphorylates soluble glycogen, which is less structurally complex than starch (17, 68). In addition, laforin does not have inherent substrate specificity or glucan-activation, two properties that are present in the plant glucan phosphatases. LSF1 is of interest because it contains the same domains as SEX4, but it has not been found to possess any phosphatase activity (86). Together, laforin and LSF1 provide an opportunity to make lateral, positive, and negative comparisons to SEX4 and LSF2 and make predictions regarding the basis of their function.

Finally, the glucan phosphatases are unique among DSPs in their ability to interact with a glucan substrate, rather than a protein, lipid, or nucleic acid. The structures of SEX4 and LSF2 provide an opportunity to define the structural requirements for glucan phosphatases activity via structural comparison with other non-glucan interacting DSPs. This information will provide further insights into the overall mechanisms of DSP activity and expand our knowledge of this class of enzymes.
Structural comparison of SEX4 and LSF2

The most obvious difference between SEX4 and LSF2 is their method for binding phosphoglucan chains. SEX4 interacts with glucan chains via a continuous CBM-DSP interface, but LSF2 interacts via three distinct glucan-binding sites (Figure 6.1). In SEX4, the CBM is primarily responsible for overall binding and loss or mutation of the CBM nearly abolishes glucan binding. LSF2 does not contain a CBM, and we determined that LSF2 binds starch via the presence of additional ancillary surface-binding sites (SBSs) associated with the CT motif. Mutation of the LSF2 SBSs decreases glucan binding with a decrease of approximately 70%, which is not as drastic as the loss observed in SEX4. These results indicate that the two enzymes depend upon the ancillary interfaces to different degrees and that the active sites have differing contributions to overall binding.

The CT motif appears to be a key variable in understanding the different glucan binding mechanisms employed in SEX4 and LSF2. The CT motif is common to both enzymes and each CT motif shares common secondary structure, but it functions in two different capacities in the two proteins (Figure 6.2). In LSF2, the CT-motif functions directly in glucan binding, with the loop region forming Site-3 and the C-terminus of the CT motif providing scaffolding for Site-2 glucan binding on the DSP. The CT motif of SEX4 lacks both of these glucan interacting sites. The glucan binding Site-3 elements are not conserved in the SEX4 CT motif loop, and the SEX4 CT-motif extends further than the LSF2 CT-motif, blocking access to the Site-2 location present in LSF2. However, we contend that the CT motif in SEX4 does still indirectly contributes to binding by maintaining the CBM-DSP interaction necessary for activity. Thus, the function of the CT motif, in both direct and indirect binding, highlights the differences in binding mechanisms used by SEX4 and LSF2.

There are also differences in the spatial relationship between the catalytic site and the ancillary binding sites in SEX4 and LSF2. The CBM binding interface in SEX4 clearly interacts with the same glucan chain as the catalytic domain. In
contrast, the glucan chains at the LSF2 SBSs are > 20Å away from the active site at their closest points. However, additional analyses suggest that the LSF2 active site and SBSs may be connected, and may engage a single glucan chain. Clues regarding possible connectivity between LSF2 and binding sites are revealed via examining the directionality of bound glucan chains in the LSF2-bound structure. Starch granules are synthesized in a unidirectional pattern, with starch synthase extending the glucan chains at the non-reducing end) (19). This process results in a radially organized starch granule, with the non-reducing end towards the outside and the reducing end towards the interior (9). Our structure reveals that the LSF2 active site and Site-2 are directionally continuous, i.e. the non-reducing ends and reducing ends are pointed in the same direction (Figure 6.3). Therefore, although the short maltohexaose chains do not reveal a continuous binding surface, it suggests that the active site and Site-2 may engage a single glucan chain. If these chains are indeed connected, then the distance between the active site chain and Site-2 chain is approximately 25Å and would accommodate 5-6 more glucose moieties. Thus, LSF2 would interact with a longer overall section of amylopectin than SEX4, which engages only 6 glucose units at a time. In addition, the two-maltohexaose chains at Site-3 are not directionally continuous with the chains at the active site and Site-2. This result suggests that Site-3 may engage a distinct glucan chain. Furthermore, the complex of two chains into a helical unit implies that perhaps Site-3 is structured to bind intact amylopectin helices within the starch microenvironment. Together, these analyses suggest that the LSF2 may interact with starch glucans with more nuance and complexity than SEX4.

The determination of the SEX4 and LSF2 structures reveals that they both contain an aromatic-rich active site pocket that permits glucan integration at the catalytic site (Figure 6.4). This characteristic is unique among DSPs. The active sites of all DSPs consist of subdomains that are tailored to the specific target substrate of the enzyme. Aromatic residues are essential for carbohydrate-interaction in a variety of contexts and different domains. Indeed most classes of
CBMs rely almost exclusively on aromatic residues to engage various carbohydrates (111). Therefore, the plant glucan phosphatases combine a glucan-interaction platform with a DSP catalytic site for glucan dephosphorylation.

SEX4 and LSF2 both contain five aromatic residues at the active site that interact with the glucan chain (Figure 6.5). Despite this commonality, there are differences in the fine topology of the active sites of SEX4 and LSF2. A D-loop phenylalanine (SEX4-F167, LSF2-F162), recognition motif tyrosine (SEX4-Y90, LSF2-Y89), and V-loop tyrosine (SEX4-Y139, LSF2-Y135) are found in both enzymes and function similarly. However, the V-loop of SEX4 and LSF2 contain different aromatic residues (SEX4-F140, LSF2-W136) at the same position. In addition, SEX4 contains a phenylalanine that is not present in the LSF2 R-motif (SEX4-F235) and LSF2 contains an additional tyrosine in its recognition motif (LSF2-Y83). In addition, the active sites contain different hydrophilic residues that contribute hydrogen bonds with the glucan chain. SEX4 contains a lysine (SEX4-K237) in its R-motif near the DSP-CBM interface and LSF2 contains an R-motif glycine (LSF2-G230) that hydrogen bonds the glucan chain. Furthermore, the LSF2 V-loop residue W136 also contributes hydrogen bonding that is not present in SEX4. Although the shape of the active site, defined as the surface formed by the R-motif, PTP-loop, and V-loop, are similar in SEX4 and LSF2, there is one key difference. SEX4 contains a ridge in its R-motif at the position of the β-carbon of F235, whereas LSF2 contains a groove at the same position caused by the lack of a side chain at G230 (Figure 5.9C). Therefore, despite the common aromatic platform, the overall shape and chemical composition of the SEX4 and LSF2 active sites are structurally distinct.

Enzymatic analyses of the SEX4 and LSF2 active sites indicate that the structural differences have consequences regarding their activity. The hydrogen-binding interfaces in LSF2 provided by W136 and G230 create a tighter binding interface than in SEX4, which lacks these features. Mutation of the SEX4 active site resulted in a minimal decrease in amylopectin binding. However, mutation of
the active site in LSF2 resulted in a comparatively considerable decrease in glucan binding and starch dephosphorylation. This evidence indicates that chemical differences between LSF2 and SEX4, i.e. hydrogen-bonding compared to van der Waals, results in the LSF2 active site to contribute more greatly to overall binding compared to the SEX4 active site. In addition, the bound glucan chains follow different paths in the two enzymes (Figure 6.6). In SEX4, the glucan chain is oriented towards the CBM, influenced by R-motif residues F235 and K237. In LSF2 this path is different, and the glucan chain is positioned directly over the R-motif at the position of G230. Moreover, the directionality of the glucan chains at the SEX4 and LSF2 active site are reversed (Figure 6.6). In SEX4, the non-reducing end of the chain is positioned at the V-loop and the reducing end is positioned at the R-motif. In LSF2, the opposite is observed; the non-reducing end is positioned at the R-motif and the reducing end is positioned at the V-loop. Thus, SEX4 and LSF2 may engage glucans from different ends of the chain within starch granules.

In addition, we determined that the SEX4 and LSF2 active sites both bind amylopectin glucan chains in a way conducive to the helical geometry of the substrate. However, we also found that the portions of the glucan chains interacting with the SEX4 or LSF2 DSP or the SEX4 CBM were linearized with respect to their $\alpha$-1,4-glycosidic bond angles. Previous studies have determined that the $\alpha$-1,4-glycosidic dihedral torsion in helical amylopectin chains have values of $(91.8^\circ, -153.2^\circ)$, $(85.7^\circ, -145.3^\circ)$, and $(91.8^\circ, -151.3^\circ)$(133). Inequality of the two angles results in torsion consistent with helical geometry. Conversely, equality in the glycosidic bond torsion angle, i.e. $(119.3^\circ, -119.3^\circ)$, would indicate a lack of torsion and a linear glucan chain. The glycosidic bonds of glucans between the SEX4 CBM and DSP binding sites (Glc3-Glc4 and Glc4-Glc5) have dihedral torsion angles consistent with a helical conformation of $(92.1^\circ, -147.8^\circ)$ and $(94.8^\circ, -143.4^\circ)$, respectively (Figure 6.7). In contrast, the glycosidic bonds of glucans at the DSP active site (Glc1-Glc2 and Glc2-Glc3) and the CBM binding site (Glc5-Glc6) have more linear dihedral torsion angles of $(107.4^\circ, -120.8^\circ)$,
(118.7°, -120.6°), and (108.2°, -129.7°), respectively. The same trend is found in LSF2. The glycosidic bonds of glucans at the DSP active site in LSF2 (Glc1-Glc2 and Glc3-Glc4) have more linear dihedral torsion angles of (93.3°, -142.0°) and (119.2, -119.4°), respectively (Figure 6.8). The glycosidic bonds of glucans outside of the LSF2 active site (Glc4-Glc5 and Glc5-Glc6) contain more helical dihedral torsion angles of (96.2°, -145.2°) and (83.7°, -154.5°), respectively. The only exception to this trend is the torsion angle between Glc2-Glc3, which is in the LSF2 active site but contains a helical conformation (93.3°, -142.0°). The influence of hydrogen bonding on the Glc2 O6 group by G230 appears to influence to adoption of this orientation. Collectively, these data show that LSF2 and SEX4 influence the geometry of helical amylopectin chains at the binding sites that may have consequences regarding the mechanism of these two enzymes.

Perhaps most importantly for the respective biological functions of SEX4 and LSF2, we have shown that differences in the active site compositions influence the orientation of the glucan chains, and consequentially the substrate specificities of each enzyme. The structures clearly show that SEX4 positions the glucan chain in a C6-specific orientation and LSF2 positions it in a C3-specific orientation. We provided evidence that the substrate specificities of each are influenced by elements within the DSP active site. Furthermore, we were able to reverse the site-specificity of SEX4 from the C6- to the C3-position via mutagenesis, but were unable to achieve the same results in LSF2. SEX4 dephosphorylates both the C6- and C3-positions of starch glucans, therefore its activity is inherently malleable. Conversely, LSF2 is strictly a C3-position specific glucan phosphatase. We established that the active site composition is a critical factor in directing substrate specificity, but were unable to address additional variables outside of the active site. Connectivity between the active site and Site-2 may provide an additional layer of complexity to glucan orientation at the active site. Furthermore, we determined that LSF2 possibly interacts with multiple glucan chains via Site-3. These additional glucan interactions may result in
localization of LSF2 within the starch granules that influence its C3-specificity; however, the precise location of C6- and C3-phosphates within the starch superstructure is unknown. Further studies into the microenvironment of phosphorylated starch and the dynamic interaction between LSF2 and glucans of varying complexity will be necessary to fully understand the complete basis of its substrate specificity.

In conclusion, SEX4 and LSF2 contain many similar structural elements, including a CT motif, ancillary binding interfaces, and an aromatic-rich active site. However, both gross and fine structural differences translate to discrepancies in the overall mechanism of glucan binding, the active site contribution to binding, and substrate specificity. The structural similarities and differences fit well within the biological context of glucan phosphatase activity. The similarities represent the fact that SEX4 and LSF2 engage a singular substrate, i.e. starch granules. The differences likely represent an evolutionary divergence allowing SEX4 and LSF2 to separately target GWD and PWD-derived phosphates, respectively. Little is known about the precise positioning of phosphate within the starch granule, but they clearly are located on opposite sides of the glucan chain and possibly are positioned in a distinct microenvironment. Additional information regarding differences in their substrate interaction may be defined by determining the structure of SEX4 with a glucan in a C3-specific orientation, the determination of LSF2 with longer glucan chains, and the determination of both structures with branched glucans bound.

**Comparison with Laforin and LSF1**

The structures of the glycogen phosphatase laforin and the inactive starch-interacting phosphatase LSF1 have not been determined, but the SEX4 and LSF2 structures provide an opportunity to make predictions about the basis of their functions. These insights may provide information helpful for understanding how amino acid substitutions in laforin result in Lafora’s Disease, and to help determine the function of LSF1 within starch metabolism.
Similar to SEX4 and LSF2, LSF1 contains a CT motif. The SEX4 and LSF2 CT motifs are 21% identical based on primary sequence and contain similar secondary structures and overall folds (Table 6.1, Figure 6.2). They are both necessary for protein expression and folding, but have different specific roles in enzymatic activity. The LSF1 CT motif is less similar, being 15% identical in sequence to both SEX4 and LSF2. We previously described that SEX4 does not contain elements of glucan binding present at LSF2 Site-3 and the SEX4 C-terminus extends further than LSF2, blocking access to the location of Site-2 (Figure 6.2). Similarly, LSF1 does not contain the Site-3 residues and its C-terminus extends even further than SEX4 (Figure 6.9). Therefore, our analyses suggest that the LSF1 CT motif functions similarly to SEX4 in integration of the CBM-DSP domains, but does not contain elements of secondary glucan binding found in LSF2.

SEX4, LSF1, and Laforin all contain CBMs, and the functionality of the CBM has been demonstrated in both SEX4 and Laforin as essential for its activity (127). A CBM is a contiguous amino acid sequence with a conserved β-strand rich tertiary fold that possesses carbohydrate-binding ability and is found within carbohydrate modifying or binding proteins (42, 83, 111, 134). CBMs are categorized according to the Carbohydrate Active enZymes (CAZy) database into 68 different families (83, 108). SEX4 and LSF1 are both categorized in the CBM48 family and Laforin is in the CBM20 family. The CBM20 and CBM48 families are closely related based upon conserved binding regions, and intermediates between the CBM48 and CBM20 families that contain conserved elements of both have been identified (47). Moreover, CBM20 and CBM48 domains do not target starch or glycogen exclusively. Both families have been found in starch-targeting enzymes and glycogen-targeting enzymes. For example, the plant enzyme PWD contains a CBM in the CBM20 family although it targets starch, and the human protein AMPKβ2 is in the CBM48 family although it targets glycogen (47). This overlap in target substrate may explain the high level of laforin activity against Arabidopsis starch and the high level of SEX4 activity.
against glycogen-like solubilized amylopectin. Laforin contains an N-terminal CBM20 followed by a DSP, while SEX4 and LSF1 possess a DSP domain followed by a CBM48. Generally speaking, the CBM usually precedes the catalytic domain, as is found in laforin, and the domain order in SEX4 and LSF1 is actually an exception to the common trend (47).

In line with their family designations, the CBM48s of SEX4 and LSF1 are most similar, with 21% identity in their primary sequences (Table 6.1, Figure 6.10). The laforin CBM20 has approximately 15% sequence identity to both SEX4 and LSF1. All three CBMs share the conserved glucan binding residues that form a Trp-Lys-Trp platform for glucan interactions, which defines both the CBM20 and CBM48 families. Our structure showed that SEX4 contains this feature (W278-K307-W314) and that it forms the central scaffold for glucan interaction at the CBM (Figure 3.2). Laforin and LSF1 share these residues suggesting that they both bind glucans via this domain (Figure 6.10).

Furthermore, the SEX4 CBM uses the additional residues H330 and N332 to form a binding complex with the glucan chain (Figure 3.2). Laforin contains the equivalent N332 residue (N115), but LSF1 lacks both of these residues. A common feature of CBM20 family members is the presence of a second glucan binding site (47). This site was first defined in the CBM20 of Aspergillus niger GH15 glucoamylase and is based on a conserved tryptophan residue (W563) (47, 135, 136). SEX4 has a tryptophan at the equivalent position (W293), but this residue did not interact with the glucan chain in our structure (Figure 6.11). Further investigation will be needed to determine the possibility of a second binding site in SEX4. LSF1 lacks this tryptophan residue, but laforin contains a tryptophan at this position (W60). Thus, laforin may contain an additional binding site. Interestingly, laforin has a higher specific activity against amylopectin and starch than SEX4 and a second glucan binding site within laforin could explain this increased activity. However, no one has defined a dissociation constant within the glucan phosphatases to date. Overall, the CBM domains in the glucan
phosphatase family represent an important domain providing scaffolding for interaction with starch and glycogen.

The DSP domain is the one feature common to SEX4, LSF2, LSF1, and laforin. We previously determined that the DSP domain of SEX4 and LSF2 are structured to promote glucan interaction via an active site pocket composed of aromatic and hydrophilic residues. The most similar DSPs, in terms of sequence, are SEX4 and LSF2 with 47% sequence identity (Table 6.1, Figure 6.12). The least similar DSPs are SEX4 and laforin, with 20% sequence identity. Not surprisingly, we find that the plant glucan phosphatases share the most sequence similarity among each other, with their sequences containing at least 35% identity. Interestingly, the DSP domain of laforin has the most similarity to LSF1, with 25% sequence identity. Further studies will be necessary to investigate possible ramifications of this similarity. Together, these analyses indicate that the DSP domains of glucan phosphatases contain a large degree of similarity that likely stems from their common targeted substrate.

We have previously established that the DSP subdomains of SEX4 and LSF2 have been modified to form a glucan-interacting platform. These DSP subdomains (PTP-loop, D-loop, recognition motif, V-loop, and R-motif) are structured to integrate a glucan substrate into the active site and are required for glucan phosphatase activity and specificity. We have identified short regions within these subdomains that are important for SEX4 and LSF2 specificity and activity. These regions should be informative in providing a basis for comparison with laforin and LSF1.

The PTP-loop, containing the HCxxGxxRS/T (Cx5R) catalytic site, is the site of catalysis of all PTP superfamily phosphatases (76). The primary sequence of the PTP-loop often provides information regarding the target substrate being integrated into the catalytic site (73). Therefore, DSPs that bind similar substrates often possess similar PTP-loop sequences. For example, PTPMT1 was identified as a PTEN-like DSP due to their similar PTP-loop sequences: PTPMT1 – HCKAGRSR, PTEN – HCKAGKGR (80). PTEN dephosphorylates the lipid
substrate PI(3,4,5)P$_3$ and, similarly, PTPMT1 dephosphorylates the lipid substrate phosphatidyglycerol phosphate (79, 80). Upon comparison, we found that the PTP-loops of SEX4 (HCTAGMGR), LSF2 (HCSAGLGR), and laforin (HCNAGVGR) all have very similar sequences (Figure 6.13A). Conversely, LSF1 contains a strikingly different PTP-loop sequence (TCTTGFD$R$) from the proven glucan phosphatases (Figure 6.13A). Interestingly, LSF1 lacks the histidine residue N-terminal to the catalytic cysteine that has been predicted to lower the pKa of the sulfhydryl microenvironment and be necessary for catalytic activity (137). This H/T change may explain its inability to dephosphorylate generic pNPP substrates. Moreover, LSF1 also contains a threonine where an alanine is located in SEX4, LSF2, and laforin and a highly dissimilar F-D motif where the other glucan phosphatases contain a short-chain hydrophobic residue and a glycine. In fact, F-D motif is reminiscent to a W-D motif found at the same position in some lipid targeting DSPs such as the myotubularin MTMR2 (HCSDGWDR) (138). These results strongly indicate that LSF1 is not a glucan phosphatase and that the PTP-loop may be suited to target a different substrate, perhaps a lipid, which has not yet been identified.

The additional DSP subdomains, D-loop, recognition motif, V-loop, and R-motif, have also been found to contribute to glucan targeting in SEX4 and LSF2. Both SEX4 and LSF2 both contain an invariant Y-Y-F motif in the recognition motif (SEX4-Y90, LSF2-Y85), V-loop (SEX4-Y139, LSF2-Y135), and D-loop (SEX4-F167, LSF2-F162). These residues form the core of the aromatic pocket in SEX4 and LSF2 (Figure 6.13B,C,D). Interestingly, we find that this core is altered to F-Q-M in laforin, despite its glucan phosphatase activity. This reveals that there are additional ways to interact with glucans at the active site. However, LSF1 contains a Y-N-A motif, losing the aromatic influence at both the V-loop and D-loop. These changes may be responsible for its inability to integrate and dephosphorylate glucan substrates. In addition, we determined that two specific residues in the SEX4 and LSF2 V-loop and R-motif (SEX4-F140/F235, LSF2-W136/G230) influence substrate specificity in these enzymes. LSF1 has a
tryptophan in its V-loop, similar to LSF2, but contains lysine in its R-motif (Figure 6.13D,E). The sequence of laforin tells the opposite story. It contains an aromatic tyrosine in its R-motif, similar to SEX4, but contains an asparagine in its V-loop (Figure 6.13D,E). Unfortunately, these results do not provide enough information to make predictions regarding the substrate specificity. In summary, these results indicate that the lack of glucan phosphatase activity in LSF1 may result from an absence in conserved aromatic residues in its DSP subdomains. However, they also indicate that laforin may have a unique DSP architecture that is nonetheless capable of interacting with glucan substrates.

In conclusion, the glucan phosphatase family is a heterogeneous group of enzymes that shares common glucan substrate and glucan binding platforms but different specific activities. The CT domain is a common feature of the plant glucan phosphatases, but it is only shown to participate directly in glucan binding in LSF2. Furthermore, the CBM domains of SEX4, LSF1, and laforin appear to be very similar, but the precise interdomain interactions that contribute to LSF1 and laforin function are unknown. The DSP domains of the glucan phosphatases are highly similar, but the structural insights derived from SEX4 and LSF2 indicate that LSF1 inactivity may stem from a breakdown in the core aromatic motif and laforin may integrate glucans into the catalytic site via a distinct mechanism. These analyses make it clear that there exists notable variability among the glucan phosphatases despite their common target substrate.

Structural comparison among glucan phosphatase and other DSPs

The glucan phosphatases are unique among DSPs in their ability to interact with a glucan substrate, rather than a protein, lipid, or nucleic acid. The DSP subdomains of SEX4 and LSF2 have been adapted to provide a glucan-binding platform at the active site to target and dephosphorylate glucan chains. This section contains a comparative structural analysis between the glucan phosphatases and other DSPs whose structures have been determined in an effort to define the glucan phosphatases within the heterogeneous DSP family.
Both SEX4 and LSF2 contain a typical $\alpha_\beta\alpha$ DSP fold, which is highly conserved in DSPs overall despite a low level of sequence identity between enzymes (72, 76). Using the DALI server (94), we determined that, after LSF2, the SEX4 DSP is most structurally similar to *Sulfolobus solfataricus*-PTP (Ss-PTP, 2I6O) (139), STYX (2R0B) (140), KAP (1FPZ) (141), VHZ (2IMG) (142), and DUSP27 (2Y96) (143) (Figure 6.14). We also determined that, after SEX4, LSF2 is most structurally similar to PTPMT1 (3RGQ) (80), VHZ, Ss-PTP, KAP, and STYX (Figure 6.15). The average root mean squared deviation between the glucan phosphatases and these other DSPs is approximately 2.5Å, despite sequence identities ranging from 10-18%.

Despite their structural similarity with the glucan phosphatases, the closest structural homologs of SEX4 and LSF2 cover a diverse array of target substrates. PTPMT1 is a lipid phosphatase that dephosphorylates phosphatidyglycerol phosphate (80), VHZ and DUSP27 are true DSPs that dephosphorylate both p-Tyr and p-Ser/Thr residues on proteins (142, 143), Ss-PTP is a p-Tyr phosphatase (139), KAP is a p-Thr phosphatase (141), and STYX is a pseudophosphatase that targets proteins but does not dephosphorylate any substrate (144). A common feature of these enzymes, with the exception of VHZ and DUSP27, is the presence of an $\alpha$-helical V-loop (Figure 6.16). An $\alpha$-helical V-loop is only found in these enzymes as the V-loop is normally a loop, as the name suggests (76). Although VHZ does not contain an $\alpha$-helical V-loop, it does contain a phenylalanine in its D-loop that is unique to SEX4 and LSF2 but is not found in other DSPs (142). We determined that this residue is crucial for glucan phosphatase activity, but its role in VHZ is unclear. Additional analysis shows that the VHZ phenylalanine is isolated, and not surrounded by a network of additional aromatic residues as is found in SEX4 and LSF2 (Figure 6.17). This suggests that, despite is peculiarity, the D-loop phenylalanine in VHZ is unlikely to provide a platform for glucan interactions. Collectively, these analyses indicate that the overall DSP fold of SEX4 and LSF2 shares close structural homology with a handful of DSPs, but the range of their target substrates demonstrates that
these gross structural similarities are not highly indicative of substrate preferences.

We have extensively discussed the role of the aromatic-rich DSP subdomains in SEX4 and LSF2 and their critical role in glucan interaction and specificity. An analysis of the sequences of 34 additional DSPs reveals that the important aromatic residues found in the glucan phosphatases are not highly conserved in other DSPs (Figure 6.18). This finding strongly establishes the aromatic active site pocket as a defining feature of the glucan phosphatases and is likely to be present in any additional glucan phosphatases discovered in the future.

Another important feature, explored earlier with SEX4 and LSF2, is the topology, or shape, of the active sites of DSPs. It has previously been established that the active site topology of PTPs and DSPs are tailored to their respective target substrates in terms of width and depth (77). PTPs, such as PTP1B generally contain a narrow and deep active site that permits access to target p-Tyr residues, but excludes shorter p-Thr/Ser residues from gaining access to the catalytic site (Figure 6.19) (145). In contrast, DSPs such as VHR generally have a shallower and wider active site that permits access of p-Tyr and p-Thr/Ser residues to the catalytic site (Figure 6.19) (73). In addition, it has been found that lipid phosphatases such as PTEN, which engages a large tri-phosphorylated phosphatidylinositol head group, contains an even wider and shallower active site than the proteinaceous phosphatases (Figure 6.19) (77). We found that the active sites of SEX4 and LSF2 are wide and shallow, similar to PTEN, which allows them to engage three glucose moieties of a long glucan chain at the catalytic site (Figure 6.19, Figure 6.20). The active site of SEX4 is extremely similar to PTEN, and LSF2 is even more shallow and wide (Figure 6.20). These results indicate that the non-proteinaceous DSPs typically contain an active site that can engage an extended substrate, rather than target individual protein side chains. Furthermore, the glucan phosphatases must be able to engage three glucose units at a time and presumably must be able to
cycle glucose moieties through the active site to identify a phosphate group. The active site topology of SEX4 and LSF2 is consistent with these enzymatic requirements.

Lastly, we found that SEX4 depends upon a CBM to engage its target starch substrate and contains extensive interdomain interaction between the CBM and DSP. Additional domains that help guide the catalytic domain to its target substrate are found in many other DSPs (75). These ancillary domains include slingshot and PDZ domains involved in protein-protein interaction, and PH-GRAM and C2 domains involved in interaction with lipid membranes (75).

There are structures for three DSPs with ancillary C2 domains: PTEN (1D5R) (77), Ciona intestinalis voltage sensing phosphatase (Ci-VSP, 3AWE) (146), and Auxilin PTEN-like protein (3N0A) (147). Moreover, there is a structure for a DSP containing an ancillary PH-GRAM domain: MTMR2 (1M7R) (148). All four of the ancillary domains are folding into roughly the same position with respect to the active site in the DSP, in between the D-loop and R-motif (Figure 6.21). This position is roughly identical to the CBM in the SEX4. This structural consistency indicates that the CBM-DSP interface in SEX4 follows an energetically favorable folding position that mimics the arrangement in these enzymes. Furthermore, this result also indicates that the position of the ancillary domain with respect to the DSP is not dictated by the order of the domains. SEX4, PTEN, Ci-VSP, and Auxilin PTEN-like proteins all contain ancillary domains C-terminal to the DSP. However, MTMR2 contains an ancillary domain N-terminal to the DSP. This suggests that the laforin CBM may also be positioned in the same location as SEX4, with respect to the active site, despite the difference in domain order. These results indicate that the CBM position of SEX4 represent the norm among DSPs.

In conclusion, SEX4 and LSF2 contain unique features that distinguish them from other DSPs. The glucan phosphatases are most unique in that their DSP subdomains contain a large amount of aromatic residues that form a glucan-interacting platform not seen in other DSPs. They both contain a
characteristic DSP fold that contains the rare elements of an $\alpha$-helical V-loop and a phenylalanine-containing D-loop. In addition, SEX4 and LSF2 contain a shallow and wide active site topology that is similar to the non-proteinaceous lipid phosphatases and distinct from the protein phosphatase active sites in general. Lastly, we found that the domain arrangement of the SEX4 CBM-DSP interface is consistent with the location of ancillary domains in lipid phosphatases. Collectively, these results indicate that the glucan phosphatases contain subtle, yet unique elements that set them apart from other DSPs.
### Table 6.1 – Sequence identity / similarity in glucan phosphatase family

<table>
<thead>
<tr>
<th></th>
<th>IDENTITY (%)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>SEX4</strong></td>
<td><strong>LSF2</strong></td>
<td><strong>LSF1</strong></td>
<td><strong>Laforin</strong></td>
</tr>
<tr>
<td>Full Length</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEX4</td>
<td>24.4</td>
<td></td>
<td>15.7</td>
<td>10.4</td>
</tr>
<tr>
<td>LSF2</td>
<td>36.2</td>
<td></td>
<td>14.6</td>
<td>13.9</td>
</tr>
<tr>
<td>LSF1</td>
<td>29.0</td>
<td></td>
<td>23.4</td>
<td>10.6</td>
</tr>
<tr>
<td>Laforin</td>
<td>20.8</td>
<td></td>
<td>28.4</td>
<td>19.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SIMILARITY (%)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>DSP</strong></td>
<td><strong>SEX4</strong></td>
<td><strong>LSF2</strong></td>
<td><strong>LSF1</strong></td>
</tr>
<tr>
<td>SEX4</td>
<td>47.0</td>
<td></td>
<td>34.7</td>
<td>20.2</td>
</tr>
<tr>
<td>LSF2</td>
<td>63.4</td>
<td></td>
<td>34.7</td>
<td>25.0</td>
</tr>
<tr>
<td>LSF1</td>
<td>53.3</td>
<td></td>
<td>53.3</td>
<td>25.0</td>
</tr>
<tr>
<td>Laforin</td>
<td>37.1</td>
<td></td>
<td>42.4</td>
<td>41.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>IDENTITY (%)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>CBM</strong></td>
<td><strong>SEX4</strong></td>
<td><strong>LSF1</strong></td>
<td><strong>Laforin</strong></td>
</tr>
<tr>
<td>SEX4</td>
<td>21.4</td>
<td></td>
<td>11.2</td>
<td></td>
</tr>
<tr>
<td>LSF1</td>
<td>34.7</td>
<td></td>
<td>13.8</td>
<td></td>
</tr>
<tr>
<td>Laforin</td>
<td>21.6</td>
<td></td>
<td>25.4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SIMILARITY (%)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>CT</strong></td>
<td><strong>SEX4</strong></td>
<td><strong>LSF2</strong></td>
<td><strong>LSF1</strong></td>
</tr>
<tr>
<td>SEX4</td>
<td>20.9</td>
<td></td>
<td>10.3</td>
<td></td>
</tr>
<tr>
<td>LSF2</td>
<td>27.9</td>
<td></td>
<td>6.9</td>
<td></td>
</tr>
<tr>
<td>LSF1</td>
<td>15.5</td>
<td></td>
<td>22.4</td>
<td></td>
</tr>
</tbody>
</table>
Figure 6.1 – Structural comparison between SEX4 and LSF2 glucan-binding methods. Side-by-side surfaces of SEX4 and LSF2 structures showing the DSP (blue) domains with the active site (red), CBM (pink), and Ct motif (tan). Glucan chains at the SEX4 and LSF2 active sites are shown in green, and glucan chains at the LSF2 secondary binding sites are shown in cyan, pink, and orange.
Figure 6.2 – CT motif in LSF2 and SEX4. A. Sequence alignment of the LSF2 and SEX4 CT motifs. Residues involved in glucan binding in LSF2 are highlighted in red. α-helices in LSF2 (yellow) and SEX4 (blue) are shown above. B. CT motif (tan) of LSF2 showing position of secondary binding Site-2 and Site-3. C. CT motif (tan) of SEX4. Extension of the C-terminus past the end point in LSF2 is noted in dotted lines.
Figure 6.3 – Continuation of directionality in LSF2 active site and Site-2 glucans. The structure of LSF2 is shown with the active site and Site-glucans in green. The directionality of the glucan chains is depicted in terms of stick width. The non-reducing end of the chain is thickest, continuing towards the thinner reducing end of each chain.
Figure 6.4 – Aromatic pocket in the SEX4 and LSF2 DSPs. A. Maltoheptaose chain (green) at the SEX4 DSP (blue). Aromatic pocket residues are shown in yellow. B. Maltohexaose chain (green) at the LSF2 DSP (blue). Aromatic pocket residues are shown in yellow.
Figure 6.5 – Aromatic pocket residues and interactions. A. Aromatic pocket residues (yellow) at the SEX4 DSP (blue) interacting with the maltoheptaose chain (green). Hydrogen bonding interaction from K237 is shown with a dotted line. B. Aromatic pocket residues (yellow) at the LSF2 DSP (blue) interacting with the maltohexaose chain (green). Hydrogen bonding interactions from W136 and G230 are shown with dotted lines.
Figure 6.6 – Directionality of glucan chains at the SEX4 and LSF2 active sites. **A.** SEX4 active site showing the maltoheptaose chain (green). Directionality is depicted via thickness of the glucose moieties, with the non-reducing end (NR) thickest and the reducing end (R) thinnest. R-motif and V-loop are labeled and colored blue. **B.** LSF2 active site showing the maltohexaose chain (green). Directionality is depicted via thickness of the glucose moieties, with the non-reducing end (NR) thickest and the reducing end (R) thinnest. R-motif and V-loop are labeled and colored blue.
Figure 6.7 – Dihedral Torsion angles about the glycosidic bonds in the SEX4-bound glucan. Helical amylopectin chains contain dihedral torsion angles about the α-1,4-glycosidic bonds of (91.8°, -153.2°), (85.7°, -145.3°), and (91.8°, -151.3°) (133). A linear glucan chain would contain torsion angles of (119.3°, -119.3°). Dihedral torsion angles about the α-1,4-glycosidic bonds in the SEX4-bound glucan (green) are indicated for A. Glc5-Glc6, B. Glc4-Glc5, C. Glc3-Glc4, D. Glc2-Glc3, and E. Glc1-Glc2. Core residues of the CBM and DSP binding sites are depicted in yellow.
Figure 6.8 – Dihedral Torsion angles about the glycosidic bonds in the LSF2 active site-bound glucan. Helical amylopectin chains contain dihedral torsion angles about the $\alpha$-1,4-glycosidic bonds of ($91.8^\circ$, $-153.2^\circ$), ($85.7^\circ$, $-145.3^\circ$), and ($91.8^\circ$, $-151.3^\circ$)\(^\text{(133)}\). A linear glucan chain would contain torsion angles of ($119.3^\circ$, $-119.3^\circ$). Dihedral torsion angles about the $\alpha$-1,4-glycosidic bonds in the LSF2 active site-bound glucan (green) are indicated for A. Glc1-Glc2, B. Glc2-Glc3, C. Glc3-Glc4, D. Glc4-Glc5, and E. Glc5-Glc6. Core residues of the DSP binding site are depicted in yellow.
Figure 6.9 – Alignment of LSF2, SEX4, and LSF1 CT motifs. ClustalW alignment of the CT motifs of LSF2, SEX4, and LSF1. Residues involved in Site-3 glucan binding in LSF2 are highlighted in red. Secondary α-helical structure of LSF2 (yellow) and SEX4 (blue) are depicted above.
Figure 6.10 – Sequence alignment of SEX4, LSF1, and laforin CBMs.

ClustalW alignment of the CBM48s from SEX4 and LSF1 and CBM20 from laforin. Conserved main glucan binding site (W-K-W) residues are highlighted in red. Second binding site residues, found in *Aspergillus niger* glucoamylase (136), are highlighted in blue. Additional glucan interacting residues from the SEX4 structure are highlighted in green. Secondary $\beta$-strands from the SEX4 structure are depicted above the sequence.
Figure 6.11 – Glucan binding sites in glucoamylase and SEX4 CBMs. A. CBM20 domain from *Aspergillus niger* glucoamylase (1AC0) (136). Two distinct glucan binding sites are depicted. B. CBM48 domain from SEX4. Only one glucan binding site is present in the structure, despite the conservation of the tryptophan from binding site 2.
Figure 6.12 – Sequence alignment of SEX4, LSF2, LSF1, and laforin DSPs.

ClustalW alignment of the DSP domains from SEX4, LSF2, LSF1, and laforin. Residues involved in glucan binding in SEX4 and LSF2 are highlighted in red. DSP subdomains are noted above the sequence.
Figure 6.13 – Sequence alignment of glucan phosphatase DSP subdomains. ClustalW alignment of the DSP subdomains **A.** PTP-loop, **B.** D-loop, **C.** recognition motif, **D.** V-loop, and **E.** R-motif from SEX4, LSF2, LSF1, and laforin.
Figure 6.14 – Structural alignment of SEX4 with 5 closest structural homologs. Structure of the SEX4 DSP (blue) aligned at the catalytic PTP-loop (red) with A. *Sulfolobus solfataricus*-PTP (SS-PTP, 216O) (139), B. STYX (2R0B) (140), C. KAP (1FPZ) (141), D. VHZ (2IMG) (142), and E. DUSP27 (2Y96) (143). The CBM and CT motif of SEX4 are colored gray.
Figure 6.15 – Structural alignment of LSF2 with 5 closest structural homologs. Structure of the SEX4 DSP (blue) aligned at the catalytic PTP-loop (red) with A. PTPMT1 (3RGQ) (80), B. VHZ (2IMG) (142), C. *Sulfolobus solfataricus*-PTP (SS-PTP, 216O) (139), D. KAP (1FPZ) (141), and E. STYX (2R0B) (140). The CT motif of LSF2 is colored gray.
Figure 6.16 – Variable loop (V-loop) in the glucan phosphatases and other DSPs. Structural alignment of SEX4 and LSF2 DSPs (blue) with KAP (orange, 1FPZ) (141), and the prototypical DSP VHR (pink, 1VHR) (73). Inset shows the variable loop (V-loop). SEX4 and LSF2 contain an α-helical V-loop. This is rare in DSPs found only in KAP, PTPMT1, SS-PTP, and STYX. Most DSPs contain a V-loop similar to VHR, depicted in the figure.
Figure 6.17 – D-loop substrate-interacting residues in DSPs. Structural alignment of LSF2 (light blue), SEX4 (darker blue), VH\(\text{Z}\) (green, 2IMG) (142), and VHR (pink, 1VHR) (73). The PTP-loop catalytic site is shown in red. The position of LSF2, SEX4, and VH\(\text{Z}\) phenylalanine residues in the D-loop is shown. A D-loop phenylalanine is rare in DSPs; normally they contain a short-chain hydrophilic residue at this position, as in the prototypical DSP VHR.
### Figure 6.18 (AB) – Alignment of glucan phosphatase subdomains with DSPs.

<table>
<thead>
<tr>
<th>A</th>
<th>PTP-loop</th>
<th>B</th>
<th>D-loop</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEX4-DSP</td>
<td>HCTAGMGR</td>
<td>SEX4-DSP</td>
<td>ETKDID</td>
</tr>
<tr>
<td>LSF2-DSP</td>
<td>HCSAGLGR</td>
<td>LSF2-DSP</td>
<td>PAKDRD</td>
</tr>
<tr>
<td>LSF1-DSP</td>
<td>TCTTGR</td>
<td>LSF1-DSP</td>
<td>PKDAD</td>
</tr>
<tr>
<td>Laforin-DSP</td>
<td>HCNAGVR</td>
<td>Laforin-DSP</td>
<td>PTTPDM</td>
</tr>
<tr>
<td>DUSP1-MKP1-hVH1-DSP</td>
<td>HCOAGISR</td>
<td>DUSP1-MKP1-hVH1-DSP</td>
<td>PVEDNH</td>
</tr>
<tr>
<td>DUSP2-PAC1-DSP</td>
<td>HCOAGISR</td>
<td>DUSP2-PAC1-DSP</td>
<td>PVEDNQ</td>
</tr>
<tr>
<td>hVHR-DSP</td>
<td>HEAGISR</td>
<td>hVHR-DSP</td>
<td>KANDTQ</td>
</tr>
<tr>
<td>DUSP4-MKP2-DSP</td>
<td>HCOAGISR</td>
<td>DUSP4-MKP2-DSP</td>
<td>PVEDNH</td>
</tr>
<tr>
<td>DUSP5-VH3-DSP</td>
<td>HCEAGISR</td>
<td>DUSP5-VH3-DSP</td>
<td>PVEDSH</td>
</tr>
<tr>
<td>DUSP6-MKP3-DSP</td>
<td>HCLAGISR</td>
<td>DUSP6-MKP3-DSP</td>
<td>PJSBHW</td>
</tr>
<tr>
<td>DUSP7-PYST2-DSP</td>
<td>HCLAGISR</td>
<td>DUSP7-PYST2-DSP</td>
<td>PJSBHW</td>
</tr>
<tr>
<td>DUSP8-hVH5</td>
<td>HCLAGISR</td>
<td>DUSP8-hVH5</td>
<td>PIGNDN</td>
</tr>
<tr>
<td>DUSP9-MKP4-DSP</td>
<td>HCLAGVR</td>
<td>DUSP9-MKP4-DSP</td>
<td>PIJSBH</td>
</tr>
<tr>
<td>DUSP10-MKP5-DSP</td>
<td>HCOAGVR</td>
<td>DUSP10-MKP5-DSP</td>
<td>PAPDNS</td>
</tr>
<tr>
<td>DUSP11-PIR1</td>
<td>HCTHGLNR</td>
<td>DUSP11-PIR1</td>
<td>VGHQVP</td>
</tr>
<tr>
<td>DUSP12-hVH1-DSP</td>
<td>HCHAGVS</td>
<td>DUSP12-hVH1-DSP</td>
<td>PALBK</td>
</tr>
<tr>
<td>DUSP13-DSP</td>
<td>HCAAGVS</td>
<td>DUSP13-DSP</td>
<td>EADNP</td>
</tr>
<tr>
<td>DUSP14-MKP6-DSP</td>
<td>HCAAGVS</td>
<td>DUSP14-MKP6-DSP</td>
<td>PLADM</td>
</tr>
<tr>
<td>DUSP15-VHY</td>
<td>HCFAGIS</td>
<td>DUSP15-VHY</td>
<td>PVDTP</td>
</tr>
<tr>
<td>DUSP16-MKP7</td>
<td>HCLAGISR</td>
<td>DUSP16-MKP7</td>
<td>PVDNSF</td>
</tr>
<tr>
<td>DUSP18-LMW DUSP20</td>
<td>HCAAGVS</td>
<td>DUSP18-LMW DUSP20</td>
<td>PVDSP</td>
</tr>
<tr>
<td>DUSP19-SKR1</td>
<td>HCNAGVR</td>
<td>DUSP19-SKR1</td>
<td>SRLDP</td>
</tr>
<tr>
<td>DUSP21-LMW DUSP21</td>
<td>HCMAGSR</td>
<td>DUSP21-LMW DUSP21</td>
<td>PVDAR</td>
</tr>
<tr>
<td>DUSP22-JSP1</td>
<td>HCLAGSR</td>
<td>DUSP22-JSP1</td>
<td>PADDSP</td>
</tr>
<tr>
<td>VH2-DSP</td>
<td>HCALLGDR</td>
<td>VH2-DSP</td>
<td>RJDGC</td>
</tr>
<tr>
<td>STYX-DSP</td>
<td>FSTAGIS</td>
<td>STYX-DSP</td>
<td>RVEDS</td>
</tr>
<tr>
<td>DUSP26-MKP8</td>
<td>HCAAGVSR</td>
<td>DUSP26-MKP8</td>
<td>EAHDSP</td>
</tr>
<tr>
<td>DUSP27-DSP</td>
<td>HCVGR</td>
<td>DUSP27-DSP</td>
<td>EADLP</td>
</tr>
<tr>
<td>TPTE2-TPIP-DSP</td>
<td>HKGGKGR</td>
<td>TPTE2-TPIP-DSP</td>
<td>MJDDHH</td>
</tr>
<tr>
<td>TPTE-Pten2-DSP</td>
<td>HCKGTD</td>
<td>TPTE-Pten2-DSP</td>
<td>MJDDHH</td>
</tr>
<tr>
<td>SSH3-slingshot-DSP</td>
<td>HKMGR</td>
<td>SSH3-slingshot-DSP</td>
<td>RLDDE</td>
</tr>
<tr>
<td>SSH1-slingshot1-DSP</td>
<td>HKMGR</td>
<td>SSH1-slingshot1-DSP</td>
<td>RVDDE</td>
</tr>
<tr>
<td>SS-TPD</td>
<td>HVGGIG</td>
<td>SS-TPD</td>
<td>PJDPG</td>
</tr>
<tr>
<td>PTPMT1-DSP</td>
<td>HKAGR</td>
<td>PTPMT1-DSP</td>
<td>STVDM</td>
</tr>
<tr>
<td>PTP4A1-PRL1-DSP</td>
<td>HCVAGLR</td>
<td>PTP4A1-PRL1-DSP</td>
<td>PFDAG</td>
</tr>
<tr>
<td>PTEN-DSP</td>
<td>HKAGGR</td>
<td>PTEN-DSP</td>
<td>PFEHNN</td>
</tr>
<tr>
<td>KAP3-DSP</td>
<td>HCYGGLR</td>
<td>KAP3-DSP</td>
<td>PIADG</td>
</tr>
<tr>
<td>CDC14-DSP</td>
<td>HKAGLGR</td>
<td>CDC14-DSP</td>
<td>EHYDLF</td>
</tr>
</tbody>
</table>
Figure 6.18 (C) – Alignment of glucan phosphatase subdomains with DSPs.
Figure 6.18 (D) – Alignment of glucan phosphatase subdomains with DSPs.
ClustalW alignment of SEX4, LSF2, LSF1, and laforin with 34 additional DSP enzymes. Alignments of the A. PTP-loop, B. D-loop, C. recognition motif, D. V-loop, and E. R-motif are shown. Catalytic residues conserved in active DSPs are highlighted in blue. Residues identified to be important for DSP-glucan interaction in SEX4 and LSF2 are highlighted in red.
Figure 6.19 – Active site topology of DSPs with different target substrates.

Cross sections of the active sites of five different DSPs are shown. The active site shapes are formed, from left to right, by the R-motif, PTP-loop, and V-loop subdomains. The trough at the center of each panel (arrow) is the PTP-loop catalytic site. PTP1B (orange, 1EEN) is a p-Tyr specific DSP and has a deep, narrow active site to exclude p-Ser/Thr residues (149). VHR (green, 1VHR) is a prototypical DSP targeting p-Tyr and p-Ser/Thr residues and has a shallow wide active site to accommodate all three residues (73). PTEN (blue, 1D5R) is a lipid phosphatase that dephosphorylates the 3-position of PI(3,4,5)P₃ and has a deep and wide active site to accommodate inositol head groups (77). SEX4 (red) and LSF2 (purple) have similar active site shapes that are shallow and wide to accommodate three glucose moieties of a polyglucan chain. A notable difference between the two is the ridge in the SEX4 R-motif where a pocket is located in LSF2.
Figure 6.20 – Overlayed Active site topologies of DSPs with different target substrates. Cross sections of the active sites of five different DSPs as shown in Figure 6.19 are overlayed for direct comparison of active site topologies. The following DSPs are shown: PTP1B (orange, 1EEN, p-Tyr specific) (149), VHR (green, 1VHR, p-Tyr/Ser/Thr specific) (73), PTEN (blue, 1D5R, PI(3,4,5)P₃ specific) (77), SEX4 (red, polyglucan specific), and LSF2 (purple polyglucan specific).
Figure 6.21 – Position of ancillary domains in DSPs with respect to the DSP domain. Structural alignment of SEX4 with four additional DSPs to identify the position of ancillary domains with respect to the DSP domain. DSP is colored gray with the PTP-loop active site in red for orientation. Ancillary domains are colored. The CBM of SEX4 is colored pink. C2 domain-containing DSPs A. PTEN (cyan, 1D5R) (77), B. Ciona intestinalis voltage-sensing phosphatase (Ci-VSP, orange, 3AWE) (146), and C. Auxilin PTEN-like protein (yellow, 3N0A) (147) are shown, aligned with SEX4. The PH-GRAM domain-containing protein MTMR2 (dark pink, 1M7R) (148) is also shown, aligned with SEX4.
7. SUMMARY AND FUTURE DIRECTIONS

Summary

Over the past decade, reversible starch phosphorylation has emerged as the main regulatory mechanism governing the transition from starch synthesis to breakdown (9). The necessity for starch phosphorylation and dephosphorylation arises from biological paradoxes inherent to starch metabolism. Starch is synthesized as a water-insoluble, crystalline structure to maximize glucose storage and to discourage breakdown during photosynthetic periods (8). However, these features prevent its catabolism during non-photosynthetic periods. Phosphorylation of starch glucose by glucan dikinases regulates the solubility of the outer starch granules and permits access to glycolytic enzymes during the night (40). However, an additional issue arises from the introduction of phosphate. β-amylase activity is impeded when it reaches a phosphate group on a glucan chain, therefore efficient catabolism cannot occur in phosphorylated starch (71). Glucan phosphatases are necessary to clear phosphate from the starch granule and permit complete breakdown of the starch granule (69). Evidence suggests that this process is dynamic, with glucan dikinases, glucan hydrolases, and glucan phosphatases all working in concert to efficiently degrade starch granules and release stored energy during non-photosynthetic periods (9).

Previously, the structural basis of interaction between reversible starch phosphorylation enzymes and phosphoglucan substrates was entirely unknown. Here, we have provided a detailed examination of the interaction between glucan phosphatases SEX4 and LSF2 and phosphoglucan ligands. This information provides insights into the mechanism of activity of SEX4 and LSF2, including the basis for their respective C6- and C3-position substrate specificities. This information bridges a significant gap whereby the structural basis of glucan phosphatase activity, unknown to science before 2006, has been elucidated with extensive detail.
It was previously known that SEX4 preferentially dephosphorylates the C6- and C3-position of starch granules, however, the basis of SEX4 phosphoglucomutase interaction and substrate specificity was previously unknown. We determined that SEX4 binds glucans via a continuous interdomain CBM-DSP pocket, maintained in part by its previously unrecognized CT motif. Furthermore, we determined that the CBM is essential for activity against insoluble substrate and the DSP is essential for positioning the glucan chain into a C6-specific orientation at the active site. These results provide an updated model for SEX4 activity, whereby SEX4 uses a concerted interdomain binding/dephosphorylation platform to target C6-position phosphates on unwound starch glucan chains (Figure 7.1).

LSF2 is a recently discovered enzyme and its ability to dephosphorylate starch glucans without a CBM ran contrary to the predicted requirements for a glucan phosphatase. We determined that LSF2 binds starch via secondary binding sites that are essential for its activity against starch granules. In addition, the basis of LSF2 C3-specificity was entirely unclear. We determined that hydrogen bonding at the LSF2 active site serves to orient the glucan chain in a C3-specific orientation. Together, these results provide a model for LSF2 activity, whereby LSF2 uses multiple binding sites to interact with glucan chains and target C3-position phosphates within the starch granule (Figure 7.2).

In addition, we have provided extensive structural and functional comparative analyses between SEX4 and LSF2, related glucan phosphatases, and DSP family members that dephosphorylate non-glucan substrates. Insights from these analyses have provided us with the ability to identify structural features common to glucan phosphatases. We determined that glucan phosphatases contain a unique network of aromatic residues that form a glucan-interacting platform at the DSP active site. Furthermore, we determined that elements within the DSP subdomains surround the active site direct substrate specificity. Moreover, we determined that the active site topology of the glucan phosphatases are structured to interact with three glucan moieties of a longer
glucan chain. Cumulatively, these results clearly establish the overall mechanism of glucan phosphatase activity.

**Future Directions**

Our structure clearly establishes the basis for interaction between SEX4 and the maltoheptaose glucan chain. However, the structure failed to illustrate how SEX4 would interact with branched glucan chains present in unwound amylopectin helices. A close examination of the glucan chain indicates that an $\alpha$-1,6-branch could be accommodated at the Glc5 or Glc6 position at the CBM. A branch point at this position would clash with the D-loop aromatic DSP residue F167 (Figure 7.3). This aromatic residue may be positioned to interact specifically with branches. Furthermore, our structure shows that a phosphate group could easily be integrated into the active site from the non-reducing end of the chain, but integration from the reducing end would result in a steric clash (Figure 7.3). This may have ramifications on the possibility of processivity in SEX4, indicating that SEX4 would have to engage phosphate groups from the non-reducing end. Determining the structure of SEX4 with branched glucans would provide additional information into the interaction between SEX4 and $\alpha$-1,6-branches. Mutagenesis of branch-interacting residues may alter the activity of SEX4 against starch granules or affect the phosphorylation pattern of granules incubated with SEX4.

We have provided details regarding the structural basis of SEX4 C6-specific activity. However, SEX4 is able to dephosphorylate C3-position phosphates and the mechanism of C3-position phosphate integration into the active site is unknown. Presumably, the glucan chain would require a different geometry to present C3-position phosphates into the active site that is not revealed by our structure. The determination of the structure of our F140W/F235G SEX4 double mutant, which targets C3-position phosphates, in complex with a phosphoglucan ligand may provide additional insights into how a C3-postiion specific phosphate could be dephosphorylated by SEX4.
Considering the extent to which the CBM and DSP of SEX4 cooperatively engage a glucan chain, it will be interesting to determine if a similar level of cooperativity exists between the glucan binding sites in LSF2. We determined that the glucan chain at the LSF2 active site and Site-2 could feasibly represent a single extended glucan chain based on their relative proximity and directional continuity. The determination of the LSF2 with a longer glucan chain (15-20 glucose units) may indicate that these two sites cooperatively engage a single glucan chain. Furthermore, Hydrogen/Deuterium eXchange Mass Spectrometry (H/DXMS) may be used to uncover glucan-interacting surfaces between the LSF2 binding sites. In addition, it will be interesting to determine if the helical complex of glucan chains at Site-3 is representative of the LSF2 starch-interacting mechanism or if it is an artifact of crystallography. If LSF2 Site-3 is structured to bind helical glucan substrates it provides a few possible models of LSF2 interaction with the complex starch superstructure (Figure 7.4). Moreover, we presented evidence that the C3-position specificity of LSF2 may be influenced by variables outside of the active site. Further analysis of specificity in LSF2 should incorporate mutagenesis of secondary binding sites to unravel the effect of these sites on specificity. The overall mechanism of LSF2 interacting with starch granules is a clear future direction and much is still unclear concerning the ramifications of its complex binding mechanism.

Although the work in this dissertation provides extensive details regarding plant glucan phosphatase activity, the structural basis of laforin activity is still unknown. The current study strongly suggests that laforin incorporates some of the same structural elements as SEX4 and LSF2, particularly an aromatic platform surrounding the active site and coordination between the CBM and DSP. However, similarities between the plant glucan phosphatases and laforin will require a structure of laforin. A catalogue of Lafora’s Disease patient mutations have been compiled over the years and for many of these mutations the precise cause of breakdown of laforin function is unknown (65). A structure of laforin,
coupled with insights from SEX4 and LSF2, will provide a clearer understanding of the link between patient mutations and laforin function within the cell.

Although LSF1 has been linked to starch metabolism, the precise function of LSF1 and its elusive target substrate is entirely unknown (86). We have provided some evidence that LSF1 may be a lipid phosphatase, therefore a comprehensive study of LSF1 activity and binding to a suite of phospholipid substrates should be conducted. In addition, LSF1 is a perfect candidate for determining the minimum structural requirements for glucan phosphatase activity. LSF1 contains a DSP, CBM, and CT motif, fulfilling all the domain requirements for glucan phosphatase activity. Therefore, we hypothesize that its inability to dephosphorylate glucans is derived from a loss of critical residues in its DSP subdomains. Site-directed mutagenesis swaps of LSF1 to SEX4 (and SEX4 to LSF1) DSP subdomains may provide evidence for the structural basis of LSF1 inactivity against glucan substrates and also reiterate the minimum structural requirements for glucan phosphatase activity (Figure 7.5). Furthermore, the inactivity of LSF1 suggests that it may protect phosphoglucans and regulate the rate of glucan dephosphorylation within the cell. Further studies, both in vitro and in vivo would be necessary to investigate the validity of this hypothesis.

This work has provided a large amount of information regarding the glucan phosphatases, but the structural basis of glucan dikinase activity is still completely unknown. The structure of the related enzyme pyruvate phosphate dikinase has been determined (1KBL) (150), but the structure of a glucan dikinase has not. These enzymes are equally as important as targets for genetic manipulation as the glucan phosphatases and understanding their mechanism of activity will provide further tools for modifying starch. It will be particularly interesting to determine if the glucan dikinases use a similar method for glucan interaction as SEX4 and LSF2 and if they contain similar mechanisms of C6- and C3-position substrate specificity.

Lastly, we have strong desire is to implement our new understanding of SEX4 and LSF2 function into biotechnological tools that will have an impact in
agriculture, medicine, and/or industrial applications of starch. The glucan phosphatases have a clear influence on the accumulation of starch in *Arabidopsis*. However, field trials using crop plants with the intent to increase storage starch yield have not yet been implemented. Recently, very promising results were obtained from the silencing of GWD in wheat plants whereby the grain yield was increased by 25-30% (151). Similarly promising results may be found in plants in which SEX4 and/or LSF2 has been silenced or mutated to alter specificity. In addition, the use of SEX4 and LSF2 in the processing of starch may help to expand the usefulness of starches for industrial purposes or decrease the economic costs of degrading starch by controlled dephosphorylation (7). It is the hope of this author that the insights gained from this research will be adopted for biotechnological purposes into useful methods to increase the yield and applicability of starch.
Figure 7.1 – Model for SEX4 interaction with starch glucans within reversible starch phosphorylation. Model of reversible starch phosphorylation and starch catabolism. GWD – Glucan Water Dikinase, PWD – Phosphoglucon Water Dikinase, Amylase - β-amylase and isoamylase, SEX4 – Starch EXcess4, LSF2 – Like Sex Four2. Inset shows the proposed model for SEX4 interaction with phosphorylated starch glucan chains at the C6-position. Phosphates are represented as red circles. SEX4 interacts with and dephosphorylates starch glucan chains via a coordinated CBM-DSP pocket maintained in part by the CT motif.
Figure 7.2 – Model for LSF2 interaction with starch glucans within reversible starch phosphorylation and starch catabolism. GWD – Glucan Water Dikinase, PWD – Phosphoglucan Water Dikinase, Amylase - β-amylase and isoamylase, SEX4 – Starch EXcess4, LSF2 – Like Sex Four2. Inset shows the proposed model for LSF2 interaction with phosphorylated starch glucan chains at the C3-position. Phosphates are represented as red circles. LSF2 interacts with and dephosphorylates starch glucan chains via three glucan binding sites with possible connectivity and with possible contacts of multiple glucan chains.
Figure 7.3 – Potential branch and main chain pathways within the SEX4 structure. SEX4 glucan interaction pocket with maltoheptaose (green) and phosphate. A. Potential pathway of α-1,6-glycosidic linked branch chain. F167 (yellow surface) would function as a branch-interacting residue in our hypothesis. B. Potential pathway of the main chain. A phosphate could more easily be integrated into the active site if engaging SEX4 from this direction.
Figure 7.4 – Potential LSF2-glucan complexes formed by multiple binding sites. LSF2 DSP domain (blue) with the active site (red) and CT motif (green) with glucan interacting residues colored orange. Glucan chains are represented as lines and phosphates as yellow circles. 

A. Helical section of unwound chain interacts with Site-3, unwound section of the same chain complex interacts with the active site, and Site-2 interacts with an adjacent unwound chain. 

B. Active site and Site-2 interact with two separate chains of the same complex and Site-3 interacts with an adjacent helical chain. 

C. Active site and Site-2 interact with two separate chains of the same complex and Site-3 interacts with an adjacent helical chain. 

D. Active site, Site-2, and Site-3 interact with the same complex, Site-3 interacts with the helical portion, and active site and Site-2 interact with different chains of the unwound portion.
Figure 7.5 – Proposed site-directed mutagenesis of SEX4 and LSF1 to define minimum requirements for glucan phosphatase activity.

Demonstration of nine residue swaps for SEX4. The mutations would possibly serve to define the minimum requirements for glucan phosphatase activity and the basis for lack of activity in LSF1. Selected residues in the V-loop, D-loop, PTP-loop and R-motif are shown. Residues known to be important for SEX4-glucan interaction are highlighted in red. Suggested swaps are indicated with arrows. Residue numbers of SEX4 are above and LSF1 are below.
References


Vita

David Adam Meekins
Birthplace: Parma, Ohio, USA

Education
B.S., Biology, Bowling Green State University (BGSU), Bowling Green, OH, 2009

Awards and Fellowships
Robert A. Lester Award, UK, 2013
Max Steckler Award, UK, 2013
Dissertation Year Fellowship, UK, 2013-14
Best Poster Presentation Award, Donald Danforth Center, 2012
Travel Fellowship, UK, 2012
2nd Place Oral Presentation, UK, 2012
Honorable Mention, NSF Graduate Research Fellowship, 2011
Travel Award, ASBMB, 2011
1st Place Oral Presentation, UK, 2011
International Travel Fellowship, Journal of Cell Science, 2010
Research Incentive Award, UK, 2009-12
Tinnappel Plant Sciences Scholarship, BGSU, 2008

Publications
