The FEBS Journal
FEBS PRESS science publishing by scientists
REVIEW ARTICLE

Structural mechanisms of plant glucan phosphatases in starch metabolism

David A. Meekins[1,2], Craig W. Vander Kooi[1] and Matthew S. Gentry[1]

1 Department of Molecular and Cellular Biochemistry and Center for Structural Biology, University of Kentucky, Lexington, KY, USA
2 Division of Biology, Kansas State University, Manhattan, KS, USA

Keywords
Arabidopsis; dual-specificity phosphatase; enzyme specificity; glucan interactions; glucan phosphatase; phosphatase; protein tyrosine phosphatase; reversible phosphorylation; starch; structural biology

Correspondence
M. S. Gentry, Department of Molecular and Cellular Biochemistry, University of Kentucky, 741 S. Limestone Avenue, Lexington, KY 40536, USA
Fax: +1 859 257 2283
Tel: +1 859 323 8482
E-mail: matthew.gentry@uky.edu

(Received 12 December 2015, revised 31 January 2016, accepted 29 February 2016)

doi:10.1111/febs.13703

Glucan phosphatases are a recently discovered class of enzymes that dephosphorylate starch and glycogen, thereby regulating energy metabolism. Plant genomes encode two glucan phosphatases, called Starch EXcess4 (SEX4) and Like Sex Four2 (LSF2), that regulate starch metabolism by selectively dephosphorylating glucose moieties within starch glucan chains. Recently, the structures of both SEX4 and LSF2 were determined, with and without phosphoglucan products bound, revealing the mechanism for their unique activities. This review explores the structural and enzymatic features of the plant glucan phosphatases, and outlines how they are uniquely adapted to perform their cellular functions. We outline the physical mechanisms used by SEX4 and LSF2 to interact with starch glucans: SEX4 binds glucan chains via a continuous glucan-binding platform comprising its dual-specificity phosphatase domain and carbohydrate-binding module, while LSF2 utilizes surface binding sites. SEX4 and LSF2 both contain a unique network of aromatic residues in their catalytic dual-specificity phosphatase domains that serve as glucan engagement platforms and are unique to the glucan phosphatases. We also discuss the phosphoglucan substrate specificities inherent to SEX4 and LSF2, and outline structural features within the active site that govern glucan orientation. This review defines the structural mechanism of the plant glucan phosphatases with respect to phosphatases, starch metabolism and protein–glucan interaction, thereby providing a framework for their application in both agricultural and industrial settings.

Introduction

Phosphatases of the protein tyrosine phosphatase (PTP) superfamily are critical regulators of a variety of cellular signaling events via dephosphorylation of specific proteinaceous and non-proteinaceous substrates [1,2]. Over the last 20 years, many PTP structures have been determined, and the relationship between structure and target substrate of both protein and lipid phosphatases has been defined in exquisite detail. The recent discovery of glucan phosphatases, PTPs within the dual-specificity phosphatase (DSP) clade that act on phosphorylated carbohydrates, highlights the striking diversity within the PTP family, which includes members that dephosphorylate proteins, lipids, nucleic acids and carbohydrates [3–5].

Abbreviations
CBM, carbohydrate-binding module; CT motif, C-terminal motif; DSP, dual-specificity phosphatase; GWD, α-glucan water dikinase; LSF1, Like Sex Four1; LSF2, Like Sex Four2; PTEN, phosphatase and tensin homolog; PTP, protein tyrosine phosphatase; PWD, phosphoglucan water dikinase; SBS, surface binding site; SEX4, Starch EXcess4.

Recent structural and biochemical studies on the plant glucan phosphatases Starch EXcess4 (SEX4) and Like Sex Four2 (LSF2) have provided detailed insights into the mechanisms by which these PTPs target and dephosphorylate starch glucans [6–8]. This review provides a comparative structural and functional analysis of SEX4 and LSF2, and outlines the novel features unique to the glucan phosphatases. Understanding these unique features will expand our comprehension of the PTP superfamily, allow us to make informed predictions about related phosphatases in different biological systems, and harness their activity for biotechnological purposes.

Glucan phosphatases in starch metabolism

Glucan phosphatases play a critical role in the regulation of transitory leaf starch metabolism during the diurnal photosynthesis cycle [3,9–11]. Starch is the primary form of glucose storage in plants, and allows partitioning of excess glucose produced during photosynthesis for both short- and long-term storage [12,13]. Starch is composed almost entirely of the glucose polymers amylose and amylopectin, with amylose being the minor constituent (10–30%) and amylopectin the major constituent (70–90%) (Fig. 1A). Trace amounts of proteins, lipids and ions are also present in the starch granule, depending on the source [10]. Both amylose and amylopectin are formed from α-1,4-glycosidic linked glucan chains, but amylopectin also contains α-1,6 branches clustered every 20–25 glucose units [10,14,15]. Adjacent amylopectin chains interact to form double helices that organize into crystalline lamellae [16]. This tightly packed arrangement of amylopectin helices results in expulsion of water, and causes the starch granules to be semi-crystalline and water-insoluble [10]. Ultimately, this property allows starch granules to be stable, energy-dense and highly effective glucose storage molecules. Cytosolic, starch-like Floridean starch has an equivalent function in red algae and during the hibernation stage of several eukaryotic protozoans [17–22]. Although starch and glycogen are each synthesized from glucose, and they utilize the same chemical linkages, the biophysical properties of starch differ from those of glycogen, as glycogen is water-soluble and available for dynamic bursts of metabolic activity [23].

Starch is adapted for the predictable diurnal cycles of photosynthetic-based metabolism in plants [10]. In chloroplasts, transitory starch is continually synthesized each day during the photosynthetic period, and the water insolubility of starch hinders its breakdown via glycolytic enzymes [24]. However, starch is almost completely degraded during the night to facilitate plant growth when photosynthesis is not occurring [25]. Therefore, during the degradative phase, hydrolytic enzymes must access the energy cache stored in this insoluble form. To overcome this obstacle, plants utilize a system of reversible phosphorylation that alters the biophysical properties of starch and increases the bioavailability of glucose chains to hydrolytic enzymes, thus permitting transition from starch synthesis to degradation [9,26–28].

While the presence of phosphate in starch was first reported in the 1890s, the role of phosphate in diurnal starch metabolism first came to light via biochemical experiments and the identification of mutant Arabidopsis plants with a starch excess phenotype [29–35]. An excellent history of starch phosphorylation was recently published [36]. *SEX* mutants have increased starch content, larger and often malformed starch granules, and show a decrease in plant growth [33,34]. Together, these features indicate an inability to efficiently degrade starch granules during non-photosynthetic periods. Biochemical experiments using these mutant lines resulted in identification of two dikinases called α-glucan water dikinase (GWD/SEX1) and phosphoglucan water dikinase (PWD) [30,35,37–40]. GWD exclusively phosphorylates the hydroxyl group at the C6 position of starch glucose, and this event triggers phosphorylation by PWD of hydroxyls at the C3 position (Fig. 1B) [30,37,39–43]. Multiple studies have proposed that introduction of covalently bound phosphate groups induces steric hindrance that results in unwinding of amylopectin helices and local solubilization of the outer starch granules (Fig. 1C) [26,42,44–47]. Thus, the glucan dikinases increase starch surface glucan solubility, and allow hydrolytic enzymes access to individual glucan chains. More recently, a role for starch phosphorylation by GWD during starch synthesis was identified [48]. The importance of these processes is highlighted by the strict conservation of the dikinases in plants and algae [49–51].

While starch phosphorylation is necessary for proper starch metabolism, it also results in a molecular dilemma that became apparent with identification of an additional *SEX* mutant called *starch excess 4* (*sex4*) [34]. Identification of the *sex4* locus in Arabidopsis resulted in discovery of a gene encoding a trimodular protein comprising a chloroplast targeting peptide, a dual-specificity phosphatase (DSP) domain, and a carbohydrate-binding module (CBM) (Fig. 2A) [50,52,53]. It was subsequently demonstrated that SEX4 is a glucan phosphatase that dephosphorylates starch-bound phosphate

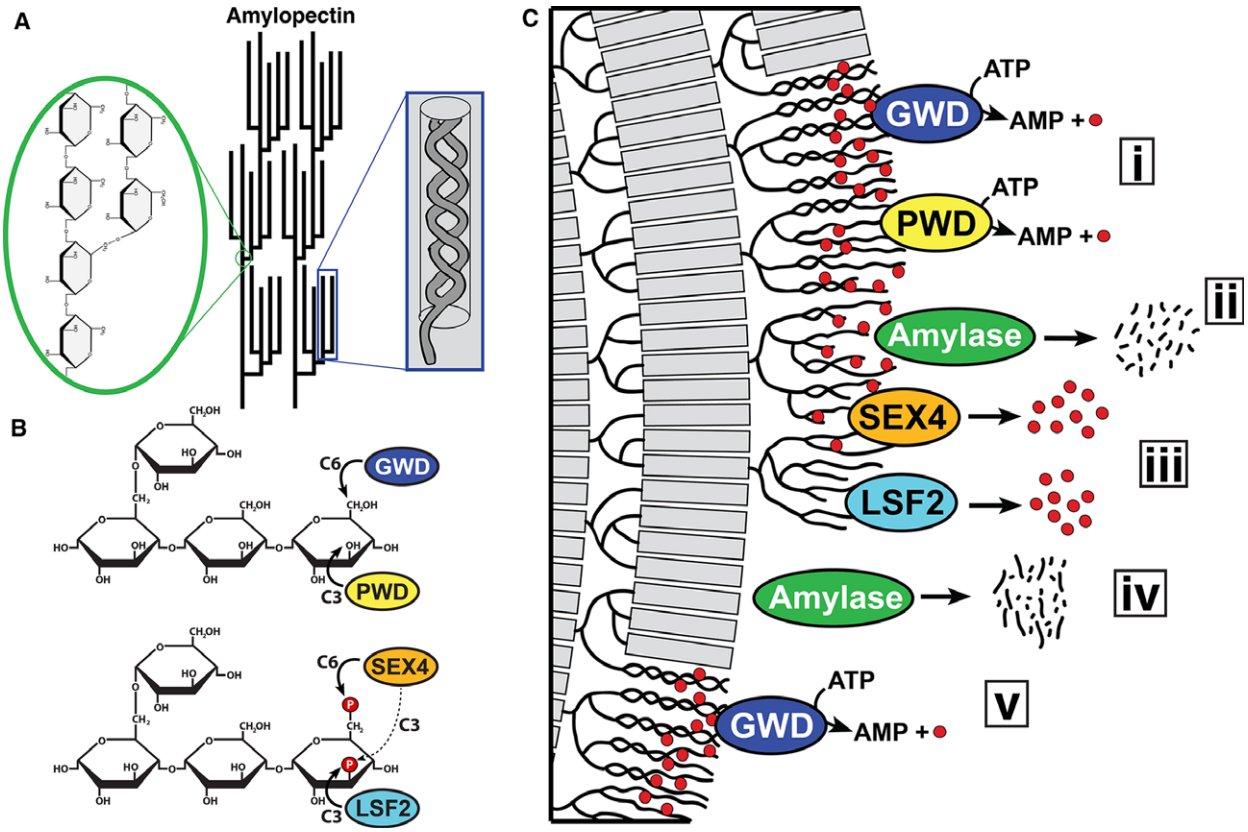

Fig. 1. Reversible phosphorylation of the plant starch granule. (A) Amylopectin is a glucose polymer and the main component of starch. It is formed from α-1,4-glycosidic linked glucose chains with α-1,6 branches that are clustered at regular intervals (green inset). Adjacent amylopectin chains form double helices (blue inset) that expel water, contributing to the water insolubility of starch. (B) Amylopectin glucans are phosphorylated and dephosphorylated by four enzymes. GWD phosphorylates glucans at the C6 position, followed by phosphorylation at the C3 position by PWD. SEX4 dephosphorylates the C6 and C3 positions, with a preference for the C6 position, and LSF2 exclusively dephosphorylates the C3 position. (C) Reversible phosphorylation in starch degradation. (i) GWD and PWD phosphorylate amylopectin helices (gray bars), causing them to unwind. (ii) Amylases, particularly β-amylase, are able to access and degrade the glucan chains, but β-amylase is unable to hydrolyze glucans past a phosphate group (red circles). (iii) SEX4 and LSF2 dephosphorylate the exposed glucan chains, allowing further degradation (iv). (v) The process continues with phosphorylation on the next layer of amylopectin helices within the starch granular lattice.

incorporated by GWD and PWD (Fig. 1B) [21,54,55]. While starch phosphorylation is necessary to solubilize surface glucans, starch dephosphorylation is necessary because starch-bound phosphate groups obstruct the movement of β-amylase, the primary enzyme that degrades starch [56]. β-amylases degrade glucan chains up to a phosphate group, but SEX4 must remove the phosphate before the amylase may proceed and fully degrade the chain (Fig. 1C) [55]. Therefore, glucan phosphatase activity essentially resets the cycle, and is an essential component of efficient starch degradation. These studies clearly demonstrated that efficient starch degradation requires the coordinated activity of dikinases, amylases and phosphatases.

An additional glucan phosphatase called Like Sex Four2 (LSF2) has been discovered based on sequence similarity with the SEX4 DSP domain [52,57,58]. Enzymatic characterization of LSF2 revealed that it possesses robust glucan phosphatase activity against starch [58]. This discovery was unexpected due to the lack of a CBM in LSF2, as a CBM was predicted to be essential for glucan phosphatase activity (Fig. 2A) [21,55,59]. An additional surprise was that LSF2 exclusively dephosphorylates the C3 position of starch glucose moieties (Fig. 1C) [58]. This specificity is in contrast to SEX4, which prefers the C6 position, but may also dephosphorylate the C3 position [6,54,60]. The characterization of LSF2 therefore established an elegant two-enzyme model for both phosphorylation

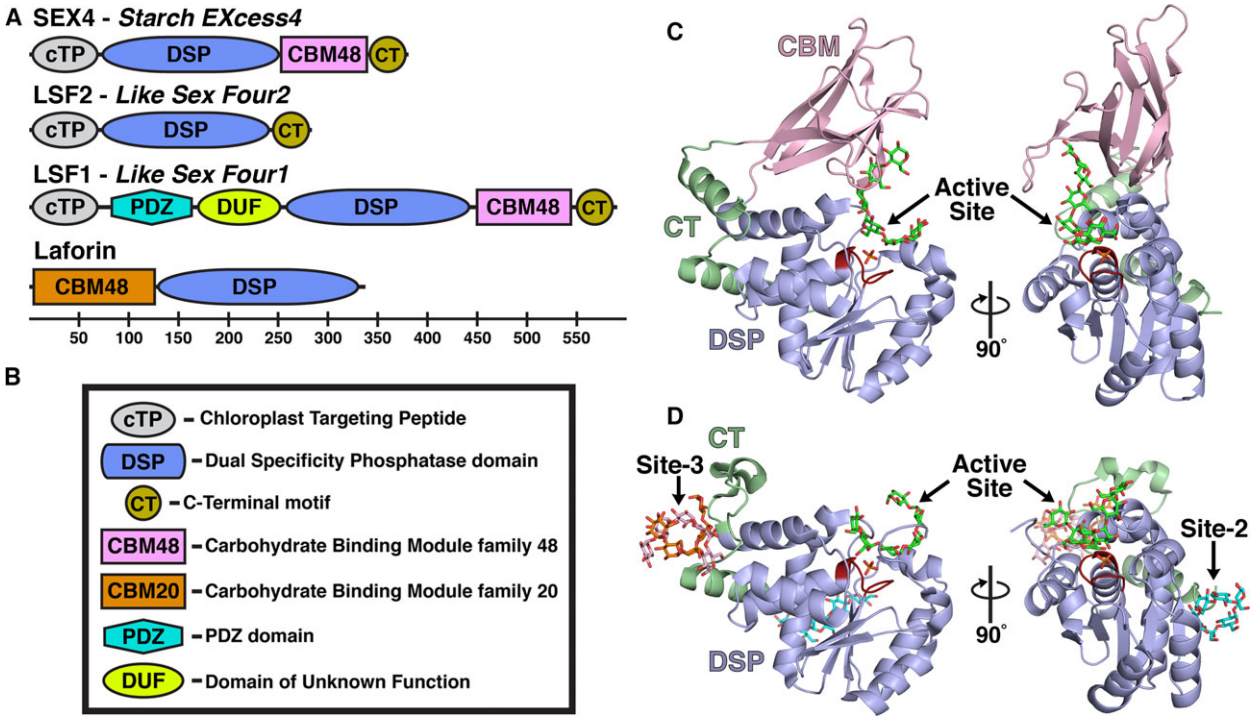

Fig. 2. The glucan phosphatase family and structure of the plant glucan phosphatases. (A) Domain outline of the plant proteins SEX4, LSF2 and LSF1, and the human protein laforin, with the residue numbers below. (B) Domain legend. (C) Glucan-bound structure of SEX4 (PDB ID 4PYH [6]), showing the DSP domain (blue) with the Cx_5R catalytic site (red), the CBM (pink), and the CT motif (CT, green). A single glucan chain (green) and phosphate (orange) were found at the active site. (D) Glucan-bound structure of LSF2 (PDB ID 4KYR [7]), showing the DSP domain, Cx_5R catalytic site, and CT motif. A single glucan chain (green) and phosphate (orange) were found at the active site. Three additional glucan chains were found at non-catalytic secondary binding sites (SBSs). A single glucan chain (cyan) was found at SBS site 2 and two glucan chains (orange and pink) were found at SBS site 3.

and dephosphorylation of starch in plants, with the activity of GWD and PWD balanced by SEX4 and LSF2. SEX4 and LSF2 are the only glucan phosphatases in plant genomes based on bioinformatics analyses [49, 58].

SEX4 and LSF2 are critical components of starch catabolism, but the structural mechanism used by these enzymes to target and dephosphorylate glucan substrates was unknown. Recently, we determined the X-ray crystal structures of SEX4 and LSF2, both with and without glucan ligands bound [6–8]. These data allowed us to establish the structural bases of their respective mechanisms, define how the glucan phosphatases incorporate phospho-glucans into their active sites, determine how the glucan chain is oriented to achieve their respective substrate specificities, and hypothesize about the activity of other phosphatases. This new understanding of glucan phosphatase activity expands our comprehension of PTP activity with respect to how substrate targeting is achieved.

Structure of SEX4 and LSF2: requirements for glucan phosphatase activity

The first structure obtained for a glucan phosphatase was of SEX4 in the absence of a glucan, and we subsequently determined the structure of SEX4 bound to a phosphoglucan product (PDB IDs 3NME and 4PYH) [6,8]. These structures revealed the interdomain interactions essential for SEX4 activity. SEX4 is 43 kDa protein containing 379 residues, and the crystallized construct (residues 90–379) contained the DSP and CBM domains. The non-glucan bound structure provided the first information regarding the fold of SEX4, and the physical relationship between domains.

The DSP domain (residues 90–252) of SEX4 contains a characteristic αβα DSP fold with a central five-stranded β-sheet flanked by eight α-helices (Fig. 2C) [8]. The catalytic site Cx_5R sequence of SEX4 is located between β5 and α6 at the base of the active-site pocket. The SEX4 CBM (residues

253–338) contains six β-strands that fold into a characteristic compact β-sandwich composed of anti-parallel sheets. The conserved binding site of the SEX4 CBM domain is positioned between the β-sandwich and an adjacent loop region comprising residues I323 to N332. Interestingly, the DSP and CBM domains contain an extended and intimate interdomain interaction with a surface area of 457 $Å^2$. This interaction is maintained by a previously unrecognized C-terminal (CT) motif (residues 338–379) consisting of two α-helices that is connected to the CBM and forms extended contacts with the DSP. As a result of this extended CBM–DSP interaction, SEX4 contains a continuous glucan-binding pocket from the CBM binding site through to the DSP catalytic site that engages a single glucan chain via both aromatic and hydrophilic residues [6].

Following determination of the SEX4 structure, the structure of LSF2 was also determined (PDB IDs 4KYQ and 4KYR) [7]. LSF2 is a 282 amino acid protein of 32 kDa containing a chloroplast targeting peptide and a DSP domain (Fig. 2A). Sequence alignments also predicted that LSF2 contained a CT motif similar to that of SEX4 [58]. The LSF2 DSP domain (residues 79–244) also possesses a characteristic αβαDSP fold consisting of a central five-stranded β-sheet region flanked by eight α-helices (Fig. 2D) [7]. The LSF2 CT motif (residues 245–282) consists of a loop region culminating in an α-helix that integrally folds into the DSP domain. Although LSF2 does not contain a CBM, the structure of LSF2 bound to the glucan ligand maltohexaose revealed the presence of two non-catalytic surface binding sites (SBSs) associated with the CT motif, in addition to a glucan ligand located at the DSP catalytic site. These results suggest that LSF2 uses these two SBSs to engage glucan ligands in lieu of a CBM [7].

Interestingly, researchers have identified an additional Arabidopsis protein called Like Sex Four1 (LSF1) that possesses a CBM and DSP, and *lsf1* mutant plants exhibit a starch excess phenotype [3,57]. LSF1 comprises a chloroplast targeting peptide, followed by a protein–protein interaction domain known as a PDZ domain, an extended domain of unknown function, a DSP domain, a family 48 CBM domain, and CT motif (Fig. 2A) [57]. As LSF1 has similar domain architecture to SEX4, it was hypothesized that LSF1 is also a glucan phosphatase. However, it has been reported that LSF1 lacks detectable glucan phosphatase activity [57]. Despite this lack of activity, LSF1 has a clear role in maintaining proper starch metabolism, but the precise nature of its function is unknown. Therefore, a comparative analysis of SEX4 and LSF2 activity should highlight features that are specific to this family of enzymes as well as providing informed predictions about the function of LSF1.

The SEX4 and LSF2 structures coupled with structure-guided mutagenesis and biochemical data established the major structural requirements for glucan phosphatase activity in these two plant glucan phosphatases. SEX4 activity requires the CBM–DSP interaction maintained by a previously unrecognized CT motif, and LSF2 activity requires its DSP glucan-binding site in addition to two previously unrecognized SBSs associated with its CT motif [6–8,60]. Thus, both proteins possess a unique glucan-binding platform that allows binding and dephosphorylation of starch glucan chains.

SEX4 and LSF2 substrate targeting

Although starch largely comprises extended glucose chains, its structure contains multiple complexities that present unique challenges to starch-modifying enzymes [10]. The starch granule is a highly condensed, semi-accessible substrate with a low frequency of individual phosphate groups, approximately one per 2000 glucose units in Arabidopsis leaf starch [39]. While this frequency is presumably higher during starch degradation, the difficulty of locating a phosphate group within the starch superstructure requires additional mechanisms for interaction with the substrate. Consequentially, glucan phosphatases invariably contain non-catalytic glucan-binding motifs, in the form of a CBM or SBSs, to efficiently interact with and dephosphorylate starch [3].

SEX4 contains a single CBM that serves as its glucan-binding motif. The maltoheptaose-bound SEX4 structure revealed a tight interaction between the CBM-binding interface and two glucose moieties of the glucan chain (Fig. 3A) [6]. The central platform for this interaction comprises a dual-tryptophan motif formed from W278 and W314. In addition, hydrogen bonds are formed between CBM residues H330, N332 and K307 and hydroxyl groups of the glucose moieties. All five of these glucose-interacting residues are highly conserved among SEX4 orthologs. Mutagenesis of these SEX4 CBM residues revealed that the CBM is critical for glucan binding and subsequent dephosphorylation of insoluble starch [6]. Single point mutants of CBM residues showed almost total abolishment of phosphatase activity in some cases, due to an inability of these mutants to bind starch glucan chains. Therefore, the CBM is necessary for substrate targeting within the SEX4 mechanism of activity.

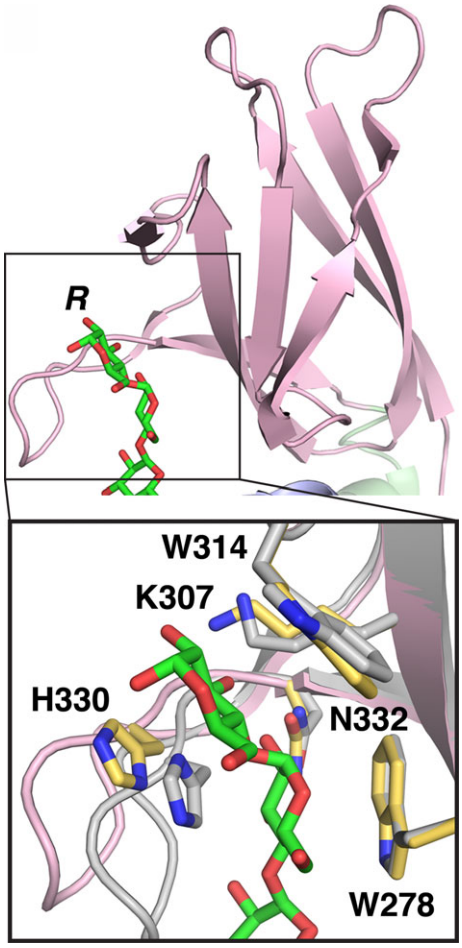

Fig. 3. SEX4 CBM-glucan interaction and conservation. Structure of the glucan-binding site of the SEX4 CBM (pink, PDB ID 4PYH [6]), showing the bound glucan chain (green). '*R*' represents the reducing end of the glucan chain. Inset: magnification of the carbohydrate-binding site at the SEX4 CBM, showing glucan-interacting side chains. The interacting residues from the glucan-bound SEX4 structure (pink, with yellow residues) are superimposed on the non-glucan bound structure (gray, with gray residues, PDB ID 3NME [8]), highlighting movement of H330.

A CBM domain 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 [61]. The most common function of these domains is to bring the catalytic portion of an enzyme in close and prolonged interaction with a mono-, oligo- or polysaccharide [62]. CBMs are common in starch-interacting enzymes, but are also found in enzymes that modify cellulose, chitin, β-glucans, glycogen, pullulan and xylan [62]. The CBM is structured to accommodate its target substrate, with the carbohydrate-binding platform adopting a flat or curved structure to target cellulose or starch, respectively [63]. This positioning is clearly apparent in SEX4, in which the dual-tryptophan platform is curved to accommodate the natural shape of the α-glucan substrate [6].

CBMs are categorized according to the Carbohydrate Active enZymes (CAZy) database into 71 families [64,65]. The SEX4 CBM belongs to the CBM48 family, and the dual-tryptophan platform and K307 represent a conserved functional motif in both CBM48 and the related CBM20 family [61,66,67]. The most similar CBM to that of SEX4 is the CBM48 in the human AMP-activated protein kinase β1 subunit (PDB ID 1Z0N), which shows an RMSD of 1.4 Å compared with the SEX4 CBM (Fig. 3B) [68]. AMP-activated protein kinase β1 contains all the SEX4 CBM glucan-interacting residues except for a threonine at the position of H330 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 (Fig. 3C) [6]. Mutation of SEX4 H330 to alanine resulted in a > 60% decrease in SEX4 starch dephosphorylation [6], highlighting the importance of this residue.

The plant protein LSF1 and the human phosphatase laforin also contain a CBM, and the need for CBM functionality has also been demonstrated in laforin [57,69,70]. SEX4 and LSF1 both contain CBM48 domains, while laforin contains a CBM20 domain [61]. The CBM20 and CBM48 families are closely related based on conserved binding regions, and intermediates between the CBM48 and CBM20 families that contain conserved elements of both families have been identified [61]. Moreover, CBM20 and CBM48 domains do not exclusively target starch or glycogen. Instead, examples of each family have been found in both starch-targeting enzymes and glycogen-targeting enzymes. For example, the plant enzyme PWD contains a CBM20 domain although it targets starch, and the human protein AMP-activated protein kinase β1 contains a CBM48 although it targets glycogen [61]. This overlap between the two families is demonstrated by the ability for SEX4 to dephosphorylate solubilized amylopectin, which is reminiscent of glycogen, and the ability of laforin to rescue some aspects of the *sex4* phenotype as well as its ability to dephosphorylate starch [21]. In humans, laforin binds and dephosphorylates glycogen [21,59,71,72]. Mutations in the human gene encoding laforin that decrease its ability to bind and/or dephosphorylate glycogen lead to accumulation of starch-like intracellular carbohydrates called Lafora bodies, resulting in the fatal neurodegenerative disease called Lafora disease [22,73,74].

In addition to differences in family designation, the glucan phosphatases also have differences in the position of the CBM with respect to the catalytic domain. Laforin contains an N-terminal CBM followed by a DSP domain, whereas both SEX4 and LSF1 possess a C-terminal CBM (Fig. 2A) [52,57,69]. Typically, the CBM precedes the catalytic domain, as in laforin, and the domain order in SEX4 and LSF1 is an exception to the trend [61]. In addition to similar domain structure, the glucan phosphatases are also similar at the primary amino acid level. The SEX4 and LSF1 CBMs are most similar, sharing 21% identity, whereas the laforin CBM20 only shares 15% sequence identity with SEX4 and LSF1 (Fig. 3B); all three CBMs share the conserved glucan-binding residues that form a Trp-Lys-Trp motif for glucan interactions, which defines both the CBM20 and CBM48 families [61].

Substrate targeting via additional domains is a common theme among dual-specificity phosphatases. These ancillary domains include slingshot and PDZ domains involved in protein–protein interactions, and PH-GRAM and C2 domains involved in interaction with lipid membranes [1,2,75]. Structures are available for three DSPs with C2 domains: phosphatase and tensin homolog (PTEN) (PDB ID 1D5R) [4], *Ciona intestinalis* voltage sensing phosphatase (PDB ID 3AWE) [76] and auxilin PTEN-like protein (PDB ID 3N0A) [77], and one structure for a DSP containing a PH-GRAM domain: myotubularin related protein2 (MTMR2) (PDB ID 1M7R) [78]. SEX4, PTEN, the *Ciona intestinalis* voltage sensing phosphatase and the Auxilin PTEN-like protein all contain ancillary domains C-terminal to the DSP, while MTMR2 contains an N-terminal PH-GRAM domain and laforin contains an N-terminal DSP. Moreover, the consistency between the structures suggests that the spatial relationship of the two domains has probable consequences with regard to coupling of the two activities, that is substrate interaction and catalysis.

In contrast to the CBM-driven mechanism for glucan targeting found in SEX4, LSF2 possesses a unique DSP active site that incorporates both a glucan-binding platform and a phosphatase catalytic site [7]. Additionally, LSF2 possesses two non-catalytic SBSs > 20 Å from the active site that involve participation from the CT motif to achieve efficient glucan binding [7]. One SBS, site 2, is located in a binding pocket formed by residues from the DSP domain and the CT motif (Fig. 4A). Specific interactions are formed between DSP residues R153, R157, M155 and W180 and the maltohexaose chain. In addition, the glucan chain wraps around the end of the CT motif itself. Another SBS, site 3, is located within the loop region of the CT motif (Fig. 4B). The LSF2 structure revealed two separate glucan chains (Hex-1 and Hex-2), located at site 3, that form a helical-like structure. LSF2 primarily interacts with Hex-1, forming hydrogen bonds via K245 and E268, and van der Waals interactions via F261. Interestingly, mutation of residues in the LSF2 SBSs had a nearly identical effect on enzymatic activity and binding as mutations in the SEX4 CBM [7]. Mutation of site 2 and site 3 residues resulted in a dramatic decrease in enzymatic activity that directly correlated with their ability to bind amylopectin.

Similarly to CBMs, SBSs are also found in starch-active enzymes, primarily in glycosyl hydrolases, and have been reported in α-amylase, β-amylase, branching enzyme, xylanase, β-agarase, β-glucosidase, galactosidase, chitinase and α-glucosidase enzymes [79]. Like CBMs, the necessity for SBSs in starch active enzymes is probably linked to the complex structure of starch granules, which have inherently poor accessibility. The main difference between the CBM and SBSs is that SBSs exist in a fixed position relative to the catalytic site, whereas CBMs are commonly connected to the catalytic domain via a flexible linker.

SBSs are difficult to identify by sequence analysis, and instead are commonly identified after co-crystallization with a glucan and structural determination, as was the case with LSF2 [79]. This mode of reverse identification is due to the absence of a unifying structural consensus among SBSs, despite the number of identified SBSs [79]. Although there are no strictly conserved SBS signatures, aromatic residues typically serve as the basis for glucan interaction in the SBSs. This is also the case for LSF2, with F261 in site 3 and W180 in site 2 [7]. The lack of a consensus signature makes preemptive identification of SBS sites in other glucan phosphatases difficult. Both of the SBSs in LSF2 are associated with the CT motif, which both SEX4 and LSF1 possess. However, neither SEX4 nor LSF1 contain the conserved residues that comprise the glucan-binding sites in the LSF2 SBSs (Fig. 4C). Moreover, the SEX4 CT motif has the additional function of maintaining the interaction between its CBM and DSP [8]. Therefore, it is likely that the use of SBSs in LSF2 is a unique mechanism in the glucan phosphatase family, despite the similar function that the SBSs impart on LSF2 function compared to the CBM in the other glucan phosphatases.

In addition to substrate targeting, both CBMs and SBSs have been postulated to participate in other related functionalities, including substrate disruption, passing on reaction products, allosteric regulation, and substrate guidance into the active site [79]. These additional functionalities have not been tested in SEX4 and LSF2, but the structural data give insights into

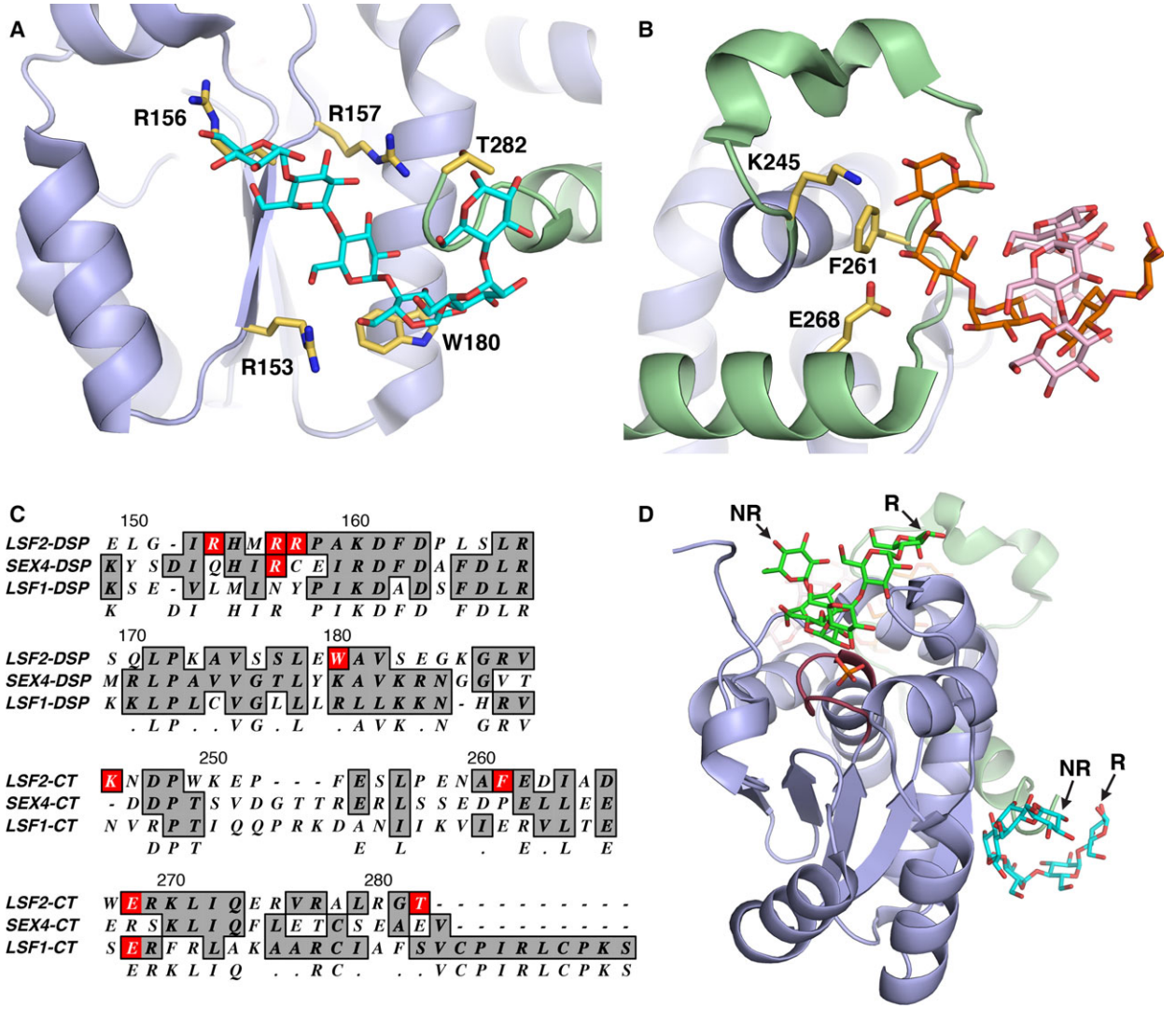

Fig. 4. Non-catalytic surface binding sites (SBSs) in LSF2. (A) LSF2 (PDB ID 4KYR [7]) SBS site 2, containing a single glucan chain (cyan) that interacts with residues (yellow) located on the DSP (blue) and the C-terminus of the CT motif (green). (B) LSF2 SBS site 3, containing two glucan chains (orange and pink) that interact with residues (yellow) located on the CT motif (green). (C) Alignment of SBS residues found on LSF2 with SEX4 and LSF1. A portion of the LSF2 DSP (residues 149–189) and the entire CT motif (residues 245–282) are shown. (D) Directional continuity between the glucan chain at the LSF2 active site (green) and site 2 (cyan). The non-reducing end (NR) and reducing end (R) are labeled.

their feasibility. Substrate disruption driven by the glucan-binding motifs of SEX4 and LSF2 assists in and increases the efficiency of their enzymatic activity, as the over-riding purpose of reversible phosphorylation is indeed substrate disruption. The multiple, disparate binding sites in LSF2 may increase the disorder of the starch granule, with each SBS engaging distinct glucan chains and interrupting glucan helices. This effect is less likely in the single continuous glucan-binding site in SEX4. Substrate guiding into the active site via the SEX4 CBM is clearly a factor in its dephosphorylation mechanism; however, the mechanism in LSF2 requires additional investigation. There may indeed be connectivity between LSF2 SBSs and the active site that was not revealed by co-crystallization with maltohexaose, which is a relatively short glucan chain. Interestingly, the glucan chains at the LSF2 active site and site 2 are directionally continuous with respect to the reducing and non-reducing ends, indicating these two sites may interact with a single glucan chain and supporting a role for substrate guidance (Fig. 4D). The possibility of processivity in glucan phosphatases has not been explored, but may involve participation of non-catalytic glucan-binding motifs.

Although the mechanisms dramatically differ, the non-catalytic glucan-binding modalities in SEX4 and LSF2 are both necessary for starch dephosphorylation. SBSs and CBMs are both found in carbohydrate-active enzymes, most relevantly in starch-active glucosyl hydrolases, and function in the same capacity as they do in the plant glucan phosphatases. Although starch-active enzymes generally contain active sites that are capable of interacting with target glucans, these non-catalytic motifs are necessary to bring the enzyme in direct and prolonged contact with the target substrate.

Glucan interaction at the SEX4 and LSF2 active sites

Based on the functional importance of the CBM and SBSs, it was initially thought that glucan phosphatases may function via connection of a CBM or SBSs to any phosphatase domain [21]. This theoretical model requires the binding domain to bring the phosphatase domain in close contact with the polyglucan substrate. However, early studies with SEX4 and laforin indicated that they contain necessary glucan-interacting elements within the DSP domain itself. A fusion protein of the SEX4 CBM with the DSP domain of the human protein phosphatase VHR possessed generic phosphatase activity, but this fusion protein did not dephosphorylate glucan substrates [59]. Indeed, specific elements unique to SEX4 and LSF2 were found within the DSP active site itself that contribute towards glucan interaction at the active site [6,7,60].

Both SEX4 and LSF2 DSP domains contain a typical αβα PTP fold that is highly conserved among DSPs despite a low level of sequence identity [7,8]. Not surprisingly, the DSP structures of SEX4 and LSF2 are most similar to each other, with an RMSD of 1.1 Å. With respect to other proteins, the SEX4 DSP is most structurally similar to the PTP of the thermophilic Archaea *Sulfolobus solfataricus* (*Ss*PTP, PDB ID 2I6O) [80], and the human phosphatases serine/threonine/tryrosine-interacting-like-protein (STYX) (PDB ID 2R0B) [81], CDK-associated phosphatase (KAP) (PDB ID 1FPZ) [82], VH1-like phosphatase Z (VHZ) (PDB ID 2IMG) [83] and dual specificity phosphatase27 (DUSP27) (PDB ID 2Y96) [84]. LSF2 is most structurally similar to human protein tyrosine phosphatse, mitochondrial (PTPMT1) (PDB ID 3RGQ) [85], VHZ, *Ss*PTP, KAP and STYX. The mean RMSD between the glucan phosphatases and these other DSPs is 2.5 Å, although sequence identities are only 10–18%. Despite their structural similarity with the glucan phosphatases, these structural homologs have a strikingly diverse array of target substrates. PTPMT1 is a lipid phosphatase that dephosphorylates phosphatidylgycerol phosphate [85], VHZ and DUSP27 are true DSPs that dephosphorylate both p-Tyr and p-Ser/Thr protein residues [83,84], *Ss*PTP is a p-Tyr phosphatase [80], KAP is a p-Thr phosphatase [82], and STYX is a pseudophosphatase that binds phospho-proteins but has no phosphatase activity [86]. A structural commonality linking SEX4 and LSF2 with most of these diverse phosphatases is the presence of an α-helical variable loop (V-loop). The V-loop is a substrate-determining PTP sub-domain that forms a loop in most phosphatases [2], but the glucan phosphatases and their structural homologs instead possess an α-helix. Despite this commonality, the global structures of the SEX4 and LSF2 DSPs are not indicative of their unique target substrates. Therefore, the glucan phosphatase DSP domains require specific features that govern their glucan-interacting ability.

Based on previous structures, a number of PTP sub-domains, including the V-loop, have been identified that function to target specific substrates at the catalytic site (Fig. 5A) [2]. The PTP-loop is the most significant sub-domain, constituting the catalytic core of all members of the PTP superfamily [1,5].The PTP-loop is characterized by an HCxxGxxRA/T (Cx_5R) motif in which the conserved catalytic cysteine generates a nucleophilic attack on the phosphate group, and the conserved arginine is necessary for integrating the phosphate within the catalytic pocket [5]. The catalytic cysteine has been shown to be sensitive to redox regulation in several PTPs [87]. Indeed, previous studies indicated that the catalytic cysteine of SEX4 (C198) is rendered inactive via formation of a disulfide bond with structurally adjacent C130 [88]. Subsequent reduction via thioredoxins probably allows SEX4 activity to be regulated within the chloroplast in a diurnal pattern without changes in protein concentration [88].

In both the glucan-bound SEX4 and LSF2 structures, a phosphate group is located within the PTP-loop directly below the glucan chain (Fig. 5B,C). Structural data from multiple DSPs have shown that the primary sequence of the PTP-loop corresponds with its target substrate [5]. For instance, PTPMT1 was identified as a lipid phosphatase due to homologous basic residues within its PTP-loop that are also present in the PTP-loop of the lipid phosphatase PTEN [85]. SEX4 and LSF2 both have very similar PTP-loop sequences (HCTAGMGR and HCSAG LGR, respectively (Fig. 5D). Defining the catalytic cysteine as residue 0, both glucan phosphatases have a short-chain hydrophilic residue at the +1 position, followed by an alanine and a glycine, then a non-aromatic hydrophobic residue at the +4 position, and a glycine before the Cx_5R arginine. This motif is

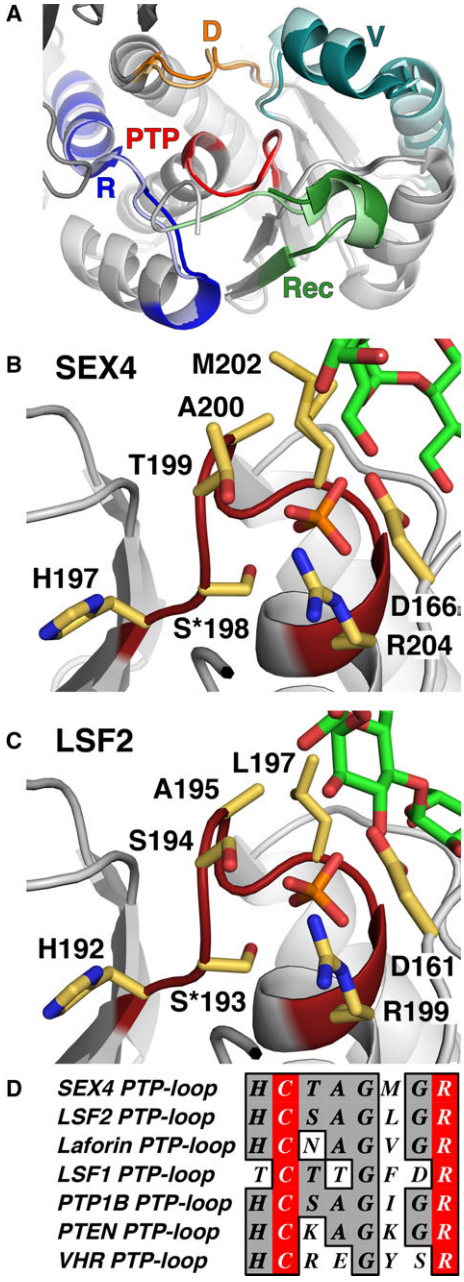

Fig. 5. DSP sub-domains and PTP-loop catalytic site. (A) Structural alignment of SEX4 (darker shade, PDB ID 4PYH [6]) and LSF2 (lighter shade, PDB ID 4KYR [7]), highlighting the DSP sub-domains. The PTP-loop (red) is the catalytic site and is located centrally, surrounded by the D-loop (orange), V-loop (teal), recognition motif (Rec, green) and R-motif (blue). (B) The PTP-loop of SEX4 (red) showing the glucan chain (green) and phosphate (orange), with PTP-loop residues in yellow. S*198 (mutated from cysteine in the crystallized construct), R204 and D166 comprise the catalytic triad. (C) PTP-loop of LSF2 (red) showing the glucan chain (green) and phosphate (orange), with PTP-loop residues in yellow. S*193, R199 and D161 comprise the catalytic triad. (D) Sequence alignment of the PTP-loops of SEX4, LSF2, laforin, LSF1 and human PTP1B, PTEN and VHR. The catalytic residues are highlighted in red.

distinct compared with other DSPs that target protein or lipid substrates. Additionally, laforin has a PTP-loop consistent with those of SEX4 and LSF2, giving a glucan phosphatase signature motif of CζAGΨGR where ζ is a hydrophilic residue and Ψ is a long-chain aliphatic residue [60].

Conversely, LSF1 has a very different PTP-loop sequence: TCTTGFDR (Fig. 5D). The residue at the -1 position in LSF1 is a threonine instead of the conserved histidine found in most PTPs [57]. This histidine residue is hypothesized to be essential for catalysis by decreasing the pKa of the adjacent catalytic cysteine and facilitating nucleophilic attack [5]. This amino acid change has been postulated to account for the inactivity of LSF1 [3,57]. There are several examples of inactive phosphatases that have regulatory functions, including STYX and members of the myotubularin family, but these inactive phosphatases contain substitutions of catalytic triad residues ($Dx_{20}Cx_5R$) [86,89,90]. Loss of activity solely due to a mutated histidine has not been reported previously. Subsequent analyses of LSF1 demonstrated that residues within the LSF1 PTP-loop sequence are also highly divergent from those found in the other glucan phosphatases [60]. LSF1 contains threonine and glutamic acid at positions 2 and 5, respectively, instead of the shorter-chain residues alanine or glycine found in glucan phosphatases. Additionally, position 4 is occupied by an aromatic residue in LSF1 and a long-chain aliphatic residue in the three glucan phosphatases. Therefore, based on the CζAGΨGR signature motif identified in SEX4, LSF2 and laforin, the LSF1 PTP-loop cannot accommodate phosphoglucan substrates [60].

Surrounding the PTP-loop are four additional sub-domains that interact with the target substrate and integrate it into the catalytic site: the D-loop, the V-loop, the R-motif and the recognition motif (Fig. 5A) [2].The combined structure of these sub-domains constitutes the overall active site of the phosphatases, and the shape and chemical composition dictate the substrates that are targeted and accommodated in the catalytic site for dephosphorylation.

The glucan-bound structures of SEX4 and LSF2 illustrate features of the DSP sub-domains that contribute specifically to glucan targeting (Fig. 6A,B) [6–8]. The D-loop contains a conserved aspartate that is part of the catalytic triad and operates in catalysis as a general acid/base to form the phospho-enzyme intermediate, followed by expulsion of the phosphate from the active site [91,92]. At the +1 position from the catalytic aspartate, the D-loops of SEX4 and LSF2 both contain a phenylalanine residue (F167 in SEX4, F162 in LSF2) that interacts with the glucan chain

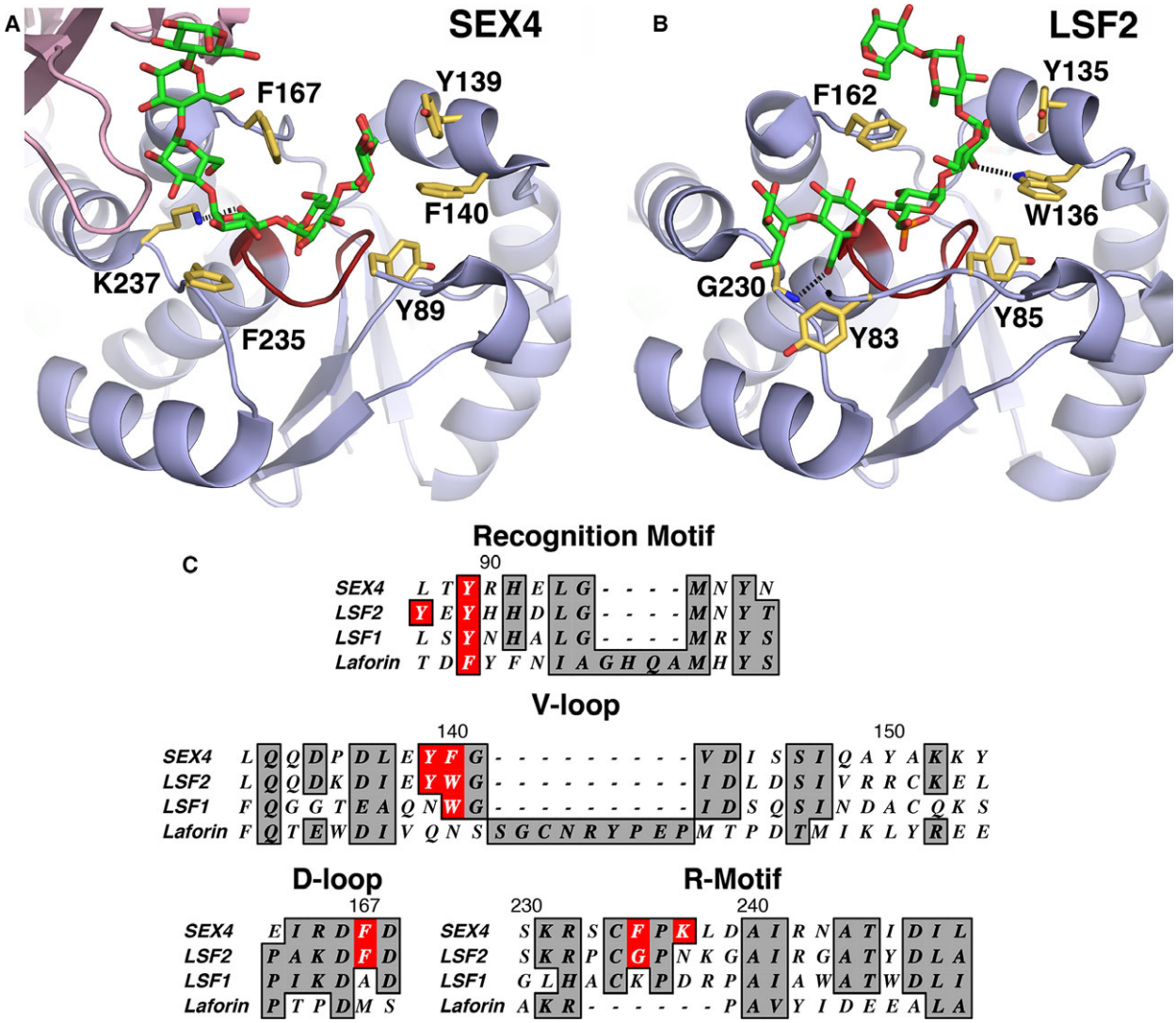

Fig. 6. Glucan phosphatase DSP interaction with the glucan chain. (A) SEX4 DSP (blue, PDB ID 4PYH [6]), with the glucan chain (green) and phosphate (orange) at the catalytic site (red). Interacting residues are shown in yellow, with a hydrogen bond between K237 and the glucan chain shown. (B) LSF2 DSP (blue, PDB ID 4KYR [7]), with the glucan chain (green) and phosphate (orange) at the catalytic site (red). Interacting residues are shown in yellow, with hydrogen bonds between the glucan chain and G230 and W136 shown. (C) Sequence alignment of DSP sub-domains between SEX4, LSF2, LSF1 and laforin. The sequence number of SEX4 residues is shown. Glucan interaction residues in SEX4 and LSF2 are highlighted in red (if applicable) with conserved residues (grey) in LSF1 and laforin.

directly over the active site. This residue caps the targeted phosphoglucose moiety, and was observed to be flexible when comparing the non-glucan bound structure with the glucan-bound SEX4 and LSF2 structures. C-terminal from the active site, the R-motif of SEX4 contains phenylalanine and lysine residues (F235 and K237) that interact with the glucan chain via van der Waals interactions and hydrogen bonding, respectively. Conversely, LSF2 contains a glycine residue (G230) in its R-motif that forms a hydrogen bond with the glucan chain upstream of the active-site moiety. N-terminal of the active site, the V-loop of SEX4 contains tyrosine and phenylalanine residues (Y139 and F140) that interact with the glucan moiety. The LSF2 V-loop contains similar tyrosine and tryptophan residues (Y135 and W136) that make equivalent glucan contacts. Finally, the recognition motifs of SEX4 and LSF2 both contain a tyrosine residue (Y89 in SEX4, Y85 in LSF2) that interacts with the glucan chain at the active site, and LSF2 possesses an additional tyrosine residue (Y83) that is not present in SEX4, which makes additional glucan contacts.

The DSP sub-domains found in SEX4 and LSF2 participate extensively in the interaction between the catalytic domain and the glucan chain.

The most obvious and functionally significant feature of the DSP sub-domains in SEX4 and LSF2 is the large population of aromatic residues that interact with the glucan chain (Fig. 6) [6,7].These aromatic residues form a network that provides an extended glucan-binding platform throughout the entire active site. Aromatic residues are essential for carbohydrate interaction in a variety of contexts via van der Waals interactions between the aromatic residues and glucose rings [61,63,79]. Indeed, most classes of CBMs rely heavily on aromatic residues to engage various carbohydrate substrates [63]. The active sites of multiple glycosyl hydrolases contain similar networks of aromatic residues, although they are typically less concentrated than is observed in the glucan phosphatase active sites [93,94]. Our analysis of the sequences of 34 additional DSPs revealed that the aromatic network found in the glucan phosphatases are novel and not conserved among DSPs. This network of aromatic residues in SEX4 and LSF2 is therefore a characteristic and defining feature of the plant glucan phosphatases.

Interestingly, a DSP alignment between SEX4, LSF2 and LSF1 suggests that this aromatic network is not conserved in LSF1 (Fig. 6C). LSF1 lacks many of the aromatic residues found in SEX4 and LSF2, and these differences probably contribute to the absence of glucan phosphatase activity (Fig. 6C). Determination of the LSF1 structure and enzymatic analysis using strategic mutagenesis would shed further light on this hypothesis.

As expected, mutation of glucan-interacting residues in the SEX4 and LSF2 DSP sub-domains had a clear effect on each enzyme's ability to dephosphorylate glucan substrates [6,7]. However, the extent to which these mutations hindered activity differed notably. In SEX4, mutation of DSP residues decreased activity 10–80%, with a mean decrease of 38% [6]. DSP domain mutation in LSF2 had a more notable effect on activity. Single point mutations of the LSF2 DSP domain resulted in a decrease in activity ranging from 40 to 95% with a mean decrease of 66% [7]. These differences appear to stem from the compensatory activity of the SEX4 CBM, and the presence of hydrogen-bonding residues in the LSF2 DSP. In SEX4, the CBM and DSP share a continuous binding pocket; therefore DSP mutations have limited effect due to glucan binding and proximity of the CBM platform [6]. Furthermore, LSF2 contains two residues (W136 and G230) that form hydrogen bonds with the glucan chain but are not present in SEX4 (Fig. 6) [7]. Their presence probably balances the loss of secondary site interaction given the distance between the active site and SBSs. Therefore, the DSP sub-domains of SEX4 and LSF2, despite having a common aromatic signature, contain distinct affinities for glucans irrespective of additional glucan-binding domains or motifs.

Conformation of glucan chains

The glucan-bound SEX4 and LSF2 structures permitted a clear study of DSP residues that interact with the glucan chain. However, α-glucan chains are directional molecules that may be engaged from multiple orientations. Considering the orientation of the glucan chain at the SEX4 and LSF2 active site is particularly relevant to determining the structural basis of their observed substrate specificity. SEX4 preferentially dephosphorylates the C6 position of starch glucans, whereas LSF2 exclusively dephosphorylates the C3 position. Understanding the structural basis of this difference may provide a possible means to manipulate glucan phosphatase substrate specificity for biotechnological purposes.

In both the SEX4 and LSF2 structures, six glucose moieties of the glucan chain were observed in the active site [6,7]. In SEX4, this chain spanned both the CBM and the DSP in a continuous binding pocket, whereas in LSF2, the entire chain was present at the DSP active site. Both enzymes also bound the glucan chains in a manner consistent with the twisting/helical geometry of the substrate, which is probably significant for optimal affinity. However, the enzymes also distorted this helical conformation in interestingly consistent ways. Previous studies showed that the α-1,4-glycosidic dihedral torsion in helical amylopectin chains has values of (91.8°, −153.2°), (85.7°, −145.3°) and (91.8°, −151.3°) [95]. The inequality of the two angles results in torsion consistent with a helical geometry. Conversely, equality in the glycosidic bond torsion angle, that is (119.3°, −119.3°), indicates a lack of torsion, that is linearity.

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 (Fig. 7) [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. The same trend is found in LSF2 (Fig. 7). The glycosidic bonds of glucans at the DSP active site in LSF2 (Glc1-Glc2 and Glc3-Glc4) have more linear dihedral torsion angles, whereas the glycosidic bonds of glucans outside of the LSF2 active site (Glc4-Glc5 and

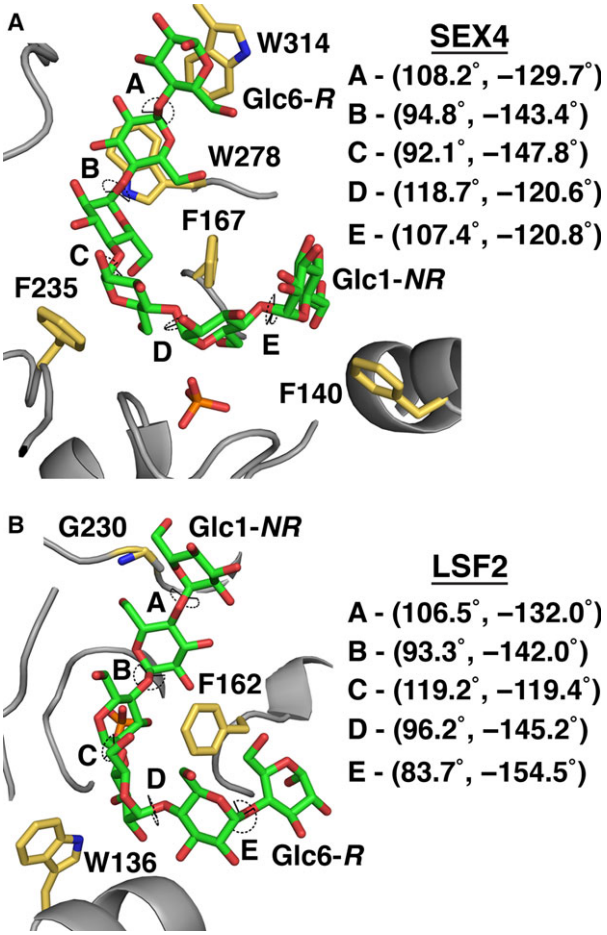

Fig. 7. Directionality and torsion angles of glucans at the SEX4/LSF2 active site. (A) Dihedral torsion angles of the glucan chain (green) at the SEX4 active site (PDB ID 4PYH [6]). Interacting residues are labeled in yellow. '*R*' represents the reducing end of the glucan chain, and '*NR*' represents the non-reducing end. (B) Dihedral torsion angles of the glucan chain (green) at the LSF2 active site (PDB ID 4KYR [7]). Interacting residues are labeled in yellow.

Glc5-Glc6) have more helical dihedral torsion angles. The only exception is the helical glycosidic bond of LSF2 Glc2-Glc3, which appears to be influenced by hydrogen bonding contributions from G230. Taken together, these results indicate that hotspots of glucan interaction in the glucan phosphatases (DSP active site and CBM) promote linearization of the glucan chain, and regions that bind more loosely inherently accommodate the helical substrate. These results suggest an interesting interplay in glucan phosphatases between accommodating the natural conformation of the substrate and modifying this conformation for enzymatic purposes.

Although engagement of the helical α-glucan chains is similar in SEX4 and LSF2, the chains are engaged with opposing directionality (Fig. 7). In SEX4, the non-reducing end of the glucan chain is positioned at the V-loop and the reducing end is positioned at the R-motif [6]. 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 [7]. Starch granules are polarized, with the reducing end towards the interior of the granule and the non-reducing end towards the outside [10,96]. Thus, the structure data suggest that SEX4 and LSF2 may engage glucans from different ends of the chain within starch granules.

Glucan phosphatase substrate specificity

One of the most significant aspects of SEX4 and LSF2 glucan chain engagement is the orientation of the glucan moiety at the catalytic site. This orientation is important for substrate specificity, as the C6 hydroxyl and the C3 hydroxyl groups are located on opposite facets of the glucan chain. Therefore, the substrate specificity of SEX4 and LSF2 must arise from integrating the glucan chain in a C6- or C3-specific orientation.

The SEX4 structure shows that the maltoheptaose chain is clearly positioned in a C6-specific orientation at the catalytic site (Fig. 8A) [6]. The O6 group of Glc2 interacts with the phosphate in the catalytic site at a distance of 2.6 Å, compared with 7.1 Å 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 catalytic triad [S(C)198, R204 and D166] is proximal to the Glc2 O6 and phosphate, and thus poised for catalysis. Interestingly, this orientation is reversed at the SEX4 CBM, where the C3 hydroxyl is positioned towards the body of the protein and the C6 hydroxyl points towards the solvent. Therefore, between the CBM and DSP, SEX4 is structured to interact with opposite facets of the glucan chain. The C6 orientation of the glucan ligand within the SEX4 structure confirms the substrate specificity indicated by the enzymatic data.

The opposite is found in the LSF2 structure, in which the maltohexaose chain is clearly positioned in a C3-specific orientation at the catalytic site (Fig. 8B) [7]. The O3 group of Glc3 interacts with the phosphate in the active site at a distance of 2.4 Å, compared with 7.0 Å 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 catalytic triad is also proximal to the Glc3 O3 group and phosphate,

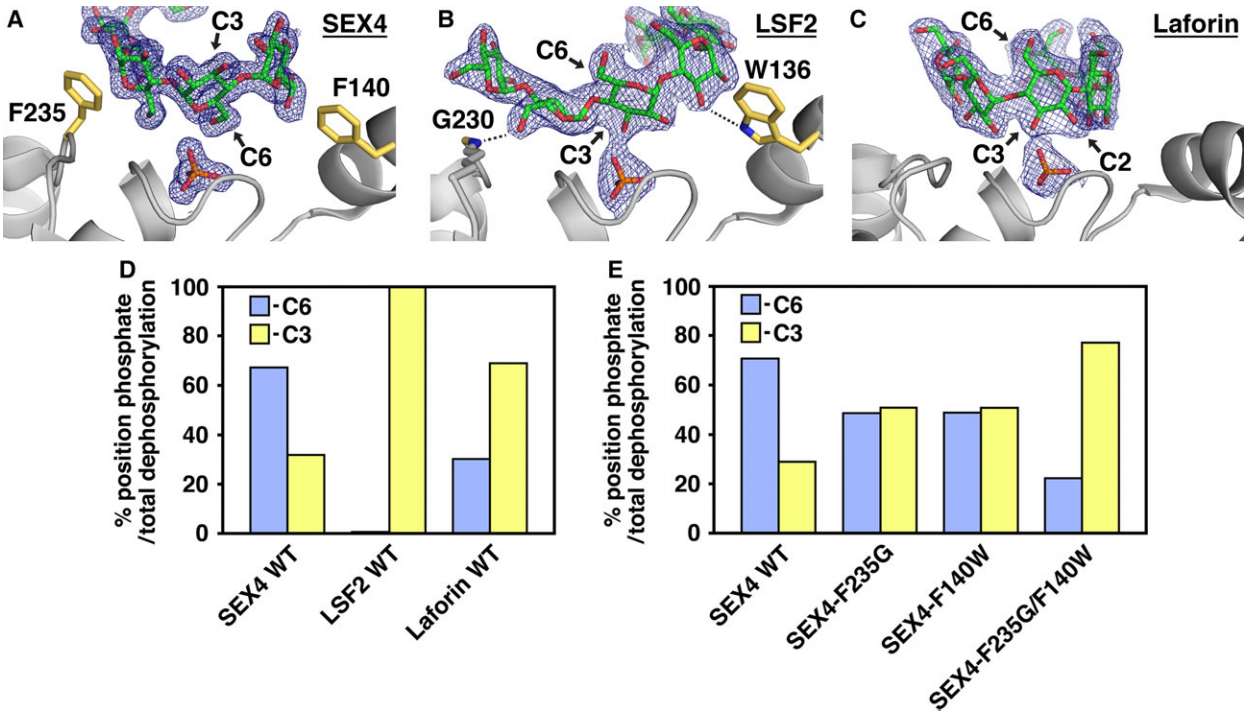

Fig. 8. Substrate specificity of plant glucan phosphatases. (A–C) Active sites of (A) SEX4 (PDB ID 4PYH [6]), (B) LSF2 (PDB ID 4KYR [7]) and (C) laforin (PDB ID 4RKK [97]), showing the 2 $F_o - F_c$ electron density (1.0σ, blue mesh) of the glucan chain (green) and phosphate (orange). The orientation of the glucan moiety at the catalytic site (C6 and C3 position) is indicated. Residues contributing to substrate specificity in SEX4 and LSF2 are shown in yellow. (D) Relative specific activity of SEX4, LSF2 and Laforin at the C6 (blue) and C3 (yellow) positions of Arabidopsis starch, represented as the percentage of total phosphate position dephosphorylation per minute per μg protein. This figure has been modified from ref. [60]. (E) Relative specific activity of SEX4 active-site mutants at the C6 (blue) and C3 (yellow) positions of Arabidopsis starch, represented as the percentage of total phosphate position dephosphorylation per minute per μg protein. This figure has been modified from ref. [6].

and is therefore poised for catalysis. As with SEX4, the LSF2 structure also provides a biophysical confirmation of the enzymatic data, indicating exclusive C3-specific activity. Additional work by members of the Zeeman laboratory has demonstrated increased C3-bound phosphate in *lsf2* plants [58]. Thus, *in planta*, biochemical and structural data all converge to define the activity of the glucan phosphatases.

Laforin was also co-crystallized in the presence of maltohexaose, and a bound maltohexaose chain and phosphate were integrated in the DSP domain active site [97]. As predicted, the laforin DSP domain is structurally most similar to SEX4 and LSF2, with RMSDs of 2.0 and 2.1 Å, respectively. The O2 and O3 groups of Glc3 are positioned 3.8 and 2.7 Å from the phosphate bound in the active site, while the O6 position is 7.7 Å away and points towards the solvent (Fig. 8C). Utilizing the position-specific assay, we determined that laforin preferentially dephosphorylates C3 hydroxyls, and possesses substantial activity towards the C6 position, almost the exact opposite of the SEX4 specificity data (Fig. 8D).

The C6-oriented glucan ligand within the SEX4 structure and the C3-oriented glucan ligand within the LSF2 structure provided an opportunity for structural analysis to determine the underlying basis for the differences in substrate specificity in the glucan phosphatases. The most obvious reason for the different substrate specificity is the fundamental difference in the glucan-binding mechanism used by SEX4 and LSF2. SEX4 uses a CBM to interact with glucan chains, and LSF2 uses SBSs, and therefore these external factors (with respect to the active site) may drive substrate specificity. However, further experiments demonstrated that the CBM and SBSs have no effect on substrate specificity [60]. Mutation of the CBM or SBSs, or removal of the CBM altogether, did not affect the ratio between C6 and C3 dephosphorylation in SEX4 and LSF2 [60].

If the binding mechanism used by the glucan phosphatases is not responsible for glucan orientation, then the DSP active site itself must define substrate specificity. The SEX4 and LSF2 glucan phosphatase active sites must have distinct structural differences that serve

to position the glucan chain in a C6- or C3-specific orientation. In the previous section, we outlined similar features of the glucan phosphatase DSP domain that account for its unique ability to bind glucans, namely its network of aromatic residues and wide/shallow topology. However, to unravel the differences in substrate specificity, differences between the active site of SEX4 and LSF2 were investigated.

In SEX4 and LSF2, different residues constitute the active-site boundaries and serve to tailor the binding capabilities and shape of the active site to accommodate their respective specificities. In SEX4, F235 and F140 form van der Waals interactions with the glucose moieties upstream and downstream of the active-site moiety, respectively (Fig. 8A) [6]. Both of these residues are altered in LSF2, in which G230 and W136 form hydrogen-bonding interactions with the O3 groups of the glucan chain (Fig. 8B) [7]. In addition, the residue differences affect the active-site topology. The absence of a side chain at G230 in LSF2 results in a small pocket, whereas the β-carbon of F235 in SEX4 results in a distinct ridge in the same area (Fig. 9) [6].

These differences create distinct shapes, and were found to contribute to substrate specificity in plant glucan phosphatases (Fig. 8E) [6]. Single mutations of F235 or F140 in SEX4 to the corresponding residues found in LSF2 abolished substrate specificity, with C6 and C3 phosphates removed at equal rates. In addition, double mutation of both residues resulted in a complete reversal of substrate specificity in SEX4 from the C6 to the C3 position, with the C3 position dephosphorylated at an approximately threefold higher rate than the C6 position. These results indicate that the substrate specificity of the glucan phosphatases is influenced by elements in the DSP active site in both enzymes. We recently utilized soluble and insoluble glucan substrates with a series of SEX4, LSF2 and laforin mutants and chimeras to address this hypothesis. We found that the glucan phosphatase DSP domain is sufficient for dephosphorylation of soluble polyglucans [60]. Conversely, the ancillary CBM domain and SBSs are required for efficient dephosphorylation of insoluble polyglucans [60].

In addition to determining the chemistry of the active site, the PTP sub-domains also dictate the overall topology or shape of the phosphatase active site. It has previously been established that the active-site topology of PTPs and DSPs is tailored to their respective target substrates in terms of width and depth [4]. This topology is normally defined by the surface of the enzyme from the R-motif to the V-loop, with the catalytic PTP-loop directly in the center of these two sub-

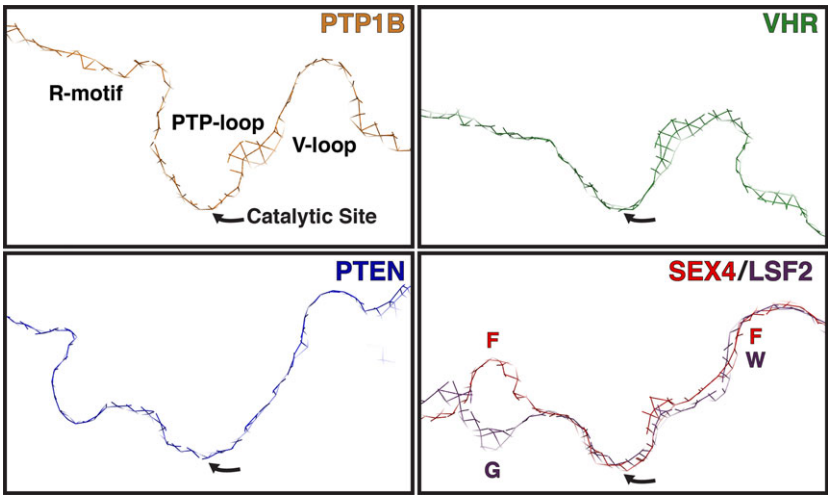

Fig. 9. Active-site topology of PTP superfamily members. The active-site topology of representative PTP superfamily members as represented by the cross-section of the active site formed by the R-motif, PTP-loop and V-loop surface, whose general positions are labeled in the PTP1B panel. The catalytic site is located at the lowest trough of the active-site pocket. PTP1B (orange, PDB ID 1EEN [104]) is a canonical protein tyrosine phosphatase that dephosphorylates p-Tyr exclusively and contains a narrow, deep active site. VHR (green, PDB ID 1VHR [5]) is a dual-specificity phosphatase that dephosphorylates p-Tyr and p-Ser/Thr residues and has a wide, shallow active site. PTEN (blue, PDB ID 1D5R [4]) is a dual-specificity phosphatase that dephosphorylates the lipid phosphatidylinositol-3,4,5-triphosphate and has a wide, deep active site. The glucan phosphatases SEX4 (red, PDB ID 4PYH [6]) and LSF2 (purple, PDB ID 4KYR [7]) have wide, shallow active sites compared to the other PTP superfamily members. The SEX4 and LSF2 active sites are relatively similar with the exception of the R-motif region, which forms a distinct ridge in SEX4 and a pocket in LSF2. The position of residues that contribute to the topology of the R-motif (F235 (F) in SEX4/G230 (G) in LSF2) and V-loop (F140 (F) in SEX4/W136 (W) in LSF2) are indicated.

domains [4]. Thus, p-Tyr-specific PTPs such as PTP1B contain a narrow and deep active site that permits access to p-Tyr residues, but excludes shorter p-Thr/Ser residues (Fig. 9) [98]. In contrast, dual-specificity phosphatases that dephosphorylate p-Tyr, p-Thr and p-Ser, such as VHR, generally have a shallower and wider active site that permits catalytic access to both short and long side chains [5]. The lipid phosphatase PTEN, which engages a large triphosphorylated phosphatidylinositol head group, contains a wide and deep active site compared to proteinaceous phosphatases [4]. Therefore, the characteristics of the target substrate closely match the topology of the phosphatases in the classical lock-and-key manner.

Comparing the active sites of SEX4 and LSF2, we found that the active sites of SEX4 and LSF2 are wide and shallow, allowing them to engage multiple glucose moieties of a long glucan chain (Fig. 9) [6]. The active site of SEX4 is most similar to that of PTEN, and that of LSF2 is even more shallow and wide. These results indicate that the non-proteinaceous DSPs typically contain an active site that may engage an extended substrate, rather than target individual protein side chains. Furthermore, the glucan phosphatases must engage multiple glucose units simultaneously and possibly 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.

An additional feature of glucan phosphatase activity that has not been explored is the influence of the LSF2 SBSs on the position of the glucan chain at the active site. It is very plausible that the three binding sites in LSF2 function cooperatively when binding complex carbohydrate substrates. The extended length of glucan chains in amylopectin and their close spatial relationship fit well with a model whereby two binding sites interact with a single glucan chain or two separate glucan chains simultaneously. Evidence for this relationship is provided by the consistent directionality, that is the position of reducing and non-reducing ends, observed between the glucan chains at the active site and site 2 (Fig. 4D). This consistency implies that these two sites are structured so as to bind a single glucan chain. Furthermore, two glucan chains were found at site 3, complexed into a helical conformation (Fig. 4B). This suggests that site 3 is capable of binding helical portions of the amylopectin superstructure that have yet to be phosphorylated or solubilized. More dynamic biophysical investigations into the LSF2 interaction with complex glucans will shed light on the more complex and inter-connected nature of its structural mechanism of activity.

Lastly, there are currently no structural or biochemical data regarding how SEX4 and LSF2 accommodate additional features of the starch granule, such as linear phosphoglucan chains on the granule surface, branches, adjacent glucan chains, residual helical structures, and the radial orientation of the starch granule, which is formed from reducing to non-reducing end. The position of phosphate with respect to branch points is also unknown, but correlation between branch points and phosphotylation may easily be imagined, possibly as a mechanism for specific phosphate positioning by GWD and/or PWD. It is likely that SEX4 and LSF2 are equipped to either accommodate branches or to circumvent them to more efficiently locate phosphate groups in the starch granule. Investigation into this area will provide a more global view of how SEX4 and LSF2 interact with some of the more complex elements of the starch granule superstructure.

In summary, the plant glucan phosphatases SEX4 and LSF2 engage glucans in a manner conducive to their natural helical conformation, and each bind the glucan chain in opposite directions and orientations. The basis for opposing orientations of the glucan chain at the active site of SEX4 and LSF2 is due to the presence of different residues in the DSP active site. This difference in active-site structure accounts for the difference in substrate specificity between the two glucan phosphatases.

Summary and future directions

Glucan phosphatases represent a new subset of PTPs that are essential for complex carbohydrate metabolism. Recent structural and enzymatic characterization of SEX4 and LSF2 provides detailed information on the physical basis for starch interaction and dephosphorylation. SEX4 binds glucans via an extended CBM/DSP domain interface that couples strong glucan binding at the active site with phosphoglucan integration into the catalytic site by the DSP. Conversely, LSF2 contains SBSs that non-catalytically interact with glucan chains, and a more strongly binding DSP active site that integrates the glucans into the catalytic site. A comparison of SEX4 and LSF2 with other DSPs indicates that the glucan phosphatase DSP domain contains a unique network of aromatic residues that function as glucan platforms, and a wide, shallow active site to accommodate three glucan moieties of a longer glucan chain. Lastly, SEX4 and LSF2 substrate specificity, which is essential for proper starch degradation, is based upon discrete elements within the DSP domain, and may be manipulated by simple mutagenesis in SEX4. This information represents a complete characterization of the fundamen-

tal enzymology of the plant glucan phosphatases. Although SEX4, LSF2, laforin and LSF1 are the only known members of the glucan phosphatase family, identification of future members will most likely reveal variations on the structural patterns described.

This initial characterization of the plant glucan phosphatases provides a framework for future research regarding finer details of their interaction with the complex starch granule. Additional insights into SEX4 and LSF2 interactions with branches, adjacent glucan chains and helical features of the starch granule are necessary to more fully understand how these enzymes deal with a complex and semi-accessible substrate. Starch, as an enzymatic target, contains multiple microenvironments that complicate binding, and understanding how glucan phosphatases are structured for efficient starch interaction may assist in the development of more efficient starch-hydrolyzing enzymes. This will also require an investigation into the possible processive mechanism used by the glucan phosphatases, which is currently not understood.

The current review defines the structural mechanisms of SEX4 and LSF2, and summarizes our current understanding of LSF1. While LSF1 is undoubtedly involved in starch metabolism, its precise role is currently unclear. What is interesting is that, while LSF1 resembles the glucan phosphatases SEX4, it lacks the glucan phosphatase signature motif and glucan phosphatase activity. Determination of the structural mechanism of SEX4 and LSF2 activity revealed that LSF1 lacks several key elements important to their function, including a PTP-loop consensus sequence, some of the aromatic network of residues in the DSP domain, and key specificity-determining residues. In addition, LSF1 contains elements that are not present in SEX4 and LSF2, namely a PDZ domain and a domain of unknown function. Determining the structure of LSF1 and further studies on its interactions with starch, and enzymes involved in starch metabolism, will reveal the exact nature of this enigmatic enzyme and possibly provide more biotechnological tools to further exert molecular control over starch metabolism.

As critical regulators of starch metabolism, the plant glucan phosphatases are promising targets for harnessing plant energy metabolism to increase starch yield in crops and improve the efficiency of starch processing for industrial purposes. Due to the recent identification of the plant glucan phosphatases, their genetic manipulation has not yet been fully explored in crops. However, silencing of GWD has recently been shown to positively affect starch yield and biomass in wheat plants [99], providing a powerful proof of principle for use of the glucan phosphatases. In addition, studies on the expression, conservation and function of SEX4 in barley (*Hordeum vulgare*) [100], rice (*Oryza sativa*) [101], maize (*Zea mays*) [102] and chestnut (Castanea sativa) [103] have been performed. However, the role of glucan phosphatases in storage starch metabolism or any link to plant pathology is currently unclear. As biotechnological tools advance in the coming years, understanding of the precise structural mechanism of the plant glucan phosphatase interaction with starch may provide even greater potential for metabolic control. In addition, the multitude of non-food applications of starch often require diverse chemical properties that require extensive processing. Using the glucan phosphatases to enzymatically alter the phosphorylation pattern on starch or to function in starch breakdown may yield industry-wide improvements.

Acknowledgements

Our work discussed in this review was supported by US National Institutes of Health grants R01NS070899, Kentucky Science and Engineering Foundation grants KSEF-2268RDE-014 and KSEF-2971-RDE-017, a Mitzutani Foundation for Glycoscience award, and US National Science Foundation grants IIA-1355438 and MCB-1252345. M.S.G. and C.W.V.K. are founders of OptiMol Enzymes LLC.

Author contribution

DAM, CVK, and MSG each analyzed data, conceived ideas, and wrote the paper. DAM generated figures.

References

1 Tonks NK (2013) Protein tyrosine phosphatases – from housekeeping enzymes to master regulators of signal transduction. *FEBS J* **280**, 346–378.
2 Alonso A, Rojas A, Godzik A & Mustelin T (2003) The Dual-Specific Protein Tyrosine Phosphatase Family. Springer, Berlin.
3 Silver DM, Kotting O & Moorhead GB (2014) Phosphoglucan phosphatase function sheds light on starch degradation. *Trends Plant Sci* **19**, 471–478.
4 Lee JO, Yang H, Georgescu MM, Di Cristofano A, Maehama T, Shi Y, Dixon JE, Pandolfi P & Pavletich NP (1999) Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association. *Cell* **99**, 323–334.
5 Yuvaniyama J, Denu JM, Dixon JE & Saper MA (1996) Crystal structure of the dual specificity protein phosphatase VHR. *Science* **272**, 1328–1331.

6 Meekins DA, Raththagala M, Husodo S, White CJ, Guo HF, Kotting O, Vander Kooi CW & Gentry MS (2014) Phosphoglucan-bound structure of starch phosphatase Starch Excess4 reveals the mechanism for C6 specificity. *Proc Natl Acad Sci USA* **111**, 7272–7277.

7 Meekins DA, Guo HF, Husodo S, Paasch BC, Bridges TM, Santelia D, Kotting O, Vander Kooi CW & Gentry MS (2013) Structure of the Arabidopsis glucan phosphatase LIKE SEX FOUR2 reveals a unique mechanism for starch dephosphorylation. *Plant Cell* **25**, 2302–2314.

8 Vander Kooi CW, Taylor AO, Pace RM, Meekins DA, Guo HF, Kim Y & Gentry MS (2010) Structural basis for the glucan phosphatase activity of Starch Excess4. *Proc Natl Acad Sci USA* **107**, 15379–15384.

9 Smirnova J, Fernie AR & Steup M (2015) Starch degradation. In Starch: Metabolism and Structure (Nakamura Y, ed.), pp. 239–290. Springer, Japan.

10 Streb S & Zeeman SC (2012) Starch metabolism in Arabidopsis. *Arabidopsis Book* **10**, e0160.

11 Hejazi M, Fettke J & Steup M (2012) Starch phosphorylation and dephosphorylation: the consecutive action of starch-related dikinases and phosphatases. In Starch: Origins, Structure and Metabolism (Tetlow IJ, ed.), pp. 279–308. Society for Experimental Biology, London.

12 Zeeman SC, Kossmann J & Smith AM (2010) Starch: its metabolism, evolution, and biotechnological modification in plants. *Annu Rev Plant Biol* **61**, 209–234.

13 Fettke J, Fernie AR & Steup M (2012) Transitory starch and its degradation in higher plant cells. In Starch: Origins, Structure and Metabolism (Tetlow IJ, ed.), pp. 309–372. Society for Experimental Biology, London.

14 Tester RF, Karkalas J & Qi X (2004) Starch – composition, fine structure and architecture. *J Cereal Sci* **39**, 151–165.

15 Bertoft E & Seetharaman K (2012) Starch structure. In Starch: Origins, Structure and Metabolism (Tetlow IJ, ed.), pp. 1–28. Society for Experimental Biology, London.

16 Buleon A, Colonna P, Planchot V & Ball S (1998) Starch granules: structure and biosynthesis. *Int J Biol Macromol* **23**, 85–112.

17 Ball SG & Morell MK (2003) From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. *Annu Rev Plant Biol* **54**, 207–233.

18 Coppin A, Varre JS, Lienard L, Dauvillee D, Guerardel Y, Soyer-Gobillard MO, Buleon A, Ball S & Tomavo S (2005) Evolution of plant-like crystalline storage polysaccharide in the protozoan parasite *Toxoplasma gondii* argues for a red alga ancestry. *J Mol Evol* **60**, 257–267.

19 Meeuse BJD, Andries M & Wood JA (1960) Floridean starch. *J Exp Bot* **11**, 129–140.

20 Viola R, Nyvall P & Pedersen M (2001) The unique features of starch metabolism in red algae. *Proc Biol Sci* **268**, 1417–1422.

21 Gentry MS, Dowen RH III, Worby CA, Mattoo S, Ecker JR & Dixon JE (2007) The phosphatase laforin crosses evolutionary boundaries and links carbohydrate metabolism to neuronal disease. *J Cell Biol* **178**, 477–488.

22 Gentry MS, Dixon JE & Worby CA (2009) Lafora disease: insights into neurodegeneration from plant metabolism. *Trends Biochem Sci* **34**, 628–639.

23 Roach PJ, Depaoli-Roach AA, Hurley TD & Tagliabracci VS (2012) Glycogen and its metabolism: some new developments and old themes. *Biochem J* **441**, 763–787.

24 Zeeman SC, Tiessen A, Pilling E, Kato KL, Donald AM & Smith AM (2002) Starch synthesis in Arabidopsis. Granule synthesis, composition, and structure. *Plant Physiol* **129**, 516–529.

25 Smith AM, Zeeman SC & Smith SM (2005) Starch degradation. *Annu Rev Plant Biol* **56**, 73–98.

26 Blennow A, Nielsen TH, Baunsgaard L, Mikkelsen R & Engelsen SB (2002) Starch phosphorylation: a new front line in starch research. *Trends Plant Sci* **7**, 445–450.

27 Kotting O, Kossmann J, Zeeman SC & Lloyd JR (2010) Regulation of starch metabolism: the age of enlightenment? *Curr Opin Plant Biol* **13**, 321–329.

28 Blennow A & Engelsen SB (2010) Helix-breaking news: fighting crystalline starch energy deposits in the cell. *Trends Plant Sci* **15**, 236–240.

29 Fernbach A (1904) Quelques observations sur la composition de l'amidon de pommes de terre. *C R Acad Sci* **138**, 428–430.

30 Ritte G, Lloyd JR, Eckermann N, Rottmann A, Kossmann J & Steup M (2002) The starch-related R1 protein is an α–glucan, water dikinase. *Proc Natl Acad Sci USA* **99**, 7166–7171.

31 Lorberth R, Ritte G, Willmitzer L & Kossmann J (1998) Inhibition of a starch-granule-bound protein leads to modified starch and repression of cold sweetening. *Nat Biotechnol* **16**, 473–477.

32 Coehn A (1897) Über elektrische Wanderung von Kolloiden. *Z Elektrochem* **4**, 63–67.

33 Caspar T, Lin TP, Kakefuda G, Benbow L, Preiss J & Somerville C (1991) Mutants of Arabidopsis with altered regulation of starch degradation. *Plant Physiol* **95**, 1181–1188.

34 Zeeman SC, Northrop F, Smith AM & Rees T (1998) A starch-accumulating mutant of *Arabidopsis thaliana* deficient in a chloroplastic starch-hydrolysing enzyme. *Plant J* **15**, 357–365.

35 Yu TS, Kofler H, Hausler RE, Hille D, Flugge UI, Zeeman SC, Smith AM, Kossmann J, Lloyd J, Ritte G *et al.* (2001) The Arabidopsis *sex1* mutant is defective in the R1 protein, a general regulator of starch degradation in plants, and not in the chloroplast hexose transporter. *Plant Cell* **13**, 1907–1918.
36 Blennow A (2015) Phosphorylation of the starch granule. In Starch: Metabolism and Structure (Nakamura Y, ed.), pp. 399–424. Springer, Japan.
37 Baunsgaard L, Lutken H, Mikkelsen R, Glaring MA, Pham TT & Blennow A (2005) A novel isoform of glucan, water dikinase phosphorylates pre-phosphorylated α-glucans and is involved in starch degradation in Arabidopsis. *Plant J* **41**, 595–605.
38 Ritte G, Steup M, Kossmann J & Lloyd JR (2003) Determination of the starch-phosphorylating enzyme activity in plant extracts. *Planta* **216**, 798–801.
39 Ritte G, Heydenreich M, Mahlow S, Haebel S, Kotting O & Steup M (2006) Phosphorylation of C6- and C3-positions of glucosyl residues in starch is catalysed by distinct dikinases. *FEBS Lett* **580**, 4872–4876.
40 Kotting O, Pusch K, Tiessen A, Geigenberger P, Steup M & Ritte G (2005) Identification of a novel enzyme required for starch metabolism in Arabidopsis leaves. The phosphoglucan, water dikinase. *Plant Physiol* **137**, 242–252.
41 Skeffington AW, Graf A, Duxbury Z, Gruissem W & Smith AM (2014) Glucan, water dikinase exerts little control over starch degradation in Arabidopsis leaves at night. *Plant Physiol* **165**, 866–879.
42 Mahlow S, Hejazi M, Kuhnert F, Garz A, Brust H, Baumann O & Fettke J (2014) Phosphorylation of transitory starch by α–glucan, water dikinase during starch turnover affects the surface properties and morphology of starch granules. *New Phytol* **203**, 495–507.
43 Hejazi M, Fettke J, Paris O & Steup M (2009) The two plastidial starch-related dikinases sequentially phosphorylate glucosyl residues at the surface of both the A- and B-type allomorphs of crystallized maltodextrins but the mode of action differs. *Plant Physiol* **150**, 962–976.
44 Hansen PI, Larsen FH, Motawia SM, Blennow A, Spraul M, Dvortsak P & Engelsen SB (2008) Structure and hydration of the amylopectin trisaccharide building blocks – synthesis, NMR, and molecular dynamics. *Biopolymers* **89**, 1179–1193.
45 Hansen PI, Spraul M, Dvortsak P, Larsen FH, Blennow A, Motawia MS & Engelsen SB (2009) Starch phosphorylation – maltosidic restrains upon 3′- and 6′-phosphorylation investigated by chemical synthesis, molecular dynamics and NMR spectroscopy. *Biopolymers* **91**, 179–193.
46 Edner C, Li J, Albrecht T, Mahlow S, Hejazi M, Hussain H, Kaplan F, Guy C, Smith SM, Steup M *et al.* (2007) Glucan, water dikinase activity stimulates breakdown of starch granules by plastidial β–amylases. *Plant Physiol* **145**, 17–28.
47 Hejazi M, Fettke J, Haebel S, Edner C, Paris O, Frohberg C, Steup M & Ritte G (2008) Glucan, water dikinase phosphorylates crystalline maltodextrins and thereby initiates solubilization. *Plant J* **55**, 323–334.
48 Hejazi M, Mahlow S & Fettke J (2014) The glucan phosphorylation mediated by α–glucan, water dikinase (GWD) is also essential in the light phase for a functional transitory starch turn-over. *Plant Signal Behav* **9**, e28892.
49 Gentry MS & Pace RM (2009) Conservation of the glucan phosphatase laforin is linked to rates of molecular evolution and the glucan metabolism of the organism. *BMC Evol Biol* **9**, 138.
50 Kerk D, Conley TR, Rodriguez FA, Tran HT, Nimick M, Muench DG & Moorhead GB (2006) A chloroplast-localized dual-specificity protein phosphatase in Arabidopsis contains a phylogenetically dispersed and ancient carbohydrate-binding domain, which binds the polysaccharide starch. *Plant J* **46**, 400–413.
51 Moorhead GB, De Wever V, Templeton G & Kerk D (2009) Evolution of protein phosphatases in plants and animals. *Biochem J* **417**, 401–409.
52 Fordham-Skelton AP, Chilley P, Lumbreras V, Reignoux S, Fenton TR, Dahm CC, Pages M & Gatehouse JA (2002) A novel higher plant protein tyrosine phosphatase interacts with SNF1-related protein kinases via a KIS (kinase interaction sequence) domain. *Plant J* **29**, 705–715.
53 Niittyla T, Comparot-Moss S, Lue W-L, Messerli G, Trevisan M, Seymour MDJ, Gatehouse JA, Villadsen D, Smith SM, Chen J *et al.* (2006) Similar protein phosphatases control starch metabolism in plants and glycogen metabolism in mammals. *J Biol Chem* **281**, 11815–11818.
54 Hejazi M, Fettke J, Kotting O, Zeeman SC & Steup M (2010) The Laforin-like dual-specificity phosphatase SEX4 from Arabidopsis hydrolyzes both C6- and C3-phosphate esters introduced by starch-related dikinases and thereby affects phase transition of α-glucans. *Plant Physiol* **152**, 711–722.
55 Kotting O, Santelia D, Edner C, Eicke S, Marthaler T, Gentry MS, Comparot-Moss S, Chen J, Smith AM, Steup M *et al.* (2009) STARCH-EXCESS4 is a laforin-like phosphoglucan phosphatase required for starch degradation in *Arabidopsis thaliana*. *Plant Cell* **21**, 334–346.
56 Takeda Y & Hizukuri S (1981) Studies on starch phosphate. Part 5. Reexamination of the action of

sweet-potato β–amylase on phophorylated (1→4)-α–D-glucan. *Carbohydr Res* **39**, 151–165.
57 Comparot-Moss S, Kotting O, Stettler M, Edner C, Graf A, Weise SE, Streb S, Lue WL, MacLean D, Mahlow S *et al*. (2010) A putative phosphatase, LSF1, is required for normal starch turnover in Arabidopsis leaves. *Plant Physiol* **152**, 685–697.
58 Santelia D, Kotting O, Seung D, Schubert M, Thalmann M, Bischof S, Meekins DA, Lutz A, Patron N, Gentry MS *et al*. (2011) The phosphoglucan phosphatase Like Sex Four2 dephosphorylates starch at the C3-position in Arabidopsis. *Plant Cell* **23**, 4096–4111.
59 Worby CA, Gentry MS & Dixon JE (2006) Laforin, a dual specificity phosphatase that dephosphorylates complex carbohydrates. *J Biol Chem* **281**, 30412–30418.
60 Meekins DA, Raththagala M, Auger KD, Turner BD, Santelia D, Kotting O, Gentry MS & Vander Kooi CW (2015) Mechanistic insights into glucan phosphatase activity against polyglucan substrates. *J Biol Chem* **290**, 23361–23370.
61 Janecek S, Svensson B & MacGregor EA (2011) Structural and evolutionary aspects of two families of non-catalytic domains present in starch and glycogen binding proteins from microbes, plants and animals. *Enzyme Microb Technol* **49**, 429–440.
62 Guillen D, Sanchez S & Rodriguez-Sanoja R (2010) Carbohydrate-binding domains: multiplicity of biological roles. *Appl Microbiol Biotechnol* **85**, 1241–1249.
63 Abbott DW & van Bueren AL (2014) Using structure to inform carbohydrate binding module function. *Curr Opin Struct Biol* **28**, 32–40.
64 Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V & Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37**, D233–D238.
65 Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM & Henrissat B (2013) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490–D495.
66 Christiansen C, Hachem MA, Glaring MA, Vikso-Nielsen A, Sigurskjold BW, Svensson B & Blennow A (2009) A CBM20 low-affinity starch-binding domain from glucan, water dikinase. *FEBS Lett* **583**, 1159–1163.
67 Machovic M & Janecek S (2006) Starch-binding domains in the post-genome era. *Cell Mol Life Sci* **63**, 2710–2724.
68 Polekhina G, Gupta A, van Denderen BJ, Feil SC, Kemp BE, Stapleton D & Parker MW (2005) Structural basis for glycogen recognition by AMP-activated protein kinase. *Structure* **13**, 1453–1462.
69 Minassian BA, Ianzano L, Meloche M, Andermann E, Rouleau GA, Delgado-Escueta AV & Scherer SW (2000) Mutation spectrum and predicted function of laforin in Lafora's progressive myoclonus epilepsy. *Neurology* **55**, 341–346.
70 Ganesh S, Tsurutani N, Suzuki T, Hoshii Y, Ishihara T, Delgado-Escueta AV & Yamakawa K (2004) The carbohydrate-binding domain of Lafora disease protein targets Lafora polyglucosan bodies. *Biochem Biophys Res Commun* **313**, 1101–1109.
71 Tagliabracci VS, Turnbull J, Wang W, Girard JM, Zhao X, Skurat AV, Delgado-Escueta AV, Minassian BA, Depaoli-Roach AA & Roach PJ (2007) Laforin is a glycogen phosphatase, deficiency of which leads to elevated phosphorylation of glycogen *in vivo*. *Proc Natl Acad Sci USA* **104**, 19262–19266.
72 Tagliabracci VS, Girard JM, Segvich D, Meyer C, Turnbull J, Zhao X, Minassian BA, Depaoli-Roach AA & Roach PJ (2008) Abnormal metabolism of glycogen phosphate as a cause for Lafora disease. *J Biol Chem* **283**, 33816–33825.
73 Gentry MS, Roma-Mateo C & Sanz P (2013) Laforin, a protein with many faces: glucan phosphatase, adapter protein, et alii. *FEBS J* **280**, 525–537.
74 Roach PJ (2015) Glycogen phosphorylation and Lafora disease. *Mol Aspects Med* **46**, 78–84.
75 Tonks NK (2006) Protein tyrosine phosphatases: from genes, to function, to disease. *Nat Rev Mol Cell Biol* **7**, 833–846.
76 Matsuda M, Takeshita K, Kurokawa T, Sakata S, Suzuki M, Yamashita E, Okamura Y & Nakagawa A (2011) Crystal structure of the cytoplasmic phosphatase and tensin homolog (PTEN)-like region of *Ciona intestinalis* voltage-sensing phosphatase provides insight into substrate specificity and redox regulation of the phosphoinositide phosphatase activity. *J Biol Chem* **286**, 23368–23377.
77 Guan R, Dai H, Harrison SC & Kirchhausen T (2010) Structure of the PTEN-like region of auxilin, a detector of clathrin-coated vesicle budding. *Structure* **18**, 1191–1198.
78 Begley MJ, Taylor GS, Kim SA, Veine DM, Dixon JE & Stuckey JA (2003) Crystal structure of a phosphoinositide phosphatase, MTMR2: insights into myotubular myopathy and Charcot–Marie–Tooth syndrome. *Mol Cell* **12**, 1391–1402.
79 Cuyvers S, Dornez E, Delcour JA & Courtin CM (2012) Occurrence and functional significance of secondary carbohydrate binding sites in glycoside hydrolases. *Crit Rev Biotechnol* **32**, 93–107.
80 Chu HM & Wang AH (2007) Enzyme–substrate interactions revealed by the crystal structures of the archaeal *Sulfolobus* PTP-fold phosphatase and its phosphopeptide complexes. *Proteins* **66**, 996–1003.

81 Almo SC, Bonanno JB, Sauder JM, Emtage S, Dilorenzo TP, Malashkevich V, Wasserman SR, Swaminathan S, Eswaramoorthy S, Agarwal R *et al.* (2007) Structural genomics of protein phosphatases. *J Struct Funct Genomics* **8**, 121–140.
82 Song H, Hanlon N, Brown NR, Noble ME, Johnson LN & Barford D (2001) Phosphoprotein–protein interactions revealed by the crystal structure of kinase-associated phosphatase in complex with phosphoCDK2. *Mol Cell* **7**, 615–626.
83 Agarwal R, Burley SK & Swaminathan S (2008) Structure of human dual specificity protein phosphatase 23, VHZ, enzyme–substrate/product complex. *J Biol Chem* **283**, 8946–8953.
84 Lountos GT, Tropea JE & Waugh DS (2011) Structure of human dual-specificity phosphatase 27 at 2.38 Å resolution. *Acta Crystallogr D Biol Crystallogr* **67**, 471–479.
85 Xiao J, Engel JL, Zhang J, Chen MJ, Manning G & Dixon JE (2011) Structural and functional analysis of PTPMT1, a phosphatase required for cardiolipin synthesis. *Proc Natl Acad Sci USA* **108**, 11860–11865.
86 Wishart MJ & Dixon JE (1998) Gathering STYX: phosphatase-like form predicts functions for unique protein-interaction domains. *Trends Biochem Sci* **23**, 301–306.
87 Salmeen A & Barford D (2005) Functions and mechanisms of redox regulation of cysteine-based phosphatases. *Antioxid Redox Signal* **7**, 560–577.
88 Silver DM, Silva LP, Issakidis-Bourguet E, Glaring MA, Schriemer DC & Moorhead GB (2013) Insight into the redox regulation of the phosphoglucan phosphatase SEX4 involved in starch degradation. *FEBS J* **280**, 538–548.
89 Amoasii L, Hnia K & Laporte J (2012) Myotubularin phosphoinositide phosphatases in human diseases. *Curr Top Microbiol Immunol* **362**, 209–233.
90 Hsu F & Mao Y (2015) The structure of phosphoinositide phosphatases: insights into substrate specificity and catalysis. *Biochim Biophys Acta* **1851**, 698–710.
91 Cho H, Ramer SE, Itoh M, Kitas E, Bannwarth W, Burn P, Saito H & Walsh CT (1992) Catalytic domains of the LAR and CD45 protein tyrosine phosphatases from *Escherichia coli* expression systems: purification and characterization for specificity and mechanism. *Biochemistry* **31**, 133–138.
92 Guan KL, Broyles SS & Dixon JE (1991) A Tyr/Ser protein phosphatase encoded by vaccinia virus. *Nature* **350**, 359–362.
93 Robert X, Haser R, Mori H, Svensson B & Aghajari N (2005) Oligosaccharide binding to barley α-amylase 1. *J Biol Chem* **280**, 32968–32978.
94 Koropatkin NM & Smith TJ (2010) SusG: a unique cell-membrane-associated α-amylase from a prominent human gut symbiont targets complex starch molecules. *Structure* **18**, 200–215.
95 Imberty A, Chanzy H, Perez S, Buleon A & Tran V (1988) The double-helical nature of the crystalline part of A–starch. *J Mol Biol* **201**, 365–378.
96 Recondo E & Leloir LF (1961) Adenosine diphosphate glucose and starch synthesis. *Biochem Biophys Res Commun* **6**, 85–88.
97 Raththagala M, Brewer MK, Parker MW, Sherwood AR, Wong BK, Hsu S, Bridges TM, Paasch BC, Hellman LM, Husodo S *et al.* (2015) Structural mechanism of laforin function in glycogen dephosphorylation and lafora disease. *Mol Cell* **57**, 261–272.
98 Barford D, Flint AJ & Tonks NK (1994) Crystal structure of human protein tyrosine phosphatase 1B. *Science* **263**, 1397–1404.
99 Ral JP, Bowerman AF, Li Z, Sirault X, Furbank R, Pritchard JR, Bloemsma M, Cavanagh CR, Howitt CA & Morell MK (2012) Down-regulation of glucan, water-dikinase activity in wheat endosperm increases vegetative biomass and yield. *Plant Biotechnol J* **10**, 871–882.
100 Ma J, Jiang QT, Wei L, Yang Q, Zhang XW, Peng YY, Chen GY, Wei YM, Liu C & Zheng YL (2014) Conserved structure and varied expression reveal key roles of phosphoglucan phosphatase gene *starch excess 4* in barley. *Planta* **240**, 1179–1190.
101 Nunez-Lopez L, Aguirre-Cruz A, Barrera-Figueroa BE & Pena-Castro JM (2015) Improvement of enzymatic saccharification yield in *Arabidopsis thaliana* by ectopic expression of the rice SUB1A–1 transcription factor. *PeerJ* **3**, e817.
102 Weise SE, Aung K, Jarou ZJ, Mehrshahi P, Li Z, Hardy AC, Carr DJ & Sharkey TD (2012) Engineering starch accumulation by manipulation of phosphate metabolism of starch. *Plant Biotechnol J* **10**, 545–554.
103 Berrocal-Lobo M, Ibanez C, Acebo P, Ramos A, Perez-Solis E, Collada C, Casado R, Aragoncillo C & Allona I (2011) Identification of a homolog of Arabidopsis DSP4 (SEX4) in chestnut: its induction and accumulation in stem amyloplasts during winter or in response to the cold. *Plant Cell Environ* **34**, 1693–1704.
104 Sarmiento M, Puius YA, Vetter SW, Keng YF, Wu L, Zhao Y, Lawrence DS, Almo SC & Zhang ZY (2000) Structural basis of plasticity in protein tyrosine phosphatase 1B substrate recognition. *Biochemistry* **39**, 8171–8179.