Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of starch synthase I from barley: insight into regulatory mechanisms of starch synthase activity

Jose A. Cuesta-Seijo, Morten M. Nielsen, Lucia Marri, Hidenori Tanaka,‡ Sophie R. Beeren and Monica M. Palcic*

Carlsberg Laboratory, Gamle Carlsberg Vej 10, 1799 Copenhagen V, Denmark

‡ Current address: Max Planck Institute of Colloids and Interfaces, Department of Biomolecular Systems, Arnimallee 22, 14195 Berlin, Germany and Freie Universität Berlin, Institute of Chemistry and Biochemistry, Arnimallee 22, 14195 Berlin, Germany.

Correspondence e-mail: monica.palcic@carlsberglab.dk

Received 24 October 2012
Accepted 14 February 2013

PDB Reference: starch synthase I, 4hln

Starch, a polymer of glucose, is the major source of calories in the human diet. It has numerous industrial uses, including as a raw material for the production of first-generation bioethanol. Several classes of enzymes take part in starch biosynthesis, of which starch synthases (SSs) carry out chain elongation of both amylose and amylopectin. Plants have five classes of SS, each with different roles. The products of the reaction of SS are well known, but details of the reaction mechanism remain obscure and even less is known of how different SSs select different substrates for elongation, how they compete with each other and how their activities are regulated. Here, the first crystal structure of a soluble starch synthase is presented: that of starch synthase I (SSI) from barley refined to 2.7 Å resolution. The structure captures an open conformation of the enzyme with a surface-bound maltooligosaccharide and a disulfide bridge that precludes formation of the active site. The maltooligosaccharide-binding site is involved in substrate recognition, while the disulfide bridge is reflective of redox regulation of SSI. Activity measurements on several SSI mutants supporting these roles are also presented.

1. Introduction

Starch is the main storage polysaccharide of plants and the major source of calories in human nutrition (Zeeman *et al.*, 2010). It is also finding increasing industrial use in applications such as papermaking (Higson & Smith, 2011) and first-generation bioethanol production (Smith, 2008). 2.4 billion tonnes of cereals are produced yearly, of which 1100 million tonnes are destined directly for food use and 800 million tonnes indirectly as animal fodder; the other 500 million tonnes are used in industry or as seeds or go to waste. Additionally, more than 700 million tonnes of roots and tubers, which are also rich sources of starch, are produced annually (Food and Agriculture Organization of the United Nations, 2012). The industrial production of pure refined starch exceeded 60 million tonnes ten years ago (Tester & Karkalas, 2002).

Starch is synthesized in plastids in higher plants, where it serves as a store of both energy and carbon. In leaves it is stored as transient starch in chloroplasts during the day and is remobilized at night when photosynthesis stops (Smith & Stitt, 2007). Nonphotosynthetic cells use amyloplasts for long-term storage of starch in preparation for future metabolic demands of the plant, *i.e.* the establishment of seedlings (Fincher, 1989). This is the source of starch in cereals, in which approximately 70% of the grains (dry weight) consists of starch (Thitisaksakul *et al.*, 2012).

Starch in plastids takes the form of starch granules with different sizes, shapes and degrees of crystallinity. The

granules themselves are composed of two types of glucose polymers: linear amylose, in which almost exclusively α-D-1,4-glycosidic linkages are found, and the highly branched amylopectin, which comprises approximately 75% of the granule weight. In amylopectin, in addition to linear chains containing between six and more than 100 glucoses, approximately 5% α-D-1,6-glycosidic linkages (branching points) are found. This results in a hierarchical cluster structure (Pérez & Bertoft, 2010) and in the formation of crystalline lamellae: crystalline layers, perpendicular to the radial axis of the granules, that alternate with more amorphous layers with a periodicity of 9 nm.

This highly complex macromolecular aggregate is the result of the coordinated action of enzymes of several families, including ADP-glucose pyrophosphorylase, which produces the donor sugar ADP-glucose, starch synthases (SSs) that use the ADP-glucose for chain elongation *via* α-1,4-glycosidic linkages, branching enzymes that create the α-1,6-linkages and some of the debranching enzymes (Zeeman *et al.*, 2010; other enzymes are only involved in starch degradation). It has recently been reported that phosphorylase can also contribute to the elongation of linear glucose chains in amylose and amylopectin (Fettke *et al.*, 2012). The exact roles of each enzyme in creating the fine molecular structure of amylose and amylopectin, which determines the physicochemical properties of starch, are only poorly understood.

Starch synthases (ADP-glucose:1,4-α-D-glucan 4-α-D-glucosyltransferases; EC 2.4.1.21) are GT-B-fold glycosyltransferases classified within family GT5 in the CAZy database (http://www.cazy.org; Coutinho *et al.*, 2003). They all catalyze the same reaction: the addition of glucose from ADP-glucose to the nonreducing end of growing maltooligosaccharide (MOS) chains exclusively *via* α-1,4-glycosidic linkages with retention of the configuration of the transferred sugar (Zeeman *et al.*, 2010). It is the starch synthases that deposit the bulk of the sugars in the starch granules.

Plant starch synthases are divided into five different gene classes: an insoluble granule-bound starch synthase (GBSS) and the soluble starch synthases I, II, III and IV (SSI, SSII, SSIII and SSIV), with differing numbers of isoforms in different plants (Jeon *et al.*, 2010). Each class has different roles in starch synthesis resulting from their different physicochemical properties and substrate specificities.

GBSS is always found associated with the starch granule (the other gene classes are soluble or at least partially soluble; Zeeman *et al.*, 2010). When considering its physical location and its product amylose, GBSS is often considered separately from the other starch synthases. It is the enzyme responsible for the synthesis of amylose (Patron *et al.*, 2002). It may also play a role in the synthesis of the extra-long-chain fraction of amylopectin (Yoo & Jane, 2002). Recently, the crystal structure of GBSSI from rice (*Os*GBSSI) has been determined (Momma & Fujimoto, 2012) and revealed broad structural similarities between plant SSs and bacterial glycogen synthases (GSs).

Of the soluble starch synthases, SSI is exceptional in that most plants only have one isoform. SSI has a preference for very small linear glucose chains as substrates, which it elongates only to moderate sizes, preferentially to a DP (degree of polymerization or chain length) of 8–12 (Fujita *et al.*, 2006). It becomes entrapped and inactive in much longer chains (Commuri & Keeling, 2001). SSI is the SS expressed to the highest level in cereal endosperm and can account for up to 70% of the total starch-synthase activity in the soluble fraction of rice endosperm (Fujita *et al.*, 2006). SSII is responsible for the synthesis of glucose chains of medium length, with a preference for DP 13–25 (Zhang *et al.*, 2004; Nakamura *et al.*, 2005). SSIII is responsible for the synthesis of linear glucose chains with a DP of >30 (Fujita *et al.*, 2007) and possibly plays a role in starch-synthesis initiation (Szydlowski *et al.*, 2007). The role of SSIV *in vivo* is not well understood, but it is involved in regulation of the number of starch granules and also in starch-synthesis initiation (Szydlowski *et al.*, 2007). No structures are available to date for any of the soluble starch synthases.

The closest counterparts of SS in the GT5 family are the archaeal and bacterial glycogen synthases (GSs). Crystal structures are available for one archaeal GS (from *Pyrococcus abyssi*; *Pa*GS; Horcajada *et al.*, 2006) and two bacterial GSs from *Agrobacterium tumefaciens* (*At*GS; Buschiazzo *et al.*, 2004) and *Escherichia coli* (*Ec*GS; Sheng, Jia *et al.*, 2009; Sheng, Yep *et al.* 2009). Eukaryotic glycogen synthases (including human) belong to the GT3 family in the CAZy classification; as such, they are only distantly related to plant SSs. They use UDP-glucose as the nucleotide donor sugar rather than ADP-glucose as used by SSs.

The mechanism of glycosyl transfer remains elusive for the GT5 family and is still a matter of debate for retaining glycosyltransferases in general (Breton *et al.*, 2006, 2012; Lairson *et al.*, 2008). There is support for double-displacement mechanisms involving two S_N2 reactions with the formation of a covalent intermediate with a catalytic nucleophile from the enzyme (Soya *et al.*, 2011) and for S_Ni-like reactions (Lee *et al.*, 2011). For glycosyltransferases it is believed that the donor substrate binds first, triggering conformational changes in flexible loops along with binding of the acceptor (Qasba *et al.*, 2005). In the case of the GT5 family and starch synthases, the order of binding of the different substrates is unknown, although structures are available of glycogen synthases bound to a hydrolyzed donor (Buschiazzo *et al.*, 2004; Sheng, Jia *et al.*, 2009; Momma & Fujimoto, 2012).

Here, we present the crystal structure of starch synthase I from barley (*Hv*SSI) refined to 2.7 Å resolution. The structure revealed the expected broad similarities at the fold level to GBSSI from rice and to bacterial GSs, but also a number of novel features of *Hv*SSI. Two details stand out in our structure: a disulfide bridge that forms spontaneously and a surface MOS-binding site. The disulfide bridge disrupts the active-centre site of *Hv*SSI, but does so in a way that is probably representative of the actual situation *in planta*, given that *Arabidopsis thaliana* SSI activity is redox-regulated (Glaring *et al.*, 2012) and *Hv*SSI activity is redox-sensitive (this work). It is a solvent-exposed disulfide bridge that surprisingly resists reduction by high concentrations of DTT in the crystal.

The surface oligosaccharide-binding site, which is situated at a distance from the catalytic centre, has broad implications for enzyme localization and substrate selectivity during enzymatic synthesis. It is also of interest to note that the substrate bound to it was not added externally to the protein stock used for crystallization, but was rather synthesized *in situ* by *Hv*SSI. The structure also inspired us to clone and express truncated versions of *Hv*SSI which display greatly enhanced levels of catalytic activity. The interplay between all SS enzymes, which essentially compete for the same acceptor chains and the same donor sugar-nucleotide molecules in the growing starch granule, is key to the ultimate fine structure of amylose and amylopectin molecules. This in turn determines the physicochemical properties of starch: solubility, crystallinity, gelatinization temperature, digestibility and long-term stability, as well as mechanical properties of industrial interest. Our structure can guide efforts to better understand the effects of mutations of SS enzymes on starch properties.

2. Materials and methods

2.1. Gene synthesis and site-directed mutagenesis

The gene for barley starch synthase I (*Hv*SSI; GenBank accession No. AAF37876.1) was codon-optimized for expression in *E. coli* using the online software tool *GENEius* (http://www.geneius.de) and was synthesized by GenScript (http://www.genscript.com). The chloroplast transit peptide (the first 31 residues) predicted by the *TargetP* server (Emanuelsson *et al.*, 2000) was not included. The synthetic gene contained additional *Nde*I and *Hin*dIII restriction sites at the 5′ and 3′ ends, respectively, which were used for recloning into the pET28a expression vector (Novagen). The expressed *Hv*SSI includes the complete mature protein sequence identified for SSI in wheat endosperm (Li *et al.*, 1999) and a 28-amino-acid N-terminal extension corresponding to a potentially cleaved sequence (seven amino acids; included in the numbering in this structure) and a thrombin-cleavable His tag (from pET28a). Mutants C126S, C506S and F538A were cloned using the QuikChange site-directed mutagenesis protocol (Stratagene/Agilent Genomics), while the mutants 'cat-dom' (containing the barley catalytic domain; residues 95–612) and 'rice-like' (barley residues 85–612; the name is based on the similarity to the length of the rice SSI protein) were constructed using a variant of this method (Wang & Malcolm, 1999) using pET28a_*Hv*SSI as the template. The gene for wheat SSI (*Ta*SSI GenBank accession No. CAB99210.1) was prepared similarly, while the gene for rice SSI (*Os*SSI; GenBank accession No. AAP56350.1) had its initial amino acid selected based on the mature kernel SSI sequence from Baba *et al.* (1993). The complete sequences of the expressed proteins and primers used in mutagenesis are listed in the Supplementary Material[1].

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: CB5023). Services for accessing this material are described at the back of the journal.

Table 1
Data-collection and refinement statistics for the *Hv*SSI crystal.

Values in parentheses are for the highest resolution shell, except for the Ramachandran plot statistics; in this case, the number of residues is indicated.

Data collection	
Space group	$P3_121$
Unit-cell parameters (Å, °)	$a = b = 153.1$, $c = 84.96$, $\alpha = \beta = 90$, $\gamma = 120$
Wavelength (Å)	0.976
Resolution (Å)	2.70 (2.80–2.70)
Total observations	143551
Unique reflections	31291
Completeness (%)	98.3 (97.5)
Multiplicity	4.5 (4.5)
$R_{\rm merge}$† (%)	7.5 (48.4)
$\langle I/\sigma(I)\rangle$	13.66 (3.24)
Refinement statistics	
Resolution range (Å)	25–2.70 (2.84–2.70)
$R_{\rm work}$‡ (%)	17.31 (26.88)
$R_{\rm free}$§ (%)	19.85 (32.25)
No. of protein atoms	3980
No. of water atoms	23
No. of ligand atoms	55
Total No. of atoms	4058
Model statistics	
R.m.s.d., bonds (Å)	0.008
R.m.s.d., angles (°)	1.27
Average B factors (Å^2)	
Protein	59.45
Water	48.95
Ligands	148.45
Ramachandran plot¶, residues in (%)	
Most favoured region	95.98 (477)
Allowed region	3.82 (19)
Outliers	0.20 (1)
PDB code	4hln

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{\rm work} = \sum_{hkl}\big||F_{\rm obs}| - |F_{\rm calc}|\big|/\sum_{hkl}|F_{\rm obs}|$. § $R_{\rm free} = \sum_{hkl}\big||F_{\rm obs}| - |F_{\rm calc}|\big|/\sum_{hkl}|F_{\rm obs}|$ calculated using a random set containing 3% of the reflections which were not included throughout structure refinement. ¶ As assessed by *MolProbity* (Chen *et al.*, 2010).

2.2. Protein expression and purification

The expression vector was transformed into *E. coli* Tuner (DE3) cells (Novagen), which were grown in 1 l LB medium containing 30 µg ml^{-1} kanamycin to an OD of ∼0.6, cooled on ice for 10 min, induced with 50 µ*M* IPTG and incubated overnight at 289 K. Cell-pellet collection, disruption and purification were performed according to standard molecular-biology techniques. Full details are given in the Supplementary Material. Briefly, protein samples were purified by nickel-affinity chromatography followed by size-exclusion chromatography for samples for crystallization or anion-exchange chromatography for samples used in activity assays.

2.3. Crystallization, structure solution and refinement

Selected *Hv*SSI fractions from the size-exclusion column were pooled and concentrated to 9.6 mg ml^{-1}. 58 µl of the concentrated *Hv*SSI stock (pH 8.0) was mixed with 1 µl 25 m*M* maltotriose (Sigma; catalogue No. M8378), 92 µl 8.7 m*M* ADP-glucose pre-adjusted to pH 8.0 (Sigma; catalogue No. A0627) and 329 µl water, resulting in a protein stock of 1.2 mg ml^{-1} with a 1:3:100 *Hv*SSI:maltotriose:ADP-glucose molar ratio and 2.5 m*M* Tris–HCl pH 8.0. This stock was stored at 277 K for 10 d prior to setting up the crystallization

plates. Initial crystallization conditions were identified using Crystal Screen from Hampton Research. After optimization, the best crystal was obtained by the sitting-drop method using Cryschem plates (Hampton Research) from a drop prepared by mixing 9 µl protein stock solution with 1 µl precipitant solution composed of 100 m*M* sodium cacodylate buffer pH 6.5, 17% PEG 8K, 0.2 *M* ammonium sulfate, equilibrated against 500 µl precipitant and stored at 277 K. The drop contained some NaCl and glycerol carried over from the initial protein stock. Under these conditions, a light precipitate initially formed, with crystals appearing after one week; the precipitate eventually disappeared (re-dissolved). After a month, crystals had grown to dimensions of 200 × 120 µm. The largest crystal in the drop was briefly soaked in cryoprotectant composed of 100 m*M* sodium cacodylate pH 6.5, 20% PEG 8K, 0.1 *M* ammonium sulfate, 20% glycerol, mounted in a nylon loop and cooled by plunging into liquid nitrogen.

2.4. Details of data collection, structure solution and refinement

Diffraction data were collected on ESRF beamline ID23-1 at a wavelength of 0.976 Å and at 100 K. 220 images were collected with an oscillation range of 0.35° per image, an exposure time of 0.2 s per frame and a crystal-to-detector distance of 406.46 mm. Diffraction data were reduced with *XDS* (Kabsch, 2010) to a resolution limit of 2.70 Å. A test set was created with 3% of the data (approximately 1000 reflections). Data-quality and final refinement statistics are reported in Table 1.

The structure was solved by molecular replacement with *MOLREP* (Vagin & Teplyakov, 2010) using a truncated version of *E. coli* glycogen synthase (PDB entry 2qzs; Sheng, Jia *et al.*, 2009) as the search model and initially searching for the N-terminal domain, followed by the C-terminal domain. The initial solution was autotraced with *ARP*/*wARP* (Langer *et al.*, 2008), which was able to trace 509 amino acids. Refinement was performed with *REFMAC* (Murshudov *et al.*, 2011) and *PHENIX* (Adams *et al.*, 2010). The final model was refined with *PHENIX* using isotropic atoms and TLS anisotropy modelling with two TLS groups: one per domain. Restraints were from the standard *PHENIX* library, including the glucose residues and linkages. Manual model building and map inspection was performed with *Coot* (Emsley *et al.*, 2010). Water molecules were added manually only where there was globular electron density that was above the 3σ level in the $F_o - F_c$ map and above the 1σ level in the $2F_o - F_c$ map as well as being within hydrogen-bonding distance of suitable polar atoms. This is a very conservative protocol for the addition of water molecules, which together with the moderate resolution results in the modelling of a relatively low number of water molecules, but helps to avoid the false interpretation of weak ligand density. Figures were rendered with *PyMOL* (http://www.pymol.org). Structural superpositions were performed with the secondary-structure matching algorithm in *Coot*. Pairwise sequence alignments were performed with *BLAST* (Johnson *et al.*, 2008) and multiple sequence alignments with *ClustalW* (Larkin *et al.*, 2007).

2.5. Activity assays

Initial rates were determined by coupling the release of ADP to NADH oxidation *via* pyruvate kinase and lactate dehydrogenase in a protocol adapted from Gosselin *et al.* (1994). Assays were performed in a final volume of 100 µl with the following final concentrations: 50 m*M* Bicine pH 8.5, 25 m*M* potassium acetate, 0.1%(*w*/*v*) bovine serum albumin, 2 m*M* $MgCl_2$, 10 m*M* reduced DTT [except when a different concentration or oxidized DTT (*trans*-4,5-dihydroxy-1,2-dithiane) was required], 0.375 m*M* NADH, 0.7 m*M* phosphoenolpyruvate tricyclohexylammonium salt, 6 U ml^{-1} pyruvate kinase, 30 U ml^{-1} lactate dehydrogenase (both from Sigma; rabbit muscle type II) with 30–800 n*M* enzyme at 310 K. 10 m*M* maltopentaose or varying concentrations of glycogen (Sigma; catalogue No. G8876, rabbit liver type III; catalogue No. G8751, oyster type II) or soluble potato starch (Sigma; catalogue No. S2630) were used as acceptors. For soluble starch, stock solutions were heated to 368 K, vortexed and allowed to cool to room temperature shortly before use. Reactions were initiated by addition of 1 m*M* ADP-glucose. Enzyme concentrations for activity assays were estimated by the method of Bradford using bovine gamma globulin as a reference. NADH oxidation was monitored by the decrease in absorbance at 340 nm. For some assays, instead of ADP-glucose from Sigma, an enzymatically synthesized ADP-glucose sample was employed (details are given in the Supplementary Material).

For activity assays in the presence of reduced or oxidized DTT, the proteins were incubated with 50 m*M* reduced or oxidized (freshly prepared) DTT, 0.1% BSA, 50 m*M* Bicine buffer pH 8.5 at 310 K for 1 h; 40 µl protein stock was then added to 50 µl of a solution containing all of the other assay components except for ADP-glucose (DTT was excluded from this particular assay solution) and this mixture was equilibrated at 310 K prior to starting the reaction by the addition of 10 µl 10 m*M* ADP-glucose. Maltopentaose (10 m*M*) was used as the acceptor and the final concentration of oxidized or reduced DTT in the reaction wells was 20 m*M*.

3. Results and discussion

3.1. Structure description and comparison with rice granule-bound starch synthase

In this paper, we present the crystal structure of starch synthase I from barley (*Hv*SSI) crystallized in the presence of maltotriose and ADP-glucose and refined to 2.7 Å resolution. The crystal belonged to space group $P3_121$ and has a very large solvent content of 70%. *Hv*SSI is a 612-amino-acid protein (as predicted by the *TargetP* server; Emanuelsson *et al.*, 2000) belonging to the GT-B superfamily of glycosyltransferases (Coutinho *et al.*, 2003). The catalytically active part spans two domains, with an additional N-terminal extension with no predicted structure or known function. The

crystal structure presented here includes residues 95–612; all of the N-terminal extension was omitted owing to lack of electron density. This N-terminal extension is oriented towards the solvent channels in the crystal and is likely to be disordered. *Hv*SSI is a monomer in the crystal with no significant intermolecular interaction surfaces, which is in agreement with the size-exclusion chromatography data (Supplementary Fig. S1) and the available data on wheat endosperm SSI (Tetlow *et al.*, 2008). *Hv*SSI adopts the characteristic GT-B fold (a double Rossmann fold) as expected for a member of the GT5 family of glycosyltransferases as categorized in the CAZy database (Coutinho *et al.*, 2003). The N-terminal domain includes residues 97–370 and the C-terminal domain includes residues 380–590. Amino acids 371–379 form a linker devoid of any secondary structure that joins the two domains together, while a C-terminal α-helix starting at residue 590 runs back towards and makes extensive interactions with the N-terminal domain. A ribbon representation of the overall structure is shown in Fig. 1(*a*). The active site of GT-B-fold glycosyltransferases is located in the cleft between the two domains, each of which contributes residues to it. Retaining GT-B enzymes are known to cycle between an open conformation (Barford & Johnson, 1992; Gibson *et al.*, 2004; Sheng, Jia *et al.*, 2009), in which the Rossmann-fold domains are relatively far apart, and a closed conformation, in which they are closer together and a catalytically competent active site is formed. In our crystal the two domains adopt the open conformation (Fig. 1*b*); hence, a competent active site is not formed.

The crystal structure of GBSSI from rice (*Os*GBSSI) has recently been determined (Momma & Fujimoto, 2012) and the structure of *Hv*SSI will largely be described here in

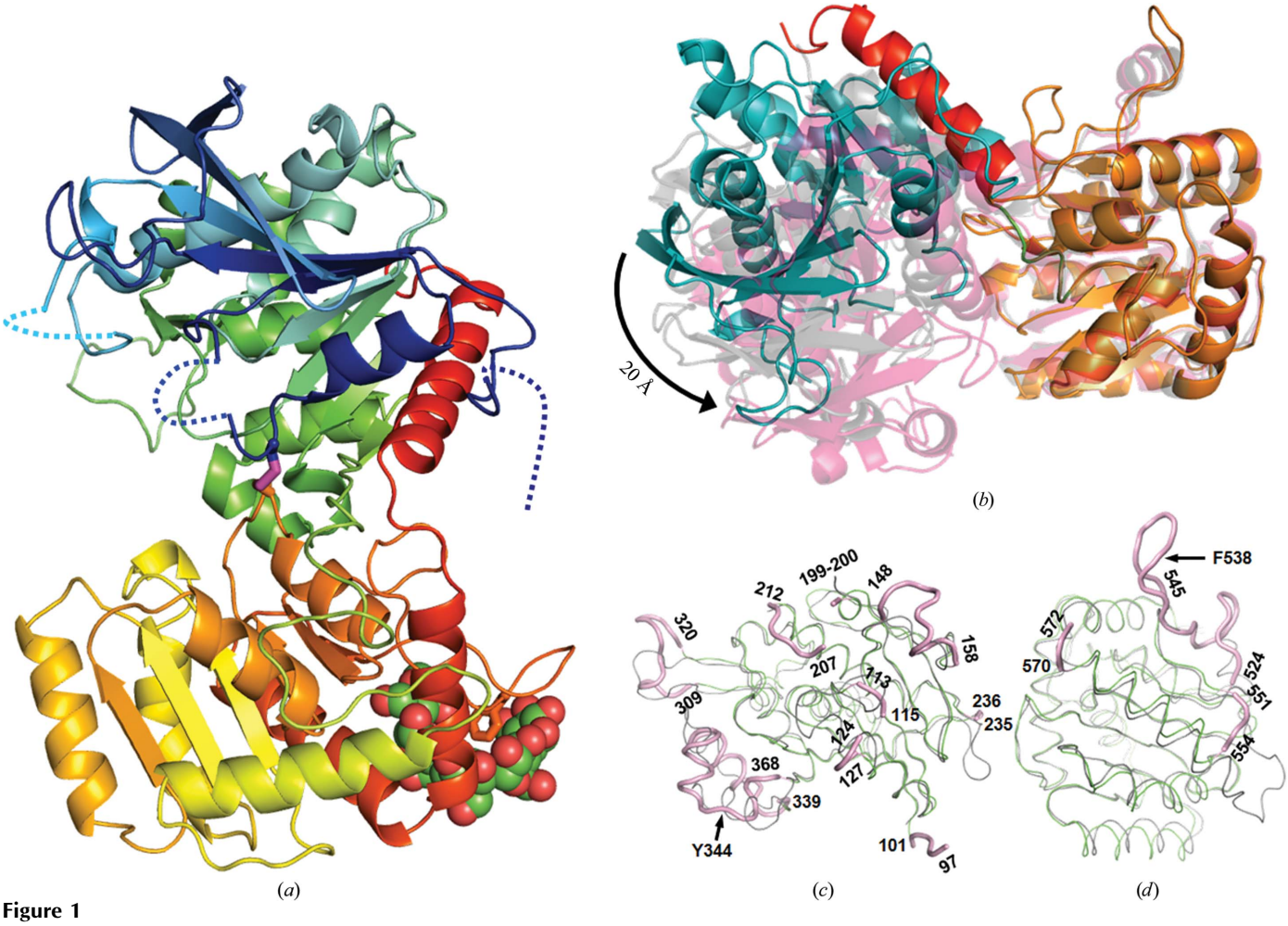


Figure 1
Crystal structure of *Hv*SSI and comparison to related GT5 proteins. (*a*) Ribbon representation of *Hv*SSI. The N-terminal domain is shown in blue and green, the C-terminal domain is shown in yellow and orange, and the final α-helix crossing over back to the N-terminal domain is shown in red. The missing N-terminus and loops are shown as dotted lines. The disulfide bridge is visible in stick representation in the centre (pink S atoms) and the bound maltopentaose is shown as spheres (green C atoms) at the back (away from the reader), with the side chain of its central Phe538 shown as sticks. (*b*) Structures of *Hv*SSI coloured similarly to as in (*a*) but with single colours for each domain; *At*GS (grey, transparent; Buschiazzo *et al.*, 2004; PDB entry 1rzu) and *Ec*GS (pink, transparent; Sheng, Jia *et al.*, 2009; PDB entry 2qzs) are superimposed by their C-terminal domains to illustrate interdomain movements. (*c*, *d*) Superpositions of the N-terminal domain (*c*) and C-terminal domain (*d*) of *Hv*SSI, shown in green, and *Os*GBSSI (Momma & Fujimoto, 2012; PDB entry 3vue), shown in grey. The core of the fold is conserved, but many loops differ. Differences are highlighted in thicker pink lines for the C^{α} trace (for the *Hv*SSI conformation). The numbers indicate amino acids at the beginning and end of these regions as well as key amino acids mentioned in the main text.

comparison to this structure. *Hv*SSI and *Os*GBSSI have 39% sequence identity (56% similarity) over 515 amino acids spanning most of the catalytic domains. *Hv*SSI also has 34% sequence identity (51% similarity) to *At*GS and 33% sequence identity (50% similarity) to *Ec*GS; hence, these structures will also be used for comparison. A sequence and structural alignment is depicted in Supplementary Fig. S2. The overall structure of *Hv*SSI better resembles that of *At*GS in complex with ADP (PDB entry 1rzu; Buschiazzo *et al.*, 2004), which is in the open conformation, than those of *Os*GBSSI (PDB entries 3vue and 3vuf; Momma & Fujimoto, 2012) or of *Ec*GS bound to ADP, glucose and an acceptor mimic (PDB entry 2qzs; Sheng, Jia *et al.*, 2009), which are in the closed conformation. For *Ec*GS a functional active site was formed; thus, it is the most reliable example of the closed conformation and the two domains are slightly more closed than in *Os*GBSSI. Relative to *Ec*GS, *Hv*SSI presents interdomain movements as large as 20.7 Å between Ile177 in *Ec*GS and the equivalent Leu295 in *Hv*SSI (Fig. 1*b*) in a movement combining a scissor-like and a sliding movement of both domains relative to each other (the equivalent distance is 18.2 Å compared with *Os*GBSSI). The structure of *Hv*SSI is even more open than that of *At*GS (Fig. 1*b*). The folds of each domain are very well conserved relative to *Os*GBSSI, with r.m.s.d. values of 1.41 Å for the N-terminal domain (Fig. 1*c*) and of 1.11 Å for the C-terminal domain (Fig. 1*d*); most large secondary-structure elements are conserved and the differences mostly lie in surface loops. Most of the residues cited below are highlighted in Figs. 1(*c*) and 1(*d*).

The large N-terminal extension of *Hv*SSI, which is not present in GBSSI or in bacterial glycogen synthases, seems to be unstructured in the large solvent channels and is not included in our model. We arbitrarily define the beginning of the N-terminal domain of *Hv*SSI as Gln97 based on this being the first residue to make close contacts to other residues in the N-terminal domain. Two other regions are disordered and hence are omitted from the final model of *Hv*SSI: the Tyr116–Gly123 loop, which includes the conserved K*x*GGL motif, likely as a consequence of the disulfide bridge including Cys126, which will be discussed below, and Gly201–Asp206. Five amino acids (207–211) could be built into our structure that were absent in the *Os*GBSSI model. The latter loop is in close proximity to the 116–123 loop. The loop is also disordered, and is seven amino acids longer, in the *Os*GBSSI structure. These disordered loops are involved in binding at the donor and acceptor sites in *Ec*GS (Sheng, Jia *et al.*, 2009) and are likely to change their conformations upon binding of the glucose donor to assist in the formation of the acceptor-binding site, as is frequent in retaining glycosyltransferases (Qasba *et al.*, 2005).

The loop between Asn148 and Lys158 includes a seven-amino-acid insertion in *Hv*SSI relative to *Os*GBSSI and is consequently longer without causing further differences in the secondary structure. Similarly, a deletion compared with *Os*GBSSI between Gly235 and Gly236 makes the corresponding loop shorter. There is no sequence similarity and a three-amino-acid insertion in Phe309–Lys320 causing this loop to be longer in *Hv*SSI and to cause a tilt in the two adjacent β-strands. This loop is involved in crystal contacts in *Hv*SSI.

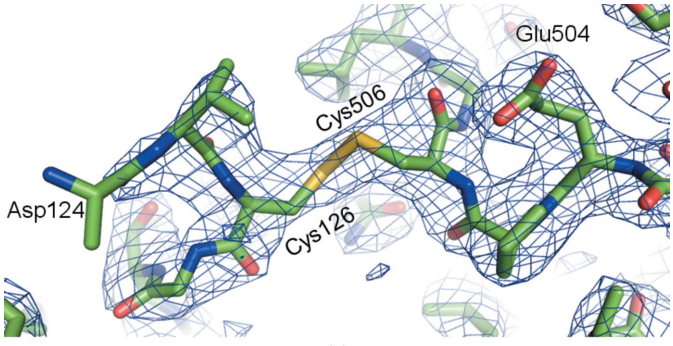
The region between Val340 and Leu368, which includes three short α-helices and a short β-strand, differs greatly from that in *Os*GBSSI. In *Os*GBSSI this region reaches out to the C-terminal domain to form a disulfide bridge with a similarly extended loop there; hence, a comparison with the bacterial glycogen synthases is more relevant in this region. They adopt a similar conformation except around a kink at residue Tyr344 that breaks a longer α-helix in two. This kink is encoded in the sequence: Tyr344 is preceded by Gln-Gly in *Hv*SSI as opposed to Pro-Thr in *Ec*GS and Pro-Ser in *At*GS (and Pro-Tyr in *Os*GBSSI); with proline defining the beginning of α-helices (Richardson & Richardson, 1988), while glycine is flexible and can disrupt helices (Pace & Scholtz, 1998). This places Tyr344 as a wedge that protrudes into the C-terminal domain. Although this residue might thus be part of the cross-talk mechanism between the two domains of SSI, it does not form any hydrogen bonds to the C-terminal domain in our structure. The residue at position −2 to Tyr344 is a proline (as in the glycogen synthases) in GBSSI, SSII, SSIIIa and SSIV from barley, while it is conserved as glutamine in most plant SSI enzymes (from wheat, rice, *A. thaliana* and potato, with a lysine in the same position in the corn enzyme, in all cases followed by a glycine); this kink thus seems to be a feature unique to SSI enzymes.

The interdomain linker between Ile371 and Asp379 displays a conformation very similar to that of *Os*GBSSI, as expected given the high sequence similarity in this area, while accommodating the large interdomain movements between the structures. The structure of the C-terminal domain of *Hv*SSI very closely resembles that of *Os*GBSSI for most of the domain (Fig. 1*d*), apart from two regions and a two-amino-acid insertion in *Hv*SSI in the 570 loop which results in an elongation of the preceding α-helix, which extends to Glu571 in *Hv*SSI, and a sharper turn to the subsequent helix. The first major difference is the region between Gly524 and Glu544. In *Os*GBSSI this corresponds to a loop–helix–loop region, while in *Hv*SSI it is a long random coil that extends away from the molecule and includes a maltopentaose bound in a site centred on Phe538, which will be discussed in greater detail below. The second major difference directly follows in this region, with a short β-strand between Gly547 and Phe550 that apparently matches a similar β-strand in *Os*GBSSI (adopting the same position) which is nine amino acids apart in the sequence alignment. This is followed by a short loop to Thr554 in *Hv*SSI, while *Os*GBSSI has a much longer loop in this position which reaches out to the N-terminal domain and includes Cys529, which forms the disulfide bridge in the *Os*GBSSI structure. Finally, the C-terminal α-helix that crosses over from the C-terminal domain to the N-terminal domain overlaps very well in all structures if it is included as part of the N-terminal domain, with no immediately obvious function for the (ordered) small post-helix C-terminal extension in *Hv*SSI which is absent from the glycogen synthase enzymes and the *Os*GBSSI structure.

*Hv*SSI was crystallized in the presence of ADP-glucose and maltotriose. There was no electron density for ADP-glucose or its hydrolysis product ADP; thus, they could not be included in our model. In the case of the sugar donor, residual electron density corresponding to a planar group is observed at the 3σ level in the $F_o - F_c$ map in the vicinity of Phe482 (not shown), possibly corresponding to the adenine moiety of ADP with partial occupancy, but it is featureless and no interpretable phosphates were present in its proximity. Similarly, no glucose molecules or interpretable electron density was found in the area in which the active site would form.

3.2. Disulfide bridge and active-site disorder

An unexpected feature of our crystal structure is a disulfide bridge over the putative active site. The Cys126–Cys506 disulfide bond (Fig. 2*a*) creates a third link (after the inter-domain peptide and the C-terminal α-helix) between the N-terminal and the C-terminal domains, with large consequences for the conformation and activity of the enzyme. A different cysteine disulfide, involving the conserved Cys337 and Cys529 in the Poaceae, was found in the crystal structure of *Os*GBSSI (Momma & Fujimoto, 2012). In *Os*GBSSI the disulfide bridge locks the enzyme in the closed conformation, but there was no consideration of its effect on the activity or its biological significance.



(*a*)

HvSSI	KSGGLGDV**C**GSLPIA	HvSSI PSRFEP**C**GLNQLY
TaSSI	KSGGLGDV**C**GSLPIA	TaSSI PSRFEP**C**GLNQLY
OsSSI	KSGGLGDV**C**GSLPIA	OsSSI PSRFEP**C**GLNQLY
ZmSSI	KSGGLGDV**C**GSLPIA	ZmSSI PSRFEP**C**GLNQLY
AtSSI	KTGGLGDV**C**GSLPIA	AtSSI PSRFEP**C**GLNQLY
StSSI	KTGGLGDV**C**GSLPIA	StSSI PSRFEP**C**GLNQLY
GBSS	KTGGLGDVLGGLPPA	GBSS TSRFEP**C**GLIQLQ
SSII	KTGGLGDVAGALPKA	SSII PSRFEP**C**GLNQLY
SSIII	KVGGLGDVVTSLSRA	SSIII PSIFEP**C**GLTQLV
SSIV	KVGGLADVISGLGKA	SSIV PSIFEP**C**GLTQMI

(*b*) (*c*)

Figure 2
Disulfide bridge between Cys126 and Cys506. (*a*) Region of *Hv*SSI around the disulfide bridge. Atoms are shown as sticks, with the cysteine S atoms coloured yellow. Electron density at the 1.0σ level of the $2F_o - F_c$ map is shown as a blue mesh. (*b*) Local sequence alignment around *Hv*SSI_Cys126 (bold and underlined) for related plant SSI enzymes (upper half) and other barley SS enzymes (lower half). (*c*) Local sequence alignment around *Hv*SSI_Cys506 (bold and underlined) as in (*b*). Sequence accession numbers are given in the Supplementary Material.

In *Hv*SSI this disulfide brings the cysteines closer to each other (and closer than they are in the *Os*GBSSI and GS structures), affecting the conformation of the first α-helix in the N-terminal domain, which becomes shorter to exclude Cys126. This results in disorder of the adjacent loop between Tyr116 and Gly123, which contains the K*x*GGL motif. This motif is conserved in the GT5 family (including plant starch synthases and bacterial glycogen synthases) and is involved in both ADP and acceptor binding in *Ec*GS (Lys118, Gly121 and Asp124 bind ADP, while Leu122 binds to the acceptor; Sheng, Jia *et al.*, 2009). While the effect of the disulfide on the loop containing Cys506 is less dramatic, it still affects the conformation of its immediate environment, including Glu504 and Gly507, which together with Cys506 would be involved in binding the glucose to be transferred. Thus, this disulfide bridge is responsible for the disorganization of the active site of *Hv*SSI. This disulfide bridge is likely to be representative of the state of *Hv*SSI *in planta*. SSI is a redox-sensitive enzyme *in vivo* (Glaring *et al.*, 2012), along with a few other enzymes in the starch-synthesizing system: SSIII, BEII and possibly isoamylase I (Glaring *et al.*, 2012). While Cys506 is conserved across the entire GT5 family, Cys126 is unique to SSI (Figs. 2*b* and 2*c*).

To determine potential roles for these two cysteines, we constructed, expressed and purified the mutants *Hv*SSI_C126S (*Hv*SSI_C126A was also prepared and behaved similarly) and *Hv*SSI_C506S which cannot form the disulfide bridge. Their activities with 10 m*M* maltopentaose as an acceptor were tested in the presence of 20 m*M* reduced or 20 m*M* oxidized DTT (see §2), with wild-type *Hv*SSI as a control. The activities for *Hv*SSI were 14.5 min^{-1} (*i.e.* the number of substrate molecules turned over per molecule of enzyme per minute) under reducing conditions and 0.38 min^{-1} under oxidizing conditions. *Hv*SSI_C126S had an activity of 13.9 min^{-1} under reducing conditions and 1.6 min^{-1} under oxidizing conditions, while *Hv*SSI_C506S had an activity of 2.6 min^{-1} under reducing conditions and 0.76 min^{-1} under oxidizing conditions. This corresponds to relative reductions of activity in oxidizing conditions of 38-fold, ninefold and threefold, respectively. The activity of the C126S mutant in the presence of reduced DTT was comparable to that of the wild-type enzyme, while the C506S mutant had 18% of the wild-type enzyme activity. A similar reduction in activity was found for the C379S and C379A mutants of *Ec*GS (equivalent to C506S; Yep *et al.*, 2004), in which Cys379 is directly involved in binding the glucose moiety of ADP-glucose. *Hv*SSI_C126S shows a lower level of sensitivity (89% activity loss *versus* 97% for the wild type) to the redox potential than the wild type. Thus, Cys126 is not critical for activity but is more likely to play a role in redox regulation. *Hv*SSI_C506S displays a reduced sensitivity to oxidizing conditions (20 m*M* oxidized DTT) compared with the wild type, with a 70% decrease in activity. Since the basal level of activity of C506S in reducing conditions is only 18% of that of the wild-type enzyme, this highly conserved residue in all SS enzymes (Fig. 2*c*) seems to be important for both redox regulation and optimal activity. While our data support the involvement of both conserved cysteines, Cys126 and Cys506,

in redox regulation of *Hv*SSI activity, it does not allow us to preclude the involvement of other residues, since none of the mutants became redox-insensitive. On the assumption that redox regulation involves only Cys126 and Cys506, we propose that this disulfide bridge might be the means by which the activity of SSI in plant plastids is reduced when exposed to oxidizing conditions (*i.e.* at night in chloroplasts; Glaring *et al.*, 2012; Geigenberger, 2011).

3.3. The solvent-exposed disulfide bridge is resistant to reduction in the crystal

Multiple soaking experiments with reducing agents support the importance of the disulfide bridge in the formation of this crystal form and maintenance of its crystal packing. Several attempts were made to crystallize *Hv*SSI after adding reducing agents (DTT or TCEP) either to the protein stock or to the mother liquor. No other crystal forms were obtained; addition of reducing agents simply delayed crystallization by about one week when the protein stock contained 1 m*M* DTT and by about a month when it contained up to 5 m*M* TCEP. Crystallization is apparently delayed until either sufficient oxygen diffuses into the sitting drops and consumes the reducing agent, or until reducing-agent decomposition (the cacodylate buffer might enhance the decomposition of TCEP since it is sensitive to neutral phosphate buffers). Once formed, crystals of this form resist soaking with reducing agents. The addition of 5 m*M* TCEP was tolerated without visual damage, while several crystals treated with cryoprotectant containing TCEP to a final concentration of 1 m*M* and incubated overnight prior to cooling showed no change in their unit cells.

One crystal was soaked for 24 h by adding 20 m*M* ADP, 25 m*M* $MgCl_2$, 50 m*M* DTT, 5 m*M* sodium cacodylate to the sitting drop and was then cryoprotected and cooled as for the crystal that produced the structure to 2.7 Å resolution. The exact volume of the sitting drop before adding the additives was not known, but we estimated the final concentration of DTT in the drop to be ∼25 m*M* after incubation. It was possible to obtain a diffraction data set to 3.5 Å resolution from this crystal on beamline ID23-1 at the ESRF. The unit cell was largely unaffected by this soaking, with unit-cell parameters $a = b = 153.51$, $c = 84.66$ Å, $\alpha = \beta = 90$, $\gamma = 120°$, corresponding to a less than a 0.5% change in the unit-cell axes after soaking. The structure was solved by molecular replacement and initial refinement showed that the disulfide bridge was still present, apparently at full occupancy. The structure appeared to be virtually identical to that reported here and refinement was stopped at this stage ($R = 16.38\%$, $R_{free} = 24.36\%$); this demonstrates that in the crystals this disulfide bridge is resistant to concentrations of reducing agents higher than those needed to reduce it in solution. Hence, its formation is at least favoured by the crystal environment. TCEP is considered to be a stronger reducing agent than DTT (Getz *et al.*, 1999), but it apparently does not affect this particular disulfide at concentrations as high as 5 m*M*, as crystal dissolution or changes in the unit cell would be expected.

To our knowledge, this is the first report of a TCEP-resistant accessible

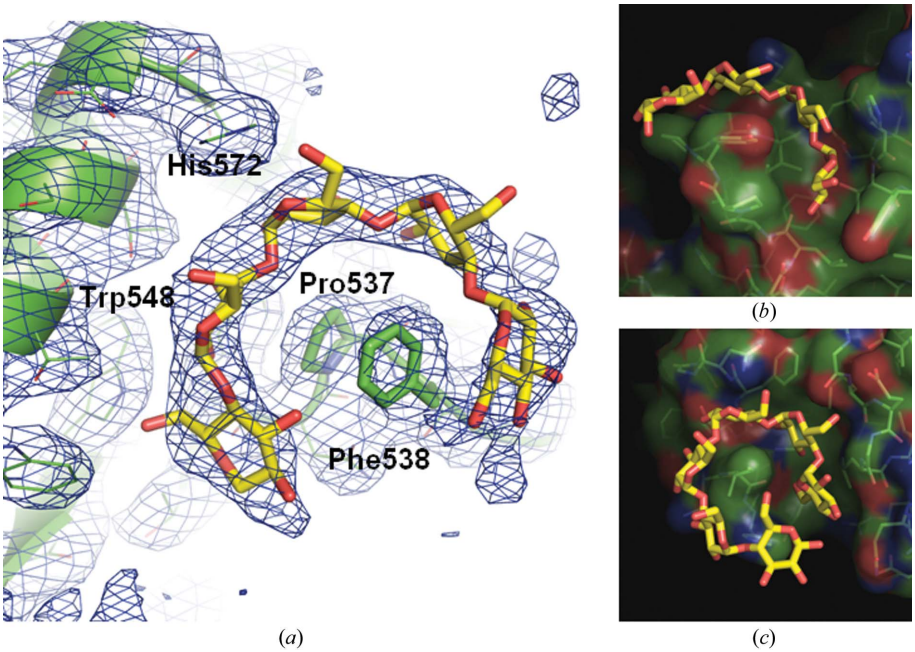


Figure 3
Maltooligosaccharide-binding site. (*a*) Maltopentaose modelled in the MOS-binding site. The protein is shown as a green ribbon and thin sticks. Pro537 and Phe538 in the centre of the site have been highlighted with thicker sticks. Maltopentaose is shown as a yellow stick model with red O atoms. The blue mesh represents electron density from the $2F_o - F_c$ map contoured at the 0.7σ level. (*b*, *c*) MOS-binding sites in (*b*) *P. abyssi* GS (PDB entry 3l01; Díaz *et al.*, 2011) and (*c*) *E. coli* GS (PDB entry 3cx4; Sheng, Yep *et al.*, 2009). The proteins are shown with green C atoms and thin sticks as well as with semitransparent surfaces. The sugars are represented as in (*a*). (*d*) Local sequence alignment in the region of *Hv*SSI_F538. The central Pro537 and Phe538 are shown in bold and underlined; Phe535, Asp543 and Trp548, which also contact the maltopentaose in our model, are shown in bold where conserved. Sequence accession numbers are listed in the Supplementary Material.

disulfide in any protein. It must be emphasized that this disulfide bridge is solvent-exposed and in a region that is expected to display structural flexibility. Its resistance to reduction cannot be attributed to inaccessibility of the reducing agents. The soluble form of *Hv*SSI is readily activated by 10 m*M* DTT on a time scale measured in seconds. Hence, the resistance to reduction must be attributable to the crystal environment.

3.4. Binding of maltooligosaccharide produced *in situ*

A maltooligosaccharide was modelled bound to the 534–546 surface loop in the C-terminal domain. Although only maltotriose was added to the crystallization cocktail, maltopentaose was included in the final model (Fig. 3*a*). Both the protein-stock and the mother-liquor conditions are compatible with SSI activity; thus, the addition of maltotriose and ADP-glucose allowed a mixture of longer MOSs to be produced *in situ*. We wished to obtain MOSs with a DP larger than eight (the longest commercially available) and to allow the enzyme to select between the different species available for binding. Starch synthase activity was confirmed by TLC in protein stocks of the same composition (Supplementary Fig. S3) and labelling with 2-aminobenzamide of a crystal extracted from the mother liquor (Supplementary Material and Supplementary Fig. S4). SS activity is expected both before crystallization and in the soluble fraction after crystal formation. Assuming that all the ADP-glucose is productively consumed, the ratios used would allow the reaction to proceed up to a 36-mer. In reality the product is expected to be a broad distribution of polymer sizes, and it is unclear whether SSI is capable of synthesizing polymers of this length (Commuri & Keeling, 2001). The maltopentaose included in the model should be considered as an ordered fragment of a larger MOS species.

Electron density for this maltopentaose unit was clear in the initial difference electron-density maps as a helical shape wrapped around Pro537 and Phe538 (Fig. 3*a*). It has high *B* factors; thus, most hydroxyl groups did not form separate individual protrusions in the electron density. Consequently, there is some ambiguity in the exact position of each atom and more than one model was possible; the deposited structure was the best of them. Still, the helical shape of the density, which is typical of α-glucans (Damager *et al.*, 2010), leaves little doubt about its nature. A plausible model was easily built in the existing density (Fig. 3*a*) with glucoses fitting each bulge and featuring the expected intermolecular hydrogen bonds between the 3-hydroxyl of each glucose and the 2-hydroxyl of the next (not included as restraints in refinement). Mutational data also support its identification as a bound MOS, but we urge caution in interpreting any fine details or particular hydrogen-bonding distances to the maltopentaose deposited in this model.

The binding site has Phe538 at its centre stacked against Pro537 (Fig. 3*a*). The MOS curls in a helical structure around these two residues while also coming into contact with Phe535 through both its main-chain carbonyl and its side chain and with the side chains of Glu543, Trp548, His572 and Ser575. Its geometry resembles that of a MOS-binding site in *P. abyssi* GS (*Pa*GS; Díaz *et al.*, 2011; Fig. 3*b*), in which Tyr174 occupies a central position inside the helix described by the MOS. Tyr174 in *Pa*GS is described as 'a reel around which the four glucose units coil'. In this structure Tyr174, at the side of the hydroxyphenyl ring that does not face glucose, is also stacked against the proline that precedes it in the sequence. In both cases the surrounding protein surface defines an arc-shaped channel around the central residues in which the oligosaccharide binds. Similar noncatalytic binding sites can be found with an isoleucine at the centre in a structure of *Ec*GS (Sheng, Yep *et al.*, 2009; Fig. 3*c*) and with a central tyrosine at the centre in barley α-amylase 1 (Robert *et al.*, 2005).

It is worth noting that this MOS-binding site is located in the C-terminal domain of *Hv*SSI. All other binding sites reported to date for bacterial glycogen synthases have been located exclusively in the N-terminal domain. This MOS-binding site is unique to starch synthases and is not present in *Ec*GS, *At*GS or *Pa*GS, which lack the central phenylalanine in their sequences (Fig. 3*d*). The binding site is conserved, at least at the sequence level, in barley for SSII (only Glu543 is missing) but not in GBSS (only Phe538 and Trp548 are conserved) or in SSIII or SSIV, in which even Phe538 is missing. Within SSI from other plant species the site is largely conserved (Fig. 3*d*), with Phe538 conservatively substituted by Tyr538 in potato and *A. thaliana* SSI.

The exact occupancy of the MOS in this site is unknown, as occupancy and *B* factors are anticorrelated in protein crystal structures at moderate resolution and thus one can compensate for the other. Considering that the MOS is largely exposed to the solvent, it is reasonable to assume that its mobility would be greater than that of the protein residues that it is anchored to. Comparison of their *B* factors suggests that the occupancy of this site should be close to 100%, or in any case higher than 50%. The concentration of maltotriose in the protein stock was 52 μM (a 3:1 ratio to *Hv*SSI) and the sitting drop was produced by mixing 9 µl protein stock solution with 1 µl precipitant solution. The exact final volume in the drop after equalization of vapour pressure between the drop and the reservoir is unknown, but is estimated as approximately 2 µl. This would result in a maltotriose concentration of 234 μM in the final drop. Although this maltotriose is converted to longer MOSs *in situ*, the total MOS concentration remains the same at 234 μM. This would be further reduced in the cryoprotection step, in which no MOS was added. Based on these assumptions, the occupancy that we observe in the crystal sets a qualitative upper limit for the K_d between the protein and MOS of approximately 200 μM, making it a high-affinity MOS-binding site comparable with previously characterized surface binding sites (Nielsen *et al.*, 2009) and carbohydrate-binding modules with affinity towards MOS and starch (Lammerts van Bueren *et al.*, 2004; Boraston *et al.*, 2006; Chou *et al.*, 2006; Giardina *et al.*, 2001). This MOS-binding site in *Pa*GS is described as having a higher affinity for MOS than the site in the active-site cleft and the same has been proposed for an equivalent site in human muscle glycogen synthase (HMGS; Díaz *et al.*, 2011). Our

MOS-binding site thus resembles that in *Pa*GS both structurally and in its affinity. The MOS-binding site identified in *Hv*SSI is noncatalytic and does not represent an acceptor MOS, with Phe538 situated 30.1 Å away from Cys506 which would be part of the active site.

3.5. Activity of a maltooligosaccharide-binding site mutant

To test the importance of the surface MOS-binding site, we tested mutant *Hv*SSI_F538A with maltopentaose, glycogen and soluble starch as substrates. *Hv*SSI_F538A was designed to disrupt interaction with the MOS by removal of the central aromatic side chain. The effects of the mutation are shown graphically in Fig. 4. With 10 m*M* maltopentaose as an acceptor, *Hv*SSI_F538A had an activity of 13.7 min^{-1} (note that this is not k_{cat} since the measurement was not carried out under saturating conditions), which is only slightly lower than for the wild-type enzyme. Thus, there should not be significant structural destabilization of the mutant. However, the activity of the F538A mutant compared with the wild-type enzyme is ∼12-fold lower with rabbit liver glycogen, sixfold lower with oyster glycogen and fourfold lower with soluble starch at high glycogen or starch concentrations, with even larger differences at lower acceptor concentrations (Fig. 4).

Since this binding site is 30 Å away from the active site, glycogen and amylopectin molecules can span both sites simultaneously, unlike maltopentaose. In the branched polymers one chain could interact with the Phe538 loop while another chain containing a nonreducing end could occupy the active-site cleft, or the same chain could occupy both sites if it is long enough. Since a lack of binding of the same chain to the high-affinity site has little effect on the active site when probed with maltopentaose, the decrease in activity for glycogen and amylopectin must be owing to geometric, not conformational, effects. It is likely that the high-affinity binding site helps to co-localize branched substrates and SSI, effectively increasing the local concentration of acceptor available to the enzyme. Indeed, in *Pa*GS and HMGS the high-affinity binding site is responsible for physically localizing GS to the glycogen particles and the enzymes can be pulled down together with glycogen (Díaz *et al.*, 2011). Disruption of the binding sites in those proteins resulted in the loss of such properties. Although SSI is typically purified from the soluble fraction of starch granules, it is partially granule-associated, at least in maize endosperm (Mu-Forster *et al.*, 1996), as part of a complex with SSIIa and branching enzyme IIb (Liu *et al.*, 2012). A high-affinity binding site can enhance the activity of the enzyme by placing it at the granule surface where its activity is needed. Alternatively, it could help to orient SSI relative to certain parts of the growing amylopectin molecules, or even act as a molecular ruler for chain-length selection.

3.6. Activity of N-terminally truncated mutants

Plant SSI enzymes include an N-terminal extension of approximately 95 amino acids (at the gene level) before their catalytic domains which is absent in GT5 glycogen synthases and in granule-bound starch synthases. In maize (Imparl-Radosevich *et al.*, 1998) and wheat (Li *et al.*, 1999) this extension remains in the mature protein in the endosperm; in barley, the first amino acid of the mature protein in the endosperm is unknown. In rice, it is expressed but hydrolyzed post-translationally such that the mature protein in the endosperm starts at amino acid Ser85 (Baba *et al.*, 1993), 12 amino acids earlier than the N-terminal domain in *Hv*SSI. These N-terminal extensions have no predicted structure or known functions. The entire extension was present in the protein stocks used for crystallization, but was disordered in the crystal; thus, we made two truncated versions to evaluate their effects on catalysis. One mutant 'rice-like' *Hv*SSI (the

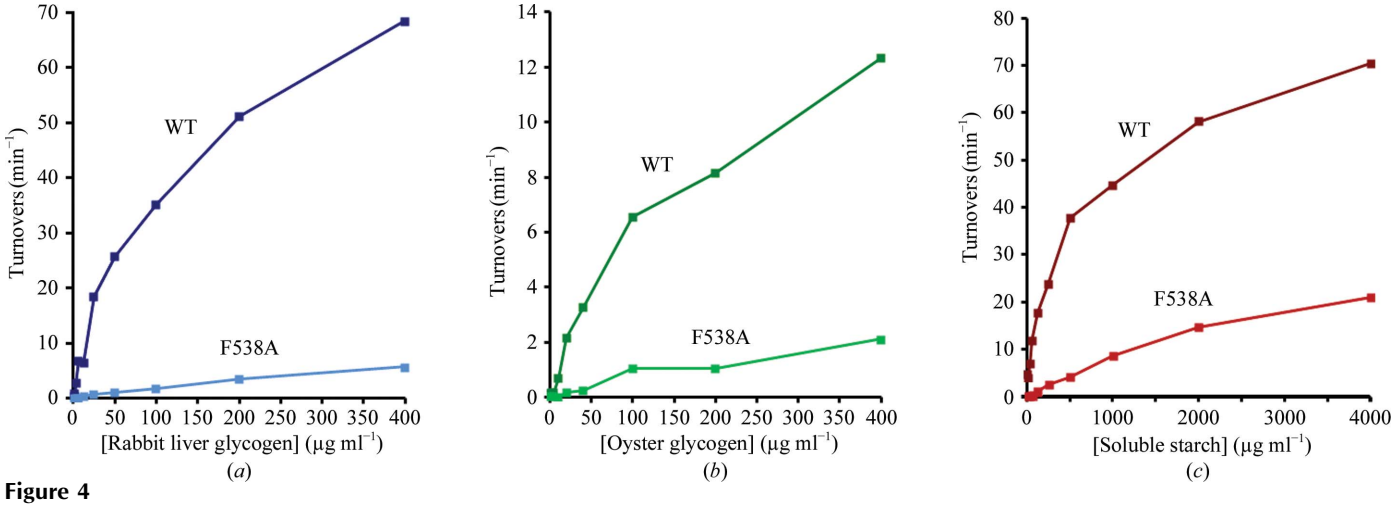


Figure 4
Activity of a surface maltooligosaccharide-binding site mutant. (*a*) Activity in substrate turnovers per molecule of enzyme per minute with increasing concentrations of glycogen (rabbit liver type III). Dark colours are for wild-type *Hv*SSI and light colours for *Hv*SSI_F538A with the same acceptor. Data points at 1, 2 and 4 mg ml^{-1} glycogen were also collected (data not shown) which continue the upward trend in activity with increasing glycogen concentrations. This is not Michaelis–Menten behaviour since glycogen is nonsaturable at the concentrations tested. The data do not represent k_{cat} since the conditions are not saturating. (*b*) As (*a*) but with oyster glycogen (type II) as the acceptor. (*c*) Activity with increasing concentrations of soluble starch as the acceptor. Colours are as in (*a*) and (*b*). The activity is nonsaturable as observed for glycogen.

Table 2
Comparison of the activity of *Hv*SSI, *Ta*SSI, *Os*SSI and truncated *Hv*SSI mutants.

Activities for the *Hv*SSI used for crystallization as well as the N-terminally truncated mutants are compared with SSI enzymes from wheat (*Triticum aestivum*; *Ta*SSI) and rice (*Oryza sativa*, indica variety; *Os*SSI). The assays were carried out with 1 m*M* ADP-glucose and 10 m*M* maltopentaose or 0.1 mg ml^{-1} glycogen as acceptors. These reactions were not carried out with saturated substrate and thus the values do not represent k_{cat}. Percentages are relative to *Hv*SSI. The sequences of the expressed constructs for *Ta*SSI and *Os*SSI are provided in the Supplementary Material.

Protein	*Hv*SSI	*Ta*SSI	*Os*SSI	*Hv* rice-like	*Hv* cat-dom
Activity (maltopentaose) (min^{-1})	16.8	13.3	48.1	42.6	39.9
Activity normalized to *Hv*SSI (%)	100	79	286	254	238
Activity (glycogen) (min^{-1})	22.9	30.6	136	74.2	68.8
Activity normalized to *Hv*SSI (%)	100	134	594	324	300

name was chosen based on the similarity to the length of the rice SSI protein) starts at Ser85, as found in mature endosperm rice SSI (Baba *et al.*, 1993). The second, 'cat-dom' *Hv*SSI (containing the catalytic domain) starts at Asn95, which was the first amino acid observed in our crystal model. Both mutants end at Met612, as in our *Hv*SSI construct. Both proteins could be expressed and purified without difficulty. Their activities with maltopentaose (10 m*M*) and rabbit liver glycogen (0.1 mg ml^{-1}) as acceptors were tested and were compared with those of the wild-type barley, wheat and rice SSI enzymes. Their specific activities are shown in Table 2.

Wheat SSI, which is very closely related to *Hv*SSI (98% identity), displayed activity levels comparable to those of the barley enzyme. The activity of both N-truncated mutants was similar and was about 2.8 times higher than the wild-type barley enzyme with maltopentaose and glycogen as acceptors. The values for the rice protein are very similar to both truncation mutants with maltopentaose as the acceptor and are twice as large (six times the activity of *Hv*SSI) with glycogen as the acceptor. Thus, it appears that truncating the N-terminal extension of barley to that which occurs naturally in rice endosperm conferred rice SSI-like properties. A similar result, with a (smaller) increase in V_{max}, has previously been reported for N-terminal truncations of SSI from maize with glycogen and amylopectin as substrates (Imparl-Radosevich *et al.*, 1998). The mechanisms responsible for the increased turnover are unclear, but considering the proximity of the first amino acids in our structure to the active-site cleft and the presence of 94 (115 including our affinity tag) extra spatially disordered amino acids at this position, steric effects impeding domain motion or substrate access could play a significant role. Since a simple truncation greatly enhanced the activity of *Hv*SSI, the extension might reduce activity for coordination with other starch synthases, control substrate selectivity or participate in interactions with other proteins.

3.7. Putative active site of *Hv*SSI

The active site of GT5 GSs includes amino-acid residues from both domains in a closed conformation. Our open-conformation structure does not have a catalytically

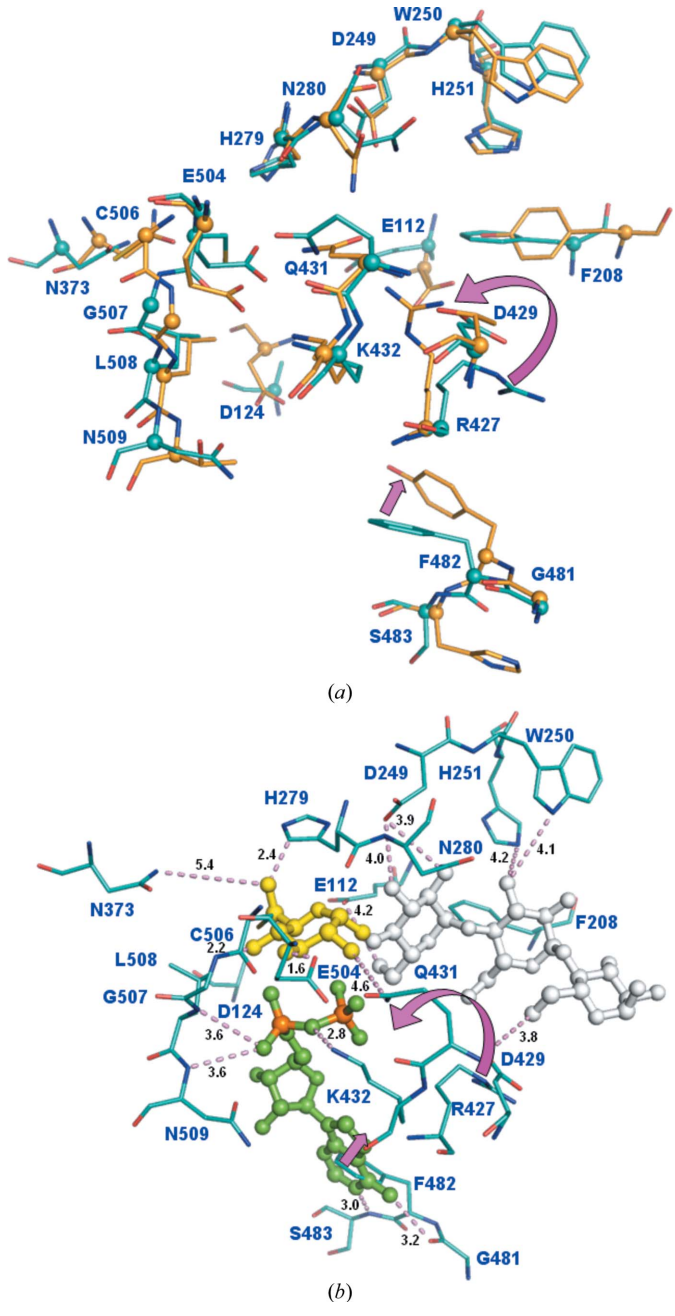


Figure 5
Comparison of residues in the active sites of *E. coli* glycogen synthase and barley SSI. (*a*) Comparison of residues common to both structures. *Ec*GS is shown with orange C atoms and *Hv*SSI with cyan C atoms; C^{α} atoms are highlighted by spheres. Numbering corresponds to *Hv*SSI. All *Ec*GS residues contacting either ligand are included except for His96 (*Ec*GS numbering), which has no equivalent in *Hv*SSI, and residues 15–18 of *Ec*GS, which correspond to a disordered loop in *Hv*SSI and are omitted for clarity. The pink arrows indicate movements that would bring side chains from *Hv*SSI into the same position that they occupy in *Ec*GS upon substrate binding. (*b*) The putative active site of *Hv*SSI with ligands taken from two different *Ec*GS structures (PDB entries 2qzs and 3cx4). ADP is coloured green with orange P atoms, glucose is coloured yellow and the acceptor maltotriose is coloured white. Dashed pink lines indicate hydrogen bonds present in *Ec*GS; they still have reasonable lengths (indicated in Å in the figure) in this model of *Hv*SSI even though the ligands are not present in our structure and no energy optimization was performed.

competent active site since the active-site residues are too far from each other. Also, the disulfide bond occupies the same space as the substrate ligands, thus precluding their binding. To compare the active site of *Hv*SSI in a closed conformation with that of *Ec*GS, which has structures in complex with an acceptor analogue, with donor and acceptor, we built a model by superimposing each domain of *Hv*SSI on their counterparts in *Ec*GS. For the linker, the entire proteins were used in the superposition. Only whole-domain rigid-body movements were used, without molecular dynamics or geometry optimization. We then analyzed the resultant positions of every amino acid involved in ADP, glucose or maltotriose (the acceptor) binding in *Ec*GS (Sheng, Jia *et al.*, 2009; Sheng, Yep *et al.*, 2009). The results are shown in Fig. 5. The structure of *Os*GBSSI including bound ADP was also superimposed and overall it overlaps well with the structure of *Ec*GS in this area. The ADP molecule occupies the same position in *Os*GBSSI and *Ec*GS except for the phosphate groups.

The modelled active site of *Hv*SSI is very similar to that of *Ec*GS. His96 in *Ec*GS is the only residue with no counterpart in *Hv*SSI, while Tyr95 of *Ec*GS is only spatially related to Phe208 of *Hv*SSI, occupying a different position in the sequence. Phe208 is the only residue where the *Os*GBSSI structure has no equivalent residue to its *Ec*GS counterpart, since this is part of the extended disordered loop in *Os*GBSSI. Amino acids 15–18 in *Ec*GS correspond to the disordered 116–123 loop in *Hv*SSI; hence, they have no equivalents in our structure (the *Os*GBSSI counterparts overlap well). *Hv*SSI Asp124 (at the end of this loop; only its main chain was modelled in the crystal) and Cys506, involved in the disulfide, occupy different positions to those in *Ec*GS by 2.2 and 2.8 Å, respectively. All of the other 18 residues involved occupy very similar positions in our model and the *Ec*GS structures, especially at the C^{α} level, and 15 of them have side chains pointing in the same direction. The other three are Ser483, which interacts with ADP through the main-chain amide, Phe482, which in the absence of the adenine moiety rotates its side chain to occupy that space, and Arg427, with a double hydrogen bond/ion pair to phosphate O atoms in *Ec*GS and pointing away in *Hv*SSI. The latter amino acids could adopt an *Ec*GS-like conformation in the presence of ligands (Fig. 5*a*). Thus, assuming that in the absence of the disulfide *Hv*SSI could adopt the closed conformation and the 116–123 region could fold as in *Ec*GS, both active sites would be virtually identical.

Without intending to perform a thorough discussion of the reaction mechanism based on a structure in which the substrates are not bound, some comments regarding the residue acting as a nucleophile can be made. In particular, His279, which is equivalent to His161 in *Ec*GS, and Glu504, which is equivalent to Glu377 in *Ec*GS, are in virtually equivalent positions, and similar spatial arrangements involving a carbonyl and a carboxylate group are found in the active sites of other retaining GT-B-fold glycosyltransferases (Lairson *et al.*, 2008). His279 in *Ec*GS was identified by Sheng, Jia *et al.* (2009) as a putative nucleophile using its main-chain carbonyl O atom by its placement relative to the anomeric C atom of glucose. Glu377 in *Ec*GS was identified as critical in positioning the glucose of the donor by Sheng, Jia *et al.* (2009), but is also considered to be a potential nucleophile (Yep *et al.*, 2004). This residue is hydrogen-bonded to O3 of the glucose to be transferred in the *Ec*GS structure, but conformational changes could plausibly bring it into close contact with the anomeric C atom of the glucose, allowing it to act as a nucleophile. Mutagenesis experiments are ongoing in our group to attempt to identify the essential amino acids and possible nucleophiles of starch synthases.

The only significant differences around the active site start at His96 of *Ec*GS, which binds the glucose at the position +3 from the transferred glucose. We conclude that the reaction mechanism of *Hv*SSI is similar to that of *Ec*GS (the details of which remain largely unknown) and the conformation of the bound ADP-glucose and acceptor molecules are likely to be the same (Fig. 5*b*). Differences affecting substrate selectivity might involve glucose residues at or beyond subsite +3, including surface binding sites rather than residues in the active site.

This work was supported by the Carlsberg Foundation. Access to synchrotron beam time at ESRF and MAX-lab was made possible by support from DANSCATT. The authors would like to thank Dr Lyann Sim and Dr Rene Jørgensen for helpful discussions and assistance with data collection.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Baba, T., Nishihara, M., Mizuno, K., Kawasaki, T., Shimada, H., Kobayashi, E., Ohnishi, S., Tanaka, K. & Arai, Y. (1993). *Plant Physiol.* **103**, 565–573.
Barford, D. & Johnson, L. N. (1992). *Protein Sci.* **1**, 472–493.
Boraston, A. B., Healey, M., Klassen, J., Ficko-Blean, E., Lammerts van Bueren, A. & Law, V. (2006). *J. Biol. Chem.* **281**, 587–598.
Breton, C., Fournel-Gigleux, S. & Palcic, M. M. (2012). *Curr. Opin. Struct. Biol.* **22**, 540–549.
Breton, C., Šnajdrová, L., Jeanneau, C., Koca, J. & Imberty, A. (2006). *Glycobiology*, **16**, 29R–37R.
Buschiazzo, A., Ugalde, J. E., Guerin, M. E., Shepard, W., Ugalde, R. A. & Alzari, P. M. (2004). *EMBO J.* **23**, 3196–3205.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Chou, W.-I., Pai, T.-W., Liu, S.-H., Hsiung, B.-K. & Chang, M. D.-T. (2006). *Biochem. J.* **396**, 469–477.
Commuri, P. D. & Keeling, P. L. (2001). *Plant J.* **25**, 475–486.
Coutinho, P. M., Deleury, E., Davies, G. J. & Henrissat, B. (2003). *J. Mol. Biol.* **328**, 307–317.
Damager, I., Engelsen, S. B., Blennow, A., Møller, B. L. & Motawia, M. S. (2010). *Chem. Rev.* **110**, 2049–2080.
Díaz, A., Martínez-Pons, C., Fita, I., Ferrer, J. C. & Guinovart, J. J. (2011). *J. Biol. Chem.* **286**, 18505–18514.
Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. (2000). *J. Mol. Biol.* **300**, 1005–1016.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Fettke, J., Leifels, L., Brust, H., Herbst, K. & Steup, M. (2012). *J. Exp. Bot.* **63**, 3011–3029.
Fincher, G. B. (1989). *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**, 305–346.

Food and Agriculture Organization of the United Nations (2012). *FAO Statistical Yearbook 2012*. http://www.fao.org/docrep/015/i2490e/i2490e00.htm.
Fujita, N., Yoshida, M., Asakura, N., Ohdan, T., Miyao, A., Hirochika, H. & Nakamura, Y. (2006). *Plant Physiol.* **140**, 1070–1084.
Fujita, N., Yoshida, M., Kondo, T., Saito, K., Utsumi, Y., Tokunaga, T., Nishi, A., Satoh, H., Park, J.-H., Jane, J.-L., Miyao, A., Hirochika, H. & Nakamura, Y. (2007). *Plant Physiol.* **144**, 2009–2023.
Geigenberger, P. (2011). *Plant Physiol.* **155**, 1566–1577.
Getz, E. B., Xiao, M., Chakrabarty, T., Cooke, R. & Selvin, P. R. (1999). *Anal. Biochem.* **273**, 73–80.
Giardina, T., Gunning, A. P., Juge, N., Faulds, C. B., Furniss, C. S., Svensson, B., Morris, V. J. & Williamson, G. (2001). *J. Mol. Biol.* **313**, 1149–1159.
Gibson, R. P., Tarling, C. A., Roberts, S., Withers, S. G. & Davies, G. J. (2004). *J. Biol. Chem.* **279**, 1950–1955.
Glaring, M. A., Skryhan, K., Kötting, O., Zeeman, S. C. & Blennow, A. (2012). *Plant Physiol. Biochem.* **58**, 89–97.
Gosselin, S., Alhussaini, M., Streiff, M. B., Takabayashi, K. & Palcic, M. M. (1994). *Anal. Biochem.* **220**, 92–97.
Higson, A. & Smith, C. (2011). *NNFCC Renewable Chemicals Factsheet: Starch.* http://www.nationalstemcentre.org.uk/elibrary/file/14774/Starch.pdf.
Horcajada, C., Guinovart, J. J., Fita, I. & Ferrer, J. C. (2006). *J. Biol. Chem.* **281**, 2923–2931.
Imparl-Radosevich, J. M., Li, P., Zhang, L., McKean, A. L., Keeling, P. L. & Guan, H. (1998). *Arch. Biochem. Biophys.* **353**, 64–72.
Jeon, J.-S., Ryoo, N., Hahn, T.-R., Walia, H. & Nakamura, Y. (2010). *Plant Physiol. Biochem.* **48**, 383–392.
Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. & Madden, T. L. (2008). *Nucleic Acids Res.* **36**, W5–W9.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. (2008). *Annu. Rev. Biochem.* **77**, 521–555.
Lammerts van Bueren, A., Finn, R., Ausió, J. & Boraston, A. B. (2004). *Biochemistry*, **43**, 15633–15642.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
Lee, S. S., Hong, S. Y., Errey, J. C., Izumi, A., Davies, G. J. & Davis, B. G. (2011). *Nature Chem. Biol.* **7**, 631–638.
Li, Z., Rahman, S., Kosar-Hashemi, B., Mouille, G., Appels, R. & Morell, M. K. (1999). *Theor. Appl. Genet.* **98**, 1208–1216.
Liu, F., Romanova, N., Lee, E. A., Ahmed, R., Evans, M., Gilbert, E. P., Morell, M. K., Emes, M. J. & Tetlow, I. J. (2012). *Biochem. J.* **448**, 373–387.
Momma, M. & Fujimoto, Z. (2012). *Biosci. Biotechnol. Biochem.* **76**, 1591–1595.
Mu-Forster, C., Huang, R., Powers, J. R., Harriman, R. W., Knight, M., Singletary, C. W., Keeling, P. L. & Wasserman, B. P. (1996). *Plant Physiol.* **111**, 821–829.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Nakamura, Y., Francisco, P. B., Hosaka, Y., Sato, A., Sawada, T., Kubo, A. & Fujita, N. (2005). *Plant Mol. Biol.* **58**, 213–227.
Nielsen, M. M., Bozonnet, S., Seo, E.-S., Mótyán, J. A., Andersen, J. M., Dilokpimol, A., Abou Hachem, M., Gyémánt, G., Naested, H., Kandra, L., Sigurskjold, B. W. & Svensson, B. (2009). *Biochemistry*, **48**, 7686–7697.
Pace, C. N. & Scholtz, J. M. (1998). *Biophys. J.* **75**, 422–427.
Patron, N. J., Smith, A. M., Fahy, B. F., Hylton, C. M., Naldrett, M. J., Rossnagel, B. G. & Denyer, K. (2002). *Plant Physiol.* **130**, 190–198.
Pérez, S. & Bertoft, E. (2010). *Starch – Stärke*, **62**, 389–420.
Qasba, P. K., Ramakrishnan, B. & Boeggeman, E. (2005). *Trends Biochem. Sci.* **30**, 53–62.
Richardson, J. S. & Richardson, D. C. (1988). *Science*, **240**, 1648–1652.
Robert, X., Haser, R., Mori, H., Svensson, B. & Aghajari, N. (2005). *J. Biol. Chem.* **280**, 32968–32978.
Robyt, J. F. & Mukerjea, R. (1994). *Carbohydr. Res.* **251**, 187–202.
Sheng, F., Jia, X., Yep, A., Preiss, J. & Geiger, J. H. (2009). *J. Biol. Chem.* **284**, 17796–17807.
Sheng, F., Yep, A., Feng, L., Preiss, J. & Geiger, J. H. (2009). *Biochemistry*, **48**, 10089–10097.
Smith, A. M. (2008). *Plant J.* **54**, 546–558.
Smith, A. M. & Stitt, M. (2007). *Plant Cell Environ.* **30**, 1126–1149.
Soya, N., Fang, Y., Palcic, M. M. & Klassen, J. S. (2011). *Glycobiology*, **21**, 547–552.
Szydlowski, N., Ragel, P., Raynaud, S., Lucas, M. M., Roldán, I., Montero, M., Muñoz, F. J., Ovecka, M., Bahaji, A., Planchot, V., Pozueta-Romero, J., D'Hulst, C. & Mérida, A. (2007). *Plant Cell*, **21**, 2443–2457.
Tester, R. F. & Karkalas, J. (2002). *Starch in Biopolymers*, p. 381. Weinheim: Wiley-VCH.
Tetlow, I. J., Beisel, K. G., Cameron, S., Makhmoudova, A., Liu, F., Bresolin, N. S., Wait, R., Morell, M. K. & Emes, M. J. (2008). *Plant Physiol.* **146**, 1878–1891.
Thitisaksakul, M., Jiménez, R. C., Arias, M. C. & Beckles, D. M. (2012). *J. Cer. Sci.* **56**, 67–80.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Wang, W. & Malcolm, B. A. (1999). *Biotechniques*, **26**, 680–682.
Yep, A., Ballicora, M. A., Sivak, M. N. & Preiss, J. (2004). *J. Biol. Chem.* **279**, 8359–8367.
Yoo, S.-H. & Jane, J. (2002). *Carbohydr. Polym.* **49**, 297–305.
Zeeman, S. C., Kossmann, J. & Smith, A. M. (2010). *Annu. Rev. Plant Biol.* **61**, 209–234.
Zhang, X., Colleoni, C., Ratushna, V., Sirghie-Colleoni, M., James, M. G. & Myers, A. M. (2004). *Plant Mol. Biol.* **54**, 865–879.