

Three-Dimensional Structures of UDP-Sugar Glycosyltransferases Illuminate the Biosynthesis of Plant Polysaccharides¹

Simon J. Charnock, Bernard Henrissat, and Gideon J. Davies*

Department of Chemistry, Structural Biology Laboratory, University of York, Heslington, York YO10 5DD, United Kingdom (S.J.C., G.J.D.); and Architecture et Fonction des Macromolécules Biologiques, Unité Mixte de Recherche 6098, Centre National de la Recherche Scientifique, Universités d'Aix-Marseille I and II, 31 Chemin Joseph Aiguier, 13402 Marseille cedex 20, France (B.H.)

Cell-free preparations of the bacterium *Acetobacter xylinum* were first reported to synthesize cellulose from UDP-Glc over 40 years ago (Glaser, 1958). Despite the elucidation of the primary sequence of the cellulose synthase from this organism in 1990 (Wong et al., 1990), a consistent mechanism to account for the biosynthesis of this and related polysaccharides has remained elusive. Opposing views of the number of catalytic centers and the molecular directionality of the synthesis have been presented (for review, see Delmer, 1999). A comprehensive classification of glycosyltransferases harnessed to the recent structural determinations of UDP-sugar dependent β -glycosyltransferases, including a cellulose synthase homolog, permits a preliminary illumination of this controversial area. The number of glycosyl transfer centers in the catalytic domain of cellulose synthase remains controversial: a two-center model has been proposed (Saxena et al., 1995), but we find it hard to reconcile with the wealth of experimental data on the three-dimensional structure of glycosyltransferases.

The sequence family classification system, originally developed for the glycoside hydrolases, has recently been extended to include the activated-sugar dependent glycosyltransferases (Campbell et al., 1997). Forty-eight families are known at the present date and may be found in a continuously updated database at <http://afmb.cnrs-mrs.fr/~pedro/CAZY/db.html> (for review, see Henrissat and Davies, 2000). Activated-sugar dependent transferases account for the vast majority of glycosyl transfer on earth. The activating group may be a phosphate, a lipid phosphate, or a nucleotide, and the reaction mechanism proceeds with either retention or inversion of the anomeric configuration of the donor sugar.

One of the features of the sequence-family classification is that the reaction mechanism is conserved within each family. Cellulose synthase, in family "GT-2," is an "invertin" glycosyltransferase, i.e. it uses α -linked UDP-sugars to generate a β -linked product. Inverting transferases are assumed to use a single displacement mechanism with nucleophilic attack by the acceptor species at the C-1 (anomeric) carbon of the donor sugar. Such a mechanism is generally believed to demand a base to activate the sugar acceptor for nucleophilic attack by deprotonation because sugar hydroxyls are in themselves quite poor nucleophiles. For most enzymes the reaction also involves an additional carboxylate or carboxylates to coordinate a divalent metal ion on the phosphate group(s) of the nucleotide (Fig. 1).

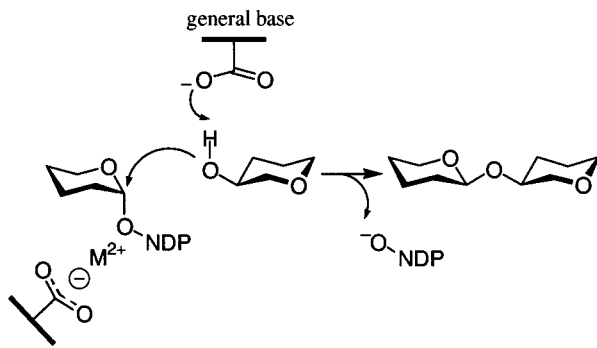
THE THREE-DIMENSIONAL STRUCTURE OF A GLYCOSYLTRANSFERASE FROM FAMILY GT-2

In the last few months, six different glycosyltransferases structures from different sequence families have been published or presented (for review, see Ünligil and Rini, 2000). This recent burst of structural activity began with the resolution of the three-dimensional structure of SpsA, a glycosyltransferase implicated in the sporulation response of *Bacillus subtilis* (Charnock and Davies, 1999). This structure is particularly relevant to those studying the biosynthesis of plant polysaccharides, such as cellulose, because it is found in glycosyltransferase family GT-2, which includes cellulose synthase and over 40 other open reading frames (ORFs) in *Arabidopsis* alone (Henrissat et al., 2001).

The three-dimensional structure of both native and Mn-UDP complexes of SpsA revealed a mixed α/β protein with two domains and featuring a central β -sheet core flanked by α -helices, Figure 2a. The N-terminal region of SpsA (residues 1–99) is a classical nucleotide-binding domain of four parallel β -strands flanked on either side by two α -helices and is the binding site for UDP. Many of the "signature motifs" of family GT-2 are contained in this

¹ This work was supported by the European Commission, the Biotechnology and Biological Science Research Council, the Wellcome Trust, the University of York, and the Centre National de la Recherche Scientifique. G.J.D is a Royal Society University Research Fellow.

* Corresponding author; e-mail davies@ysbl.york.ac.uk; fax 44-1904-410519.



carboxylate(s) to coordinate the divalent metal ion

Figure 1. The putative mechanism for an inverting nucleotide-sugar glycosyltransferase. Such a mechanism most likely requires at least two catalytic carboxylates, one to function as a base to activate the acceptor species and at least one more to coordinate a divalent metal ion associated with the UDP-sugar. In many enzymes the metal-ion coordination involves two carboxylates in what is misleadingly (because it is often seen as DDX, xDD, DDD) termed the "D(x) D" motif.

N-terminal UDP binding domain, which equates to the "A-domain" described for cellulose synthase, Figure 2b. The Rib and Mn phosphate are coordinated by Asp-98 and Asp-99 which correspond to the "D(x) D" motif found in many families of glycosyltransferase. The second, C-terminal, domain is the site for binding of the acceptor species, and we predicted that Asp-191, in this acceptor-binding region, would function as the catalytic base in the inverting mechanism.

THREE-DIMENSIONAL STRUCTURES OF OTHER INVERTING GLYCOSYLTRANSFERASES

In addition to SpsA, three additional glycosyltransferase structures of relevance to the synthesis of plant polysaccharides have also recently been determined: the β -galactosyl transferase from family GT-7 (Gastinel et al., 1999), the rabbit GalNAc transferase from family GT-13 (Ünlügil et al., 2000), and the human β -1,3-glucuronyltransferase I from family GT-43 (Pedersen et al., 2000). The striking feature of these four structures is their structural and mechanistic similarity, despite insignificant sequence identity (Ünlügil and Rini, 2000). The N-terminal UDP-binding domain is well-conserved and the critical amino acids "D... D(x) D" that interact with UDP are invariant. Biochemical confirmation of the roles of these residues also comes from site-directed mutagenesis on another family GT-2 enzyme, ExoM from *Sinorhizobium mellioli* (Garinot-Schneider et al., 2000). All four structures present a C-terminal acceptor-binding domain and it is particularly important that the residue predicted to be the catalytic base in SpsA (Charnock and Davies, 1999), Asp-191, has correspondence in

all four enzyme families. In the structures from families 2, 7, and 13 this residue is an Asp, in family GT-43 it is a Glu. In the family GT-43 human β -1,3-glucuronyltransferase structure, the sugar acceptor species is seen hydrogen bonding to the base in position for nucleophilic attack on the UDP-sugar donor. Therefore, these enzyme families use a minimum of four carboxylates to form a single catalytic center: Three aspartates are involved in UDP coordination whereas a fourth residue in the acceptor domain, Asp or Glu, acts as the catalytic base.

THE BIOSYNTHESIS OF CELLULOSE: CONFLICTING PROPOSALS?

The mode of action of polymerizing family GT-2 enzymes such as cellulose synthase has been plagued by contradicting speculations (for review, see Davies and Charnock, 1999). Are these polymers extended at their reducing or non-reducing ends, does the mechanism involve lipid pyrophosphate intermediates or direct transfer from the nucleotide-sugar donor, and are there one or more catalytic centers per peptide chain?

^{14}C -pulse-chase experiments with *Acetobacter xylinum* favored extension at the reducing end (Han and Robyt, 1998). [^{14}C]UDP work suggested a similar mechanism for hyaluronan synthase, a related family GT-2 enzyme (for review, see DeAngelis, 2000). However, numerous experiments have demonstrated recently that elongation by family 2 enzymes occurs by polymerization at the non-reducing end of the growing polysaccharide. Electron crystallography provided the first experimental evidence for cellulose synthesis (Koyama et al., 1997), and this has been followed by numerous reports unambiguously demonstrating non-reducing end elongation by direct transfer from the nucleotide-sugar (DeAngelis, 1999; Kamst et al., 1999; Cartee et al., 2000).

THE "TWO-CENTER" MODEL FOR CELLULOSE SYNTHASE

Cellulose is a β -1,4-linked polymer of D-Glc, in which adjacent monosaccharides are rotated through 180° with respect to one another, and thus the "structural" repeating unit is considered to be cellobiose. One can speculate that in the case of cellulose, this alternating orientation of monomers provides a "torsional" problem for catalysis. Likewise, for the synthesis of hyaluronan two different sugars, with two different linkages, need to be added sequentially. Attempts to relieve these perceived problems led to the proposal of a dual active-center model for "processive" or polymerising enzymes (Saxena et al., 1995).

Hydrophobic cluster analysis of family GT-2 β -glycosyltransferases correctly identified that these glycosyltransferases were two domain proteins. The A-domain contained the conserved family GT-2 motifs "D...D(x)D," which were common to both polymerizing and single-addition transferases. The B-domain contained a fourth invariant Asp, which in the polymerizing enzymes only, was followed by a characteristic QxxRW motif (Saxena et al., 1995). The two-domain architecture, coupled to observation of four conserved aspartates spread over these two domains, led to the development of a two catalytic center model in which the A and B domains were each proposed to function as glycosyl transfer centers. The effective addition of a disaccharide occurred via the dual addition of monosaccharides.

The double addition of monosaccharides provided an extremely seductive solution to potential problems associated with the synthesis of cellulose. Furthermore, it could also explain the synthesis of hyaluronan (a polymer of alternating β -1,4-linked *N*-acetylglucosamine and β -1,3-linked GlcA) by class-I hyaluronan synthases. The two-center model initially featured reducing-end elongation with three UDP-sugar binding sites and two glycosyl transfer centers (Saxena et al., 1995). It later evolved to feature two centers for glycosyl transfer and two UDP-sugar binding sites when direct addition to the non-reducing end became favored (Koyama et al., 1997; Saxena and Brown, 1997; Carpita and Vergara, 1998).

IS THE TWO CENTER MODEL CONSISTENT WITH RECENT THREE-DIMENSIONAL STRUCTURES?

The release of single, monosaccharide adducts by root nodulation factor, NodC (another polymerising family GT-2 enzyme) is not consistent with the dual addition mechanism (Kamst et al., 1999). Furthermore, on many systems mutations of the conserved aspartates, in either domain, tends to abolish catalytic activity completely, consistent with the notion that together these residues form just a single transfer center (Yoshida et al., 2000). What do the three-dimensional structures of glycosyltransferases tell us about this confusing area?

The two catalytic-center model is based upon the assumption that four aspartates, spread over two domains in a 230-amino-acid fragment, are sufficient to constitute two discrete glycosyl transfer centers (Saxena et al., 1995). Yet, the family 2, 7, 13, and 43 glycosyltransferase structures reveal that four such residues are required to form a single viable catalytic center. These proteins all display two domains, one of which binds the UDP-sugar and the other the acceptor. The N-terminal UDP binding domain of these proteins (approximately 100 amino acids) contains the "D...

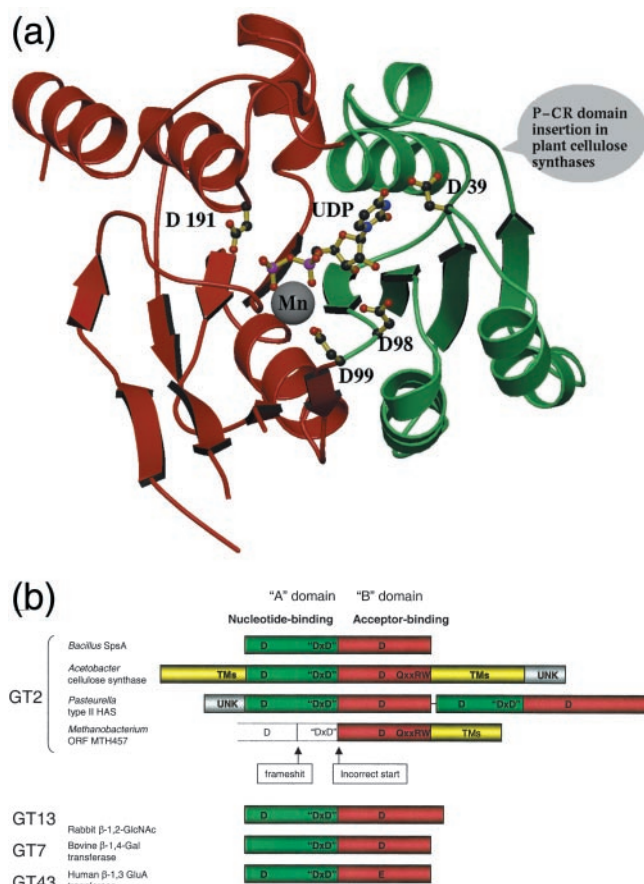


Figure 2. A, The three-dimensional structure of a family GT-2 UDP-sugar-dependent glycosyltransferase, SpsA, from *Bacillus subtilis*. The conserved N-terminal nucleotide-binding domain is shown in green and the C-terminal acceptor domain in red. Four aspartates, 39, 98, 99, and 191 discussed in the text are shown in ball-and-stick representation as is the UDP molecule. The Mn^{2+} ion is shown as a shaded sphere. Plant sequences also contain insertions in the basic core structure. Within the N-terminal UDP-binding domain there is a "P-CR" insertion of approximately 115 residues in the Arabidopsis RSW1 and IRX3 cellulose synthases and the length of this region varies from approximately 70 to approximately 190 residues for other plant cellulose synthase homologs. This insertion occurs between strand β -3 and β -4 in the three-dimensional structure as indicated. Plant cellulose synthases also have an insertion in the "B" domain termed an "HVR" domain (not shown; Pear et al., 1996). There are no extensive insertions in the bacterial cellulose synthases such as the *Acetobacter* enzyme whose catalytic core domain is essentially the same size as SpsA (Saxena et al., 1995). This figure was drawn with MOLSCRIPT/BOBSCRIPT (Kraulis, 1991; Esnouf, 1997). B, Schematic representation of family GT-2 and related glycosyltransferases (from families 7, 13, and 43 whose three-dimensional structures are known) indicating the position of important catalytic and UDP-binding residues. "TM" indicates the presence of extensive transmembrane segments. ORF MTH457 from *Methanobacterium thermoautotrophicum* has been cited as an enzyme containing just a "B" domain. Close examination of the nucleotide sequence, however, shows that this is not the case. The apparent lack of the "A" domain results purely from an incorrect start codon in the genome-annotation, compounded by a frame-shift error. Errors associated with this ORF are therefore shown for reference.

D(x) D'' motif. It is equivalent to the complete A-domain of cellulose synthase. It is extremely hard for these authors to reconcile demonstration that this domain is a UDP-binding domain and thus uses the three aspartates merely to bind UDP with proposals that it may function as a discrete catalytic entity, the A-domain, in a two catalytic-center protein.

We therefore propose that the catalytic core domain of cellulose synthase (and related polymerizing family GT-2 enzymes) is similar to these four glycosyltransferases. They all consist of a two-domain protein in which the N-terminal A-domain binds the UDP-sugar and the C-terminal B-domain forms the acceptor binding region and completes the single transfer center by provision of the catalytic base. A few enzymes in family GT-2, such as the type-II hyaluronan synthases do possess two repeats of both of the UDP and acceptor domains in a genuine two-center enzyme whose two catalytic centers are consequently spread over 6 to 700 amino acids (Jing and DeAngelis, 2000). In contrast with the results on cellulose synthase or the type-I hyaluronan synthase (Yoshida et al., 2000) mutation of the aspartates in this genuine two-center type-II hyaluronan synthase abolish just one transfer activity.

How a polypeptide with a single active-site could account for the synthesis of an alternating polysaccharide such as hyaluronan is not understood. A general feature of polymerizing glycosyltransferases is that the reaction product of one addition becomes the acceptor for a subsequent events. One can speculate that the last residue added on the growing hyaluronan chain (either a GlcNAc or a GlcA), once in the acceptor site, may tune the affinity of the donor site for the complementary sugar-nucleotide. This may be controlled in part by the negatively charged carboxylate of the GlcA moiety. Small changes in the active center environment are known to change the specificity of transferases. The specificity "switch" from blood group α -Gal to α -GalNAc transferase demands just a single amino acid substitution (Seto et al., 1999) and the *Campylobacter jejuni* enzyme, Cst-II, changes its regioselectivity depending on the nature of the acceptor species (Gilbert et al., 2000).

These recent results on the three-dimensional structures and mechanisms of inverting glycosyltransferases begin to reveal the intimate details of oligosaccharide biosynthesis. Although these enzymes appear unrelated to plant biochemistry, they actually have great relevance for plant polysaccharides. Many of the unresolved issues may now be placed on a stronger experimental foundation. In particular, the roles of many of the conserved motifs in cellulose synthase and related enzymes may now be assigned. By far the vast majority of data point to polymer elongation at the non-reducing end by direct transfer from the nucleotide sugar donor. Fur-

thermore, we propose that most family GT-2 transferases, including cellulose synthase, use the A and B domains, and the four conserved aspartates to form a single center for glycosyl transfer. The three-dimensional structure of the GT-2 enzyme SpsA does not rule out all possible two-center models such as dimerization or the recruitment of other domains for catalysis. If the majority of family 2 enzymes use a single center, perceived problems with torsional stress and addition of alternating sugars require new explanations. Only when the three-dimensional structure of a polymerizing family GT-2 glycosyltransferase is reported will these issues be fully resolved.

Received September 20, 2000; returned for revision October 10, 2000; accepted November 9, 2000.

LITERATURE CITED

- Campbell JA, Davies GJ, Bulone V, Henrissat B (1997) *Biochem J* **326**: 929–942
- Carpita N, Vergara C (1998) *Science* **279**: 672–673
- Cartee RT, Forsee WT, Schutzbach JS, Yother J (2000) *J Biol Chem* **275**: 3907–3914
- Charnock SJ, Davies GJ (1999) *Biochemistry* **38**: 6380–6385
- Davies GJ, Charnock SJ (1999) *In* HJ Gilbert, GJ Davies, B Henrissat, B Svensson, eds, *Recent Advances in Carbohydrate Bioengineering*. Royal Society of Chemistry, Cambridge, UK, pp 132–143
- DeAngelis PL (1999) *J Biol Chem* **274**: 26557–26562
- DeAngelis PL (2000) *Cell Mol Life Sci* **56**: 670–682
- Delmer DP (1999) *Annu Rev Plant Physiol Plant Mol Biol* **50**: 245–276
- Esnouf RM (1997) *J Mol Graphics* **15**: 133–138
- Garinot-Schneider C, Lellouch AC, Geremia RA (2000) *J Biol Chem* **275**: 31407–31413
- Gastinel LN, Cambillau C, Bourne Y (1999) *EMBO J* **18**: 3546–3557
- Gilbert M, Brisson JR, Karwaski MF, Michniewicz J, Cunningham AM, Wu Y, Young NM, Wakarchuk WW (2000) *J Biol Chem* **275**: 3896–3906
- Glaser L (1958) *J Biol Chem* **232**: 627–636
- Han NS, Robyt JF (1998) *Carbohydr Res* **313**: 125–133
- Henrissat B, Couthino P, Davies GJ (2001) *Plant Mol Biol Rev* (in press)
- Henrissat B, Davies GJ (2000) *Plant Physiol* **124**: 1515–1519
- Jing W, DeAngelis PL (2000) *Glycobiology* **10**: 883–889
- Kamst E, Bakkers J, Quaedvleig NEM, Pilling J, Kijne JW, Lugtenberg BJJ, Spaik HP (1999) *Biochemistry* **38**: 4045–4052
- Koyama M, Helbert W, Imai T, Sugiyama J, Henrissat B (1997) *Proc Natl Acad Sci USA* **94**: 9091–9095
- Kraulis PJ (1991) *J Appl Cryst* **24**: 946–950

- Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP** (1996) *Proc Natl Acad Sci USA* **93**: 12637–12642
- Pedersen LC, Tsuchida K, Kitagawa H, Sugahara K, Darden TA, Negishi M** (2001) *J Biol Chem* **275**: 34580–34585
- Saxena IM, Brown RM** (1997) *Cellulose* **4**: 33–49
- Saxena IM, Brown RM, Fevre M, Geremia RA, Henrissat B** (1995) *J Bacteriol* **177**: 1419–1424
- Seto NO, Compston CA, Evans SV, Bundle DR, Narang SA, Palcic MM** (1999) *Eur J Biochem* **259**: 770–775
- Ünlügil UM, Rini JM** (2000) *Curr Opin Struct Biol* **10**: 510–517
- Ünlügil UM, Zhou S, Yuwaraj S, Sarkar M, Schachter H, Rini JM** (2000) *EMBO J* **19**: 5269–5280
- Wong HC, Fear AL, Calhoon RD, Eichinger GH, Mayer R, Amikam D, Benziman M, Gelfand DH, Measde JH, Emerick AW, Bruner R, Ben-Bassat A, Tal R** (1990) *Proc Natl Acad Sci USA* **87**: 8130–8134
- Yoshida M, Itano N, Yamada Y, Kimata K** (2000) *J Biol Chem* **275**: 497–506