PERGAMON

Phytochemistry 57 (2001) 1135–1148

PHYTOCHEMISTRY

www.elsevier.com/locate/phytochem

Review

# Structure–function characterization of cellulose synthase: relationship to other glycosyltransferases

Inder M. Saxena[a], R. Malcolm Brown Jr.[a,*], Thomas Dandekar[b]

[a]*Section of Molecular Genetics and Microbiology, School of Biological Sciences, University of Texas at Austin, Austin, TX 78712, USA*
[b]*European Molecular Biology Laboratory, Postfach 102209, D-69012 Heidelberg, Germany*

Received 18 September 2000, received in revised form 2 February 2001

**Abstract**

A combined structural and functional model of the catalytic region of cellulose synthase is presented as a prototype for the action of processive β-glycosyltransferases and other glycosyltransferases. A 285 amino acid segment of the *Acetobacter xylinum* cellulose synthase containing all the conserved residues in the globular region was subjected to protein modeling using the genetic algorithm. This region folds into a single large domain with a topology exhibiting a mixed alpha/beta structure. The predicted structure serves as a topological outline for the structure of this processive β-glycosyltransferase. By incorporating new site-directed mutagenesis data and comparative analysis of the conserved aspartic acid residues and the QXXRW motif we deduce a number of functional implications based on the structure. This includes location of the UDP–glucose substrate-binding cavity, suggestions for the catalytic processing including positions of conserved and catalytic residues, secondary structure arrangement and domain organization. Comparisons to cellulose synthases from higher plants (genetic algorithm based model for cotton CelA1), data from neural network predictions (PHD), and to the recently experimentally determined structures of the non-processive SpsA and β4-galactosyltransferase retest and further validate our structure-function description of this glycosyltransferase. 

*Keywords*: Cellulose; Cellulose synthase; *Acetobacter xylinum*; Glycosyltransferases; Genetic algorithm

## Contents



* Corresponding author. Tel.: +1-512-471-3364; fax: +1-512-471-3573.
*E-mail address*: rmbrown@mail.utexas.edu (R.M. Brown).

0031-9422/01/$ - see front matter 
PII: S0031-9422(01)00048-6

## 1. Introduction

Cellulose is an abundant polysaccharide produced most notably by plants and in varying degrees by several other organisms including algae, bacteria, social amoeba, and tunicates. The polymerization reaction is catalyzed by the enzyme cellulose synthase, a glycosyltransferase that uses UDP–glucose as the substrate. Most glycosyltransferases transfer monosaccharides from a nucleotide sugar donor to hydroxyl groups of acceptor carbohydrates. A glycosyltransferase may catalyze the transfer of a single sugar residue to an acceptor (non-processive) or transfer more than one sugar residue to the acceptor (processive). Glycosyltransferases exhibit specificity for both the donor and the acceptor molecules and are involved in making a specific linkage (α- or β-linkage) between specific atoms of the sugar and the acceptor molecule. Since each linkage is catalyzed by a specific glycosyltransferase, most compounds that either have different kinds of sugars and/or different linkages would require a number of glycosyltransferases for their synthesis. Alternatively, a single enzyme may recognize more than one sugar donor and catalyze more than one type of linkage (for example, hyaluronan synthase in the synthesis of hyaluronic acid). Based upon the linkage at the anomeric carbon, glycosyltransferases can be classified as retaining glycosyltransferases or inverting glycosyltransferases. Since the catalytic mechanism in the two types (retention or inversion) of glycosyl transfer reactions is different, it is generally believed that the two classes of glycosyltransferases would have different sequence and structure characteristics.

The catalytic machinery of glycosyltransferases involves a general acid-base catalysis mechanism that involves acidic residues (aspartate and/or glutamate) whose side chains have the property to act as the general base for acceptor activation or as the nucleophile for formation of a glycosyl–enzyme intermediate. These acidic residues are also involved in the catalytic mechanisms of glycosyl hydrolases where once again they act either as a general base or as a nucleophile (Davies and Henrissat, 1995). Active site glutamate residues have also been shown to be essential for the ADP-ribosyltransferase activity of certain bacterial toxins (Barth et al., 1998) as well as the glycosylphosphatidylinositol (GPI)-anchored and secretory ADP–ribosyltransferases of eucaryotic cells (Okazaki and Moss, 1998). The carboxylate group of an aspartic acid residue (Asp 100) in the active site of T4 β-glucosyltransferase has been shown to be in a position where it activates the 5-hydroxymethyl group on the DNA for nucleophilic attack on the glucose of the donor UDP–glucose (Moréra et al, 1999). Enzymes that catalyze the different glycosyl transfer reactions and that use NDP–sugar as substrates were classified into 26 families based upon sequence similarity (Campbell et al., 1997) and the list now includes 47 families (http://afmb.cnrs-mrs.fr/~pedro/CAZY/gtf.html). Family 2 β-glycosyltransferases include both the processive and the non-processive enzymes. The processive enzymes (for example, cellulose synthase, chitin synthase, hyaluronan synthase, NodC protein) are generally membrane-proteins that synthesize a polysaccharide/oligosaccharide chain. None of these proteins have been characterized and compared structurally; however, the different proteins are related based on sequence characteristics (especially identification of putative catalytic residues). The presence of the conserved catalytic residues in members of the same glycosyltransferase family suggests that they catalyze glycosyl transfer reactions by similar mechanisms and in all likelihood have very similar active-site architecture. Processive glycosyltransferases stay attached to the growing end of a glucan chain, and there is considerable interest in identifying the amino acid residues and the structure of the region that is involved in holding the glucan chain.

Cellulose biosynthesis offers a simple and elegant system for studies on the synthesis of polysaccharides, especially as studied in the bacterium *A. xylinum*. Although the detailed features of cellulose biosynthesis may not be common to the synthesis of other polysaccharides, the structure of the catalytic region of cellulose synthase and other glycosyltransferases may show some conserved features. A prediction of the structure of the globular region of cellulose synthase based on the available data will therefore be helpful in understanding the structure and function of glycosyltransferases, especially the processive β-glycosyltransferases. Moreover it will be important to confirm the relevance of the conserved residues in these glycosyltransferases and to determine their spatial relationship within the three-dimensional structure. More recently the structures of a few β-glycosyltransferases have been determined experimentally (Vrielink et al., 1994; Charnock and Davies, 1999; Gastinel et al., 1999). Although these structures show some similarities, they also exhibit sufficient variations making it all the more important to predict or determine the three-dimensional structures of as many glycosyltransferases for understanding the catalytic properties of this group of enzymes (Breton and Imberty, 1999).

### *1.1. Sequence relationship between cellulose synthases and other glycosyltransferases*

Genes for cellulose synthase were first cloned from the bacterium *A. xylinum* (Saxena et al., 1990; Wong et al., 1990). Sequence comparisons of the derived amino acid sequence of the bacterial cellulose synthase with other proteins revealed the presence of conserved aspartic acid residues and a QXXRW motif in the processive β-glycosyltransferases (Saxena et al., 1995). Since the catalytic machinery of glycosyltransferases is likely to involve residues whose side chains have the proper reactivity to act as

a general base or a nucleophile, the conserved aspartic acid residues are predicted to be involved in the catalytic reaction as acid-base catalysts. The QXXRW motif is present in all the processive β-glycosyltransferases analyzed so far, and this sequence motif is believed to be involved in the processivity function of the enzyme, possibly required for holding the growing glycan chain in the active site. The 'D, D, D35QXXRW' motif has also been identified in a large number of transmembrane proteins involved in the synthesis of a variety of polysaccharides/oligosaccharides (Saxena and Brown, 1997). This motif is present in a large and contiguous globular region that is predicted to be localized in the cytoplasm. Apart from the identification of this motif in proteins whose functions are known (for example, cellulose synthase, chitin synthase, hyaluronan synthase, NodC protein) this motif has been identified in a number of other sequences in the sequence databases. The precise function of a number of these proteins remains to be determined, as in the case of the *E. coli* YhjO polypeptide (Sofia et al., 1994) whose sequence is very similar to the sequence of the *A. xylinum* cellulose synthase.

A substantial number of sequences identified in higher plants also contain the 'D, D, D35QXXRW' motif, and these proteins are characterized either as cellulose synthase or cellulose synthase-like proteins (http://cellwall.stanford.edu/cellwall/). In at least three reports, proteins with the 'D, D, D35QXXRW' motif have been identified as cellulose synthases (CelA1, Rsw1, and Irx3) based on experimental evidence (Pear et al., 1996; Arioli et al., 1998; Taylor et al., 1999). The globular region that contains the 'D, D, D35QXXRW' motif is much larger in these higher plant cellulose synthases (the globular region of CelA1 and Rsw1 is predicted to be 534 and 538 amino acids in length, respectively) in comparison to the globular region of the *A. xylinum* AcsAB cellulose synthase (285 amino acids). Analysis of the regions present between the three conserved aspartic acid residues and the QXXRW motif shows that while the number of amino acids between these residues is similar in the higher plant cellulose synthases, they differ from that of the *A. xylinum* cellulose synthase which is much more similar to the cellulose synthase-like proteins from higher plants and the *E. coli* YhjO polypeptide.

Table 1
Cellulose synthase activity in *Acetbacter xylinum* membrane fractions

| Strain (plasmid) | Amino acid substitution | Cellulose synthase specific activity[a] |
|---|---|---|
| AY41-34-895 (pIS311-9) | none | 1159.72 |
| AY41-34-895 (pIS311-D188P) | D188P | 1.16 |
| AY41-34-895 (pIS311-D188N) | D188N | 65.29 |
| AY41-34-895 (pIS311-D333R) | D333R | 3.46 |
| AY41-34-895 (pIS311-Q369M) | Q369M | 1.33 |
| AY41-34-895 (pIS311-R370Q) | R370Q | 203.15 |
| AY41-34-895 (pIS311-R370P) | R370P | 3.55 |
| AY41-34-895 (pIS311-R372A) | R372A | 5.19 |

[a] Specific activity is defined as picomoles of glucose incorporated into alkali-insoluble product per minute per milligram protein.

### 1.2. Functional analysis of conserved residues in cellulose synthase and other glycosyltransferases

Although different numbers of amino acids may separate them, the conserved residues in cellulose synthases are predicted to be present at structurally similar positions in the active site. To determine the role of the conserved aspartic acid residues (D189, D236, and D333) and residues in the QXXRW motif (Q369, R370, V371, R372, and W373) of the *A. xylinum* cellulose synthase (AcsAB), we substituted the D189, D236, D333, Q369, R370, and R372 residues with other amino acids using site-directed mutagenesis. Mutants of cellulose synthase with the following non-conservative substitutions- D189Y, D236Y, D333R, Q369M, R370Q, R370P, and R372A were obtained and analyzed for in vitro cellulose synthase activity using *A. xylinum* membrane preparations. The D189Y and D236Y mutations have been described earlier (Saxena and Brown, 1997) and the effect of the remaining mutations on cellulose synthase activity is shown in Table 1. Apart from substitutions in these conserved positions and R370, we also obtained mutations in which the aspartic acid residue at position 188 was replaced with a more conservative (D188N) and a non-conservative (D188P) amino acid. As shown in Fig. 1, cellulose synthase activity in mutants with substitutions at the conserved positions D189, D236, D333, Q369, and R372 was barely detectable, suggesting that these residues are involved in the enzyme activity. Moreover, amino acid substitutions at the non-conserved positions (D188 and R370) adjacent to the conserved residues also led to a reduction in enzyme activity. The level of expression of the mutated cellulose synthase protein in the *A. xylinum* membrane preparations was examined by western blot analysis using antibody raised against the 93 kDa cellulose synthase polypeptide (Chen and Brown, 1996). The 93 kDa cellulose synthase polypeptide was not detected by western analysis in the *A. xylinum* mutant strains AY41-34 and AY41-34-895 (Fig. 2, lanes 1 and 2) (Saxena and Brown, 1995). AY41-34-895 was therefore used as the host strain for expression of the wild-type and mutant cellulose synthase genes cloned in the plasmid pRK311 (Ditta et al., 1985). The amount of cellulose synthase in strains expressing the wild-type protein, and the D188P, D188N, D333R, Q369M, R370Q, R370P, and R372A substitutions was found to be at comparable levels as determined by western analysis (Fig. 2). Cellulose synthase was also detected at similar levels in a strain where

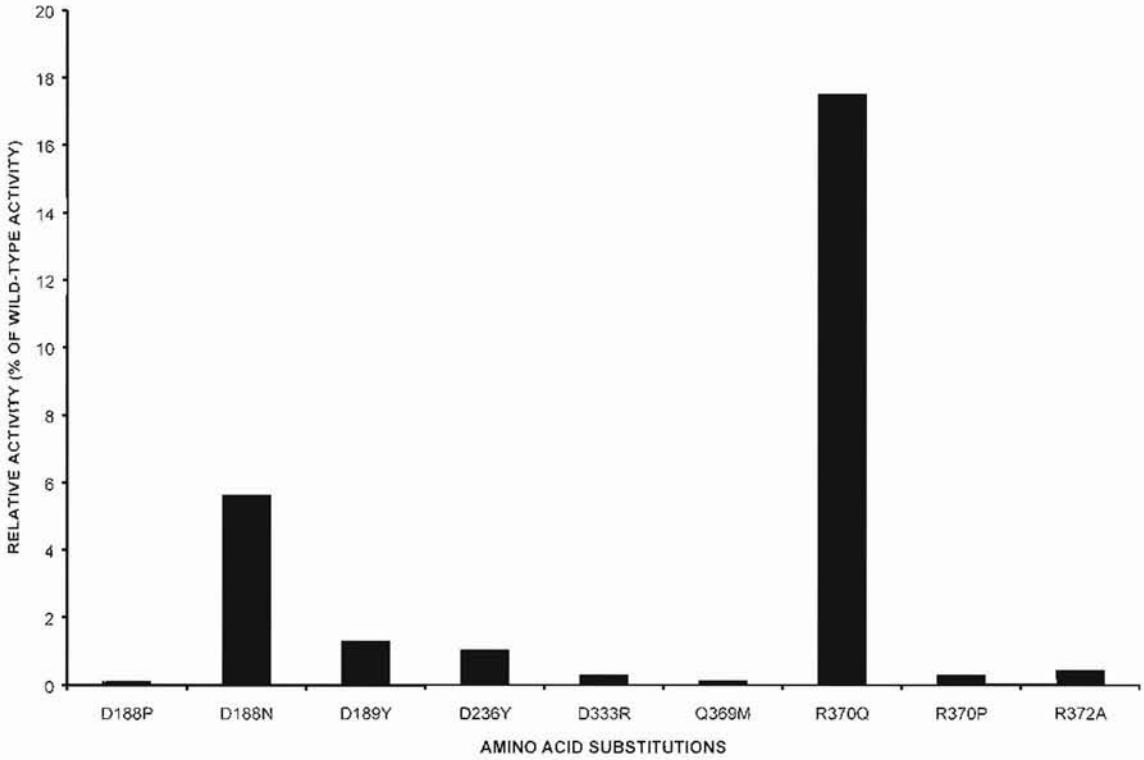


Fig. 1. Effect of amino acid substitutions on the *A. xylinum* cellulose synthase (AcsAB) activity determined in vitro.

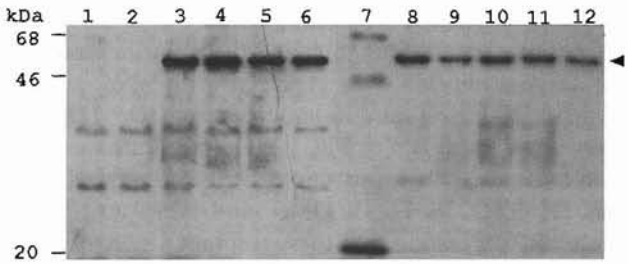


Fig. 2. Western blot analysis of membrane preparations of *A. xylinum* using anti-93 kDa antibody to detect the cellulose synthase. Lanes 1 and 2 contain membrane preparations from *A. xylinum* strains AY41-34 and AY41-34-895 respectively. Lanes 3–6, and 8–12 contain membrane preparations from *A. xylinum* AY 41-34-895 with plasmids expressing wild type cellulose synthase (lane 3), a chimeric cellulose synthase in which the globular region of the *A. xylinum* AcsAB cellulose synthase is replaced by a globular region encoded by *Arabidopsis thaliana* EST T88271 (lane 4), and cellulose synthase with amino acid substitutions D188P (lane 5), D188N (lane 6), D333R (lane 8), Q369M (lane 9), R370Q (lane 10). R370P (lane 11), and R372A (lane 12). Lane 7 contains the molecular weight markers. The cellulose synthase polypeptide fragment detected using the anti-93 kDa antibody is shown by an arrowhead.

the *A. xylinum* AcsAB globular region was substituted with the globular region of an *Arabidopsis thaliana* β-glycosyltransferase (Fig. 2, lane 4).

Whether the conserved residues are involved as catalytic residues, in substrate-binding, in maintaining the architecture of the active site or in some other function has not been determined for the *A. xylinum* cellulose synthase. Support for the idea that the conserved residues function in enzyme activity is also obtained from analysis of a few other glycosyltransferases (Fig. 3). Substrate-binding experiments done using a plasmid construct that expresses the globular region of the cotton CelA1 cellulose synthase in *E. coli* showed that the expressed protein bound to UDP–glucose (Pear et al., 1996). However, when the *N*-terminal part that includes the first conserved aspartic acid residue (corresponding to D189 of the *A. xylinum* AcsAB protein) was missing from this protein, no UDP–glucose binding was observed. These observations suggest that this region, which includes the conserved aspartic acid residue, is essential for binding the substrate. Site-directed mutagenesis experiments using the chitin synthase 2 of yeast showed that the conserved aspartic acid residues and the conserved residues in the QXXRW motif are required for chitin synthase activity (Nagahashi et al., 1995). A conserved aspartic acid residue identified in the β-glycosyltransferases also has been identified in a number of other glycosyltransferases, including the yeast α-1,3-mannosyltransferase (Mnn1p) (Wiggins and Munro, 1998) and large clostridial cytotoxins (Busch et al., 1998). This conserved aspartic acid residue is present in the DXD motif in a variety of glycosyltransferases, and the first aspartic acid in this motif corresponds to D236 of the *A. xylinum* cellulose synthase. In almost all cases the DXD

| PROTEIN | REGIONS | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| | | MOTIFS | | |
| | | DXD | | QXXRW |
| | * | * | * | * ** |
| AcsAB | 188 DDG | 231 YILIFDCDHV | 330 VTEDAH | 366 HIGQRVRWAR |
| CelA1 | DDG | FILNLDCDHY | VTSDIL | RLHQVLRWAL |
| SacChs2 | SDG | VVTLVDVGTR | LAEDRI | FISQRRRWLN |
| Mnn1p | | EFVFLDIDAI | | |
| LT | | GGVYLDVDIL | | |
| KfiC | DDC | FITFQDADDL | | |

Fig. 3. Proteins in which the conserved residues identified in β-glycosyltransferases (Saxena et al., 1995) have been analyzed by site-directed mutagenesis- AcsAB, *A. xylinum* cellulose synthase (Saxena and Brown, 1997; this work); SacChs2, *S. cerevisiae* chitin synthase 2 (Nagahashi et al., 1995); Mnn1p, *S. cerevisiae* α-1,3-mannosyltransferase (Wiggins and Munro, 1998); LT, *C. sordellii* lethal toxin (Busch et al., 1998), KfiC, *E. coli* glycosyltransferase (Griffiths et al., 1998), and by assaying a deleted globular region in which one of the conserved residues (conserved aspartic acid in region 1) is missing- CelA1, *G. hirsutum* cellulose synthase (Pear et al., 1996). The regions in which these conserved residues are present and the motifs identified in these regions are indicated. Residue numbers are shown at the start of each amino acid segment for AcsAB. Asterisks indicate the conserved residues identified in the β-glycosyltransferases. Residues analyzed by site-directed mutagenesis for each of the proteins are underlined. Positions in which substitution of the conserved residue resulted in loss of enzyme activity or substrate binding are indicated in bold.

motif is flanked by four hydrophobic residues on the *N*-terminal side, with the third of these often being an aromatic residue (Wiggins and Munro, 1998). Site-directed mutagenesis of the conserved aspartic acid residues and the flanking residues in Mnn1p showed that the aspartic acid residues are essential for α-1,3-mannosyltransferase activity. Furthermore the aspartic acid residues in the DXD motif were shown to have no effect on protein multimerization, and were suggested to be involved in catalysis. Substitution of the aspartic acid residues in the DXD motif of the *Clostridium sordellii* lethal toxin (LT) led to a loss of the binding of the substrate UDP–glucose, suggesting that these residues are important for UDP–glucose binding (Busch et al., 1998). One possible function of the DXD motif in the catalytic site could be in binding a metal ion that is used to coordinate the phosphates of the NDP–sugar in the active site. A DXD motif is also conserved in DNA polymerases where the aspartates are in the active site coordinating manganese and magnesium ions that hold the phosphates of the nucleotide triphosphate substrate (Pritchard and McHenry, 1999).

Although three aspartic acid residues have been identified at conserved positions in the processive β-glycosyltransferases, at present there is insufficient evidence to suggest which of these aspartic acid residues actually functions as a base during the catalytic reaction. Our prediction that at least two glycosidic linkages are formed simultaneously or sequentially during the synthesis of cellulose and a few other polysaccharides would require at least two aspartate residues to function as bases. There is only a single conserved aspartic acid residue in regions 3 and 4 (Fig. 3) present in domain B of the processive β-glycosyltransferases and this aspartic acid residue probably functions as a base. On the other hand, regions 1 and 2 (Fig. 3) present in domain A of both the processive and the non-processive β-glycosyltransferases and a number of other proteins, contains two conserved aspartic acid residues. The catalytic mechanism for glycosyl transfer with inversion of configuration requires a single base and so a single aspartic acid residue is sufficient for this function. However, where glycosyl transfer reaction takes place with retention of configuration, at least two amino acid residues are required, with one residue functioning as a nucleophile and the other as an acid/base. Interestingly, two amino acid residues are conserved in regions 1 and 2 of β-glycosyltransferases, and only one conserved residue is observed in the other proteins, including the α-1,3-mannosyltransferases and α-1,3-fucosyltransferases. Moreover this conserved residue present in region 2 is part of the DXD motif that has been identified in both α- and β- glycosyltransferases. The role of the aspartic acids in the DXD motif is suggested to be in the binding of the NDP–sugar substrate. Therefore, the aspartic acid residue present in region 1 of the β-glycosyltransferases (Fig. 3) may function as a catalytic residue in this group of enzymes. However, in the structure determined for the SpsA protein (Charnock

and Davies, 1999), the conserved aspartic acid residues that align with regions 1 and 2 of the β-glycosyltransferases are shown to contact the UDP portion of the UDP-sugar donor (the nature of the sugar is not known at present). Furthermore, it is speculated that the aspartate that functions as a catalytic base resides about 100 residues C-terminal to the conserved aspartate in region 2. This aspartate could be in a position similar to the aspartate present in region 3 of the β-glycosyltransferases. However, there is no QXXRW motif in the SpsA sequence and not much sequence similarity in the C-terminal portion of SpsA and regions 3 and 4 of the processive β-glycosyltransferases.

### *1.3. Architecture of the active site of cellulose synthase: a structural model for the active-site region of processive β-glycosyltransferases*

Cellulose synthase is a transmembrane protein that has not been sufficiently purified from any source for structural analysis. Since the conserved residues predicted to be involved in catalysis and processivity are present in a globular region, we have attempted to predict the folding of this region from the *A. xylinum* cellulose synthase and determine if this structure is conserved in the other β-glycosyltransferases. To this end, a 285 amino acid segment (residues 128 to 412 of AcsAB) which contained all the conserved amino acid residues in a globular region of cellulose synthase was subjected to modeling. A BLAST search of the PDB protein sequences database using the 285 amino acid segment of *A. xylinum* cellulose synthase as a query did not identify any significant sequence matches in the database, nor did, incidentally, the related sequences YhjO, CelA1, and Rsw1. Similarly, no structural relatives were found applying the PHD server (Rost and Sander, 1996), iterative searches [PSI-BLAST (Altschul et al., 1997)] or sequence-based three dimensional structural homology searches [sophisticated search techniques after Huynen et al. (1998), as well as the direct SCOP blast server (Hubbard et al., 1999)]. Furthermore, there were no similarities to known domains using the latest version of SMART (Schultz et al., 1998) or ProDom (Corpet et al., 1998).

In the absence of any structural information for modeling, we decided to obtain an ab initio model using the genetic algorithm (Dandekar and Argos, 1994). The algorithm evolves protein structure predictions in parallel in a computer evolution, recombining and selecting "fittest" structures according to basic protein building principles. These include heavy selection against structural clashes and globular packing and more complex ones such as cooperative growth of secondary structure and sheet formation. Detailed criteria for these and their respective selection weights have been described previously and validated on a test battery of 20 proteins with known structure and a wide range of topologies, yielding correct topology predictions and an average RMSD error of 4-6 Å (Dandekar and Argos, 1996). As an additional structure selection criterion derived from experimental data (see Dandekar and Argos, 1997 for such selection parameters), catalytic aspartic acid residues were rewarded when being close together. The combination of experimental data with the genetic algorithm model has previously been shown to be an effective modeling tool including blind tests for topology description and prediction (Dandekar and Leippe, 1997; Saxena et al., 1997; Tribble et al., 2000). The algorithm folded the region of the cellulose synthase starting from sequence and secondary structure (using several available standard packages for secondary structure prediction) with the assumption that it has a globular solution structure and is an independent folding unit. The fittest and also the topologically most satisfying structure obtained from all the simulations is shown in Fig. 4a. The less fit structures had similar topologies but were not packed as well. The dimensions of the predicted structure of the globular region of the *A. xylinum* cellulose synthase are roughly 60 × 60 × 70 Å (less than half of this box is filled by the protein) and the topology of the folded structure shows 10 strands and 8 alpha-helices (Fig. 4b). The globular region has a mixed topology of beta-strands and alpha-helices (Table 2) and does not exhibit a true alpha-beta barrel as observed in the structure of some other proteins. As seen in Fig. 4a and b, the helices and strands in the globular region of cellulose synthase as predicted by the genetic algorithm do not alternate regularly and are interrupted by long loop regions. Note that the exact arrangement of the strands is only roughly indicated by the algorithm and has not been refined further (Fig. 4a and b). However the topological sketch is sufficient to indicate and compare their overall connectivity and arrangement (Fig. 6). In a similar way, the nucleotide-binding fold is delineated, but the arrangement of strands is again only approximately given.

The predicted structure of the globular region of the *A. xylinum* cellulose synthase shows the presence of a central cavity between the conserved aspartic acid residues. The distances between the conserved aspartic acid residues are estimated to be 20 Å between D189 and D236, 7 Å between D189 and D333, and 18 Å between D236 and D333. From these estimates, the central, elongated cavity can accommodate two UDP–glucose residues (each UDP–glucose residue having a dimension of about 8 Å) as shown in Fig. 4c and predicted in the model for cellulose biosynthesis (Saxena et al., 1995). We note that a suitable cavity size to accommodate the UDP–glucose was by no means predefined by the selection criteria. For instance, having the three catalytic residues maximally close together would have satisfied best the additional structure selection criterion. However, optimizing among all selection criteria by the genetic

algorithm achieved, nevertheless, the correct, suitable cavity size, indicating one internal control for the prediction. Moreover, the residues in the QXXRW motif are present in a region close to the central cavity that is believed to be the catalytic pocket. Our model suggests that helix H6 with the conserved aspartate D333 is

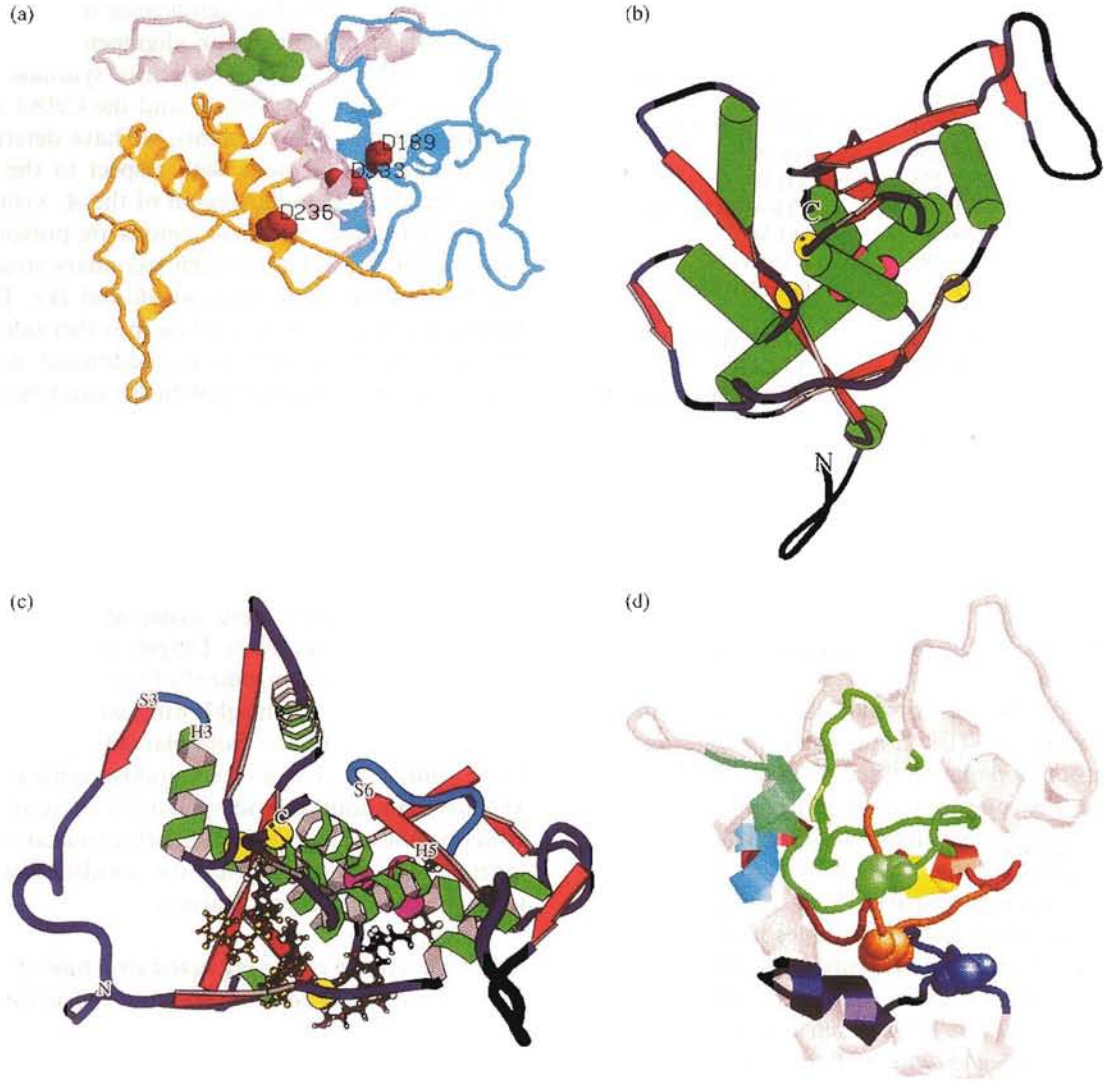


Fig. 4. (a) A ribbon model of the globular region (amino acid residues 128–412) of the *A. xylinum* cellulose synthase (AcsAB) as determined by genetic algorithm. Helical regions are indicated as such. The chain trace of the 285 amino acid globular region is shown starting from cyan (AcsAB residues 128–227) to gold (AcsAB residues 228-327) and grey (AcsAB residues 328–412). The catalytic aspartic acid residues are labeled and shown in red and spacefill representation. The QRVRW conserved motif is shown in green and spacefill. The model was drawn using the program Rasmol by Roger Sayle; (b) a schematic representation of the topology of the globular region of *A. xylinum* cellulose synthase. The alpha-helices are shown as cylinders and the beta-strands as arrows. Small balls indicate the aspartic acid residues (yellow) and the QXXRW motif (purple); (c) ribbon diagram of the globular region of the *A. xylinum* cellulose synthase showing the presence of two UDP-glucose molecules in the central cavity. The conserved aspartic acid residues are shown as yellow balls and the QXXRW region is shown in purple. The loop regions between H3 and S3 and S6 and H5 (shown in cyan) are the regions where insertions are observed in the cellulose synthases from plants. The glucose residues are arranged in a way that they could contact the catalytic aspartic acid residues and would form a cellobiose unit after removal of the UDP moieties during chain initiation. To prevent cluttering of the view in the catalytic cleft, the UDP moieties are a bit more spread out; the position of the right UDP moiety indicates also that the helix with the QXXRW motif can easily be reached (shown as an arrow in the general model in Fig. 7) after further growth of the carbohydrate chain. A glucan chain of 6 glucose residues is shown in the extended catalytic site. The reducing end of the glucan chain is on the top (probably close to the QXXRW motif) and the non-reducing end (the growing end of the glucan chain) is shown close to the second conserved aspartic acid residue (D236CD in the DXD motif); (d) a ribbon model of the globular region (534 residues; amino acids 225-758) of the *G. hirsutum* cellulose synthase (CelA1) as determined by genetic algorithm. This is a coarse model in which every third residue was modeled (a total of 178 residues were modeled) to confirm the general model for processive β-glycosyltransferases. The model depicts the large insertions in the cellulose synthases from plants (shown in white) and the spatial arrangement of the catalytic aspartates (shown thick in spacefill representation) and residues in the QXXRW motif (shown in yellow). The chain trace is color-coded- N-terminus of the chain is blue, C-terminus is red.

Table 2
Secondary structures predicted by genetic algorithm in the globular region of *Acetobacter xylinum* cellulose synthase

| Pattern of helices and strands | Amino acid residues[a] | Amino acid sequence |
|---|---|---|
| H1 | 144–146 | NPD |
| S1 | 150–157 | TVDIFVPT |
| H2 | 159–173 | NEELSIVRLTVLGSL |
| S2 | 182–187 | VRVHIL |
| H3 | 192–201 | RPEFAAFAAE |
| S3 | 205–208 | NYIA |
| S4 | 231–235 | YILIF |
| H4 | 241–251 | PTRAFLQLTMG |
| S5 | 259–262 | IALM |
| S6 | 289–295 | NLFYGVV |
| H5 | 311–323 | CAILRRTAIEQIG |
| S7 | 327–329 | TQT |
| H6 | 333–344 | DAHTALKMQRLG |
| S8 | 348–353 | AYLRIP |
| H7 | 356–381 | GGLATERLILHIGQR VRWARGMLQIF |
| S9 | 391–393 | GLS |
| S10 | 396–401 | QRLCYL |
| H8 | 402–411 | SAMTSFLFAV |

[a] The amino acid residues correspond to the complete AcsAB sequence.

central and close to the cleft. Moreover, this helix or helix H7 that contains the QXXRW motif are involved dynamically (perhaps with a structural change) in the enzymatic reaction. The tryptophan residue in the QXXRW motif is perhaps required in glucan binding. Furthermore, the long loop flanked by beta-strands S3 and S4 may help to accommodate the growing glucan chain. Residues making contact with the substrate include the conserved aspartic acid residues. A key region making contact includes residues I234, F235, and D236 while Y231 and I232 are flanking residues. Besides D333, H335 is also in close contact to the substrate. Similarly, besides D189, R191 could also be important. Further away residues I319 to V330 make up the flanking region. The structure determined by the genetic algorithm shows a globular model in accordance to the information from sequence and experiment. The loop regions apparent from the model agree as an independent test well with the accessibility predictions available for the sequence from PHD, an independent prediction algorithm using a neuronal network (Rost and Sander, 1996). Moreover, experiments done with the globular region of the cellulose synthase (CelA1) from plants suggest that the N-terminus of this region is required for UDP–glucose binding (Pear et al., 1996) and the model positions it not far away from the catalytic cleft.

As mentioned earlier, the globular region of cellulose synthases from plants is much larger than the *A. xylinum* cellulose synthase with large insertions between the conserved aspartic acid residues. In CelA1 these regions have been identified as P-CR (plant-conserved region) between the first and the second conserved aspartic acid residues and HVR (hypervariable region) between the second and the third conserved aspartic acid residues (Pear et al., 1996). Similar regions have now been identified in other plant cellulose synthases as well (Arioli et al., 1998; Taylor et al., 1999). The significance of these regions is not known at present. Upon alignment of the globular regions of the *A. xylinum* cellulose synthase (AcsAB), the *E. coli* YhjO polypeptide, and the CelA1 and Rsw1 cellulose synthases from plants, we have determined the position of these regions with respect to the predicted structure of the globular region of the *A. xylinum* cellulose synthase (Fig. 5). These regions are present in loops with the P-CR present between secondary structures H3 and S3 and the HVR between S6 and H5. These loop regions are on the surface and on the other side such that they can easily accommodate additional amino acid residues without interfering with the catalytic activity.

To further examine and compare the structure predicted for the globular region of the *A. xylinum* cellulose synthase, a coarse model of the globular region of cotton (*Gossypium hirsutum*) cellulose synthase CelA1 (Pear et al., 1996) is also predicted. To derive this model, every third residue was modeled by the genetic algorithm (residues in between were assumed to have identical secondary structural state). Despite long sequence insertions, the spatial arrangement of the conserved aspartic acid residues and the neighboring secondary structures in the cellulose synthase from plants (Fig. 4d) was found to be similar to that in the cellulose synthase from *A. xylinum*. Without providing details in resolution, the structure shown in Fig. 4d is another indication from an independent simulation that our model (Fig. 4a) captures the correct general features.

### *1.4. Relationship of the predicted structure of cellulose synthase with structures of other β-glycosyltransferases*

Although the structure and mechanisms for a number of glycosyl hydrolases was known for some time (Davies and Henrissat, 1995), the structure for only a few β-glycosyltransferases have been determined so far. The structure of the T4 β-glucosyltransferase was first determined by Vrielink et al. (1994) and more recently the substrate binding and catalytic mechanism for this enzyme has been proposed (Moréra et al., 1999). This enzyme transfers a glucose residue from UDP–glucose to hydroxymethylcytosine in modified DNA. The sequence of the T4 enzyme does not show similarity to any of the known glycosyltransferases and so far has not been classified into any glycosyltransferase family (Campbell et al., 1997). However, the structure of the T4 enzyme was found to exhibit similarity to the structure of glycogen phosphorylase. Though no sequence similarities were observed between these two proteins, the similarity in the spatial arrangement and in the

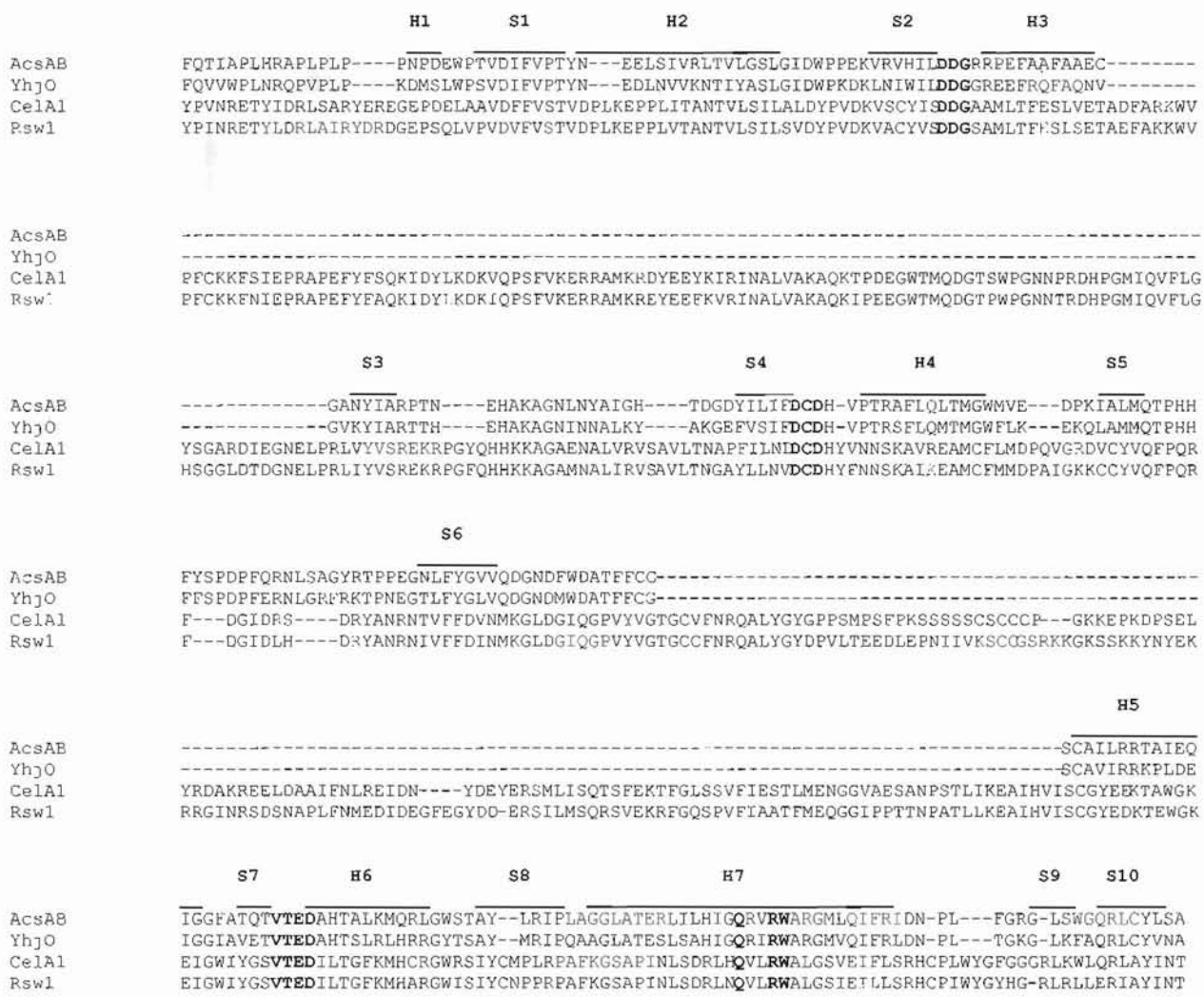


Fig. 5. Multiple alignment of the globular regions of the *A. xylinum* cellulose synthase (AcsAB), *E. coli* YhjO protein, *G. hirsutum* CelA1, and *A. thaliana* Rsw1. The secondary structures identified in the AcsAB globular region by genetic algorithm are shown above the AcsAB sequence.

chemical nature of the substrates suggested an evolutionary relationship between these two proteins based on a structurally conserved catalytic core (Holm and Sander, 1995). The predicted fold of the α-1,3-fucosyltransferases has also been suggested to show similarity to the structure of the T4 β-glucosyltransferase (Breton et al., 1998). Recently, the structures of two more β-glycosyltransferases have been determined experimentally. The structure of SpsA, a glycosyltransferase from *Bacillus subtilis* has been solved (Charnock and Davies, 1999). The substrate and acceptor specificity for this protein is not known, but it is presumably involved in the synthesis of the spore coat. This protein shares sequence homology with the glycosyltransferases in family 2 which includes enzymes that catalyze sugar transfer by an inversion mechanism. This family includes both non-processive as well as processive enzymes. From sequence analysis, SpsA is a non-processive enzyme that transfers a single sugar residue from a nucleotide–diphospho–sugar donor to a specific acceptor. The structure of SpsA shows that it is a two-domain protein with a nucleotide-binding domain and an acceptor binding domain. On the other hand, the structure of β4Gal-T1 reveals a single large domain and the overall fold was found to be different from any fold previously determined (Gastinel et al., 1999).

The modeled topology of the cellulose synthase as a prototype of a processive β-glycosyltransferase fits well in the range of structures experimentally described (Charnock and Davies, 1999; Gastinel et al., 1999) or predicted for other glycosyltransferases (Breton and Imberty, 1999). None of these structures were used for our prediction. They also cannot be used, as there is not sufficient homology for a homology prediction. At the same time, the genetic algorithm prediction is an *ab initio* prediction and does not provide a homology model. The

connectivity of the folds determined for SpsA and β4Gal-T1, and that predicted for cellulose synthase is shown in Fig. 6. Specifically comparing the structure of SpsA (Charnock and Davies, 1999) with the cellulose synthase structure as modeled by the genetic algorithm, we note that the first 100 *N*-terminal residues in SpsA provide the nucleotide binding fold and consist of four parallel β-strands flanked on either side by two helices. In our predicted structure there is a similar arrangement of helix alternating with strand for the first three helices and four strands (no helix was found between strands 3 and 4). Strands 1, 2 and 4 participate in the nucleotide-binding region and strand 8 (later in the chain trace) joins this. Furthermore, in our model strand 3 and helix 3 loop out (which is not the case in the SpsA structure but has been noted for strand 1 and helix 1 of the β4Gal-T1 structure). A more refined model (including refinement of strand positions to each other) may be expected to lead to a better topology of the nucleotide-binding site. On the other hand, re-screening the complete cellulose synthase region modeled here for hidden regular nucleotide-binding domains using IMPALA (Schaffer et al., 1999) showed that none of these are present, indicating by an independent method that cellulose synthase seems to be different in this respect and therefore the need for an *ab initio* model. The C-terminal part of the SpsA structure is a mixed β-sheet flanked by three helices on one side and one helix on the other. In our model of the cellulose synthase, the C-terminal part shows a similar but less regular succession of helices and strands as compared to the N-terminal part. Strands 9, 10 and 6 form a sheet region with helices interspersed and the flanking strand 8 adds to the nucleotide-binding region. The cellulose synthase fold is more integrated than the SpsA structure, which shows a clear split between the N-terminal nucleotide-binding part and the C-terminal part. In fact the compactness of the cellulose synthase is closer in relationship to the β4Gal-T1 structure.

The β4Gal-T1 structure (Gastinel et al., 1999) reveals eleven strands and eight helices (two of them being short $3_{10}$ helices) in comparison to the cellulose synthase structure that shows ten strands and eight helices. The nucleotide-binding region of β4Gal-T1 is part of the central eight-stranded mixed and twisted β-sheet surrounded by two α-helices on one side and by four α-helices on the other. This bigger sheet is neither predicted for the cellulose synthase nor found in the SpsA structure. However, the sequence of helices and strands in the nucleotide-binding region of β4Gal-T1 looks similar to our predicted cellulose synthase fold and is not as regular as it is in SpsA. The C-terminal part of β4Gal-T1 contains a small sheet region with a similar arrangement as found in our model of cellulose synthase. The β4Gal-T1 fold is again different from the structure of the T4 phage β-glucosyltransferase (Vrielink et al., 1994).

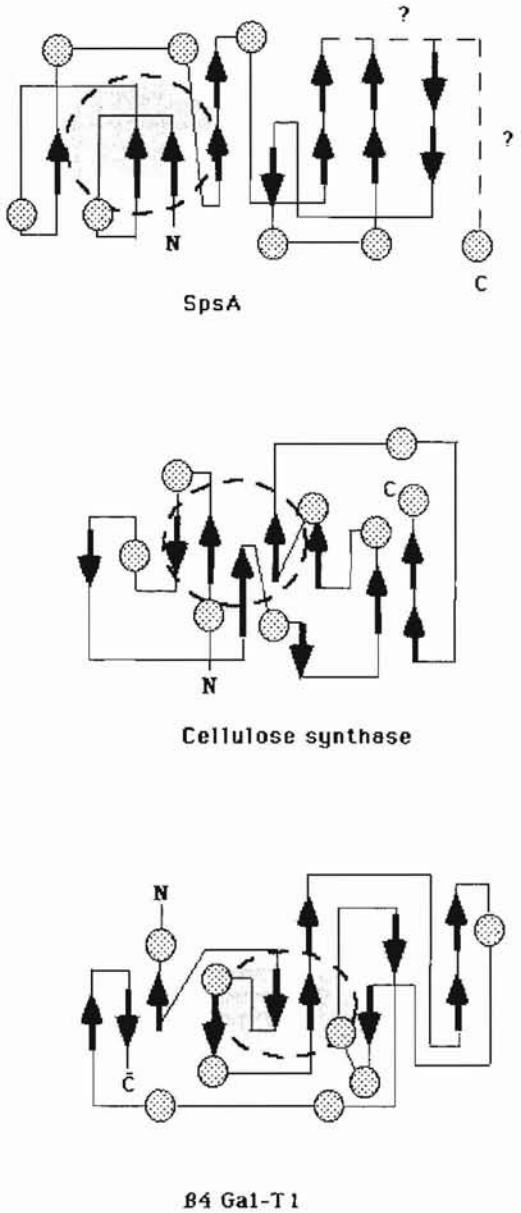


Fig. 6. Topological comparison of the *B. subtilis* SpsA protein, bovine β4Gal-T1 catalytic domain, and the *A. xylinum* cellulose synthase globular region. The circle represents the approximate nucleotide-binding site in each protein.

### *1.5. Possible mechanisms for synthesis of cellulose and other polysaccharides*

Since every glucose unit in the cellulose chain is oriented 180° with respect to its neighbor, we had proposed a model in which two UDP–glucose molecules are positioned in the active site of cellulose synthase at an 180° orientation. Upon catalysis, the glucose units are linked simultaneously or sequentially to extend the cellulose chain two units at a time (Saxena et al., 1995). This model proposed that the active site of the enzyme has an extended region that carries not only the catalytic

residues, but also a region that holds the growing chain (glucan-binding region). Since the catalytic region is predicted to have two UDP–glucose binding sites and as no primer requirement is observed, initiation of cellulose synthesis probably takes place by the formation of a cellobiose unit in the catalytic pocket. In such a case, the UDP-portion of the first UDP-glucose is cleaved, leaving an exposed reducing end in the first glucose (terminal residue) and this glucose is linked via its non-reducing end to the second glucose by a β-1,4–linkage. The cellobiose unit thus formed moves, as it has no bonding in the catalytic pocket, into the glucan-binding region, but does not leave the enzyme because it is in a twisted conformation. In this conformation the cellobiose unit is associated with the glucan-binding region and does not move or slide. In the next step of chain elongation, two new UDP–glucose molecules enter the catalytic pocket and two new β-1,4-linkages are formed either sequentially or simultaneously. The formation of a covalent bond at the non-reducing end of the cellobiose unit present in the glucan-binding region makes this cellobiose unit to adopt an extended conformation and allows it to slide further in the glucan-binding region. This model of the catalytic and glucan-binding region of cellulose synthase is similar to the one proposed, based on direct structural data, for the catalytic domain of CBHI (Divne et al., 1998). Both cellulose synthase and CBHI are processive enzymes that have to be tightly associated with the cellulose chain. In CBHI, the cellulose chain is passaged through a tunnel-like structure in which the protein interacts with at least 10 glucose residues. For the glycosyl hydrolases, the tunnel-like structure probably imposes a steric confinement for the stabilization of a sugar conformation that is able to undergo catalysis, and in this the ring distortion is thought to play an important role (Divne et al., 1998). A similar structure can also be visualized for the cellulose synthase and possibly other processive glycosyltransferases where a glycan chain is passed through a tunnel-like structure where it is held in a conformation that allows both its tight association as well as allows it to slide. The amino acid residues involved in the sugar-protein interaction are usually aromatic residues, such as tryptophan with its indole ring and tyrosine, and we expect them to be conserved in the glucan-binding region of the processive glycosyltransferases.

Is there a need for two catalytic centers in cellulose synthase and possibly other processive β-glycosyltransferases? The requirement for two catalytic centers in these enzymes was proposed to account for the two-fold symmetry of the glucose residues in a cellulose chain where every residue is rotated 180° with respect to its neighbor (Saxena et al., 1995). The absence of duplicated active sites in the cellulose synthase catalytic subunit has led others to suggest that there can be as much as 120 rotation about the β-1,4-linkage. Accordingly all glucose residues can be added in a particular orientation at a single site and the alternating residues can relax into opposing orientations after they have exited the catalytic site (Delmer, 1999). This leads into the nature of the glucan chain in the active site of the enzyme. Do the glucose residues in the glucan chain at this point of synthesis have intramolecular H-bonds to stabilize the chain or do these H-bonds form at a later point? The intramolecular H-bonds between O3′ and O5 exist in crystalline cellulose, and it will be interesting to determine if they are also present in the glucan chain in the active site. In the structure of CBHI from *Trichoderma reesei* bound to cellulose, the cellulose chain has torsion angles at the linkages that are similar to other cellobiosyl linkages (Divne et al., 1998).

The two catalytic centers were also thought to be required for synthesis of polysaccharide chains that have alternating sugar residues that are different and that are linked by different linkages as in the case of hyaluronan (DeAngelis, 2000) and type 3 capsular

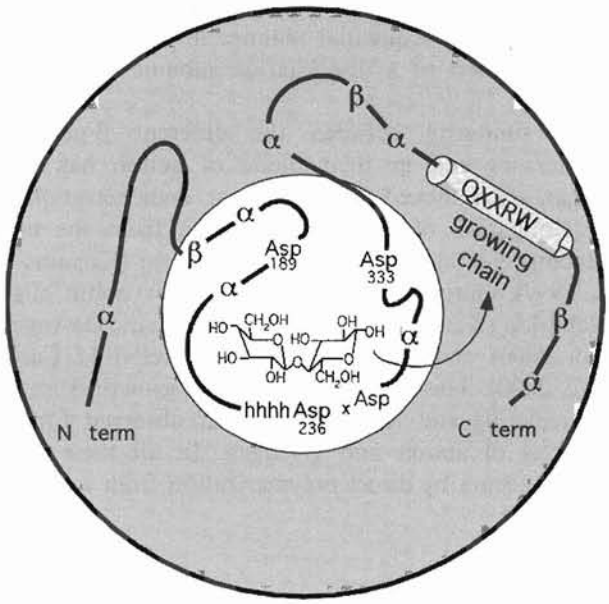


Fig. 7. A cartoon presenting a general model for the globular region of processive glycosyltransferases. The black line represents the main chain trace of the globular region of the *A. xylinum* cellulose synthase. The globular topology of this region is indicated by the grey circle and the globular arrangement of the catalytic residues is shown in the center. The three conserved catalytic residues are indicated by Asp (aspartate, sometimes also glutamate residues; the numbers are only given to identify the corresponding residues in the *A. xylinum* cellulose synthase). α and β indicate stable secondary structure elements that provide a framework for the catalytic residues. The QXXRW motif (alpha-helix with these residues; see text) holds the growing glucan chain (chain itself is only written as text in the cartoon). The processive assembly of the glucan chain takes place by the sequential or simultaneous addition of two glucose residues to the growing end. Functional motifs flank the key catalytic residues (for instance the DXD motif with accompanying *N*-terminal hydrophobic and aromatic residues for Asp 236). h represents hydrophobic residues.

polysaccharide of *Streptococcus pneumoniae* (Cartee et al., 2000). These polysaccharides have a disaccharide repeat and current evidence suggests that the enzymes required for their synthesis are capable of forming all of the glycosidic linkages present in these polysaccharides. Moreover, it is envisaged that a single polypeptide catalyzes the transfer of two distinct sugars in the case of hyaluronan and type 3 capsular polysaccharide. The transfer of the two different sugars to the growing polysaccharide chain is probably catalyzed by a mechanism similar to that proposed for cellulose biosynthesis. Whether catalysis involves a single active site or two independent sites for transfer of the two sugars to the growing chain is not very well understood, although two separate glycosyltransferase activities have been implicated for the synthesis of hyaluronan and other polysaccharides. The *Pasteurella multocida* hyaluronan synthase (PmHasA) shows the presence of duplicated domains A (A1 and A2) instead of domains A and B, and mutational analysis suggests the presence of two relatively independent glycosyltransferase activities on a single polypeptide chain (Jing and DeAngelis, 2000). Moreover, the transfer of sugars by this enzyme has been shown to take place in a sequential manner instead of simultaneous addition of a disaccharide subunit (DeAngelis, 2000).

The similarity between the different β-glycosyltransferases and in their mode of action has been further strengthened by the recent demonstration of the elongation of the growing chain from the non-reducing end as in the case of cellulose (Koyama et al., 1997), chitin (Sugiyama et al., 1999), chitin oligosaccharide (Kamst et al., 1999), hyaluronan (DeAngelis et al., 2000), and type 3 capsular polysaccharide (Cartee et al., 2000). This mechanism of chain growth from the non-reducing end is similar to that observed for the synthesis of starch and glycogen. In all these cases growth occurs by direct polymerization from a nucleotide sugar.

## 2. Significance

A combination of functional and modeling studies offers an insight into the folding pattern and the arrangement of conserved amino acids in the globular region of the *A. xylinum* cellulose synthase and other processive β-glycosyltransferases. The protein structure calculated by the genetic algorithm was carefully optimized and is the overall fittest from a total of 11 million solution trials. The fitness criteria applied had been validated on 20 proteins with known crystal structure and different topologies (Dandekar and Argos, 1996). The structure calculated is supported by the predicted spatial constraints (size of the UDP–glucose binding pocket; transfer of the growing chain to the QXXRW region), the available experimental data (requirement of the first conserved aspartic acid and *N*-terminus in UDP–glucose binding experiments is supported in the model where the *N*-terminus is flanking the catalytic cleft; requirement of the conserved aspartic acid residues for enzyme activity as determined by site-directed mutagenesis and activity measurements), and sequence analysis (identification of the conserved residues and the QXXRW motif in the globular region). As an independent test, the distribution of accessible loop regions agrees well with that derived by neuronal network prediction (Rost and Sander, 1996). Nevertheless, we suspect the RMSD error in several areas of the model to be still high (higher than 5 Å). However, the structure is useful to get a first approximation view at the spatial arrangement of the catalytic residues, the glucan holding residues (QXXRW region) for the growing cellulose chain and, also importantly, of all the neighboring residues acting in concert with these central ones. These features as well as the alpha/beta mixed topology without a regular succession of helices and strands predict a simplified and general model for the structure and action of processive β-glycosyltransferases (Fig. 7). We hope that this first structure-function characterization of cellulose synthase will stimulate further research and will be an incentive for new studies.

## Acknowledgements

This work was supported by grants from the USDA (96-35301-3176) and DOE (DE-FG03-94ER20145) to I.M.S. and R.M.B. T.D. thanks DFG and BMBF for support.

## References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Arioli, T., Peng, L., Betzner, A.S., Burn, J., Wittke, W., Herth, W., Camilleri, C., Höfte, H., Plazinski, J., Birch, R., Cork, A., Glover, J., Redmond, J., Williamson, R.E., 1998. Molecular analysis of cellulose biosynthesis in *Arabidopsis*. Science 279, 717–720.

Barth, H., Preiss, J.C., Hofmann, F., Aktories, K., 1998. Characterization of the catalytic site of the ADP-ribosyltransferase *Clostridium botulinum* C2 toxin by site-directed mutagenesis. J. Biol. Chem. 273, 29506–29511.

Breton, C., Imberty, A., 1999. Structure/function studies of glycosyltransferases. Curr. Op. Structural Biol. 9, 563–571.

Breton, C., Oriol, R., Imberty, A., 1998. Conserved structural features in eukaryotic and prokaryotic fucosyltransferases. Glycobiology 8, 87–94.

Busch, C., Hofmann, F., Selzer, J., Munro, S., Jeckel, D., Aktories, K., 1998. A common motif of eukaryotic glycosyltransferases is essential for the enzyme activity of large clostridial cytotoxins. J. Biol. Chem. 273, 19566–19572.

Campbell, J.A., Davies, G.J., Bulone, V., Henrissat, B., 1997. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem. J. 326, 929–939.

Cartee, R.T., Forsee, W.T., Schutzbach, J.S., Yother, J., 2000. Mechanism of type 3 capsular polysaccharide synthesis in *Streptococcus pneumoniae*. J. Biol. Chem. 275, 3907–3914.

Charnock, S.J., Davies, G.J., 1999. Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. Biochemistry 38, 6380–6385.

Chen, H.P., Brown Jr., R.M., 1996. Immunochemical studies of the cellulose synthase complex in *Acetobacter xylinum*. Cellulose 3, 63–75.

Corpet, F., Gouzy, J., Kahn, D., 1998. The ProDom database of protein domain families. Nucleic Acids Res. 26, 323–326.

Dandekar, T., Argos, P., 1994. Folding the main chain of small proteins with the genetic algorithm. J. Mol. Biol. 236, 844–861.

Dandekar, T., Argos, P., 1996. Identifying the tertiary fold of small proteins with different topologies from sequence and secondary structure using the genetic algorithm and extended criteria specific for strand regions. J. Mol. Biol. 256, 645–660.

Dandekar, T., Argos, P., 1997. Applying experimental data to protein fold prediction with the genetic algorithm. Protein Engineering 10, 877–893.

Dandekar, T., Leippe, M., 1997. Molecular modeling of amoebapore and NK-lysin: a four-alpha-helix bundle motif of cytolytic peptides from distantly related organisms. Fold Des. 2, 47–52.

Davies, G., Henrissat, B., 1995. Structure and mechanisms of glycosyl hydrolases. Structure 3, 853–859.

DeAngelis, P.L., 2000. Molecular directionality of polysaccharide polymerization by the *Pasteurella multocida* hyaluronan synthase. J. Biol. Chem. 274, 26557–26562.

Delmer, D.P., 1999. Cellulose biosynthesis. Exciting times for a difficult field of study. Annual Review of Plant Physiology. Plant Mol. Biol. 50, 245–276.

Ditta, G., Schmidhauser, T., Yakobson, E., Lu, P., Liang, X., Finlay, D.R., Guiney, D., Helinski, D.R., 1985. Plasmids related to the broad host range vector, pRK290, useful for gene cloning and for monitoring gene expression. Plasmid 13, 149–153.

Divne, C., Ståhlberg, J., Teeri, T.T., Jones, T.A., 1998. High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei*. J. Mol. Biol. 275, 309–325.

Gastinel, L.N., Cambillau, C., Bourne, Y., 1999. Crystal structure of the bovine β4galactosyltransferase catalytic domain and its complex with uridine diphoshogalactose. EMBO J. 18, 3546–3557.

Griffiths, G., Cook, N.J., Gottfridson, E., Lind, T., Lidholt, K., Roberts, I.S., 1998. Characterization of the glycosyltransferase enzyme from the *Escherichia coli* K5 capsule gene cluster and identification and characterization of the glucuronyl active site. J. Biol. Chem. 273, 11752–11757.

Holm, L., Sander, C., 1995. Evolutionary link between glycogen phosphorylase and a DNA modifying enzyme. EMBO J. 14, 1287–1293.

Hubbard, T.J.P., Ailey, B., Brenner, S.E., Murzin, A.G., Chothia, C., 1999. SCOP: a structural classification of proteins database. Nucleic Acids Res. 27, 254–256.

Huynen, M., Doerks, T., Eisenhaber, F., Orengo, C., Sunyaev, S., Yuan, Y., Bork, P., 1998. Homology-based fold predictions for *Mycoplasma genitalium* proteins. J. Mol. Biol. 280, 323–326.

Jing, W., DeAngelis, P.L., 2000. Dissection of the two transferase activities of the *Pasteurella multocida* hyaluronan synthase: two active sites exist in one polypeptide. Glycobiology 10, 883–889.

Kamst, E., Bakkers, J., Quaedvlieg, N.E.M., Pilling, J., Kijne, J.W., Lugtenberg, B.J.J., Spaink, H.P., 1999. Chitin oligosaccharide synthesis by rhizobia and zebrafish embryos starts by glycosyl transfer to O4 of the reducing-terminal residue. Biochemistry 38, 4045–4052.

Koyama, M., Helbert, W., Imai, T., Sugiyama, J., Henrissat, B., 1997. Parallel-up structure evidences the molecular directionality during biosynthesis of bacterial cellulose. Proc. Natl. Acad. Sci. USA 94, 9091–9095.

Moréra, S., Imberty, A., Aschke-Sonnenborn, U., Rüger, W., Freemont, P.S., 1999. T4 Phage β-glucosyltransferase: Substrate binding and proposed catalytic mechanism. J. Mol. Biol. 292, 717–730.

Nagahashi, S., Sudoh, M., Ono, N., Sawada, R., Yamaguchi, E., Uchida, Y., Mio, T., Takagi, M., Arisawa, M., Okabe, H.Y., 1995. Characterization of chitin synthase 2 of *Saccharomyces cerevisiae*. J. Biol. Chem. 270, 13961–13967.

Okazaki, I.J., Moss, J., 1998. Glycosylphosphatidylinositol-anchored and secretory isoforms of mono-ADP-ribosyltransferases. J. Biol. Chem. 273, 23617–23620.

Pear, J.R., Kawagoe, Y., Schreckengost, W.E., Delmer, D.P., Stalker, D.M., 1996. Higher plants contain homologs of the bacterial *celA* genes encoding the catalytic subunit of cellulose synthase. Proc. Natl. Acad. Sci. USA 93, 12637–12642.

Pritchard, A.E., McHenry, C.S., 1999. Identification of the acidic residues in the active site of DNA polymerase III. J. Mol. Biol. 285, 1067–1080.

Rost, B., Sander, C., 1996. Bridging the protein sequence-structure gap by structure predictions. Ann. Rev. Biophys. Biomol. Str. 25, 113–136.

Saxena, I.M., Brown Jr., R.M., 1995. Identification of a second cellulose synthase gene (acsAII) in *Acetobacter xylinum*. J. Bacteriol. 177, 5276–5283.

Saxena, I.M., Brown Jr., R.M., 1997. Identification of cellulose synthase(s) in higher plants: Sequence analysis of processive β-glycosyltransferases with the common motif 'D,D,D35Q(R,Q)XRW.' Cellulose 4, 33–49.

Saxena, I.M., Brown Jr., R.M., Fevre, M., Geremia, R.A., Henrissat, B., 1995. Multidomain architecture of β-glycosyl transferases: Implications for mechanism of action. J. Bacteriol. 177, 1419–1424.

Saxena, I.M., Lin, F.C., Brown Jr., R.M., 1990. Cloning and sequencing of the cellulose synthase catalytic subunit gene of *Acetobacter xylinum*. Plant Mol. Biol. 15, 673–683.

Saxena, P., Whang, I., Voziyanov, Y., Harkey, C., Argos, P., Jayaram, M., Dandekar, T., 1997. Probing Flp: a new approach to analyze the structure of a DNA recognizing protein by combining the genetic algorithm, mutagenesis and non-canonical DNA target sites. Biochimica et Biophysica Acta 1340, 187–204.

Schaffer, A.A., Wolf, Y.I., Ponting, C.P., Koonin, E.V., Aravind, L., Altschul, S.F., 1999. IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices. Bioinformatics 15, 1000–1011.

Schultz, J., Milpetz, F., Bork, P., Ponting, C.P., 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. USA 95, 5857–5864.

Sofia, H.J., Burland, V., Daniels, D.L., Plunkett III, G., Blattner, F.R., 1994. Analysis of the *Escherichia coli* genome. V. DNA sequence of the region from 76.0 to 81.5 minutes. Nucleic Acids Res. 22, 2576–2586.

Sugiyama, J., Boisset, C., Hashimoto, M., Watanabe, T., 1999. Molecular directionality of β-chitin biosynthesis. J. Mol. Biol. 286, 247–255.

Taylor, N.G., Scheible, W.-R., Cutler, S., Somerville, C.R., Turner, S.R., 1999. The *irregular xylem3* locus of *Arabidopsis* encodes a cellulose synthase required for secondary cell wall synthesis. Plant Cell 11, 769–779.

Tribble, G., Ahn, Y.T., Lee, J., Dandekar, T., Jayaram, M., 2000. DNA recognition, strand selectivity, and cleavage mode during integrase family site-specific recombination. J. Biol. Chem. 275, 22255–22267.

Vrielink, A., Ruger, W., Driessen, H.P.C., Freemont, P.S., 1994. Crystal structure of the DNA modifying enzyme β-glucosyl-

transferase in the presence and absence of the substrate uridine diphosphoglucose. EMBO J. 13, 3413–3422.

Wiggins, C.A.R., Munro, S., 1998. Activity of the yeast MNN1 α1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. Proc. Natl. Acad. Sci. USA 95, 7945–7950.

Wong, H.C., Fear, A.L., Calhoon, R.D., Eichinger, G.H., Mayer, R., Amikam, D., Benziman, M., Gelfand, D.H., Meade, J.H., Emerick, A.W., Bruner, R., Ben-Bassat, A., Tal, R., 1990. Genetic organization and regulation of the cellulose synthase operon in *Acetobacter xylinum*. Proc. Natl. Acad. Sci. USA 87, 8130–8134.