**Glycobiology** vol. 9 no. 1 pp. 31–41, 1999

# Detection of two loci involved in (1→3)-β-glucan (curdlan) biosynthesis by *Agrobacterium* sp. ATCC31749, and comparative sequence analysis of the putative curdlan synthase gene

**Stan J.Stasinopoulos[3], Paul R.Fisher[1], Bruce A.Stone and Vilma A.Stanisich[1,2]**

Department of Biochemistry and [1]Department of Microbiology, La Trobe University, Bundoora 3083, Australia

*Received on March 9, 1998; revised on May 6, 1998; accepted on May 14, 1998*

[2]To whom correspondence should be addressed

[3]Present address: Department of Biochemistry, Mt. Sinai School of Medicine, New York, NY 10029

**Genes essential for the production of a linear, bacterial (1→3)-β-glucan, curdlan, have been cloned for the first time from *Agrobacterium* sp. ATCC31749. The genes occurred in two, nonoverlapping, genomic fragments that complemented different sets of curdlan(*crd*)-deficient transposon-insertion mutations. These were detected as colonies that failed to stain with aniline blue, a (1→3)-β-glucan specific dye. One fragment carried a biosynthetic gene cluster (locus I) containing the putative curdlan synthase gene, *crdS*, and at least two other *crd* genes. The second fragment may contain only a single *crd* gene (locus II). Determination of the DNA sequence adjacent to several locus I mutations revealed homology to known sequences only in the cases of *crdS* mutations. Complete sequencing of the 1623 bp *crdS* gene revealed highest similarities between the predicted CrdS protein (540 amino acids) and glycosyl transferases with repetitive action patterns. These include bacterial cellulose synthases (and their homologs), which form (1→4)-β-glucans. No similarity was detected with putative (1→3)-β-glucan synthases from yeasts and filamentous fungi. Whatever the determinants of the linkage specificity of these β-glucan synthases might be, these results raise the possibility that (1→3)-β-glucans and (1→4)-β-glucans are formed by related catalytic polypeptides.**

*Key words*: *Agrobacterium* sp. ATCC31749/(1→3)-β-glucan/repetitive β-glycosyl transferases/β-glucan synthases/biosynthesis/curdlan/hydrophobic cluster analysis/protein sequence analysis

## Introduction

Extracellular homo- and hetero-polysaccharides are produced by many bacterial species (Leigh and Coplin, 1992; Sutherland, 1993). The best known is the (1→4)-β-glucan, cellulose, from *Acetobacter xylinum*, also produced by *Agrobacterium tumefaciens*, *Rhizobium* spp., *Sarcina ventriculi*, and other species (Ross *et al*., 1991). Members of the Rhizobiaceae, including *Agrobacterium*, *Bradyrhizobium*, and *Rhizobium*, synthesize linear and cyclic (1→2)-β-glucans and cyclic (1→3, 1→6)-β-glucans (Breedveld and Miller, 1994).

A linear (1→3)-β-glucan, known as curdlan, is produced by an *Agrobacterium* sp. and some other bacteria (Harada and Harada, 1996). This polysaccharide is obtained in good yields in an insoluble, microfibrillar form when the *Agrobacterium* is grown on glucose-rich medium, and its synthesis is induced on depletion of the nitrogen source (Phillips and Lawford, 1983). The (1→3)-β-glucan molecules in native curdlan probably exist as an association of single helical chains (Kasai and Harada, 1980). When suspensions of native curdlan are heated at 55–70°C, a low set gel is produced. In this state, x-ray diffraction (Okuyama *et al*., 1991) and $^{13}$C NMR (Saito *et al*., 1977) studies have demonstrated that most of the molecules are single chain helices but some triple-stranded helices may also occur. On heating to higher temperatures, the strength of the gel increases and the (1→3)-β-glucan molecules are present as triple-stranded helices (Kasai and Harada, 1980). The unique gel forming properties of curdlan have led to its use as a food texture modifier (Harada *et al*., 1993; Harada and Harada, 1996) and as an enhancer of flow properties of concrete (Sakamoto *et al*., 1991).

Structurally comparable (1→3)-β-glucans are found in the walls of yeasts and certain groups of filamentous fungi (Stone and Clarke, 1992; Cid *et al*., 1995) and as storage polysaccharides in brown algae (laminarin), euglenoids (paramylon), chrysophytes (leucosin) and some fungi (cellulin, mycolaminarin, pachyman; Stone and Clarke, 1992). Higher plants also produce a linear (1→3)-β-glucan, callose, in the cell plate at the earliest stage of cell wall deposition, and in special walls in cells of reproductive tissues (e.g., pollen mother cells, pollen tubes) as well as forming deposits on the plasma membrane in abiotic and biotic stress (Stone and Clarke, 1992).

Curdlan is specifically stained by the triphenylmethane dye aniline blue (Nakanishi *et al*., 1974) and by the aniline blue fluorochrome (Evans *et al*., 1984) as are other (1→3)-β-glucans such as yeast (*Saccharomyces cereviseae*) glucan, pachyman (*Poria cocus*; Nakanishi *et al*., 1974), and callose (*Vitis vinifera*; Aspinall and Kessler, 1957). Aniline blue has thus been used both to detect curdlan production by bacteria grown on agar medium (Nakanishi *et al*., 1974) and to identify mutants unable to produce curdlan (Nakanishi *et al*., 1976).

A considerable body of biochemical and molecular genetic information is available concerning bacterial cellulose synthesis, including the intriguing indication that the mechanisms may differ between *A.xylinum* (Saxena *et al*., 1991; Saxena and Brown, 1995) and *A.tumefaciens* (Matthysse *et al*., 1995a) even though these bacteria encode highly homologous cellulose synthases. Nothing is known about the molecular genetics of curdlan biosynthesis in bacteria or callose synthesis in higher plants, but there is growing information about the genes required for (1→3)-β-glucans produced by yeasts and filamentous fungi (Douglas *et al*., 1994; Inoue *et al*., 1995; Mazur *et al*., 1995; Kelly *et al*., 1996; Mio *et al*., 1997). Thus, knowledge of the synthesis and secretion of curdlan is also of considerable interest from a

comparative viewpoint. In addition, this information might prove useful in the isolation of plant (1→3)-β-glucan synthase gene homologues as has been achieved for plant cellulose synthase genes (Pear *et al*., 1996). As a first stage in such a study of curdlan, we have cloned several essential curdlan genes from the industrially used strain, *Agrobacterium* sp. ATCC31749 (formerly *Alcaligenes faecalis* var. *myxogenes*; Harada *et al*., 1966). The genes occur in two nonoverlapping genomic clones that complement the mutations in curdlan-deficient insertion-mutants. The nucleotide sequence of one gene, *crdS*, was determined and, based on its homology with known β-glycan synthases, is proposed to encode the (1→3)-β-glucan (curdlan) synthase catalytic protein.

## Results

### *Isolation of curdlan-deficient mutants of* Agrobacterium

Curdlan-producing ($Crd^+$) strains of bacteria form colonies that stain dark blue on agar medium containing aniline blue whereas curdlan nonproducers form nonstaining colonies (Nakanishi *et al*., 1976). This phenotypic difference permitted the isolation of curdlan-deficient mutants of LTU50 (a $Cm^r$ mutant of *Agrobacterium* sp. ATCC31749) after mutagenesis with transposon Tn*phoA* (Manoil and Beckwith, 1985). This was achieved by conjugation experiments in which a suicide plasmid containing Tn*phoA*($Km^r$) (i.e., pRT733; Taylor *et al*., 1986) was mobilized from *E.coli* SM10 to LTU50. Kanamycin-resistant transconjugant colonies were isolated at frequencies of about $5 \times 10^{-3}$ per donor cell and some (about 0.15%) stained poorly or not at all with aniline blue. When curdlan production by such mutants (LTU61-LTU113) was assessed from the amount of alkali-soluble polysaccharide that they produced (Nakanishi *et al*., 1976), it was significantly less than that from LTU50 (6 g $l^{-1}$). Moreover, the amount correlated with the staining intensity of the colonies (i.e., <0.2–0.5 g $l^{-1}$ from white-colony mutants ($Crd^-$) and 1–4 g $l^{-1}$ from mutants that formed weakly staining colonies ($Crd^{+/-}$)). Southern analysis of the mutants showed that each had Tn*phoA* inserted in one of nine distinguishable *Stu*I or *Pvu*I fragments (from 1–9 kb in size). Thus, it was likely that several different genes required for curdlan synthesis had been affected.

### *Cloning of curdlan synthesis genes from* Agrobacterium

The broad host range plasmid RP1 (=RK2) (Pansegrau *et al*., 1994) was found to transfer efficiently from an HB101 derivative to *Agrobacterium* LTU50, and to be stably maintained in this strain (data not shown). The cosmid vector pLAFR1, which is derived from RK2 (Friedman *et al*., 1982), was therefore chosen for the construction of an LTU50 gene library in *E.coli*. Wild type curdlan genes in the library were detected using probes containing disrupted *crd* gene sequences from $Crd^-$ mutants carrying distinguishable insertion mutations. The probes were obtained by cloning the $Km^r$-containing *Eco*RI-junction fragments from five mutants (LTU61–LTU64 and LTU87) into pUC19 (forming pVS1500–pVS1504, respectively). Each junction fragment contained a different length of *Agrobacterium* DNA (from 0.2–4.0 kb) but the same portion of Tn*phoA* (including the *Hpa*I (0.18 kb) site; Figure 1A) allowing the recovery of the *Agrobacterium* sequences from pVS1500-pVS1504 on *Hpa*I(Tn*phoA*)–*Eco*RI(*Agrobacterium*) fragments (Figure 1B). These fragments were used as probes (P1500–P1504, respectively).

When 4200 colonies containing the gene library were screened with the probes, 25 colonies (Group I) hybridized with P1501–P1504, whereas 12 colonies (Group II) hybridized only

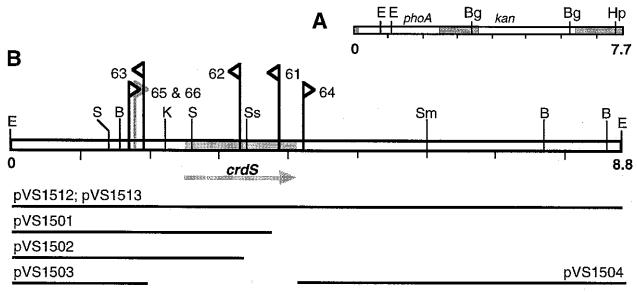


**Fig. 1.** (**A**) Simplified map of transposon Tn*phoA* (Manoil and Beckwith, 1985) showing the location of IS50L containing the *phoA* gene, the *kan* gene, IS50R, and various restriction sites. The IS50 elements are shaded. (**B**) Map of curdlan locus I of *Agrobacterium* sp. LTU50 in an 8.8 kb *Eco*RI fragment cloned in pUC19 (pVS1512) and pLAFR1 (pVS1513). The location of $Crd^-$ Tn*phoA*-insertion mutations (in LTU61–LTU66) is shown relative to the putative curdlan synthase gene, *crdS*, and various restriction sites. The orientation of insertion of each insertion is indicated by the flag which points towards the *Hpa*1 site The direction of transcription of *crdS* is from left to right. The bars below the map represent *Eco*RI-junction fragments that were cloned from LTU61–LTU66 into pUC19 (pVS1501–pVS1504). *Hpa*I(Tn*phoA*)–*Eco*RI(LTU) fragments recovered from pVS1501–pVS1504 were used as probes (P1501–P1504). Restriction site abbreviations: B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hpa*I; K, *Kpn*I; S, *Sal*I; Sm, *Sma*I; Ss, *Sst*II.

with P1500. This suggested that the recombinant cosmids in the two groups of bacteria contained nonoverlapping genomic clones and hence that at least one different gene required for curdlan synthesis must be present in each clone. The first of these conclusions was confirmed by *Eco*RI-digestion of a genomic clone from each of Group I (pVS1506) and Group II (pVS1511). In each case, both the number and the sizes of all of the fragments differed (five fragments from pVS1506 (of 1.2–9.5 kb) and more than nine from pVS1511 (from about 0.2–5.5 kb)). The second conclusion, that each clone carried different *crd* genes, was confirmed by complementation analysis of the set of $Crd^-$ mutants (LTU61–LTU88) that formed white colonies on ABA medium. Plasmids pVS1506 and pVS1511 were each transferred individually into these mutants by mobilization with the helper plasmid (an RK2-derivative) located in the chromosome of *E.coli* S17–1 (Simon *et al*., 1983). Curdlan production was restored in 26 mutants (LTU61–LTU86) by pVS1506 and in the remaining two mutants (LTU87 and LTU88) by pVS1511.

### *Identification of two regions (locus I and locus II) in the* Agrobacterium *genome encoding* crd *genes*

The region of pVS1506 responsible for complementation of the mutations in LTU61–LTU86 was found to be located in an 8.8 kb *Eco*RI fragment. This fragment was identified initially because it was the only one from pVS1506 that hybridized with probes P1501–P1504 (derived from pVS1501–pVS1504, the clones recovered from mutants LTU61–LTU64). When the 8.8 kb fragment was cloned into pLAFR1, the resulting plasmid (pVS1513; Figure 1B) complemented the mutations in LTU61–LTU64 as expected, and also those in LTU65–LTU88. Thus, pVS1513 mimics the behavior of the cosmid recombinant, pVS1506. A physical map of the 8.8 kb fragment was constructed (Figure 1B) based on restriction analysis of pVS1513 and pVS1501–pVS1504 which carry various parts of this fragment. The locations of Tn*phoA* in LTU61–LTU64 were deduced from the restriction analysis and confirmed by Southern analysis in

which *Eco*RI-digested genomic DNAs were probed individually with P1501–P1504 (data not shown). The map locations of Tn*phoA* in LTU65 and LTU66 were also deduced by Southern analysis. All six of the mapped mutations clustered in a 2.5 kb region of the 8.8 kb fragment. This region defines the position of locus I.

The Crd⁻ mutations in LTU87 and LTU88 that were complemented only by the second cosmid recombinant (pVS1511) were found by Southern analysis to be located in the same 2.5 kb *Pvu*I- and 8 kb *Stu*I-fragments but different *Eco*RI-fragments (an 0.4 kb fragment in LTU87 and another, unidentified, fragment in LTU88; data not shown). These mutations are therefore closely linked and define the position of locus II.

### *Sequencing and analysis of the locus I gene,* crdS

The plasmids pVS1501–pVS1504 carry junction fragments cloned from four different locus I insertion mutants. This permitted a Tn*phoA*-specific primer to be used to determine the sequence of the *Agrobacterium* DNA adjacent to the transposon in each plasmid. Of the sequences obtained, only those from LTU61 (in pVS1501) and LTU62 (in pVS1502) shared homology with known genes (see below). The single *crd* gene that had been inactivated in these two mutants was then sequenced and named *crdS*. This 1623 nt gene commences with an ATG start codon and terminates with a TGA stop codon. A putative ribosome binding site (GCGAGGT) (Shine and Dalgarno, 1974) occurs 8 nt upstream of the initiation codon, but no consensus -10 and -35 *E.coli*-like promoter sequences (Staden, 1984) were detected in the upstream (120 nt) sequences. The G+C content of *crdS* is 57% and is thus within the range found for *Agrobacterium* species (57–63%; Kersters and DeLey, 1984).

The *crdS* gene encodes a predicted 540 amino acid protein (59 kDa) that contains a high proportion of hydrophobic amino acids (50.4%) suggestive of a membrane-associated protein. This possibility was supported by hydropathy analysis (Kyte and Doolittle, 1982) which revealed stretches of hydrophobic residues, four of which are predicted by the ALOM program (Klein *et al*., 1985) to be transmembrane domains (residues 9–33, 35–58, 379–395, and 482–498; Figure 2). The first of these domains contains a signal sequence with a probable cleavage site between Ala27 and Leu28 (von Heijne, 1986). Hence, CrdS may be anchored in the cytoplasmic membrane by transmembrane domains II, III, and IV and contains an extensive region (residues 59–378) between the first two of these.

The CrdS protein was found by a BLASTX search (Atshul *et al*., 1990; Gish and States, 1993) to share homology with members of the HasA family of β-glycosyl transferases with repetitive action patterns (Keenleyside and Whitfield, 1996). Further analysis using LFASTA (Pearson and Lipman, 1988) revealed that homology over regions exceeding 300 amino acid residues was highest (ca. 46% similarity, 26% identity) with bacterial cellulose synthases, namely, with *Acetobacter xylinum* AcsA (Saxena *et al*., 1994), AcsAII (Saxena and Brown, 1995) and BcsA (Wong *et al*., 1990), *Agrobacterium tumefaciens* CelA (Matthysse *et al*., 1995b) and with cellulose synthase homologs from *E.coli* (YhjO; Sofia *et al*., 1994) and a chlorella virus (A473L; Kutish *et al*., 1996). Lower homology was found with other repetitive glycosyl transferases from bacteria (e.g., with *Rhizobium meliloti* NodC (30% similarity, 22% identity); Jacobs *et al*., 1985) and plants (i.e., with a partial gene sequence from *Arabidopsis thaliana* T88271 (ca. 36% similarity, 27% identity); Saxena and Brown, 1997) and with cotton CelA1 and CelA2

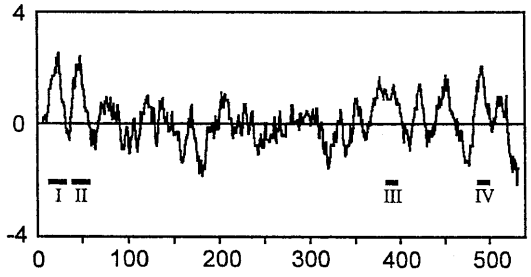


**Fig. 2.** Hydropathy profile of CrdS. The plot (Kyte and Doolittle, 1982) is of the derived amino acid sequence of CrdS, using a window of 12 residues. The number of amino acids is plotted on the x-axis, and hydropathy is plotted on the y-axis. Of the six hydrophobic stretches, four (indicated by bars I–IV) are predicted by the ALOM program (Klein *et al*., 1985) to be transmembrane domains. Domain I is also a putative cleavable signal sequence (von Heijne, 1986).

cellulose synthases (ca. 40% similarity, 23% identity; Pear *et al*., 1996) when the cotton-specific regions of these proteins were omitted from the comparison]. Interestingly, no significant homology was detected with closely related proteins believed to be components of fungal and yeast (1→3)-β-glucan synthases (Cid *et al*., 1995; Inoue *et al*., 1995; Mazur *et al*., 1995; Kelly *et al*., 1996, Mio *et al*., 1997) nor with bacterial proteins involved in cyclic (1→3,1→6)-β-glucan or (1→2)-β-glucan production (Zorrigueta *et al*., 1988; Ielpi *et al*., 1990; Bhagwat *et al*., 1996).

A CLUSTALW (Thompson *et al*., 1994) alignment of CrdS and the bacterial cellulose synthase proteins, which are all about 200 residues longer, showed that homology was confined mainly to the central regions corresponding to CrdS residues 118 to 353 (Figure 3). The proportion of completely conserved amino acids was highest within the group AcsA, AcsAII, BcsA, and YhjO (54%) but decreased if CelA (35%) or CrdS (25%) was included in the comparison. CrdS had highest sequence identity with CelA (33%). These sequence homologies and a dendrogram constructed by the method of maximum parsimony (Hein, 1990) (Figure 4) indicated that CrdS is a distant relative of the bacterial cellulose synthases, albeit one that produces a β-glucan comprised of (1→3)- rather than (1→4)-linkages. These analyses also showed that the cellulose synthases and CrdS were distinct from other repetitive transferases (e.g., HasA and NodC) whose products contain monosaccharides other than glucose.

Computer analyses have revealed that β-glycosyl transferases share conserved sequences and structural features (based on hydrophobic cluster analysis (HCA); Saxena *et al*., 1995; Saxena and Brown, 1997). The repetitive transferases in this group contain a D,D,D35QXXRW motif distributed over two domains, A and B (Saxena *et al*., 1995; Keenleyside and Whitfield, 1996). Non-repetitive transferases have domain A but not B, and so have only the first two Asp residues of the motif (Saxena and Brown, 1995; Keenleyside and Whitfield, 1996). Inspection of the CrdS sequence (Figure 3) and its HCA plot (Figure 5) shows that the entire motif and both domains are present and align with corresponding elements in AcsA, a type member of the repetitive transferases (Saxena and Brown, 1995, 1997), and CelA (Matthysse *et al*., 1995b) which shares greatest homology with CrdS. The entire motif also occurs in the cotton and chlorella virus homologs (Table I), extending the number of proteins in which it has been detected (Saxena and Brown, 1997). An additional noteworthy structural feature revealed by the CrdS HCA plot is a Pro-Leu-rich cluster at the boundary of transmembrane domain II and domain A.

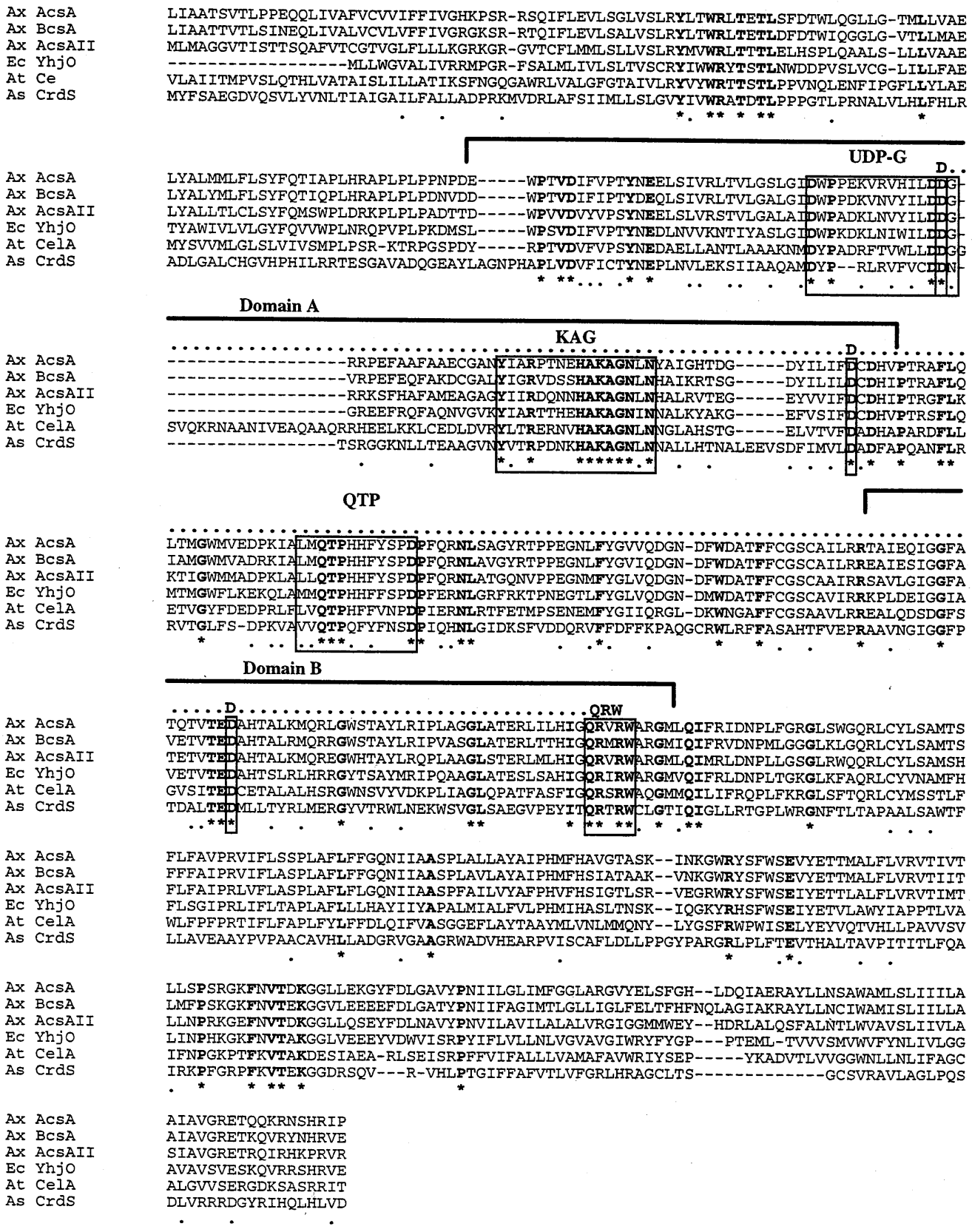


**Fig. 3.** Multiple alignment of the entire CrdS amino acid sequence (residues 1–540) and portions of the polypeptides of *A.xylinum* (residues 41–589), BcsA (residues 41–591), and AcsAII (residues 37–587), *E.coli* YhjO (residues 1–531), and *A.tumefaciens* CelA (residues 17–583), using the CLUSTALW program (v. 1.60). Each protein name is preceded by an acronym for the name of the bacterium from which the sequence was derived. Domains A and B previously identified in repetitive β-glycosyl transferases are indicated by the horizontal lines above the sequence (Saxena *et al*., 1995; Keenleyside and Whitfield, 1996). Boxed regions enclose other conserved motifs described in Table I including the components of the D,D,D,QXXRW motif that are distributed across domains A and B (Saxena and Brown, 1997). Identical amino acid residues in all six sequences shown here are highlighted in boldface and marked with asterisks, whereas conservative amino acid replacements are indicated with dots.

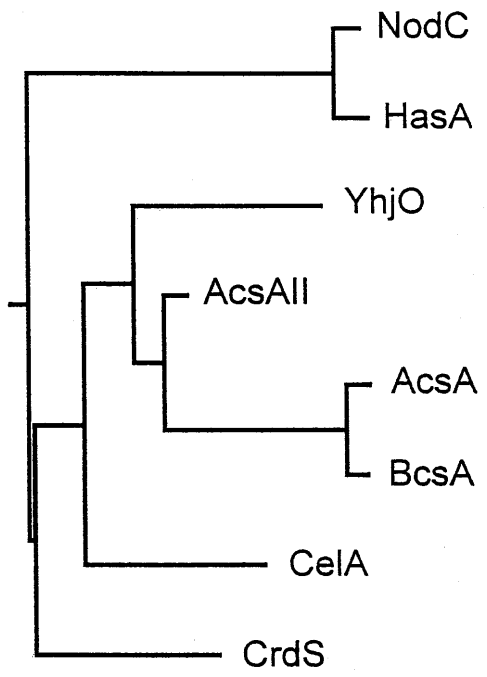


**Fig. 4**. Phylogenetic tree constructed by the maximum parsimony algorithm (TREEALIGN; Hein, 1990) using the amino acid sequence of CrdS (Figure 3) and other sequences retrieved from GenBank.

This feature, with some variations, is also apparent in the HCA plots of AcsA (Saxena *et al*., 1995; Saxena and Brown, 1997) and CelA (Figure 5).

Domain A also contains an incompletely conserved sequence proposed to be a potential UDP-glucose/UDP-N-acetylglucosamine binding motif (Delmer and Amor, 1995). A modified motif derived from the Delmer sequences (Delmer and Amor, 1995; Kawagoe and Delmer, 1997) was detected in all cellulose synthases and several other β-glycosyl transferases as well as in domain A of CrdS (residues 146–157; Table I), and was located in the expected position (Figure 3). CrdS also carried a second motif (residues 387–401), but this was located in putative transmembrane domain III and contained only one Asp of the highly conserved pair (i.e., that corresponding to the first Asp of the D,D,D35QXXRW motif; Saxena and Brown, 1997). Two additional conserved sequences, one in domain A (KAG) and the other in the region between domains A and B (QTP), were confined to CrdS, the cellulose synthases and their homologues YhjO and A473L (Table I). Three potential N-glycosylation sites (Asn-X-Ser/Thr) were detected in CrdS commencing at position 16 (Asn-Leu-Thr), 157 (Asn-Thr-Ser) and 365 (Asn-Phe-Thr). Two such sites also occur in AcsA (Saxena *et al*., 1990), which is known to be a glycoprotein (Lin and Brown, 1989), and four occur in the cotton CelA1 (Pear *et al*., 1996).

## Discussion

The experiments described here are the first report of the cloning of genes required for the biosynthesis of the bacterial (1→3)-β-glucan, curdlan. These genes occur in two unlinked loci (I and II) and were identified in nonoverlapping genomic clones of *Agrobacterium* sp. ATCC31749 by their ability to complement transposon (Tn*phoA*)-insertion mutations that blocked curdlan synthesis. The *crdS* gene, which was fully sequenced, occurs in locus I together with at least two other curdlan genes. This was suggested by the finding that mutations within *crdS*, and on either side of it, prevented curdlan production (Figure 1B**),** and that locus I mutations were detected more frequently than those at locus II (26 versus 2) in the random collection of mutants that was studied. It thus seems likely that locus I is a biosynthetic gene cluster comparable with those found in other systems for bacterial polysaccharide production (Leigh and Coplin, 1992). On the other hand, locus II may contain only a single *crd* gene that is separated from locus I by as much as 25 kb, the average size of the cosmid clones that were isolated. Such wide separation of the genes involved in polysaccharide production is unusual but occurs, for example, in the alginate system of *Pseudomonas aeruginosa* where the biosynthetic gene cluster is distant from its two regulatory genes (Deretic *et al*., 1994). The role of the locus II gene(s) is unknown but it may also be involved in the regulation of curdlan production.

Evidence that locus I contains a gene required for curdlan synthesis rests on the characteristics of *crdS*, the curdlan gene that was sequenced. Mutations in this gene (in LTU61 and LTU62; Figure 1B) abolished curdlan production. Moreover, sequence analysis of the derived CrdS protein revealed homology with β-glycosyl transferases with repetitive action patterns. Amongst these, homology was highest with known bacterial cellulose synthases (e.g., AcsA (Saxena *et al*., 1994), BcsA (Wong *et al*., 1990), CelA (Matthysse *et al*., 1995b)) and with their prokaryotic, eukaryotic and viral homologs (e.g., YhjO (Sofia *et al*., 1994), CelAI and CelA2 (Pear *et al*., 1996), A473L (Kutish *et al*., 1996)). Homologies with the prokaryotic proteins are mainly confined to the proposed cytoplasmic domain in the central portion of each protein (Figure 3) where sequence alignment was possible over about 235 amino acids. Homology was also evidenced by the presence of three similarly located Asp residues and other conserved sequences detected in repetitive glycosyl transferases and implicated in catalysis and UDPGlc substrate binding (Delmer and Amor, 1995; Saxena *et al*., 1995; Saxena and Brown, 1997). Two conserved motifs, designated KAG and QTP (Figure 3, Table I), of unknown significance were identified and found to be confined to CrdS, the known cellulose synthases and the YhjO and A473L homologs. Structural conservation was also evident in HCA comparisons of CrdS with AcsA, a type member of the repetitive transferases (Saxena *et al*., 1995), and CelA (Matthysse *et al*., 1995b), which shares greatest homology with CrdS (Figure 5). This included the two domains, A and B, that are proposed to be characteristic of repetitive glycosyl transferases (Saxena *et al*., 1995; Keenleyside and Whitfield, 1996; Saxena and Brown, 1997) and a Pro-Leu-rich cluster at the start of the proposed cytoplasmic region (Figure 5). From these various similarities we deduced that *crdS* encodes the curdlan synthase catalytic protein.

Although computer-based comparisons have provided a means for identifying (Saxena *et al*., 1995) and classifying glycosyl transferases (Keenleyside and Whitfield, 1996; Campbell *et al*., 1997; Saxena and Brown, 1997), experimental evidence supporting functional roles for the conserved sequences is still limited. Amino acids in the D,D,D35QXXRW motif of *Saccharomyces cerevisiae* chitin synthase (Chs2p; Nagahashi *et al*., 1995) and *A.xylinum* cellulose synthase (AcsAB; Saxena and Brown, 1997) have been implicated in the catalytic mechanism by site-directed mutagenesis studies. In AcsAB, replacement of

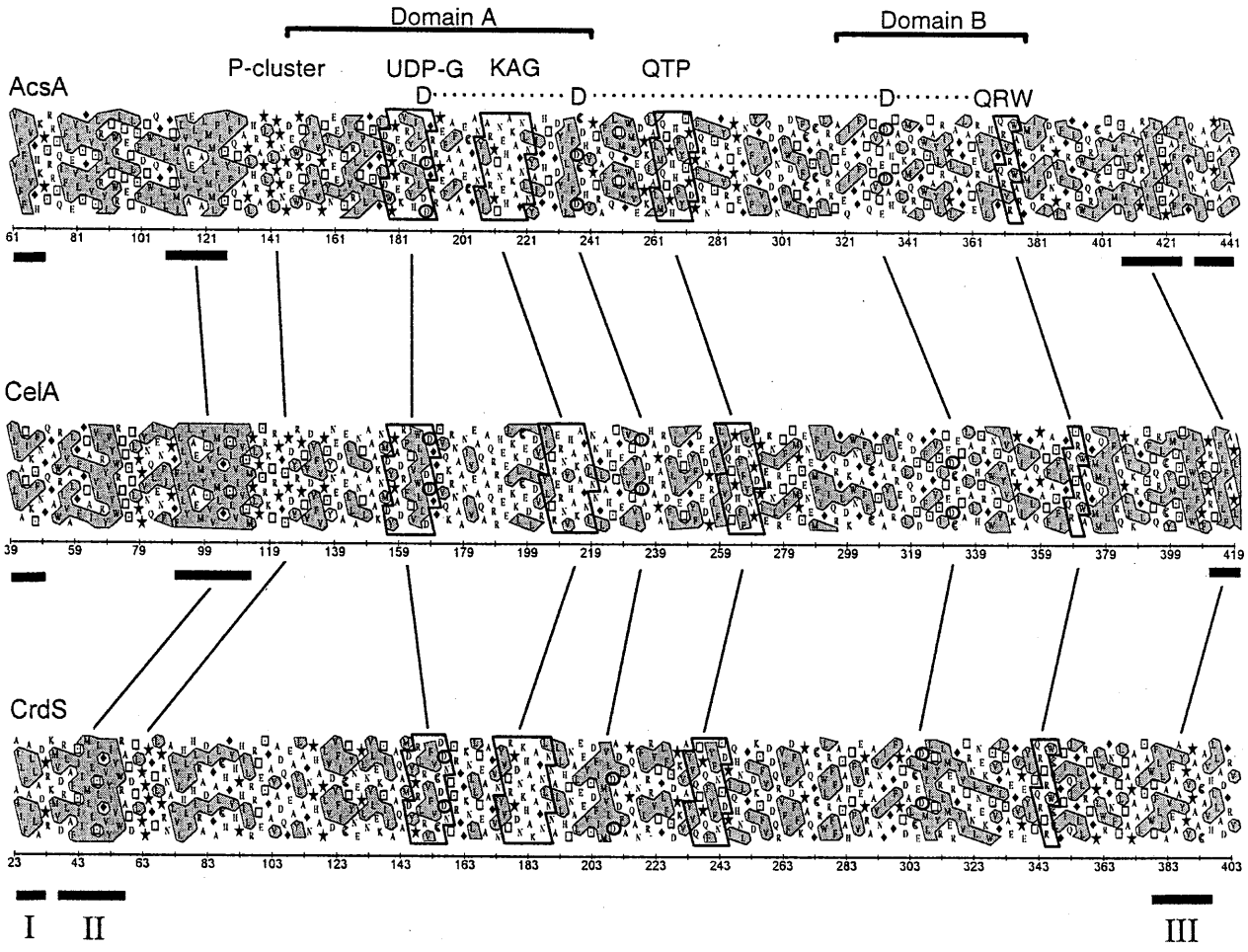


**Fig. 5.** HCA plots (Lemesle-Varloot *et al*., 1990) of CrdS, CelA, and AcsA for the amino acid sequence indicated under each plot. The analysis was performed using the HCA-Plot program (Doriane Informatique, Le Chesnay, France). The sequences are written on a classical α-helix (3.6 amino acids per turn) smoothed on a cylinder. To make the three-dimensional representation easier to handle, the cylinder is cut parallel to its axis and then spread out. Since some adjacent amino acids are now separated by the unfolding of the cylinder, their representation is duplicated to restore the full connectivity of the amino acids. Clusters of adjacent hydrophobic residues (Ala, Val, Leu, Ile, Met, Phe, Gln) are within contours. Amino acids are denoted with the one-letter code except for proline (stars), glycine (diamonds), serine (squares with dot), and threonine (squares). Vertical lines are drawn to indicate structurally conserved features including transmembrane domains (indicated by bars but numbered only for CrdS), domains A and B identified in AcsA (Saxena *et al*., 1995), and the various conserved motifs (enclosed in boldface outline).

the first or second Asp residue impaired *in vitro* enzyme activity, as did replacement in Chs2p of the first or third Asp, or of Gln, Arg, or Trp in the QXXRW sequence. Even conservative replacements of the third Asp, or of Gln, Arg, or Trp in Chs2p were not tolerated, suggesting that these residues are involved in the catalytic reaction. The two residues flanking the third Asp in Chs2p were also essential for enzyme activity. This is of interest as one of these, Glu, is conserved in CrdS and the bacterial cellulose synthases (Figure 3), and in NodC proteins (Nagahashi *et al*., 1995).

The binding motif for UDPGlc, the presumptive curdlan synthase substrate, was initially recognized on the basis of photoaffinity labeling which showed that a peptide from spinach (*Spinacea oleracea*) sucrose phosphate synthase (EC 2.4.1.13) could bind this monosaccharide donor (Salvucci and Klein, 1993). Part of this peptide sequence occurs in the UDPGlc-binding motif recognized by Delmer and Amor (1995) in plant and other glycosyl transferases (Kawagoe and Delmer, 1997), and includes the first Asp of the D,D,D35QXXRW motif (Figure 3). A fragment of the cotton CelA protein containing the binding motif has been expressed in *E.coli* and found to bind UDPGlc in a $Mg^{2+}$-dependent manner (Pear *et al*., 1996), as is the case for bacterial cellulose synthase (Mayer *et al*., 1991). UDPGlc-binding did not occur with a corresponding fragment that lacked the motif. Taken together, all this evidence is consistent with the proposed importance of the D,D,D35QXXRW and UDPGlc-binding motifs in the catalytic events mediated by repetitive glycosyl transferases.

The detection of two potential UDPGlc-binding motifs in CrdS was based on a search using a modified motif derived from sequences identified by Delmer and Amor (1995) and Kawagoe and Delmer (1997). Both CrdS motifs differ from the consensus sequence at three positions, but the first aligns with the corresponding single motif in the cellulose synthases (Figure 3, Table I). The second is unique to CrdS and is exceptional in that it lacks the highly conserved DD residues (Table I). Moreover,

**Table I.** Characteristic sequence motifs in β-glycan synthases and other β-glycosyl transferases

| Protein[a] | Motif[b] | | | |
|---|---|---|---|---|
| | **UDP-G**[c] D(WY)PX$_{3:6}$(IV)(WY)(ILV)(LV)DDG [≤3 mismatches] | **KAG**[d] Y(IV)XRX$_{4:8}$H(AK)KAG(NA)LN [≤3 mismatches] | **QTP**[d] (LV)(LMV)QTP(HQ)X(FY)X$_{3:4}$D [≤3 mismatches] | **DDD$_{35}$QRW**[e] DX$_{40:130}$DX$_{90:140}$DX$_{30:40}$Q(QR)XRW [0 mismatches] |
| Axy **AcsA** | DWPPEKVR.V*H*ILDDG | YIARPTNE....HAKAGNLN | LMQTPHHFYSP.D | DX$_{46}$DX$_{96}$DX$_{35}$QRVRW |
| Axy **BcsA** | DWPPDKVN.VYILDDG | YIIRDQNN....HAKAGNLN | LMQTPHHFYSP.D | DX$_{46}$DX$_{96}$DX$_{35}$QRMRW |
| Axy **AcsAII** | DWPADKLN.VYILDDG | YIIRDQNN....HAKAGNLN | LLQTPHHFYSP.D | DX$_{46}$DX$_{96}$DX$_{35}$QRVRW |
| Atu **CelA** | DYPADRFT.VWLLDDG | Y*LT*RERNV....HAKAGNLN | LVQTPHFFVNP.D | DX$_{65}$DX$_{96}$DX$_{35}$QRSRW |
| Eco **YhjO** | DWPKDKLN.IWILDDG | YIARTTHE....HAKAGN*I*N | *M*MQTPHHFFSP.D | DX$_{46}$DX$_{96}$DX$_{35}$QRIRW |
| Asp **CrdS** | DYPRLR...V*F*V*C*DD*N* *A*YPVPAACA*V*H*L*LADG | YVTRPDNK....HAKAGNLN | VVQTPQFYFNS.D | DX$_{52}$DX$_{96}$DX$_{35}$QRTRW |
| Ghi **CelA1** | DYPVDKVS.*C*YI*S*DDG | YVSREKRPGYQHHKKAGA*E*N | *Y*VQ*F*PQRFDGI.D | DX$_{165}$DX$_{212}$DX$_{37}$Q*V*LRW [g] |
| Ghi **CelA2** | - [f] | YVSREKRPGYQHHKKAGA*E*N | *Y*VQ*F*PQRFDGI.D | -X$_{148}$DX$_{236}$DX$_{37}$Q*V*LRW [g] |
| PbCV **A473L** | DWPKLN...V*H*VLDDG | YITRPNNH....*M*KKAGNL*R* | *I*VQTPQFFEVRPD | DX$_{46}$DX$_{93}$DX$_{35}$QQYRW |
| Ath **T88271** | VWPSSRLV.VQVVDDS | - | - | DX$_{58}$DX$_{93}$DX$_{35}$QQHRW |
| Spy **HasA** | *T*YPLSE...IYIVDDG | - | - | DX$_{55}$DX$_{99}$DX$_{35}$QQNRW |
| Rme **NodC** | DYPGELR..VYVVDDG | - | - | DX$_{51}$DX$_{100}$DX$_{35}$QQLRW |
| Sce **Chs1** | - | - | - | DX$_{101}$DX$_{114}$DX$_{38}$QRRRW |

[a]The protein name is preceded by an acronym for the name of the organism from which the encoding DNA was derived. AcsA, AcsAII, and BcsA are cellulose synthases from strains of *Acetobacter xylinum*. CelA is cellulose synthase from a strain of *Agrobacterium tumefaciens*. CelA1 and CelA2 are cellulose synthases from *Gossypium hirsutum* (cotton). YhjO and A473L are cellulose synthase homologs from *Escherichia coli* and *Paramecium bursaria* Chlorella virus, respectively. CrdS is curdlan synthase from *Agrobacterium sp.* ATCC31749. T88271 is a cellulose synthase homolog from *Arabidopsis thaliana*. HasA is hyaluronate synthase from *Streptococcus pyogenes* WF14. NodC is Nodulation C protein from *Rhizobium meliloti* 1021. Chs1 is chitin synthase 1 from *Saccharomyces cerevisiae*.

[b]Boldface text indicates matches to the motif, italics indicate mismatches and roman text indicates redundancies (i.e., X in motif). Also shown are how many mismatches were allowed in sequences scored as containing the motif. The GCG find patterns program was used to search the SwissProt database at the Australian National Genome Information Service for sequences containing the motifs. The database contained 59,021 protein sequences at the time of the search, including 13 NodC (N-acetylglucosaminyltransferase), 3 putative cellulose synthase (AcsA, BcsA, YhjO), 4 fungal putative (1→3) β-glucan synthase subunit, 40 chitin synthase, and 14 glycogen synthase but no hyaluronate synthase sequences. Of the proteins listed in this table only AcsA, BcsA, YhjO, NodC, and Chs1 were present. None of the motifs were found in any of the glycogen synthase or putative (1→3) β-glucan synthase sequences.

[c]As well as AcsA, BcsA, YhjO, and NodC, the SwissProt database included 45 sequences carrying this motif, including 5 other NodC proteins, all of which contained the DD pair. Of these 49 sequences with the motif as defined, 26 lacked the DD pair.

[d]The SwissProt database included only three sequences carrying this motif: AcsA, BcsA, and YhjO.

[e]As well as AcsA, BcsA, YhjO, NodC, and Chs1, the SwissProt database contained 27 sequences carrying this motif, including 13 other chitin synthase and 6 other NodC genes. The minimal motif Q(QR)XRW with no mismatches was found in 88 SwissProt sequences, including 8 NodC and 13 chitin synthase sequences. All sequences in the table contained the minimal Q(QR)XRW motif except the cotton cellulose synthase genes (CelA1 and CelA2).

[f]This sequence is incomplete and lacks the N-terminal portion where this motif is expected.

[g]The motif is detected only if the additional cotton-specific sequences are excluded and a mismatch at the valine residue is permitted. The first aspartate is expected in the missing N-terminal portion of the CelA2 sequence.

it is located in a putative transmembrane domain and in this situation would be unlikely to be accessible to the glucosyl donor.

Perhaps the most intriguing finding of this study is that CrdS, which catalyses (1→3)-β-linkage formation, shares considerable homology in both sequence and structure with cellulose synthases which specifically form (1→4)-β-linkages. This homology also extends to two bifunctional enzymes that catalyze the formation of both (1→3) and (1→4) linkages between two different monomeric units. These are the hyaluronan synthase (HasA) from *Streptococcus pyogenes* (Dougherty and van de Rijn, 1994) and Type 3 capsular polysaccharide synthase (Cps3S) from *S.pneumoniae* (Arrecubieta *et al*., 1995; Dillard *et al*., 1995). HasA and Cps3S also display the sequence and structural characteristics of repetitive glycosyl transferases (Keenleyside and Whitfield, 1996) and have been grouped into the same family as the bacterial cellulose synthases (Campbell *et al*., 1997). On the basis of the results presented here, this family would also include CrdS. The important issue of which features determine linkage (and donor) specificity in these enzymes remains to be resolved. One proposal is that linkage specificity is determined by subtle features of the active site rather than by the global folding reflected by HCAs (Campbell *et al*., 1997). There are two examples in which limited replacement of amino acids over a restricted sequence determines specificity. One example is in the family of human α-fucosyltransferases which share 85% sequence homology and discriminate between acceptor substrates to form either (1→3)- or (1→4)-α-fucosyl linkages (Legault *et al*., 1995). The other example is the (1→3)- and (1→3, 1→4)-β-glucan hydrolases (EC 3.2.1.39 and EC 3.2.1.73, respectively) which hydrolyze (1→3)- or (1→4)-β-glucosidic linkages in their respective substrates. Because their polypeptide backbones are structurally very similar, the differences in their substrate specificity is acquired primarily by substitutions in amino acids lining a deep substrate-binding groove on the enzyme surface (Varghese *et al*., 1994; Høj and Fincher, 1995). Alternatively, linkage specificity may be provided by ancillary proteins. An analogy would be the β-galactosyl transferases (EC 2.4.1.22 and EC 2.4.1.90) which, in the presence of α-lactalbumin, use glucose as an acceptor to form Galβ(1→4)Glc (lactose) but in its absence the acceptor is GlcNAc on N-linked oligosaccharides forming a Galβ(1→4)GlcNAc structure (Hill and Brew, 1975).

**Table II.** Bacterial strains and plasmids

| Strain or plasmid | Relevant characteristics | Source or reference |
|---|---|---|
| *E.coli* | | |
| DH5α | *recA1 hsdR17* | Gibco-BRL |
| SM10 | *recA* chr-RP4–2-Tc::Mu $Km^r$ $Tra^+$ | Simon *et al.* (1983) |
| S17–1 | *recA hsdR* chr-RP4–2-Tc::Mu-Km::Tn7 $Tp^r$ $Sm^r$ $Tra^+$ | Simon *et al.* (1983) |
| HB101 | *recA13 hsdR hsdM rpsL20* | Sambrook *et al.* (1989) |
| LE392 | *hsdR* | Stratagene Cloning Systems |
| *Agrobacterium* | | |
| ATCC31749 | Type strain; $Crd^+$ | ATCC |
| LTU50[a] | Spontaneous $Cm^r$ derivative of ATCC31749 | This work |
| LTU61 & LTU62 | LTU50*crdS*::Tn*phoA* | This work |
| LTU63-LTU86 | LTU50::Tn*phoA* locus I $Crd^-$ mutants | This work |
| LTU87 & LTU88 | LTU50::Tn*phoA* locus II $Crd^-$ mutants | This work |
| LTU89-LTU113 | LTU50::Tn*phoA* $Crd^{+/-}$ | This work |
| Plasmids | | |
| RP1 | $Ap^r$ $Tc^r$ $Km^r$ $Tra^+$ IncP | Pansegrau *et al.* (1994) |
| pLAFR1 | Broad host range cosmid vector; $oriV_{RK2}$ $oriT_{RK2}$;$Tc^r$ | Friedman *et al.* (1982) |
| pUC19 | $Ap^r$, cloning vector | Yanisch-Perron *et al.* (1985) |
| pRT733 | Suicide plasmid; $oriV_{R6K}$ $oriT_{RP4}$ *TnphoA*; $Ap^rKm^r$ | Taylor *et al.* (1989) |
| pVS1500 | pUC19 with 6.8 kb *Eco*RI-junction fragment of LTU87; $Km^r$ $Ap^r$ | This work |
| pVS1506 | LTU50 locus I clone in pLAFR1; $Tc^r$ | This work |
| pVS1511 | LTU50 locus II clone in pLAFR1; $Tc^r$ | This work |

[a] Used as the wild type strain in this work.

It is remarkable that homology was detected between CrdS and cellulose synthases (which specify (1→4)-β-linkages) but not, as might have been anticipated, with the proposed (1→3)-β-glucan synthases detected in *S.cerevisiae* (Douglas *et al.*, 1994; Inoue *et al.*, 1995), *Candida albicans* (Mio *et al.*, 1997) and *Aspergillus nidulans* (Kelly *et al.*, 1996). The uniqueness of this group of closely related fungal proteins is also reflected in their absence from the 26 families of glycosyl transferases classified by Visual BLAST and HCA comparisons (Campbell *et al.*, 1997). This apparent difference between fungal (1→3)-β-glucan synthases and CrdS and the bacterial and plant cellulose synthases suggests that the two groups of proteins have evolved from different ancestors and are products of convergent evolution. This is not without precedent. The (1→3, 1→4)-β-glucan hydrolases (EC 3.2.1.73) from plants (Høj and Fincher, 1995) on the one hand and bacteria (Heinemann *et al.*, 1996) and protists (Chen *et al.*, 1997), on the other, belong to distinct folding groups (Henrissat, 1990; Henrissat and Bairoch, 1993; Davies and Henrissat, 1995), but their specificities and catalytic mechanisms are identical (Høj and Fincher, 1995).

## Materials and methods

### *Bacterial strains, plasmids, and media*

The bacteria and plasmids used in this study are listed in Table II. Additional plasmids and their modes of construction are shown in Figure 1B. Strains of *E.coli* or *Agrobacterium* were grown in nutrient broth (NB) and nutrient agar (NA; Palombo *et al.*, 1989). *Agrobacterium* was also grown in defined broth (ADB) prepared according to Phillips and Lawford (1983) except that 20 mM $KNO_3$ replaced $NH_4Cl$, and on aniline blue agar medium (ABA) containing 4% (wt/vol) glucose, 0.5% Oxoid yeast extract, 0.005% (wt/vol) aniline blue (Fischer), and 2% Difco agar. The pH of these media were adjusted to 7.0 before autoclaving. When required, antibiotic supplements (mg/l) in NA were ampicillin (Ap, 100), kanamycin (Km, 15), and tetracycline (Tc, 15); those in ABA were ampicillin (100), chloramphenicol (Cm, 50), kanamycin (100), and tetracycline (2.5).

### *Detection and assay of curdlan*

Curdlan production by bacteria was detected on ABA medium which contains the (1→3)-β-glucan-specific dye, aniline blue (Nakanishi *et al.*, 1976). After incubation at 28°C for 5 d, curdlan-producing colonies are stained blue by the dye. Curdlan production in broth was assayed quantitatively by the method of Nakanishi *et al.* (1976) and involved stirring a 50 ml ADB culture of *Agrobacterium* (cell density ~$5 \times 10^8$/ml) together with an equal volume of 1 M NaOH. After 15 min the mixture was centrifuged and the supernatant neutralized (pH 7.0) by adding 4 M HCl. The precipitate that formed was collected by centrifugation at $20,000 \times g$ for 10 min, washed three times with distilled water, then frozen at -20°C before being freeze-dried and weighed.

### *Conjugation procedures for transposon mutagenesis and complementation studies*

Bacterial cultures in 10 ml NB were grown to about $5 \times 10^8$ cells/ml either at 37°C for *E. coli* donors or 28°C for *Agrobacterium* recipients. Conjugations were performed at 28°C for 3–4 h on

nitrocellulose filters as described by Palombo *et al*. (1989) and then the transconjugants recovered on selective medium incubated at 28°C for 3–5 d. Transposon mutagenesis of *Agrobacterium* was performed using the suicide plasmid pRT733 which carries Tn*phoA* and is mobilizable from *E.coli* SM10 (Taylor *et al*., 1989). To ensure the isolation of non-sibling mutants, separate filter matings between *E.coli* SM10 carrying pRT733 and LTU50 were performed and the transconjugants selected on ABA medium containing Cm and Km (the *phoA* characteristic was not utilized). Colonies that failed to stain with aniline blue were purified and the inheritance of Tn*phoA* and loss of pRT733 confirmed by the $Km^r$ $Ap^s$ phenotype of the bacteria. Complementation studies involved conjugation between *E.coli* S17–1 carrying a pLAFR1-recombinant and recipient cultures (Tn*phoA*-induced curdlan mutants). The transconjugants were selected on ABA medium containing Tc and Cm and after growth of the bacteria at 28°C for 3–5 d, complementation was detected by the blue staining of the transconjugant colonies.

### *DNA techniques and preparation of an* Agrobacterium *genomic library*

Methods for DNA isolation, manipulation, cloning, and transformation of electrocompetent DH5α cells were primarily adapted from the manual by Sambrook *et al*. (1989). DNA fragments were purified from low melting agarose gels using the Bresa-Clean DNA isolation kit (Bresatec) and then used for cloning or as probes in Southern or colony hybridizations. DNA–DNA hybridizations were performed by the method of Sambrook *et al*. (1989) for the detection of low abundance sequences. Probes were prepared from the 2.8 kb *Bgl*II fragment of Tn*phoA* (Figure 1A) or the *Hpa*I(Tn*phoA*)-*Eco*RI(*Agrobacterium*) junction fragments of pVS1500–1505 (Figure 1B) and labeled with [α-$^{32}$P]dCTP using the Megaprime DNA labeling kit (Amersham). Hybridization to DNA on nitrocellulose filters was in 6× SSPE, 5× Denhardt's reagent, 0.5% SDS and 100 μg $ml^{-1}$ denatured, fragmented salmon sperm DNA for 18 h at 65°C. Filters were washed twice (20 min each time at 65°C) in 2× SSC, 0.5% SDS and once (30 min at 65°C) in 0.2× SSC, 0.1% SDS, then exposed to x-ray film. Genomic DNA was prepared by the procedure of Ma *et al*. (1982) from 50 ml cultures grown overnight in NB. The *Agrobacterium* genomic library was prepared from LTU50 genomic DNA that was partially digested with *Eco*RI and size-fractionated in a linear 10–40% sucrose gradient. Fragments in the size range of 15–35 kb were ligated to pLAFR1($Tc^r$) that had been digested with *Eco*RI and treated with calf intestine alkaline phosphatase. The ligation mixture was packaged *in vitro* using Gigapack II (Stratagene) and the phage particles obtained used to transduce *E.coli* LE392 to Tc-resistance. Twenty transductants chosen at random all carried pLAFR1-recombinants with DNA inserts averaging 25 kb.

### *DNA sequencing and sequence analysis*

The DNA sequence of *crdS* was determined using a Tn*phoA*-based primer (5′-CATGAACGTTACCATGTTAG-3′) to sequence the *Agrobacterium* DNA adjoining Tn*phoA* in pVS1501 and pVS1502 (derived from LTU61 and LTU62; Figure 1B). Thereafter, the primer walking strategy was used to complete the sequencing of both strands employing 15 oligonucleotide primers. The DNA sequencing reactions were performed by the dideoxy-chain termination method (Sanger *et al*., 1977), employing the ABI PRISM Ready DyeDeoxy Termination Cycle Sequencing System with the AmpliTaq DNA polymerase (Perkin-Elmer Corporation). The DNA extension products that were produced were analyzed in an ABI Model 373A DNA Sequencer. Nucleotide and amino acid sequences were analyzed with the sequence analysis package (version 7.0; Genetics Computer Group, Madison, WI) and other software provided by the Australian National Genome Information Service (ANGIS, University of Sydney).

### *Nucleotide sequence accession numbers*

The nucleotide sequence of *crdS* was submitted to GenBank under the accession number AF057142. GenBank accession numbers of other sequences used in this study (see Table I) were AcsA (X54676), BcsA (M96060), AcsAII (U15957), CelA (L38609), CelA1 (U58283), CelA2 (U58284), A472L (U42580), HasA (L21187), NodC (M11268), and Chs1 (M14045). The YhjO accession number was from SwissProt P37653 and accession number T88271 was from the EMBL EST database.

## Acknowledgments

We thank Dr. B. T. O. Lee and Dr. S. R. Long for providing SM10 and pLAFR1, respectively; Dr. J. Beckwith for the nucleotide sequence of Tn*phoA*; and Dr. B. Henrissat for the HCA program.

## Abbreviations

ATCC, American Type Culture Collection; HCA, hydrophobic cluster analysis; UDPGlc, uridine diphosphate-α-D-glucose.

## References

Altshul,S.F., Gish,W., Miller,W., Meyers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol*., **215,** 403–410.

Arrecubieta,C., Garcia,E. and Lopez,R. (1995) Sequence and transcriptional analysis of a DNA region involved in the production of capsular polysaccharide in *Streptococcus pneumoniae* type 3. *Gene*, **167,** 1–7.

Aspinall,G.O. and Kessler,G. (1957) The structure of callose from the grape vine. *Chem. Ind. (Lond.)*, 1296.

Bhagwat,A.A., Gross,K.C., Tully,R.E. and Keister,D.L. (1996) β-glucan synthesis in *Bradyrhizobium japonicum*: characterisation of a new locus (*ndvC*) influencing β-(1→6) linkages. *J. Bacteriol*., **178,** 4635–4642.

Breedveld,M.W. and Miller,K.J. (1994) Cyclic β-glucans of members of the family *Rhizobiaceae*. *Microbiol. Rev.*, **58,** 145–161.

Campbell,J.A., Davies,G.J., Bulone,V. and Henrissat,B. (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarity. *Biochem. J*., **326,** 929–942.

Chen,H., Li,X.-L. and Ljungdahl,L.G. (1997) Sequencing of a 1,3–1,4-β-D-glucanase (lichenase) from the anaerobic fungus *Orpinomyces* strain PC-2: properties of the enzyme expressed in *Escherichia coli* and evidence that the gene has a bacterial origin. *J. Bacteriol*., **179,** 6028–6034.

Cid,V.R., Duran,A., del Rey,F., Snyder,M.P., Nombela,C. and Sanchez,M. (1995) Molecular basis of cell intergity and morphogenesis in *Saccharomyces cerevisiae*. *Microbiol. Rev.*, **59,** 345–386.

Davies, G. and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure*, **3,** 853–959

Delmer,D. and Amor,Y. (1995) Cellulose biosynthesis. *The Plant Cell*, **7,** 987–1000.

Deretic,V., Schurr,M.J., Boucher,J.C. and Martin.D.V. (1994) Conversion of *Pseudomonas aeruginosa* to mucoidy in cystic fibrosis: environmental stress and regulation of bacterial virulence by alternative sigma factors. *J. Bacteriol*., **176,** 2773–2780.

Dillard,J.P., Vanderse,M.W. and Yother,J. (1995) Characterisation of a cassette containing genes for Type 3 capsular polysaccharide biosynthesis in *Streptococcus pneumoniae J. Exp. Med*., **181,** 973–983.

Dougherty,B.A. and van de Rijn,I. ( 1994) Molecular characterisation of *hasA* from an operon required for hyaluronic acid synthesis in Group A Streptococci. *J. Biol. Chem*., **269,** 169–175.

Douglas,C.M., Foor,F., Marrinan,J.A., Morin,N., Nielsen,J.B., Dahl,A.M., Mazur,P., Baginsky,W., Li,W., El-Sherbeini,M., Clemas,J.A., Mandala,S.M., Frommer,B.R. and Kurtz,M.B. (1994) The *Saccharomyces cerevisiae FKS1*(*ETG1*) gene encodes an integral membrane protein which is a subunit of 1,3-β-D-glucan synthase. *Proc. Natl. Acad. Sci. USA*, **91,** 12907–12911.

Evans,N.A., Hoyne,P.A. and Stone,B.A. (1984) Characteristics and specificity of the interaction of a fluorochrome from aniline blue (Sirofluor) with polysaccharides. *Carbohydr. Polym*., **4,** 215–230.

Friedman,A.M., Long,S.R., Brown,S.E., Buikema,W.J. and Ausubel,F.M. (1982) Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of *Rhizobium* mutants. *Gene*, **18,** 289–296.

Gish,W. and States,D.J. (1993) Identification of protein coding regions by database similarity search. *Nature Genet*., **3,** 266–272.

Harada,T. and Harada,A. (1996) Curdlan and succinoglycan, In Dimitrio,S. (ed.), *Polysaccharides in Medical Applications*. Marcel Dekker, New York, pp. 21–57.

Harada,T., Masada,M., Fujimori,K. and Maeda,I. (1966) Production of a firm, resilient gel-forming polysaccharide by a mutant of *Alcaligenes faecalis* var *myxogenes* IOC3. *Agric. Biol. Chem*., **30,** 196–198.

Harada,T., Terasaki,M. and Harada,A. (1993) Curdlan. In Whistler,R.L. and BeMiller,J.N. (eds), *Industrial Gums*. Academic Press, New York, pp. 427–445.

Hein,J. (1990) A unified approach to alignment and phylogenies. *Methods Enzymol*., **183,** 626–645.

Heinemann,U., Ay,J., Gaiser,O., Muller,J.J. and Ponnuswamy,M.N. (1996) Enzymology and folding of natural and engineered bacterial β-glucanases studied by x-ray crystallography. *Biol. Chem. Hoppe-Seyler*, **377,** 447–454.

Henrissat,B. (1990) A classification of glycosyltransferases based on amino acid sequence similarities. *Biochem. J*., **280,** 309–316.

Henrissat,B. and Bairoch,A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J*., **293,** 781–788.

Hill,R.L. and Brew,K. (1975) Lactose synthetase. *Adv. Enzymol. Relat. Areas Mol. Biol*., **43,** 411–517.

Høj,P.B. and Fincher,G.B. (1995) Molecular evolution of plant β-glucan endohydrolases. *Plant J*., **7,** 367–379.

Ielpi,L., Dylan,T., Ditta,G.S., Helinski,D.R. and Stanfield,S.W. (1990) The *ndvB* locus of *Rhizobium meliloti* encodes a 319-kDa protein involved in the production of β-(1→2)-glucan. *J. Biol. Chem*., **265,** 2843–2851.

Inoue,S.B., Takewaki,N., Takasuka,T., Mio,T., Adachi,M., Fujii,Y., Miyamoto,C., Arisawa,M., Furuichi,Y. and Watanabe,T. (1995) Characterisation and gene cloning of 1,3-β-D-glucan synthase from *Saccharomyces cerevisiae*. *Eur. J. Biochem*., **231,** 45–854.

Jacobs,T.W., Egelhoff,T.T. and Long,S.R. (1985) Physical and genetic map of a *Rhizobium meliloti* nodulation gene region nucleotide sequence of NodC. *J. Bacteriol*., **162,** 469–476.

Kasai,N. and Harada,T. (1980) Ultrastructure of curdlan. In French,A.D. and Gardner,K.H. (eds.), *Fiber Diffraction Methods*. American Chemical Society Symposium Series, No. 141. American Chemical Society, Washington, DC, pp. 363–383.

Kawagoe,Y. and Delmer,D.P. (1997) Pathways and genes involved in cellulose biosynthesis. In Setlow,J.W. (ed.), *Genetic Engineering*. Plenum Press, New York, pp. 63–83.

Keenleyside,W.J. and Whitfield,C. (1996) A novel pathway for O-polysaccharide biosynthesis in *Salmonella enterica* serovar Borreze. *J. Biol. Chem*., **271,** 28581–28592.

Kelly,R., Register,E., Hsu,M.-J., Kurtz,M. and Nielsen,J. (1996) Isolation of a gene involved in β-glucan synthesis in *Aspergillus nidulans* and purification of the corresponding protein. *J. Bacteriol*., **178,** 4381–4391.

Kersters,K. and De Ley,J. (1984) Genus III *Agrobacterium*. In Krieg,N.R. and Holt,J.G. (eds.), *Bergey's Manual of Determinative Bacteriology*, 8th ed. Williams & Wilkins, Baltimore, pp. 244–254.

Klein,P., Kanehisha,M. and DeLisi,C. (1985) The detection and classification of membrane-spanning proteins. *Biochim. Biophys. Acta*, **815,** 468–476.

Kutish,G.F., Li,Y., Lu,Z., Furuta,M., Rock,D.L. and van Etten,J.L. (1996) Analysis of 76 kb of the Chlorella virus PBCV-1 330-kb genome: map positions 182 to 258. *Virology*, **223,** 303–317.

Kyte,J. and Doolittle,R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol*., **157,** 105–132.

Legault,D.J., Kelly,R.J., Natsuka,Y. and Lowe,J.B. (1995) Human α(1,3/1,4)-fucosyltransferases discriminate between different oligosaccharide acceptor substrates through a discrete peptide fragment. *Biol. Chem*., **270,** 20987–20996.

Leigh,J.A. and Coplin,D.L. (1992) Exopolysaccharides in plant-bacterial interactions. *Annu. Rev. Microbiol*., **46, 3**07–346.

Lemesle-Varloot,L., Henrissat,B., Gaboriaud,C., Bissery,V., Morgat,A. and Mornon,J.P. (1990) Hydrophobic cluster analysis: procedures to derive structural and functional information from 2D-representation of protein sequences. *Biochemie*, **72,** 555–574.

Lin,F.C. and Brown,R.M. (1989) Purification of cellulose synthase from *Acetobacter xylinum*. In Schuerch,C. (ed.), *Cellulose and Wood—Chemistry and Technology*. John Wiley and Sons, New York, pp. 473–492.

Ma,Q.-S., Johnston,A.W.B., Hombrecher,G. and Downie,J.A. (1982) Molecular genetics of mutants of *Rhizobium leguminosarum* which fail to fix nitrogen. *Molec. Gen. Genet*., **187,** 166–171.

Manoil,C. and Beckwith,J. (1985) Tn*phoA*: a transposon probe for protein export signals. *Proc. Natl. Acad. Sci. USA*, **82,** 8129–8133.

Matthysse,A.G., Thomas,D.L. and White,A.R. (1995a) Mechanism of cellulose synthesis in *Agrobacterium tumefaciens*. *J. Bacteriol*., **177,** 1076–1081.

Matthysse,A.G., White,S. and Lightfoot,R. (1995b) Genes required for cellulose synthesis in *Agrobacterium tumefaciens*. *J. Bacteriol*., **177,** 1069–1075.

Mayer,R., Ross,P., Weinhous,H., Amikam,D., Volman,G., Ohana,P., Calhoon,R.D., Wong,H.C., Emerick,A.W. and Benziman,M. (1991) Polypeptide composition of bacterial cyclic diguanylic acid-dependent cellulose synthase and the occurrence of immunologically crossreacting proteins in higher plants. *Proc. Natl. Acad. Sci. USA*, **88,** 5472–5476.

Mazur,P., Morin,N., Baginsky,W., El-Sherbeini,M., Clemas,J.A., Nielsen,J.B. and Foor,F. (1995) Differential expression and function of two homologous subunits of yeast 1,3-β-D-glucan synthase. *Molec. Cell Biol*., **15,** 5671–5681.

Mio,T., Adachi-Shimizu,M., Tachibana,Y., Tabuchi,H., Inoue,S.B., Yabe,T., Yamada-Okabe,T., Arisawa,M., Watanabe,T. and Yamada-Okabe,H. (1997) Cloning of the *Candida albicans* homologue of *Saccharomyces cerevisiae* GSC1/FKS1 and its involvement in β-1,3-glucan synthesis. *J. Bacteriol*., **179,** 4096–4105.

Nagahashi,S., Sudoh,M., Ono,N., Sawada,R., Yamaguchi,E., Uchida,Y., Mio,T., Takagi,M., Atisawa,M. and Yamada-Okabe,H. (1995) Characterisation of chitin synthase 2 of *Saccharomyces cerevisiae*. *J. Biol. Chem*., **270,** 13961–13967.

Nakanishi,I., Kimura,K., Kusui,S. and Yamazaki, E. (1974) Complex formation of gel-forming bacterial (1→3)-β-D-glucans (curdlan-type polysaccharides) with dyes in aqueous solution. *Carbohydr. Res*., **32,** 47–52.

Nakanishi,I., Kimura,K., Suzuki,T., Ishikawa,M., Banno,I., Sakare,T. and Harada,T. (1976) Demonstration of curdlan-type polysaccharide and some other β-1,3-glucan in microorganisms with aniline blue. *J. Gen. Appl. Microbiol*., **22,** 1–11.

Okuyama,K., Otsubo,A., Fukazawa,Y., Ozawa,M., Harada,T. and Kasai,N. (1991) Single helical structure of native curdlan and its aggregation state. *J. Carbohydr. Chem*., **10,** 645–656.

Palombo,E.A., Yusoff,K., Stanisich,V.A., Krishnapillai,V. and Willetts,N.S. (1989) Cloning and genetic analysis of *tra* cistrons of the Tra2/Tra3 region of plasmid RP1. *Plasmid*, **22,** 59–69.

Pansegrau,W., Lanka,E., Barth,P.B., Figurski,D. H., Guiney,D.G., Haas,D., Helinski,D.R., Schwab,H., Stanisich,V.A. and Thomas,C.M. (1994) Complete nucleotide sequence of Birmingham IncPα plasmids: compilation and comparative analysis. *J. Mol. Biol*., **239,** 623–663.

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

Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA*, **85,** 2444–2448.

Phillips,K.R. and Lawford,H.G. (1983) Curdlan: its properties and production in batch and continuous fermentations. In Bushell,M.E. (ed.), *Progress in Industrial Microbiology* 18. Elsevier Press, Amsterdam, pp. 201–229.

Ross,P., Mayer,R. and Benziman,M. (1991) Cellulose biosynthesis and function in bacteria. *Microbiol. Rev*., **55,** 35–58.

Saito,H., Ohki,T. and Sasaki,T. (1977) A $^{13}$C nuclear magnetic resonance study of gel-forming (1→3)-β-D-glucans. Evidence of the presence of single-helical conformation in a resilient gel of a curdlan-type polysaccharide 13140 from *Alcaligenes faecalis* var. *myxogenes* IFD 13140. *Biochem*., **16,** 908–914.

Sakamoto,J., Matsuoka,Y., Shindoh,T. and Tangtermsirikuh,S. (1991) An application of superworkable concrete to construction of actual structures. *Trans. Japan Concrete Instit*., **13,** 41–78.

Salvucci,M.E. and Klein,R.R. (1993) Identification of the uridine-binding domain of sucrose-phosphate synthase. *Plant Physiol*., **102,** 529–536.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sanger,F., Nicklen,S. and Coulsen,A.R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, **74,** 5463–5467.

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

Saxena,I.M., and Brown,R.M. (1997) Identification of cellulose synthase(s) in higher plants: sequence analysis of processive β-glycosyl transferases with the common motif "D,D,D35Q(RQ)XRW." *Cellulose*, **4,** 33–49.

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

Saxena,I.M., Lin,F.C. and Brown,R.M. (1991) Identification of a new gene in an operon for cellulose biosynthesis in *Acetobacter xylinum*. *Plant Mol. Biol*., **16,** 947–954.

Saxena,I.M., Kudlincka,K., Okuda,K. and Brown,R.M. (1994) Characterisation of genes in the cellulose-synthesising operon (*acs* operon) of *Acetobacter xylinum*: implications for cellulose crystallisation. *J. Bacteriol*., **176,** 5735–5752.

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

Shine,J. and Dalgarno,L. (1974) The 3′-terminal sequence of *E.coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA*, **71,** 1342–1346.

Simon,R., Priefer,U. and Puhler,A. (1983) A broad host range mobilisation system for *in vivo* genetic engineering: transposon mutagenesis in gram negative bacteria. *Bio/Technology*, **1,** 784–791.

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

Staden,R. (1984) Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Res*., **12,** 505–519.

Stone,B.A. and Clarke,A.E. (1992) *The Chemistry and Biology of (1→3)-β-Glucans*. La Trobe University Press, Australia.

Sutherland,I.W. (1993) Biosynthesis of extracellular polysaccharides (exopolysaccharides). In Whistler,R.L. and BeMiller,J.N. (eds), *Industrial Gums*, 3rd ed. Academic Press, New York, pp. 69–85

Taylor,R.K., Manoil,C. and Mekalanos,J.J. (1989) Broad-host-range vectors for delivery of Tn*phoA*: use in genetic analysis of secreted virulence determinants of *Vibrio cholerae*. *J. Bacteriol*., **171,** 1870–1878.

Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through weighing, position specific gap penalties and weight matrix choice. *Nucleic Acids Res*., **22,** 673–4680.

Varghese,J.N., Garrett,T.P.J., Coleman,P.M., Chen,L., Høj,P.B. and Fincher,G.B. (1994) Three-dimensional structures of two plant β-glucan endohydrolases with distinct substrate specificities. *Proc. Natl. Acad. Sci. USA*, **91,** 2785–2789.

von Heijne,G. (1986) The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J*., **5,** 3021–3027.

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., Brunder,R., Ben-Bassat,A. and Tal,R. (1990) Genetic organisation of the cellulose synthase operon in *Acetobacter xylinum*. *Proc. Natl. Acad. Sci. USA*, **87,** 8130–8134.

Yanisch-Perron,C. Vieira,J. and Messing,J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, **33,** 103–119.

Zorrigueta,A., Geremia,R.A., Cavaignac,S., Cangelosi,G.A., Nester,E.W. and Ugalde,R.A. (1988) Identification of the chromosomal virulence gene. *Mol. Plant Microb. Interact*., **1,** 12–127.