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

Stan J.Stasinopulos 3, Paul R.Fisher 1, Bruce A.Stone and Vilma A.Stanisch 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\(\rightarrow\)3)-\(\beta\)-glucan, curdlan, have been cloned for the first time from *Agrobacterium* sp. ATCC31749. The genes occurred in two, nonoverlapping,rogenic 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\(\rightarrow\)3)-\(\beta\)-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\(\rightarrow\)4)-\(\beta\)-glucans. No similarity was detected with putative (1\(\rightarrow\)3)-\(\beta\)-glucan synthases from yeasts and filamentous fungi. Whatever the determinants of the linkage specificity of these \(\beta\)-glucan synthases might be, these results raise the possibility that (1\(\rightarrow\)3)-\(\beta\)-glucans and (1\(\rightarrow\)4)-\(\beta\)-glucans are formed by related catalytic polypeptides.

**Key words:** *Agrobacterium* sp. ATCC31749/(1\(\rightarrow\)3)-\(\beta\)-glucan/ repetitive \(\beta\)-glycosyl transferases/\(\beta\)-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\(\rightarrow\)4)-\(\beta\)-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\(\rightarrow\)2)-\(\beta\)-glucans and cyclic (1\(\rightarrow\)3, 1\(\rightarrow\)6)-\(\beta\)-glucans (Breedveld and Miller, 1994).

A linear (1\(\rightarrow\)3)-\(\beta\)-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\(\rightarrow\)3)-\(\beta\)-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\(^\circ\)C, a low set gel is produced. In this state, x-ray diffraction (Okuyama et al., 1991) and 13C 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\(\rightarrow\)3)-\(\beta\)-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\(\rightarrow\)3)-\(\beta\)-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 (cellulitin, mycolaminarin, pachymann; Stone and Clarke, 1992). Higher plants also produce a linear (1\(\rightarrow\)3)-\(\beta\)-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\(\rightarrow\)3)-\(\beta\)-glucans such as yeast (Saccharomyces cerevisiae) glucan, pachymann (Portia coccus: Nakamishi 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.xylina* (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\(\rightarrow\)3)-\(\beta\)-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 Crd- mutant of Agrobacterium sp. ATCC31749) after mutagenesis with transposon TnphoA (Manoil and Beckwith, 1985). This was achieved by conjugation experiments in which a suicide plasmid containing TnphoA(Km+) (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 × 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 × 10⁻¹). Moreover, the amount correlated with the staining intensity of the colonies (i.e., <0.2–0.5 g l⁻¹ from white colonies 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 TnphoA inserted in one of nine distinguishable SstI or PvuI 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 distinguishing insertion mutations. The probes were obtained by cloning the Km⁻ containing EcoRI-junction fragments from five mutants (LTU61–LTU66 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 TnphoA (including the HpaI (0.18 kb) site; Figure 1A) allowing the recovery of the Agrobacterium sequences from pVS1500–pVS1504 on HpaI(TnphoA)–EcoRI(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 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 EcoRI-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 EcoRI 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 TnphoA in LTU61–LTU64 were deduced from the restriction analysis and confirmed by Southern analysis in
which EcoRI-digested genomic DNAs were probed individually with P1501–P1504 (data not shown). The map locations of TnophoA 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 PvuII and 8 kb StuI-fragments but different EcoRI-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 TnphoA-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 (GGCAGGTT) (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 De Ley, 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 hydrophy prediction 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 residues 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 De Ley, 1984).

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 TnphoA-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 (GGCAGGTT) (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 De Ley, 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 hydrophy prediction 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 residues 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 De Ley, 1984).

The *crdS* protein was found by a BLASTX search (Ashshul 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* AcS (Saxena et al., 1994), AcSAll, 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 in 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)-β-glucan (data not shown). 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).

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, AcsAll, 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 in 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)-β-glucan (data not shown). 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).

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.
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 (TphoA)-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.xylimum cellulose synthase (AcsAB; Saxena and Brown, 1997) have been implicated in the catalytic mechanism by site-directed mutagenesis studies. In AcsAB, replacement of

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.
the first or second Asp residue impaired \textit{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 \textit{(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 \textit{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,
One proposal is that linkage specificity is determined by subtle (and donor) specificity in these enzymes. The particular issue of which features determine linkage specificity is acquired primarily by substitutions in amino acids lining a deep substrate-binding groove on the enzyme surface. 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 different monomer units. These are the hyaluronan synthase (HasA) from Streptococcus pyogenes (Dougherty and van de Rijn, 1994) and Type 3 capsular polysaccharide synthase (Cps3S) from Pseudomonas aeruginosa (Arrechieta et al., 1995; Dillard et al., 1995). HasA and Cps3S also display the sequence and structural characteristics of repetitive glycosyl transferases (Keenleyside and Whiffin, 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)-α-fucosyl linkages (Legault et al., 1995). The other example is the (1→3)- and (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 are reflected by HCAs (Campbell et al., 1997). 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 I. Characteristic sequence motifs in β-glycan synthases and other β-glycosyl transferases

<table>
<thead>
<tr>
<th>Proteina</th>
<th>UDG b</th>
<th>KACc</th>
<th>Motifd</th>
<th>DDDQRW e</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axy AcSA</td>
<td>DWPEKVR, VHI2DDG</td>
<td>YIARPTNE, ... HAGKNL</td>
<td>IMQTPHPFTS, ...D</td>
<td>DX0DX0DX0QRWR</td>
</tr>
<tr>
<td>Axy BcsA</td>
<td>DWPEKVR, VHI2DDG</td>
<td>YIIRSDQNN, ... HAGKNL</td>
<td>IMQTPHPFTS, ...D</td>
<td>DX0DX0DX0QRWR</td>
</tr>
<tr>
<td>Axy AcsAll</td>
<td>DWPAKDRN, VY12DDG</td>
<td>YITREMNV, ... HAGKNL</td>
<td>IMQTPHPFTS, ...D</td>
<td>DX0DX0DX0QRWR</td>
</tr>
<tr>
<td>Atu CelA</td>
<td>DYPADRT, VML12DDG</td>
<td>YITTTHRE, ... HAGKNL</td>
<td>IMQTPHPFTS, ...D</td>
<td>DX0DX0DX0QRWR</td>
</tr>
<tr>
<td>Eco YhjO</td>
<td>DYPDKQMR, W112DDG</td>
<td>YTVFPRDK, ... HAGKNL</td>
<td>IMQTPHPFTS, ...D</td>
<td>DX0DX0DX0QRWR</td>
</tr>
<tr>
<td>Asp CrdS</td>
<td>AYPVPA, CAV1L12DDG</td>
<td>YVYREKGRQY, ... HAKKAG, ...</td>
<td>YMPPQPRFDG, ...D</td>
<td>DX0DX0DX0QRWR</td>
</tr>
<tr>
<td>Gbi CelA</td>
<td>DYPKVR, V1I12DDG</td>
<td>YYRSSKEKPRGNY, ... HKKAG, ...</td>
<td>YMPPQPRFDG, ...D</td>
<td>DX0DX0DX0QRWR</td>
</tr>
<tr>
<td>Gbi CelA2</td>
<td>DYPKVR, V1I12DDG</td>
<td>YYRSSKEKPRGNY, ... HKKAG, ...</td>
<td>YMPPQPRFDG, ...D</td>
<td>DX0DX0DX0QRWR</td>
</tr>
<tr>
<td>PbCV A473L</td>
<td>DWPKNL, VIV12DDG</td>
<td>YITRPNNH, ... HKKAGNLR</td>
<td>YITPQFPEVF, ...D</td>
<td>DX0DX0DX0QRWR</td>
</tr>
<tr>
<td>Ath T88271</td>
<td>VMPSRKLV, VJVVDSS</td>
<td>-</td>
<td>-</td>
<td>DX0DX0DX0QRWR</td>
</tr>
<tr>
<td>Spy HasA</td>
<td>TYPKLL, ... V1Y12DDG</td>
<td>-</td>
<td>-</td>
<td>DX0DX0DX0QRWR</td>
</tr>
<tr>
<td>Rme NodC</td>
<td>DYPGRER, VVVVDSS</td>
<td>-</td>
<td>-</td>
<td>DX0DX0DX0QRWR</td>
</tr>
<tr>
<td>Sce Chs1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>DX0DX0DX0QRWR</td>
</tr>
</tbody>
</table>

aThe protein name is preceded by an acronym for the name of the organism from which the encoding DNA was derived. AcsA, AcsAll, 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 Pseudomonas 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.
bBoldface text indicates matches to the motif, italics indicate mismatches and roman text indicates redundancies (i.e., X in motif). Also shown are how many sequences were allowed in those 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-galactosyl transferase (Legault et al., 1995; Dougherty and van de Rijn, 1994; 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).
Relevant characteristics of the bacteria (Heinemann 3.2.1.73) from plants (Høj and Fincher, 1995) on the one hand and ancestors and are products of convergent evolution. This is not that the two groups of proteins have evolved from different and CrdS and the bacterial and plant cellulose synthases suggests According to Phillips and Lawford (1983) except that 20 mM KNO₃ replaced NH₄Cl, 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 required, antibiotic supplements (mg/l) in NA were ampicillin (Ap, 100), kanamycin (Km, 15), and tetracycline (Tc, 15); those

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., 1997), Candida albicans (Mio et al., 1994; Inoue 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

### 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 KNO₃ replaced NH₄Cl, 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 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 × 10⁹/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 × 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 × 10⁸ 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 TnphoA and is mobilizable from E.coli SM10 (Taylor et al., 1989). To ensure the isolation of non-sibling mutants, separate filter matrices between E.coli SM10 carrying pRT733 and LTU50 were performed and the transconjugants selected on ABA medium containing Km and Km. transconjugants confirmed by the Km and Km phenotype TnphoA A characteristic was not utilized. Colonies that failed to stain with aniline blue were purified and the inheritance of TnphoA and loss of pRT733 was 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).

**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 BglII fragment of TnphoA (Figure 1A) or the HpaI(TnphoA-EcoRI/Agrobacterium) junction fragments of pVS1500–1505 (Figure 1B) and labeled with [α-32P]dCTP using the Megaprime DNA labeling kit (Amersham). Hybridization to DNA on nitrocellulose filters was in 6x SSPE, 5x 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 2x SSC, 0.5% SDS and once (30 min at 65°C) in 0.2x 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 overnight in NB. The procedure of Ma and Henrissat (1989) for the detection of low abundance sequences. Probes were prepared from the 2.8 kb BglII fragment of TnphoA (Figure 1A) or the HpaI(TnphoA-EcoRI/Agrobacterium) junction fragments of pVS1500–1505 (Figure 1B) and labeled with [α-32P]dCTP using the Megaprime DNA labeling kit (Amersham). Hybridization to DNA on nitrocellulose filters was in 6x SSPE, 5x 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 2x SSC, 0.5% SDS and once (30 min at 65°C) in 0.2x 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 EcoRI and size-fractionated in a linear 10–40% sucrose gradient. Fragments in the size range of 15–35 kb were ligated to pLAFR1(Tc) that had been digested with EcoRI 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 cdS was determined using a TnphoA-based primer (5′-CATGAGCTGTTTCCATGTTAG-3′) to sequence the Agrobacterium DNA adjoining TnphoA 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).


A (1→3)-β-glucan synthase gene from an Agrobacterium sp.