Topological characterization of an inner membrane (1→3)-β-D-glucan (curdlan) synthase from Agrobacterium sp. strain ATCC31749

Tara Karnezis², V. Chandana Epa³, Bruce A. Stone², and Vilma A. Stanisich¹,4

¹Department of Biochemistry, La Trobe University, Melbourne, Victoria 3086, Australia; ²CSIRO Division of Health Sciences and Nutrition, 343 Royal Parade, Parkville, Victoria 3052, Australia; and ³Department of Microbiology, La Trobe University, Melbourne, Victoria 3086, Australia

Received on April 13, 2003; revised on June 5, 2003; accepted on June 10, 2003

The crdS gene of Agrobacterium sp. strain ATCC31749 encodes the curdlan synthase (CrdS) protein based on the homology of the derived CrdS protein sequence with those of β-glycosyl transferases with repetitive action patterns (Stasinopoulos et al. [1999] Glycobio, 9, 31–41). Here we show that chemical (NTG) mutagenesis of crdS abolishes curdlan production and the induced mutations can be complemented by a cloned crdS amplicon, thus providing genetic confirmation that crdS is essential for curdlan production. When expressed in the native Agrobacterium or in Escherichia coli, the largely hydrophobic CrdS protein exhibited an Mₐ of ~60 kDa (compared with the predicted mass of 73,121 Da) and was located in the inner membrane of both bacteria. By analyzing reciprocal fusions between crdS and the reporter genes, lacZ and phoA, and assessing the sensitivity of CrdS in spheroplasts to proteinase K, CrdS was shown to be an integral membrane protein with seven transmembrane helices and an N_mar-C_in disposition. A central large and relatively hydrophilic cytoplasmic region carries the substrate-binding and catalytic D,D,D35QxxRW motif. The amino acid sequence of this domain of CrdS was threaded onto the 3D structure of the comparable domain of the SpsA protein, a member of the GT-2 glycosyl transferases (Coutinho and Henrissat, 1999), that encodes a membrane-bound phosphatidylinerine synthase responsible for the formation of phosphatidylserine, the precursor of the quantitatively important membrane phospholipid phosphatidylethanolamine, is also necessary for maximal high-molecular-mass curdlan production (Karnezis et al., 2002). Homologs of each of these genes occur in the genome of Agrobacterium tumefaciens C58 (Wood et al., 2001) where the crdASC gene cluster is located on the linear chromosome and the crd gene cluster is located on the circular chromosome.

The various crd genes were originally detected by analysis of random TnphoA insertion mutants of the Agrobacterium strain that failed to stain blue on agar containing the (1→3)-β-D-glucan-specific dye aniline blue (Stasinopoulos et al., 1999). The essential nature of the crdS gene is thus based on the abolition of curdlan production following insertional inactivation of this gene. This assumption awaits verification because the organization of crdASC in an operon raises the possibility that the curdlan-deficiency of crdS::TnphoA mutants is due to polarity of the mutations on the downstream gene, crdC, rather than to inactivation of crdS.

Here we report experiments that confirm that crdS is essential for curdlan production and that the curdlan synthase (CrdS) protein is located in the inner membrane of the bacterial envelope and may be part of a multimeric complex. The topology of CrdS is also investigated, revealing the presence of seven transmembrane helices and a large cytoplasmic domain bearing the substrate-binding and catalytic motifs. Components of the conserved D,D,D35QxxRW motif (Saxena and Brown, 1997) that have been implicated in the catalytic activity of family GT-2 glycosyl transferases are evaluated for their position in a 3D model of the catalytic domain of CrdS based on the crystallographic structure of SpsA, a soluble glycosyl transferase belonging to the GT-2 family.
Results

The 1965-bp crdS gene is essential for curdlan production

A number of observations made during this research suggested that the crdS nucleotide sequence reported by Stasinopoulos et al. (1999) was not completely accurate. We reevaluated the region containing crdS and detected an open reading frame (ORF) of 1,965 bp that commences at the ATG start codon reported for crdS but is longer by 342 bp. The deduced CrdS protein sequence (654 amino acid residues; 73,121 Da) shares significant homology with β-glycan synthases in family GT-2, as previously described (Stasinopoulos et al., 1999), validating the conclusion that CrdS is the curdlan synthase protein. CrdS is predicted to be an integral membrane protein with seven or eight trans-membrane helices (Figure 1).

To demonstrate experimentally that the 1965-bp crdS sequence encodes a membrane protein, an ampiclon of the crdS ORF was cloned downstream of the strong ribosome-binding site in pET23(a). The plasmid was named pVS1575. A similarly constructed plasmid, pVS1574, contained the 1.64-kb 5'-portion of crdS (i.e., crdSA) encompassing the crdS sequence described by Stasinopoulos et al. (1999). Expression of the crdS component of each plasmid in E. coli yielded a protein of larger size from pVS1575 than from pVS1574 (Figure 2A). Moreover, when lysates of bacteria carrying pVS1575 were fractionated prior to sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE), the larger protein was detected in the membrane fraction but not in the soluble fraction (Figure 2B). These data indicated that the crdS gene is longer than initially reported by Stasinopoulos et al. (1999) and encodes a membrane protein. The discrepancy between the measured and predicted sizes of CrdS (M_r ~60 versus ~73 kDa) and of the truncated derivative (M_r ~50 versus 61 kDa) is attributed to the overall hydrophobicity of the proteins (Figure 1), a feature that can increase the protein mobility in SDS–PAGE (Sarsero and Pittard, 1995; Debarbieux et al., 1997; Tlapak-Simmons et al., 1998; Alexeyev and Winkler, 1999).

To establish experimentally that the 1965-bp crdS gene is required for curdlan production, the crdS ORF was cloned into the broad host range vector pBBR1MCS-3. This enabled the resulting plasmid (pVS1570) to be conjugally transferred from E. coli S17-1 to a collection of chemically induced (with N-methyl-N'-nitro-N-nitrosoguanidine, NTG), locus I, curdlan-deficient (Crd'-) mutants of the wild type Agrobacterium (LTU50) (i.e., those with mutations that were complemented by pVS1506). Such a transfer resulted in the restoration of curdlan production in four of the mutants, indicating complementation of the respective mutations by the crdS gene of pVS1570. In contrast, complementation failed to occur in parallel experiments using pVS1572, a derivative of pVS1570 that contains only the 820 bp 5'-portion of crdS. Together these findings provide genetic evidence that crdS is essential for curdlan production.

CrdS is an inner-membrane protein that occurs in protein complexes

To further investigate the subcellular location of the CrdS protein, polyclonal antibodies were raised to a

![Fig. 1. Hydropathy profile of CrdS. The plot (Kyte and Doolittle, 1982) is for the derived 654 amino acid residues of CrdS, using a window of 17 residues. Eight hydrophobic stretches, denoted by Roman numerals, are putative TM helices based on the prediction programs of Klein et al. (1985), von Heijne (1992), Hofmann and Stoffel (1993) and Cserzo et al. (1997). The bracket indicates the location of the region containing the conserved motifs shared by CrdS (amino acid residues 97–375) and family GT-2 β-glucan synthases.](image)

![Fig. 2. Expression of crdS in E. coli and subcellular location of CrdS. (A) Autoradiogram of proteins in lysates prepared from [35S]-methionine/cysteine-labeled E. coli BL21(DE3)pLysS carrying pVS1575(crdS'), pVS1574(crdSA), or the pET23(a) vector, and resolved by SDS–PAGE (7.5–12%). (B) Autoradiogram of proteins in the soluble (S) and membrane (M) fractions prepared from E. coli strain carrying pVS1575(crdS') or the pET23(a) vector. (C) Upper panel: Analysis of proteins in the inner (IM), outer (OM), and total membrane (TM) fractions prepared from E. coli strain carrying pVS1575(crdS'). The proteins were resolved by SDS–PAGE (10%) and stained with Coomassie blue. Lower panel: Western immunoblot of the resolved proteins showing bands that reacted with GST-ΔCrdS antibodies.](image)
glutathione-S-transferase (GST)-fusion protein that contains an internal portion of CrdS (amino acid residues 84–362) at its C-terminus (see Materials and methods). This portion of CrdS (designated ΔCrdS) was chosen because it is relatively hydrophilic (Figure 1) and encompasses a domain of CrdS that shares greatest homology (up to 46% similarity) with the sequences of glycan synthases in family GT-2 with a repetitive action pattern. The GST-ΔCrdS antibodies were used to determine the subcellular location of CrdS in the native Agrobacterium strain. Periplasmic, cytoplasmic, and membrane fractions were prepared from disrupted cells of LTU50 and examined by SDS-PAGE and western analysis (Figure 3A and B). The identity of each fraction was verified using antibodies to representative periplasmic (maltose-binding protein), cytoplasmic (SecA protein A), and membrane (signal peptidase) proteins (Figure 3C). When GST-ΔCrdS antibodies were used, an immunoreactive protein band of $M_r$ ~60 kDa was found only in the membrane fraction (Figure 3B). We conclude that this protein, of the same size and cellular location as that expressed in crdS$^+$ E. coli (Figure 2B), is CrdS.

To determine whether CrdS is located in the inner or outer membrane of Agrobacterium, total membranes from LTU50 were separated in a continuous sucrose gradient on the basis of their different densities. Figure 4A shows the absorbance profile of the gradient fractions that revealed two major bands, whose peak fractions had buoyant densities of 1.16 and 1.25 g/cm$^3$. These values are in the ranges observed for the inner and outer membranes of A. tumefaciens (Thorstenson et al., 1993) and the related bacterium Rhizobium leguminosarum (de Maagd and Lugtenberg, 1986). Additionally, the distribution of the outer membrane marker, 2-ketodeoxyoctonate (KDO), and the inner membrane marker, NADH oxidase (NOX), indicated efficient separation of the two membranes (Figure 4A), as did the different protein compositions of these membranes (fractions 2–5 versus 15–17; Figure 4B).

Western analysis using GST-ΔCrdS antibodies (Figure 4C) detected an immunoreactive protein of $M_r$ ~60 kDa, concluded to be CrdS, that occurred only in the inner membrane fractions. Western analysis was also performed using separated inner and outer membranes from E. coli carrying pVS1575 (crdS$^+$) and, again, a protein of $M_r$ ~60 kDa was found exclusively in the inner membrane fraction (Figure 2C). Thus the inner membrane location of CrdS is not compromised when crdS is the sole crd gene expressed in E. coli, an unrelated, Gram-negative bacterium.

The total membrane fractions prepared from Agrobacterium LTU50 were further analysed using blue native (BN)-PAGE (Schägger et al., 1994), a procedure that has been used to identify native membrane protein complexes in mammalian cells (e.g., Dekker et al., 1998) and more recently in Gram-negative bacteria (Bailey et al., 1999; Aufurth et al., 2000). Membrane proteins from LTU50 solubilized in 1% (w/v) digitonin and subjected to BN-PAGE formed a number of prominent bands of $M_r$ from ~550 kDa to ~80 kDa (Figure 5A). Western analysis using GST-ΔCrdS antibodies detected immunoreactive protein

![Fig. 3. Subcellular location of CrdS in Agrobacterium LTU50.](https://academic.oup.com/glycob/article-abstract/13/10/693/554324)

**A** Analysis of proteins in the periplasmic (P), cytoplasmic (C), and membrane (M) fractions prepared from the wild-type Agrobacterium LTU50. The proteins were resolved by SDS-PAGE (10%) and stained with Coomassie blue. **B** Western immunoblot of the resolved proteins showing a band that reacted with GST-ΔCrdS antibodies. **C** Western immunoblot of the resolved proteins showing bands that reacted with antibodies to periplasmic (maltose-binding protein), cytoplasmic (SecA secretory protein), and membrane (signal peptidase) protein markers.

![Fig. 4. Detection of CrdS protein in the inner-membrane fraction of Agrobacterium LTU50.](https://academic.oup.com/glycob/article-abstract/13/10/693/554324)

**A** Isopycnic centrifugation in a sucrose density gradient of cell membranes prepared from wild-type Agrobacterium LTU50. The graphs show sucrose density (circles) and protein content (diamonds). Fractions (0.5 ml) containing the outer and inner membranes were identified from the distribution of the outer-membrane component KDO (triangles) and the inner-membrane NOX activity (squares). **B** Analysis of the protein contents of fractions 2–5 (containing the inner membrane) and fractions 15–17 (containing the outer membrane), using SDS-PAGE (4–12%) and staining with Coomassie blue. **C** Western immunoblot of the resolved proteins using GST-ΔCrdS antibodies.
bands at $M_r \sim 420$ kDa and $M_r \sim 500$ kDa, compared with a $M_r \sim 60$ kDa for the CrdS polypeptide (Figure 3B) indicative of protein complexes.

Hydrophobicity analysis of the CrdS sequence predicts seven or eight transmembrane helices

Hydropathy analysis of the deduced CrdS sequence using the ALOM (Klein et al., 1985), TMPred (Hofmann and Stoffel, 1993), TopPredII (von Heijne, 1992), and DAS (Cserzo et al., 1997) prediction programs revealed eight potential transmembrane (TM) helices (Figure 1). Three (TMI to -III) are located in the N-terminal portion of CrdS and are separated from five (TMIV to -VIII) in the C-terminal region end by a large (278 amino acid residues), more hydrophilic region between TMIII and TMIV. However a comparison of the average hydrophobicity of each TM showed that TMVI (amino acids 434±457) has the lowest score (800 versus 1472±2134; Hofmann and Stoffel, 1993) so that its designation as a bona fide TM helix is ambiguous. TMVI may, instead, be a membrane-associated helix that is partially buried in the membrane, such as has been noted in another polytopic membrane protein, the hyaluronan synthase of *Streptococcus pyogenes* (Heldermon et al., 2001). Based on these predictions, a number of membrane topologies for CrdS are possible, two of which are shown in Figure 6. In these two models, the membrane topology is identical over the N-terminal portion of CrdS but differs in the C-terminal portion depending on whether the putative TMVI helix does (model A) or does not (model B) span the membrane. Two experimental approaches were taken to determine the topology of CrdS in the inner membrane.

The large hydrophilic region of CrdS is intracellular

The large hydrophilic region of CrdS between TMIII–IV (amino acid residues 97–375) contains residues and motifs that are shared by family GT-2 glycosyl transferases with repetitive action patterns. Based on site-directed mutagenesis studies, this region constitutes the active site of these enzymes (Nagahashi et al., 1995; Saxena and Brown, 1997; Cos et al., 1998; Garinot-Schneider et al., 2000; Saxena et al., 2001; Ruiz-Herrara et al., 2002). Furthermore, it is located in the cytoplasm as revealed by experimental topological studies in two instances (Barny et al., 1996; Heldermon et al., 2001). We investigated the cellular disposition of the large hydrophilic region of CrdS by assessing its *in vivo* susceptibility to proteinase K. In this procedure it is assumed that TM helices buried in the membrane and regions that reside in the cytoplasm are inaccessible to proteolysis in intact but not lysed spheroplasts.

When *E. coli* carrying pVS1575(crdS<sup>−</sup>) were induced with isopropyl-D-thiogalactopyranoside (IPTG) and converted to spheroplasts, they produced the full-length CrdS protein ($M_r \sim 60$ kDa) detectable by western analysis (Figure 7,}

---

**Fig. 5.** Analysis of *Agrobacterium* LTU50 total membranes using BN-PAGE. (A) Coomassie blue-stained gel (6–15%) of digitonin-solubilized membrane proteins. (B) Western immunoblot of the resolved proteins using GST-A-crds antibodies.

**Fig. 6.** Alternative topological models for CrdS. (A, B) Two possible topological models predicted by TMPred (Hofmann and Stoffel, 1993) and other programs.

**Fig. 7.** Susceptibility of CrdS in spheroplasts of *E. coli* to digestion by proteinase K. Western analysis of proteins prepared from either intact or lysed spheroplasts of *E. coli* BL21(DE3)(pLysS) carrying pVS1575(crdS<sup>−</sup>) treated with proteinase K and separated by SDS-PAGE (10%): Lane 1, intact spheroplasts; lanes 2 and 3, intact spheroplasts with proteinase K added to 200 μg/ml and 400 μg/ml, respectively; lane 4, lysed spheroplasts with proteinase K added to 200 μg/ml.
lane 1). This protein was susceptible to proteinase K digestion following ultrasonic lysis of the spheroplasts (lane 4). In contrast, intact spheroplasts that were treated with proteinase K produced a truncated CrdS protein of Mr ~29 kDa (lanes 2 and 3). The mass of this protein is in accord with that predicted for the peptide encompassing TMIII–IV (residues 77–399; 36.4 kDa) when the typically higher mobility of wild-type and truncated CrdS proteins (Figure 2A) is taken into account. It was concluded that the failure of proteinase K to digest the large hydrophilic region of CrdS is due to its intracellular location, confirming a significant feature of the predicted topology of the CrdS protein that is common to both models A and B (Figure 6).

CrdS-LacZ and CrdS-PhoA fusion analysis supports a CrdS topology with seven TM helices

The topology of CrdS was studied in E. coli using protein fusions to alkaline phosphatase (PhoA) or to β-galactosidase (LacZ). These two reporter enzymes have been used widely as sensors of the subcellular location of portions of membrane proteins (Silhavy and Beckwith, 1985; Manoil, 1990). Their utility arises from the general finding that PhoA lacking its own export signal is active when fused to a periplasmic domain of a membrane protein, whereas LacZ is active when fused to a cytoplasmic domain. Using polymerase chain reaction (PCR)–mediated approaches, we generated PhoA fusion proteins to 21 different N-terminal fragments of CrdS and, in 18 of these cases, also produced reciprocal fusions by substituting the PhoA moiety with LacZ (see Materials and methods and Table I). The subcellular locations of the LacZ and PhoA reporters (i.e., cytoplasmic or periplasmic, respectively) was assessed quantitatively by enzymic assays of permeabilized E. coli CC118 carrying the respective gene fusion on a plasmid vector (pVS1549). The data obtained are summarized in Table I. A broad range of activities was observed (9–1320 U for PhoA and 17–356 U for LacZ), and the validity of the strategy was confirmed by the clear reciprocal responses obtained for most of the fusion pairs studied (i.e., normalized activity ratios of ~1:10 to 1:100; Table I).

Over the N-terminal segment of CrdS, strong reciprocal enzymic activities were observed for eight of the ten fusions located between residues R38 and T438 (Figure 8). This distribution of sites was designed to test the disposition of CrdS over the topological region that is common to both models, that is, the region from TMI–V (Figure 6). The

| Table I. Fusion sites and activities of CrdS-PhoA and CrdS-LacZ hybrid proteins |
|-----------------|-----------------|-----------------|-----------------|
| Plasmid(pVS . . .) | crdS-phoA | crdS- lacZ | Predicted location of reportera Deducetion in CrdS |
| Fusion site | | Model A Model B | Normalized units |
| | | | PhoA LacZ NAR b |
| 1579 | — | L17 | TMI TMI | 5.3 — 1:32 C |
| 1580 | 1600 | R38 | C C | 1.9 59.5 1:32 C |
| 1581 | — | Y52 | TMII TMII | 24.5 — — P |
| 1582 | — | V54 | P P | 30.4 — — P |
| 1583 | 1601 | L68 | P P | 100 4.8 21:1 P |
| 1584 | 1602 | T75 | P P | 19.6 84.8 1:4 P? |
| 1585 | 1603 | L119 | C C | 0.9 53.6 1:60 C |
| 1586 | 1604 | L191 | C C | 1.1 56.1 1:51 C |
| 1587 | 1605 | I245 | C C | 0.8 77.5 1:103 C |
| 1588 | 1606 | E323 | C C | 0.9 51.9 1:58 C |
| 1589 | 1607 | L408 | P P | 30.0 55.3 1:2 P? |
| 1590 | 1608 | K412 | TMV TMV | 36.4 8.4 4:1 P |
| 1591 | 1609 | T438 | C C | 6.6 62.6 1:10 C |
| 1592 | 1610 | T450 | TMVI C/MA | 10.7 42.9 1:4 C/MA |
| 1593 | 1611 | K466 | P C | 0.6 100 1:167 C |
| 1594 | 1612 | V477 | P C | 0.9 62.0 1:69 C |
| 1595 | 1613 | D506 | C P | 4.2 25.8 1:6 ? |
| 1596 | 1614 | R514 | C P | 5.7 28.6 1:5 ? |
| 1597 | 1615 | S558 | P C | 22.1 83.7 1:4 C |
| 1598 | 1616 | A578 | P C | 1.7 39.9 1:24 C |
| 1599 | 1617 | K630 | P C | 1.3 42.9 1:33 C |

aC, cytoplasm; P, periplasm; TM, transmembrane segment; MA, membrane-associated segment; ?, unassigned location.
bNAR, the normalized activity ratio, was calculated as the ratio of PhoA:LacZ activities. The highest and lowest values for PhoA and LacZ activities were 1320 and 356 units, respectively, which were normalized to 100 to calculate the NAR values.
Enzymic activities of the strong fusions at R38 (high LacZ) and L68 (high PhoA) suggested, respectively, that CrdS is cytoplasmic and periplasmic at these positions; this pattern was true also of the fusions at T438 (high LacZ) and K412 (high PhoA). In each of these two situations, the fusion sites are separated by ~30 residues so that their different enzymic responses suggest that the fusion sites occur on either side of a TM helix, namely, TMII and TMV respectively, according to the models (Figure 6). All four of the remaining strong reciprocal fusions (from L119 to E323) suggested that CrdS is cytoplasmic at these fusion sites.

These various findings, together with the observation that the peptide from residues 77–399 is insensitive to proteinase K treatment (Figure 7), support a model in which the N-terminus of CrdS is periplasmic and is separated from a large cytoplasmic region by three TM helices (TMI–III) on the N-terminal side and two TM helices (TMIV and -V) on the C-terminal side. This topology thus conforms with predictions from both models (Figure 6) and is further supported by the high PhoA activities observed for the single fusions at Y52 and V54. The two exceptional reciprocal fusions (T75 and L408) yielded high PhoA activities, also in accord with the models, but anomalously high LacZ activities. The latter responses might be attributable to a subpopulation of proteins that formed tetramers of the LacZ moiety prior to membrane insertion to produce an

---

**Fig. 8.** Deduced topology of CrdS. The locations of CrdS-PhoA and CrdS-LacZ replacement fusions in the CrdS polypeptide are indicated on the two topological models A and B. The alkaline phosphatase activity of each PhoA fusion is indicated by a box, as follows: filled, >90 units; hatched, 50–90 units; open, <50 units. The β-galactosidase activity of each LacZ fusion is indicated by an oval, as follows: filled, >90 units; hatched, 50–90 units; open, <50 units. The locations of TM helices predicted by TMPred (Hofmann and Stoffel, 1993) are as follows: TMI (amino acids 14–30), TMII (amino acids 35–55), TMIII (amino acids 77–96), TMIV (amino acids 376–399), TMV (amino acids 408–428), TMVI (amino acids 434–457), TMVII (amino acids 484–505), and TMVIII (amino acids 519–537).
active β-galactosidase (Fulkerson and Mobley, 2000; van Geest and Lolkema, 2000).

Over the C-terminus of CrdS, four of the eight fusion pairs located between residues T450 and K630 yielded strong reciprocal responses. These included two fusions (A578 and K630) close to the C-terminus that exhibited high LacZ activity, suggesting that this region of CrdS is cytoplasmic. This finding is in keeping with model B (Figure 6). Model B is also consistent with the high LacZ activities of the remaining two strong reciprocal fusions at K466 and V477. In the alternative model A, fusions at all four of these sites would be expected to display high PhoA activity, not high LacZ activity. In contrast to the strong reciprocity of the fusions at A578 and K630, those at position S558 in the C-terminal region exhibited high LacZ activity and also anomalously high PhoA activity. Boyd and Beckwith (1989) have shown that loss of positive charged residues can lead to loss of topological information. Positively charged residues are especially concentrated in the region after position S558 and may have led to the retention of the CrdS moiety with the fused PhoA in the cytoplasm.

Of the remaining reciprocal fusions, those at D506 and R514 were unusual in that they exhibited intermediate PhoA activities (~60 U) rather than the high levels predicted by model B, and significant LacZ activities (~49 U). Similar phenomena have been attributed (Calamia and Manoil, 1992; van Geest and Lolkema, 2000) to a charged residue, arginine, in the case of CrdS, in the hydrophobic stretch that precedes these fusion sites, destabilizing the formation of the TM helix TMVII. This would lead, on one hand, to an inefficient export of the PhoA moiety into the periplasm and, on the other hand, to significant retention of the LacZ moiety in the cytoplasm. Finally, the fusion pair at T450 is of interest as it occurs in the middle of a hydrophobic region that, according to model B, does not constitute a bona fide TM helix (Figure 6). It has been shown (Bibi and Béja, 1994; Heldermon et al., 2001) that fusions positioned after a start-transfer segment for a fragment that inserts in the membrane without traversing it can produce fusion proteins that orient the PhoA moiety extracellularly. This would account for the anomalously high PhoA activity associated with the T450 fusion, a feature that is in marked contrast with the low PhoA activities exhibited by the more distal K466 and V477 fusions, both of which are located in a strongly hydrophilic region of the same segment. Overall these studies suggest that CrdS has an N_out–C_in topology with seven TM helices, as suggested by model B.

Model of the donor-binding region of CrdS

The three-dimensional structure of only one nucleotide monosaccharide-dependent glycosyltransferase currently classified in the GT-2 family has been solved. This is SpsA, a glycosyltransferase from Bacillus subtilis (Charnock and Davies, 1999). Although no substrate or product has yet been specified for SpsA, its crystal structure has been solved both with and without bound UDP-Mn²⁺ and UDP-Mg²⁺. Thus the active site structure of SpsA may serve as a prototype for the organization of the UDP-binding site in other family GT-2 inverting synthases, including CrdS. The programs Threader v.3 (Jones et al., 1992) and Modeler 6v2 (Sali and Blundell, 1993) were used to build a 3D model of a 241-amino-acid portion of the CrdS sequence. This portion of CrdS (i.e., amino acids 115–356), located between TMIII and -IV, contains the conserved UDP-binding motif and the conserved D,D,D₃QxxRW motif of the GT-2 glycosyltransferases with repetitive action (Saxena et al., 1995; Keenleyside and Whitfield, 1996; Saxena and Brown, 1997). Figure 9A shows this model in a ribbon form with some of the key active-site amino acid residues depicted in stick form. The predicted secondary structure (using PSIPRED, available online at http://bioinf.cs.ucl.ac.uk/psipred) for CrdS and the experimental secondary structure of SpsA are more similar in the N-terminal region (i.e., the donor region) than in the C-terminal region (i.e., the acceptor region), and hence the topology of the N-terminal region of CrdS could reliably be expected to be as depicted in the model. This was confirmed by ProsaII (Sippl, 1993), a program that estimates the compatibility of the sequence with a given fold. The unreliability of the C-terminal topology was also indicated by ProsaII. This was not unexpected because the C-terminal region confers acceptor specificity and also carries the QxxRW motif believed to be involved in action of repetitive GT-2 glycosyltransferases and is absent from SpsA.

The modeled structure of CrdS also shows the presence of a central cavity occupied by the conserved aspartic acid residues (Asp156, 208, and 304 identified with D¹, D², and D³ in the D¹,D²,D³₃QxxRW motif) that have been implicated in substrate binding (Charnock and Davies, 1999) and/or catalysis (Nagahashi et al., 1995; Saxena et al., 1995, 2001; Cos et al., 1998; Garnot-Schneider et al., 2000). In the model this cavity is spanned by a loop region and has a predominantly negative electrostatic potential, as was seen in the SpsA structure. Furthermore, in SpsA, Asp98 is involved in hydrogen bonding to the ribose of UDP, whereas in the CrdS model, the equivalent amino acid is Asp208. The carboxylate oxygens of Asp208 interact with the O3' atom of the ribose in UDP when the UDP moiety from SpsA is directly superimposed on to this CrdS model. In SpsA, Tyr11 is positioned to allow effective stacking of the uracil moiety of UDP in the binding site region. The equivalent residue in the CrdS model is Tyr127.

A noteworthy component in many glycosyltransferase families, including GT-2, is a DxD sequence that incorporates the D² of the D,D,D₃QxxRW motif and has been implicated in metal binding (Wiggins and Munro, 1998). In SpsA, Asp99 sits adjacent to the distal phosphate of UDP in the binding site and coordinates with the leaving group divalent metal (Mn²⁺), providing support for the conjectured role of this aspartate in metal binding (Charnock and Davies, 1999). In the CrdS model, the equivalent amino acid is Asp210. Lys183 from the conserved HAKAGNLNN motif in CrdS (see Discussion) is positioned to interact with UDP in the model. Charnock and Davies (1999) speculate that Asp191 of SpsA may be the catalytic base in the proposed inverting enzyme mechanism. The equivalent in CrdS is the conserved Asp304. The model does not shed light on the possible function of the conserved QxxRW motif, because this is located in the (unreliable) C-terminal.
region. The positions of the amino acids identified as contacts with the UDP portion of the substrate are indicated by arrows on the sequence of CrdS in Figure 9B.

Discussion

Here we demonstrate that crdS (1965 bp) is essential for curdlan production by showing that a cloned crdS amplicon (2010 bp) restored this property in chemically (NTG) induced, curdlan-deficient mutants, whereas a 3'-truncated crdS gene did not. This also eliminates the formal possibility that curdlan deficiency of crdS::TnphoA mutants studied by Stasinopoulos et al. (1999) was due to polarity on the downstream crdC gene. We have not been able to test crdS mutants biochemically because the membranes of Agrobacterium do not support the synthesis of curdlan from UDPGlc in vitro.

Hydropathy analysis predicts that both prokaryotic and eukaryotic polysaccharide synthases of family GT-2 are membrane proteins (see Barny et al., 1996; Kawagoe and Delmer, 1998). We have established immunochemically that CrdS is an inner membrane protein of Agrobacterium and of E. coli when expressed in this host (Figures 2 and 4).
Membrane locations have also been established experimentally for synthases for hyaluronan from Group A streptococci (Stoolmiller and Dorfman, 1969), including \textit{S. pyogenes} (Heldermon et al., 2001), and for \textit{S. pneumoniae} III polysaccharide (Smith et al., 1960; Cartee et al., 2000) and inner membrane locations demonstrated for NodC chito-oligosaccharide synthase from \textit{Rhizobium leguminosarum} (Barny and Downie, 1993) and for the \textit{E. coli} glucosyltransferase (MdoH) for periplasmic glucans (Debarbieux et al., 1997).

The membrane topology of a number of family GT-2 bacterial polysaccharide synthases have been predicted (see Barny et al., 1996; Kawagoe and Delmer, 1998), but experimentally determined membrane topologies are limited to the \textit{R. leguminosarum} NodC (Barny et al., 1996), \textit{E. coli} MdoH (Debarbieux et al., 1997), and HasA hyaluronan synthase from \textit{S. pyogenes} (Heldermon et al., 2001). Some information on the topology of the large hydrophilic region of \textit{Acetobacter xylinus} cellulose synthase, AcsA, has also been obtained experimentally (Saxena et al., 1995). From our fusion analyses we concluded that CrdS has seven membrane-spanning helices rather than the eight predicted by model A, and an N\textsubscript{out}--C\textsubscript{in} topology (Figure 6). The large hydrophilic region is oriented to the cytoplasm, a conclusion reinforced by experiments showing that a polypeptide encompassing this region remained intact when \textit{E. coli} spheroplasts were treated with proteinase K (Figure 7). The N\textsubscript{out}--C\textsubscript{in} topology led to the prediction that a portion of CrdS in the C-terminal region occurred as a nonmembrane-spanning amphipathic helix (Figure 8). Such a feature also occurs in the HasA protein from \textit{S. pyogenes} (Heldermon et al., 2001), but its functional significance is not understood. The two other GT-2 enzymes whose membrane topologies have been investigated, NodC (Barny et al., 1996) and HasA (Heldermon et al., 2001), also have a large cytoplasmic region. In CrdS, NodC, and HasA, this region carries the D,D,D\textsubscript{3}QxxRW motif characteristic of family GT-2 transferases with repetitive action patterns (Saxena et al., 1995; Saxena and Brown, 1997; Keenleyside and Whitfield, 1996). All residues of this motif are required for the catalytic activity based on site-directed mutagenesis of \textit{Saccharomyces cerevisiae} (Nagahashi et al., 1995; Cos et al., 1998; Ruiz-Herrera et al., 2002) and \textit{A. xylinus} cellulose synthase (Saxena and Brown, 1997; Saxena et al., 2001).

Sequence alignment of the CrdS cytoplasmic region with bacterial cellulose synthase sequences shows that the three Asp residues in the D\textsubscript{2}D\textsubscript{2}D\textsubscript{3}QRXRW motif occur in DD\textsubscript{2}G/T, D\textsubscript{2}C/AD, and T/SED\textsubscript{2} submotifs, respectively, and that there are additional common motifs, namely, YX\textsubscript{2}RX\textsubscript{1}HAKAGNL/IN, QTPTH/Q (Stasinopoulos et al., 1999), and PXVD\textsubscript{2}YxeE (Römling, 2002) (Figure 9B). All of these motifs also occur in \textit{Bradyrhizobium japonicum} (1\textarrow{}3)-\beta-glucan synthase (NdvB; accession number AF047687) involved in cyclic (1\textarrow{}3,1\textarrow{}6)-\beta-glucan production (Bhagwat and Keister, 1995, Bhagwat et al., 1996) and in a \textit{B. japonicum} hypothetical protein (B112448) (Kaneko et al., 2003). The hypothetical protein, in fact, has higher homology with CrdS than does NdvB (i.e., 40% sequence identity and 54% similarity over \textasciitilde{}300 residues versus 26% identity and 44% similarity over \textasciitilde{}300 residues, both including the cytoplasmic region). Two other motifs, FFCGS and RX\textsubscript{2}FLX\textsubscript{2}PL, found in the bacterial cellulose synthases (Römling, 2002) are poorly conserved in CrdS (Figure 9B), NdvB, and the hypothetical protein, giving support to the suggestion by Römling (2002) that they may have a role in the determination of (1\textarrow{}4)-linkage specificity.

A model of the donor-binding domain of the CrdS protein was constructed by threading the sequence of the CrdS protein containing the UDP-binding site motif onto the SpsA structure. Amino acids in CrdS involved in binding UDP were predicted to correspond to those in SpsA. The topology of the UDP binding region found in the crystal structure of SpsA and deduced from the model of CrdS is also a feature of glycosyltransferases from families GT-7, -13, and -43, whose crystallographic structures have been determined (Tarbouriech et al., 2001).

The preliminary evidence from BN-PAGE experiments suggests that CrdS is probably a component of a large complex that may be a homomultimer or a heteromultimer, perhaps involving CrdA and/or CrdC. These two proteins show no homology with transferases, hydrolases, nucleotidase monosaccharide-metabolizing proteins, or transport proteins that have been found associated with other polysaccharide synthases, for example, for polysialic acid (Bliss et al., 1996) and the \textit{G. xylinus} polysaccharide synthase (Barny et al., 1996, 2000), or for cellulose from \textit{A. tumefaciens} (Matthysse et al., 1995) and \textit{A. xylinus} (Saxena et al., 1994). The unique CrdA and CrdC if in a multimeric complex with CrdS perhaps assist transit of curdlan across the two membranes of the cell envelope. In rhizobia, the lipo-oligosaccharide Nod factors are synthesized at the cytoplasmic face of the inner membrane, and it is proposed that they are secreted by a functional complex encoded by the nodI genes (Vazquez et al., 1993; Fernandez-Lopez et al., 1996). In the case of the \textit{E. coli} K5 capsular polysaccharide, four proteins are involved in its biosynthesis and there is evidence for the involvement of a stable heterooligomeric complex on the cytoplasmic membrane (Hodson et al., 2000, Arrecubieta et al., 2001).

In contrast to these examples, the polytopic HasA membrane protein of \textit{S. pyogenes} (Heldermon et al., 2001) is able to synthesize and secrete to the cell surface the high-molecular-weight hyaluronan product without the involvement of other proteins. HasA function is, however, dependent on cardiolipin that in association with the synthase protein, perhaps creates a porelike passage for chain translocation (Thap-Simmons et al., 1998, 1999). In \textit{Agrobacterium}, maximal curdlan production requires phosphatidylethanolamine (Karnezis et al., 2002), which may be needed for correct folding or for maintenance of folding of one or more of the essential Crd proteins or perhaps may also be involved in creation of a porelike structure.

Materials and methods

Bacterial strains and plasmids

The bacteria and plasmids used in this study are listed in Table II. Additional plasmids carrying various \textit{crdS}-gene fusions are listed in Table I. The plasmids pVS1575 (\textit{crdS}+)
and pVS1574 (crdSΔ) carry PCR-derived NdeI-BamHI clones produced from pVS1512 using the primers SF1 (5'-GGAATTCATATGATGTTATCATGCTG-3') and, respectively, SRI (5'-CGGGATCCATGTTTGAGG-3') and SR2 (5'-CGGAATTCCGGATCCACCCGGG-3') with the phoA cassette. Correctly oriented clones were identified by restriction analysis, and specific primers (5'-GGGA- TTAAGTTGGGTA and 5'-TCTGATCACCCGTTAT) were used to verify the in-frame alignment of crdS with lacZ and phoA, respectively.

**Construction of crdS-lacZ and crdS-phoA gene fusions**

Various lengths of the 5' end of the crdS gene were generated by PCR from pVS1512 using the SF1 forward primer and different reverse primers. The resulting amplicons were each cloned into pVS1549 using the NdeI and BamHI sites. In a second cloning step, the various BamHI-digested recombinant plasmids were ligated with a 3-kb lacZ-containing BamHI fragment recovered from pMC1871. To convert the crdS-lacZ fusions into corresponding crdS-phoA fusions, a PCR-derived phoA fragment was generated from pSWFII using the primers 5'-TCCTGGAC- GGGATCCTTTCC and 5'-GGGAATTCATGTTTGAGG-3', the latter providing a stop codon for the phoA sequence. The terminal BamHI cleavage sites introduced by amplification enabled replacement of the lacZ cassette with the phoA cassette. Correctly oriented clones were identified by restriction analysis, and specific primers (5'-GGGA- TTAAGTTGGGTA and 5'-TCTGATCACCCGTTAT) were used to verify the in-frame alignment of crdS with lacZ and phoA, respectively.

**Growth media**

Strains of *E. coli* were grown at 37°C and *Agrobacterium* at 28°C. Nutrient broth (NB), nutrient agar (NA), *Agrobacterium* defined broth (ADB), and aniline blue agar (ABA) was as described by Stasinopoulos et al. (1999) and Luria agar (LA) was as described by Sambrook et al. (1989). Supplements (mg/L) in NA were ampicillin (Ap, 100) and tetracycline (Tc, 15); those in ABA were chloramphenicol (Cm, 50) and tetracycline (10); and those in LA were 40 XP and 40 X-Gal.

**Conjugal mobilization of plasmids and complementation studies**

The mobilization of plasmids from *E. coli* S17-1 to *Agrobacterium* mutants was performed according to Stasinopoulos et al. (1999). For complementation analysis,
the Agrobacterium transconjugants were selected on ABA containing Tc and Cm and were assessed after growth for 5–7 days. Blue-staining colonies indicated complementation (i.e., restoration of curdlan production), whereas unstained (white) colonies indicated the absence of complementation.

Expression of the Agrobacterium crdS gene in E. coli

The crdS and crdSD genes of pVS1575 and pVS1574, respectively, were expressed (Tabor and Richardson, 1985) from the T7φ10 promoter following induction of the chromosomal T7 RNA polymerase gene of the E. coli host, BL21(DE3)(pLysS). The procedure was similar to that described (Ausubel et al., 1994) except that the bacteria were induced with 1 mM IPTG for 30 min prior to the addition of rifampicin (200 μg/ml for 30 min at 37°C) and subsequent exposure to [35S]methionine-cysteine mix (10 μCi/ml for 5 min). Whole cell lysates or cell fractions were analyzed by SDS–PAGE.

Preparation of cell fractions from Agrobacterium and E. coli

Agrobacterium cells harvested from a 5-day 200-ml ADB culture were washed three times in potassium phosphate buffer (50 mM, pH 7.4), resuspended in 30 ml 20% (w/v) sucrose gradient prepared in 10 mM Tris–HCl, 1 mM MgCl2, 10 mM EDTA (pH 7.4), layered on a continuous 25–60% sucrose buffer (50 mM Tris–HCl, pH 7.4), and treated on ice for 30 min. The bacteria were washed three times in potassium phosphate buffer (10 mM Tris–HCl [pH 7.4], 1 mM MgCl2, 10 mM EDTA, 2 mg/ml lysozyme). After 30 min incubation at 4°C, 1 μg/ml DNase was added together with 10 mM MgCl2 and 1 mM MnCl2; incubation continued for a further 30 min. Lysis of the cells was completed by the addition of 16 ml detergent buffer (1% [w/v] deoxycholic acid, 1% [v/v] Nonidet P-40, 20 mM Tris–HCl [pH 7.5], 2 mM EDTA). The lysate was centrifuged at 12,000 × g for 10 min at 4°C, and the pellet, which contained most of the GST–CrdS fusion protein, was collected, resuspended in 5 ml Triton X-100, and recentrifuged. This process was repeated until a compact pellet was obtained. The final pellet was resuspended in 20 ml 8 M urea; after 30 min at room temperature, the sample was centrifuged at 20,000 × g for 1 h at 4°C and the clarified protein solution filtered (pore size 0.22 μm) and stored at −70°C. Antibodies were prepared by subcutaneous injection of a rabbit with 150 μg of the GST–CrdS protein administered first with Freund’s complete adjuvant and at weeks 6 and 10 with Freund’s incomplete adjuvant. Blood was collected at week 12 and the recovered serum stored at −20°C.

Electrophoresis, western blotting, and immunostaining

Extracts from Agrobacterium and E. coli were resolved by SDS–PAGE (Ausubel et al., 1994) or BN-PAGE (Schägger et al., 1994) and the proteins were transferred onto nitrocellulose or polyvinylidinedifluoride filters using a semidyrid western transfer cell (BioRad, Hercules, CA). Immunoreactions were performed using the appropriate specific rabbit anti-serum anti-GST–CrdS, anti-LacZ (gift of T. Lithgow) and anti-PhoA (from Eppendorf-5Prime, Inc., Boulder, CO) and horseradish peroxidase–coupled anti-rabbit antibodies (Sigma, St. Louis, MO), and were detected by the enhanced chemiluminescence method according to the manufacturer’s instructions (Roche Diagnostics, Australia). Sizes of proteins were estimated by comparison with SeaBlue + pre-stained molecular mass markers (Novagen, Madison, WI). Antibodies to maltose-binding protein, SecA secretory protein, and signal peptidase were a gift of T. Lithgow.
Enzyme assays

Alkaline phosphatase and β-galactosidase were measured according to Brickman and Beckwith (1975) and Miller (1972), respectively, using mid-log phase (OD₆₆₀nm of 0.6–0.8) cultures of plasmid-carrying E. coli CC118 that were grown in LB supplemented with 4% (w/v) glucose prior to induction with 1 mM IPTG for 1 h.

Spheroplast preparation and proteolysis with proteinase K

Ten-milliliter NB cultures of E. coli (pVS1575) in exponential phase (OD₆₆₀nm of 0.6) were induced with 1.5 mM IPTG for 2 h and harvested; the cells were resuspended in 10 ml ST buffer (20 mM Tris–HCl, pH 8.5) containing 20% w/v sucrose. Spheroplasts were prepared by the addition of EDTA (2 mM final concentration) and lysozyme (10 mg/ml) and incubated at 4°C for 20 min. The spheroplasts (80–90% of the treated bacteria) were used immediately. A lysate of the spheroplasts was prepared by the addition of MgSO₄ (15 mM final concentration) and sonication of the sample. Aliquots (1 ml) of spheroplasts and the spheroplast lysate were treated with proteinase K (0.2 or 0.4 mg/ml; Sigma) at 4°C for 30 min, and proteolysis was stopped by the addition of 5 mM phenylmethylsulphonyl fluoride. Deoxycholate was added to give a final concentration of 0.01% (w/v) and the proteins were precipitated by adding one-fifth volume of ice-cold 72% (w/v) TCA. The precipitated proteins were washed twice with acetone (4°C), air-dried and prepared for SDS–PAGE.

Construction of a 3D model of the catalytic domain of CrdS

The fold recognition program Threader v.3 (Jones et al., 1992) was used to thread a 241-amino-acid portion of the CrdS sequence onto the X-ray crystallographic structure of SpsA (Protein Data Bank ID: 1QGQ), using a value of 1000 for the maximum number of considered paths and default values for all other parameters. The threaded sequence–structure alignment, after some manual readjustment, was used as input to the program Modeller 6v2 to build a 3D model.

This Modeler program (Sali and Blundell, 1993) builds models by satisfaction of spatial restraints derived from the template. Twenty structural models of CrdS were built using the following parameter values for all other parameters. The threaded sequence–structure alignment, after some manual readjustment, was used as input to the program Modeller 6v2 to build a 3D model.

The model with the lowest value of the Modeller objective function (i.e., −1.0 × ln, the molecular probability density function) was selected as the best model structure. The structural quality of this model was checked with the programs PROSAII v. 3.0 (Sippl, 1993) and Profiles-3D (Lüthy et al., 1992). The 3D-1D self-compatibility score from Profiles-3D for this model was 80.3, compared with an expected score of 109.5 for a protein structure of this length and with incorrect structures having scores below 49.3. The electrostatic potential was calculated and mapped onto the molecular surface of the model structure with the program

GRASP v. 1.3.6 (Nicholls et al., 1991) using formal charges on Asp, Glu, Arg, and Lys side chains and partial charges on backbone atoms.

Nucleotide sequence accession number

The nucleotide sequence of crdS was submitted to GenBank under the accession number AF057142.

Acknowledgments

We thank Trevor Lithgow for helpful advice and the generous gift of anti-LacZ serum and antibodies to membrane markers. This work was supported in part by an Australian Research Council grant (AO9925079).

Abbreviations

ABA, aniline; blue agar; ADB, Agrobacterium defined broth; Ap, ampicillin; BN, blue native; CtN, chloramphenicol; Crd, curdlan; EDTA, ethylenediamine tetra-acetic acid; GST, glutathione-S-transferase; IPTG, isopropyl-D-thiogalactopyranoside; KDO, 2-ketodeoxyoctonate; LA, Luria agar; LB, Luria broth; NA, nutrient agar; NB, nutrient broth; NOX, NADH oxidase; NTG, N-methyl-N′-nitro-N-nitrosoguanidine; ORF, open reading frame; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; PhoA, alkaline phosphatase; SDS, sodium dodecyl sulfate; Tc, tetracycline; TCA, trichloroacetic acid; TM, transmembrane.

References


Downloaded from https://academic.oup.com/glycob/article-abstract/13/10/693/554324 by guest on 03 January 2019


Stasinopoulos, S.J. (1997) Molecular genetics of (1→3)-β-glucan (curdlan) biosynthesis by *Agrobacterium* sp. ATCC31749. Bundoora, Australia: LaTrobe University.


