J. Mol. Biol. (1994) 236, 590–600

Nucleotide Sequence and X-ray Structure of Cyclodextrin Glycosyltransferase from *Bacillus circulans* Strain 251 in a Maltose-dependent Crystal Form

Catherine L. Lawson[1]†, Rob van Montfort[1], Boris Strokopytov[1] Henriëtte J. Rozeboom[1], Kor H. Kalk[1], Gert E. de Vries[2], Dirk Penninga[2] Lubbert Dijkhuizen[2] and Bauke W. Dijkstra[1]‡

[1]*Laboratory of Biophysical Chemistry and BIOSON Research Institute University of Groningen, Nijenborgh 4, 9747 AG Groningen The Netherlands*

[2]*Laboratory of Microbiology, University of Groningen Kerklaan 30, 9751 NN Haren, The Netherlands*

The cyclodextrin glycosyltransferase (CGTase, EC 2.4.1.19) gene from *Bacillus circulans* strain 251 was cloned and sequenced. It was found to code for a mature protein of 686 amino acid residues, showing 75% identity to the CGTase from *B. circulans* strain 8. The X-ray structure of the CGTase was elucidated in a maltodextrin-dependent crystal form and refined against X-ray diffraction data to 2·0 Å resolution. The structure of the enzyme is nearly identical to the CGTase from *B. circulans* strain 8. Three maltose binding sites are observed at the protein surface, two in domain E and one in domain C. The maltose-dependence of CGTase crystallization can be ascribed to the proximity of two of the maltose binding sites to intermolecular crystal contacts. The maltose molecules bound in the E domain interact with several residues implicated in a raw starch binding motif conserved among a diverse group of starch converting enzymes.

Keywords: nucleotide sequence; crystal structure; cyclodextrin glycosyltransferase; *Bacillus circulans*; starch binding

1. Introduction

Many bacteria and fungi excrete enzymes that degrade starch to facilitate the uptake of carbohydrates into the cell. A small group of bacteria produce enzymes that are able to convert starch *via* an intramolecular transglycosylation reaction into cyclic compounds called cyclodextrins. These enzymes, the cyclodextrin glycosyltransferases (CGTases§; EC 2.4.1.19), are functionally related to α-amylases, which hydrolyse starch into linear products. CGTases and α-amylases share about 30% amino acid sequence identity. The three-dimensional structures of α-amylases (Matsuura *et al.*, 1984; Buisson et al., 1987; Boel *et al.*, 1990; Brady *et al.*, 1991; Swift *et al.*, 1991) and the CGTases from *Bacillus circulans* strain 8 (Hofmann *et al.*, 1989; Klein & Schulz, 1991) and *Bacillus stearothermophilus* (Kubota *et al.*, 1990, 1991) also show clear structural similarity in an N-terminal domain of approximately 400 residues, which folds into a $(\beta/\alpha)_8$-barrel structure. However, compared to α-amylases, CGTases have two additional C-terminal domains that fold into β-pleated sheets. The active site is present at the wide end of the $(\beta/\alpha)_8$-barrel, with catalytic residues Asp229, Glu257 and Asp328 (CGTase numbering) (Klein *et al.*, 1992; Nakamura *et al.*, 1992).

CGTases usually produce a mixture of α, β and γ-cyclodextrins, corresponding to six, seven and eight circularly linked α(1→4) glycosidic units, the relative proportions of which depend on the bacterial source of the enzyme (Schmid, 1989). Cyclodextrins are able to form inclusion complexes with a wide variety of organic molecules (Saenger,

† Current address: Biology Department, Brookhaven National Laboratory, Upton, NY 11973, U.S.A.

‡ Author to whom all correspondence should be addressed.

§ Abbreviations used: CGTase, cyclodextrin glycosyltransferase; *cgt*, cyclodextrin glycosyltransferase gene; LB, Luria Broth; MPD, 2-methyl-2,4-pentanediol; Hepes, *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid; $UO_2(Ac)_2$, uranyl acetate; SIRAS, single isomorphous replacement with anomalous scattering; DM, density modification; ORF, open reading frame.

0022-2836/94/070590-11 $08.00/0

1980; Szejtli, 1982) and are therefore of interest for research and industrial applications. They are produced commercially by means of CGTases (Bergsma *et al.*, 1988). This industrial process might be improved by isolation of mutant CGTases with improved product specificity, decreased product inhibition, and improved stability at extremes of temperature and pH. We have embarked on a project to rationally design such mutants of the CGTase from *B. circulans* strain 251, which produces mainly β-cyclodextrin. Rational design of mutants with improved properties requires detailed knowledge of the three-dimensional structure of the enzyme, as well as elucidation of the transglycosylation reaction pathway.

Here we report the *cgt* nucleotide sequence, the derived amino acid sequence and the crystal structure at 2·0 Å resolution of the CGTase from *B. circulans* strain 251. Klein & Schulz (1991) have reported the X-ray structure of a CGTase from a separately isolated strain of *B. circulans* (strain 8) at 2·0 Å resolution. Although there is a 25% difference in amino acid sequence between the two *B. circulans* enzymes, their three-dimensional structures are essentially the same. In contrast to Klein & Schulz (1991), who obtained their crystals in the absence of carbohydrate, our crystal form absolutely requires the presence of a high concentration of maltose (Lawson *et al.*, 1990). This is reflected in the crystal structure, where we observe three occupied maltose binding sites, two of which are close to intermolecular crystal contacts.

2. Materials and Methods

(a) *DNA manipulations and sequencing*

Bacillus circulans strain 251 (RIV nr. 11115; Rijksinstituut voor Volksgezondheid, Bilthoven, The Netherlands) was routinely grown on medium with 0·5% (w/v) modified potato starch (Paselli SA2, AVEBE, The Netherlands) at pH 6, 48°C and at a dissolved oxygen tension equal to 50% of the air saturation. *Escherichia coli* PC1990 (Lazzaroni & Portalier, 1979) served as host strain and was grown in Luria Broth (LB) or M9 medium. DNA handling, modification, cloning and transformation of *E. coli* was carried out as described by Sambrook *et al.* (1989). A *Bacillus circulans* genomic clonebank was constructed from partially *Sau*3A digested DNA, which was sized on a sucrose gradient. The 3 to 5 kb fractions were pooled, ligated to *Bam*HI-digested pBS$^+$ (Stratagene, La Jolla, CA) plasmid DNA, and transformed to *E. coli* strain PC1990, known to leak periplasmic proteins into the medium because of a mutation in the *tol*B locus (Lazzaroni & Portalier, 1979). Clones expressing CGTase activity were scored on LB agar plates containing 1% starch and 0·1 mM isopropyl-β-D-thiogalactoside. Halo formation was observed either directly or after iodine staining. CGTase activity was determined *via* the phenolphtalein method (Vikmon, 1982) as described previously (Lawson *et al.*, 1990). DNA sequence determination was performed by the chain termination technique (Sanger *et al.*, 1977) on supercoiled plasmid DNA. Custom primers were synthesized by Eurosequence (Groningen, The Netherlands). The sequence determination of the *cgt* gene was initiated from the plasmid fusion sites, using commercially available M13 DNA sequence primers.

(b) *Crystallization, soaking and data collection*

Purification of CGTase from *B. circulans* strain 251 was performed essentially as described in Lawson *et al.* (1990). Crystals of *B. circulans* CGTase were grown at room temperature by vapour diffusion against 60% (v/v) 2-methyl 2,4-pentanediol (MPD) buffered with 100 mM Hepes (pH 7·55), in the presence of 10 to 20 mg/ml α-cyclodextrin as described (Lawson *et al.*, 1990). The crystals have $P2_12_12_1$ space group symmetry with unit cell dimensions $a = 120{\cdot}4$ Å, $b = 110{\cdot}9$ Å and $c = 66{\cdot}4$ Å, and one molecule of CGTase per asymmetric unit. During and after CGTase crystallization, α-cyclodextrin is slowly degraded to glucose plus an unidentified compound that we believe to be glycosylated MPD; CGTase crystals lose diffraction quality when stored at room temperature and finally dissolve after about 3 weeks, near the end-point of cyclodextrin degradation. Suitable crystals for data collection were therefore transferred to 4°C and stabilized in a mother liquor containing fresh α-cyclodextrin (1 mg/ml, in 60% MPD with 100 mM Hepes (pH 7·55)). Derivative crystals were prepared by soaking crystals for 2 to 5 days at 4°C in mother liquor containing, in addition, 2·5 to 5·0 mM of the heavy atom compound. Diffraction data for 2 native and 2 $UO_2(Ac)_2$ derivative crystals were collected at room temperature with an Enraf–Nonius FAST area detector. The data were processed with MADNES (Messerschmidt & Pflugrath, 1987) and the profile fitting programs of Kabsch (1988). Data statistics are summarized in Table 1.

(c) *Structure determination*

The method of single isomorphous replacement with inclusion of anomalous differences (SIRAS) was used to obtain initial phases. Two uranyl ion binding sites, 4 Å apart, were determined by isomorphous and anomalous difference Patterson syntheses for the $UO_2(Ac)_2$ derivatives. Refinement of heavy atom parameters and calculation of phases were carried out with a version of the program PHARE (provided by Z. Otwinowski) modified for improved error treatment (Einstein, 1977). The overall figure of merit of the resulting phase set was 0·51 for 16,530 reflections between 3·0 and 15·0 Å. An improved map was obtained by density modification (DM) with both solvent flattening and negative density truncation using the program package of Wang (1985). Solvent masks were calculated by the method of Leslie (1987). The resulting DM map had an overall figure of merit of 0·80, with an r.m.s. phase change of 37·8° from the SIRAS map.

A partial (73%) model of CGTase incorporating the amino acid sequence derived from the *B. circulans* gene (Fig. 2) was built from the DM map with the program FRODO (Jones, 1985). After refinement with GROMOS (Fujinaga *et al.*, 1989), phases from this partial model and the DM procedure were combined in the program SIGMAA (Read, 1986) to generate an improved map. The refinement and phase combination procedure was repeated with an 85% complete model, after which all but the final carboxy-terminal residue of the 686 residue polypeptide chain were incorporated. Finally, least-squares refinement against diffraction data between 5·0 Å and 2·0 Å was performed with the TNT-package (Tronrud *et al.*, 1987). Solvent molecules were included in the model

Table 1
Statistics of data collection

Dataset[a]	Native-I	Native-II	Native-III[f]	$UO_2(Ac)_2$-I	$UO_2(Ac)_2$-II
Resolution limit (Å)	2·5	2·0	2·0	3·7	3·0
Unique reflections[b]	28566	41840	48567	13081	49480
R_{merge} (%)[c]	5·6	6·1	—	4·8	7·7
Completeness (%)[d]	91·5	68·8	78·9	72·3	98·2
Phasing power[e]				2·46	1·41
Number of sites				2	2
Residue nearest site				D328 (42·6)	D328 (26·9)
(Relative occupancy)				E257 (25·2)	E257 (14·7)

[a] Each dataset was collected from a single CGTase crystal.
[b] Bijvoet reflections $I(hkl)$ and $I(-h-k-l)$ were treated separately for the $UO_2(Ac)_2$ derivatives.
[c] The R_{merge} indicates agreement of individual reflection measurements over the set of unique averaged reflections:

$$R_{merge} = (\Sigma\Sigma|I_j(hkl) - I_{avg}(hkl)|)/(\Sigma\Sigma I_{avg}(hkl)).$$

[d] The completeness of the data to the indicated resolution is given.
[e] The phasing power is defined as the mean heavy-atom structure factor amplitude divided by the mean lack-of-closure error.
[f] Merged dataset of Native I and Native II.

at positions of positive peaks that were at least 2·5 σ above the mean in $(F_{obs} - F_{calc})\exp i\alpha_{calc}$ difference Fourier maps and within 2·4 to 3·5 Å from hydrogen bond donor or acceptor atoms.

3. Results and Discussion

(a) *Cloning, expression and sequence determination*

A chromosomal genebank of *B. circulans* strain 251 DNA in the pBS$^+$ vector was transformed to *E. coli* strain PC1990. Screening of a collection of 25,000 transformants on LB-starch agar medium yielded several halo-forming colonies. A single CGTase expressing clone, PC1990/pDV58, with a 2197 bp insert, was chosen for further study. Analysis of its nucleotide sequence revealed a single large ORF of 2139 bp in reverse orientation with respect to the *lacZ* promoter of the cloning vector pBS$^+$, beginning with an ATG initiation codon at nucleotides 1 to 3 and ending with a TAA termination signal at nucleotides 2140 to 2142 (Fig. 1). A sequence that could be the ribosome binding site is located 7 to 11 bp upstream of the ATG start codon. The 55 bp region preceding the *cgt* ORF does not contain promoter-like structures; expression of the *cgt* gene in *E. coli* therefore appears to depend on a cryptic promoter residing in the plasmid *lacZ* coding sequence. The N-terminal sequence of the purified extracellular CGTase from *B. circulans*, determined by automated Edman degradation as Ala-Pro-Asp-Thr-Ser-Val-Ser-Asn-Lys-Gln, was found to correspond to the nucleotide sequence

```
          10        20        30        40        50        60        70        80        90       100
          |         |         |         |         |         |         |         |         |         |
-54                                          GGATC CGGCGACAGG GACAAGCCTG GAATTCAAAC GATTACATAG GAGGTATAAC
  1 ATGAAGAAAT TTCTGAAATC GACAGCTGCG CTTGCCCTGG GATTATCGCT GACGTTCGGG CTTTTCAGCC CTGCCCAGGC CGCGCCGGAT ACCTCGGTAT
101 CCAACAAGCA AAATTTCAGC ACCGACGTCA TCTATCAAAT TTTCACCGAC AGGTTTTCGG ACGGCAATCC CGCCAACAAT CCGACCGGCG CGGCGTTTGA
201 CGGAACCTGC ACGAACCTCC GGCTGTATTG CGGCGGCGAC TGGCAGGGCA TCATCAACAA AATCAACGAC GGTTACCTGA CCGGGATGGG CGTTACCGCC
301 ATCTGGATCT CCCAGCCGGT CGAAAACATC TACAGCATCA TCAATTATTC CGGCGTAAAC AACACGGCCT ATCACGGCTA CTGGGCCCGG GACTTCAAGA
401 AGACGAATCC GGCCTACGGC ACGATTGCGG ACTTCCAGAA CCTGATCGCC GCCGCGCATG CAAAAAACAT CAAAGTCATT ATCGACTTTG CCCCGAACCA
501 TACGTCGCCC GCCTCGTCCG ACCAGCCTTC CTTTGCGGAA AACGGCCGGC TGTACGATAA CGGCACGCTG CTCGGGGGAT ACACGAACGA TACGCAGAAC
601 CTGTTCCACC ATAACGGCGG CACGGACTTT TCCACGACCG AAAACGGCAT CTACAAAAAC CTGTACGATC TCGCCGACCT GAACCATAAC AACAGCACCG
701 TGGACGTCTA CTTGAAGGAC GCGATCAAAA TGTGGCTGGA CCTCGGCATC GACGGCATCC GCATGGATGC GGTGAAGCAT ATGCCGTTCG GCTGGCAGAA
801 GAGCTTTATG GCTGCCGTCA ACAACTATAA GCCGGTCTTT ACCTTCGGCG AATGGTTCCT GGGCGTAAAT GAAGTGAGCC CGGAAAACCA TAAGTTTGCC
901 AACGAATCCG GCATGAGCCT GCTTGATTTC CGTTTTGCCC AAAAGGTGCG GCAGGTGTTC CGGGACAACA CCGACAATAT GTACGGCCTG AAGGCGATGC
1001 TGGAGGGCTC CGCAGCCGAT TACGCCCAGG TGGATGACCA GGTGACGTTC ATCGACAACC ATGACATGGA GCGTTTCCAC GCAAGCAATG CAAACCGCCG
1101 GAAGCTGGAG CAAGCGCTTG CGTTCACGCT GACCTCGCGC GGCGTCCCCG CCATTTATTA CGGCACCGAG CAGTACATGT CGGGCGGGAC CGATCCGGAC
1201 AACCGGGCGC GGATCCCTTC CTTCTCCACG TCGACGACCG CCTATCAGGT CATTCAAAAG CTGGCGCCGC TGCGCAAGTG CAACCCGGCC ATCGCCTACG
1301 GATCGACGCA GGAGCGCTGG ATCAACAACG ACGTGCTCAT TTATGAGCGC AAATTCGGCA GCAACGTTGC CGTCGTTGCC GTCAACCGCA ATTTAAACGC
1401 GCCGGCTTCC ATTTCGGGAC TTGTCACTTC CCTGCCGCAA GGCAGCTACA ACGACGTCCT TGGCGGCCTT CTGAACGGCA ACACGTTATC GGTAGGCTCC
1501 GGCGGGGCCG CCTCCAATTT CACGCTTGCG GCCGGCGGCA CGGCGGTGTG GCAGTACACC GCGGCTACGG CGACGCCGAC CATCGGGCAT GTCGGGCCGA
1601 TGATGGCCAA GCCGGGCGTG ACGATCACGA TCGACGGCCG CGGCTTCGGC TCTAGCAAAG GCACCGTCTA CTTCGGTACG ACGGCGGTGA GCGGGGCGGA
1701 CATCACGTCT TGGGAAGACA CGCAGATCAA AGTGAAAATT CCGGCCGTCG CAGGCGGCAA CTACAACATT AAAGTCGCAA ACGCTGCCGG AACGGCAAGC
1801 AATGTGTATG ACAACTTCGA GGTATTGTCC GGAGACCAGG TCAGCGTCCG CTTCGTGGTC AACAACGCGA CGACGGCCCT TGGGCAAAAT GTGTACCTGA
1901 CGGGCAGTGT CAGCGAGCTG GGGAACTGGG ACCCGGCAAA AGCAATCGGG CCGATGTACA ATCAGGTCGT TTACCAATAT CCGAACTGGT ATTATGACGT
2001 CAGCGTTCCG GCCGGCAAAA CGATCGAGTT CAAGTTTTTG AAAAAACAAG GCTCCACCGT CACGTGGGAA GGCGGCAGCA ACCACACCTT CACCGCGCCG
2101 TCCAGCGGCA CCGCGACCAT TAACGTGAAT TGGCAGCCAT AA
```

Figure 1. Nucleotide sequence of the *Bacillus circulans* strain 251 *cgt* gene. Underlined nucleotides form the ribosome binding site (GGAGG; nucleotides −7 to −11), the start codon (ATG; nucleotides 1 to 3) and the signal sequence processing site (GCCGCG; nucleotides 79 to 84).

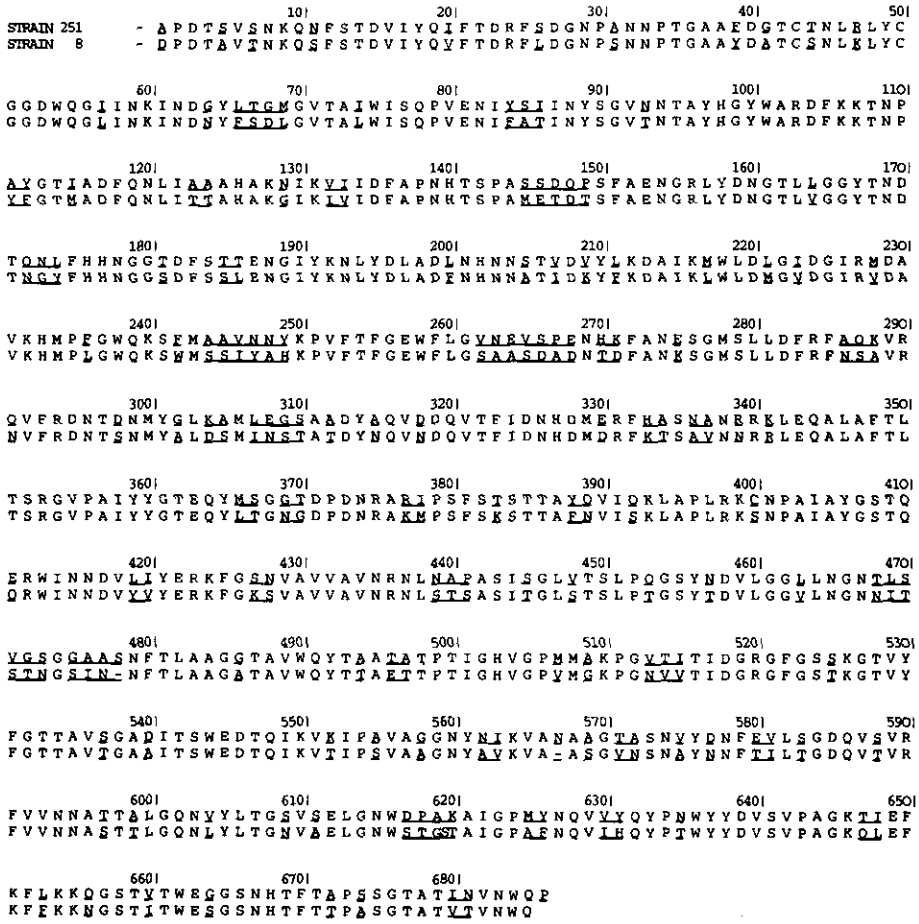


Figure 2. Amino acid sequence of the CGTase from *Bacillus circulans* strain 251, derived from the nucleotide sequence and compared with the sequence of *Bacillus circulans* strain 8 CGTase (Nitschke *et al.*, 1990). Sequence differences are underlined.

starting at position 82 in Figure 1. Accordingly, the preceding 27 amino acid residues (bp positions 1 to 81 in Fig. 1) correspond to a signal peptide that is removed during export. The extracellular CGTase thus contains 686 amino acid residues with a molecular mass of 74,514 encoded by 2058 bp of the *cgt* gene. A comparison of the CGTase amino acid sequences of *B. circulans* strains 251 and 8 (Nitschke *et al.*, 1990) shows that 178 of the 686 amino acid residues (25·9%) are different (Fig. 2). None of these differences, however, occur in regions that are normally conserved in other CGTases. The raw-starch-digesting amylase from *Bacillus* sp. B1018 (Itkor *et al.*, 1990) is even more similar in sequence to our CGTase. Only seven residues are different.

(b) *Crystallographic refinement*

The final refined model contains all 686 amino acid residues, 471 water-molecules, two calcium ions and three maltose molecules. Refinement statistics are given in Table 2. The average co-ordinate error in the model is estimated to be 0·26 Å from a Luzzati-plot (Luzzati, 1952). The estimate from a σ_A plot (Read, 1986) is 0·30 Å. The stereochemical correctness of the main-chain conformation was also analysed with a Ramachandran plot (Ramakrishnan & Ramachandran, 1965). Ala152 and Tyr195 have energetically unfavourable ϕ/ψ angle combinations (Fig. 3). Tyr195 is a residue in the active site; Ala152 is in a reverse turn at the protein surface. Both residues have well-defined density. Their main-chain conformations are stabilized by water-mediated hydrogen bonds with neighbouring residues. The unusual ϕ/ψ combinations, as well as the four *cis*-peptide bonds in our structure, which occur before residues Pro372, Pro506, Pro624 and Pro634, were also observed in

Table 2
Statistics of refinement

Parameter	Value
R-factor‡	16·4%
Average *B*-factor	24·3 Å^2
r.m.s. bond lengths†	0·016 Å
r.m.s. bond angles†	2·49°
r.m.s. torsion angles†	24·67°
r.m.s. *B*-factor	2·17 Å^2

† Deviations from ideal geometry.

‡ Crystallographic *R*-factor is defined as $R=\frac{\Sigma||F_{obs}|-|F_{calc}||}{\Sigma|F_{obs}|}$.

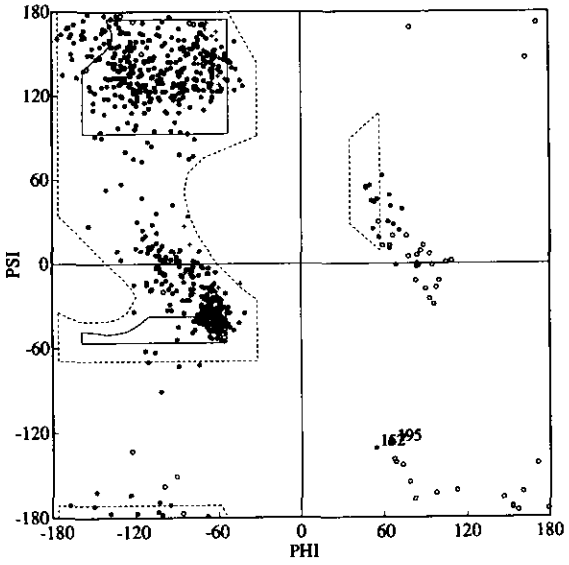


Figure 3. Ramachandran plot of the main-chain (ϕ/ψ)-conformational angles of our CGTase model. Glycine (ϕ/ψ)-pairs are represented by open circles, non-glycine (ϕ/ψ)-pairs are represented by filled circles. Continuous lines define the areas of preferred (ϕ/ψ)-combinations and dashed lines indicate the sterically allowed regions in conformational space (Ramakrishnan & Ramachandran, 1965). Non-glycine residues with energetically unfavourable (ϕ/ψ)-values are labelled.

the *B. circulans* strain 8 CGTase (Klein & Schulz, 1991). There are eight surface loops with relatively high temperature factors (see Fig. 4) indicating flexibility, comprising residues 35 to 36, 86 to 93, 145 to 152, 169 to 172, 335 to 336, 473 to 474, 496 to 497 and 599 to 601. The atomic co-ordinates of the final model have been deposited with the Brookhaven Protein Data Bank (entry number 1CDG) (Bernstein *et al.*, 1977).

(c) *Three-dimensional structure of CGTase*

CGTase from *B. circulans* strain 251 consists of a single polypeptide chain of 686 residues that is subdivided into five domains, as in the *B. circulans* strain 8 CGTase (Hofmann *et al.*, 1989; Klein & Schulz, 1991). The largest of these is the A domain with $(\beta\alpha)_8$ topology. Domain B is an extended loop between the third β-strand and third α-helix of the A domain with very little secondary structure. The other three domains have β-sheet structure. The detailed features of these domains are essentially identical to those of the *B. circulans* strain 8 CGTase. The overall fold of the enzyme is presented *in* Fig. 5.

Two separate divalent cation binding sites are observed in the *B. circulans* strain 251 CGTase structure. Because no divalent cations are introduced in the last stages of enzyme purification or in the crystallization, the sites must have high binding affinities. The co-ordination geometry of the first site (Asn139 Oδ1, Ile190 O, Asp199 Oδ1 and Oδ2, His233 O and 3 water molecules) is exactly as observed in *B. circulans* strain 8 CGTase and in the *Aspergillus oryzae* and *Aspergillus niger* amylase structures (Boel *et al.*, 1990). The second calcium binding site has no structural homology with amylases. The co-ordinating protein ligands (1 carboxyl oxygen each of Asp27 and Asp53, the amide oxygen atoms of Asn32 and Asn33 and the carbonyl oxygen atom of Asn29) are located in the loop connecting β-strand 1 and α-helix 1 of the A domain, which also contains the only disulphide bridge (connecting Cys43 and Cys50) of the enzyme. In *B. circulans* strain 8, the second calcium binding site has as a sixth ligand, the carbonyl oxygen of Gly51 (at 2·45 Å from the calcium). In our structure the carbonyl oxygen of Gly51 is a weakly bound ligand at 2·6 Å and a water molecule, at 2·3 Å, is a seventh ligand.

Comparison of the two *B. circulans* CGTase sequences with the three-dimensional structure reveals that nearly all differences are on the surface of the molecule. Moreover, none of the sequence differences between the two *B. circulans* strains occur in regions conserved in other CGTase sequences (e.g. *Klebsiella pneumonia* (Binder *et al.*, 1986); *B. macerans* (Takano *et al.*, 1986); *B. stearothermophilus* (Fujiwara *et al.*, 1992); *Bacillus* sp. 1011 (Kimura *et al.*, 1987); alkalophilic *Bacillus* sp. 38-2 (Kaneko *et al.*, 1988); alkalophilic *Bacillus* sp. 1-1 (Schmid *et al.*, 1988) and *B. licheniformis* (Hill *et al.*, 1990)).

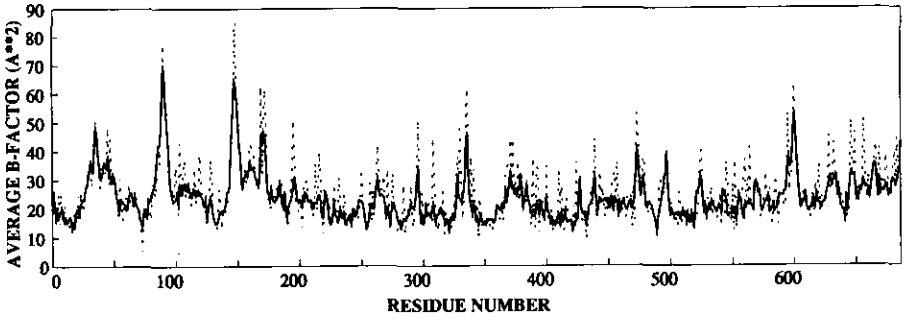


Figure 4. Variation of the atomic temperature factors along the polypeptide chain, averaged for main-chain atoms (thick line) and for side-chain atoms (broken line).

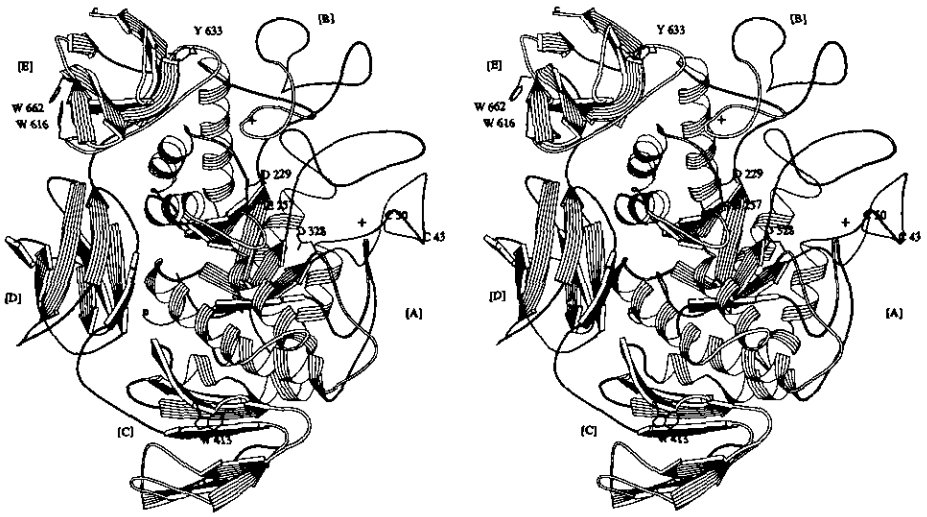
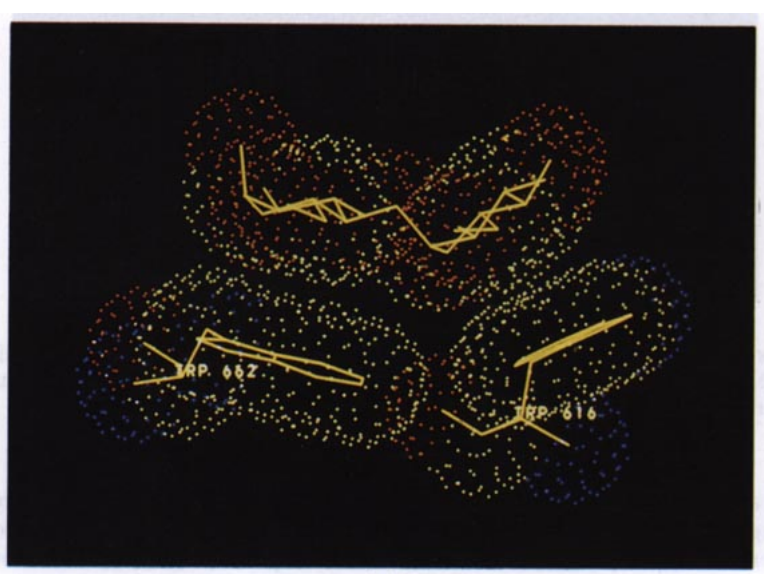


Figure 5. Stereo ribbon drawing of CGTase from *Bacillus circulans* strain 251. The program RIBBON (Priestle, 1988) was employed to produce a simplified drawing of the polypeptide fold of CGTase from α-carbon co-ordinates and a list of the secondary structural elements (determined with the program DSSP from Kabsch & Sander, 1983). Arrows represent β-strands; coils represent α-helices. The amino and carboxyl termini are labelled n and c, respectively. The domains of CGTase are labelled [A] through [E]. Selected side-chains are drawn to indicate relative positions of sites mentioned in the text: Cys43 and Cys50 (disulphide bridge), Asp229, Glu257 and Asp328 (active centre), Trp616 and Trp662 (maltodextrin binding site 1), Tyr633 (maltodextrin binding site 2), Trp413 (maltodextrin binding site 3). The 1st calcium site is indicated by the cross between the A and B domains; the 2nd cation binding site is indicated by a cross near the disulphide bridge.

(d) *Maltodextrin binding sites and the starch binding domain*

Our CGTase crystals must be grown and stored in solutions that have a large molar excess of α-cyclodextrin. The cyclodextrin is slowly degraded into smaller maltodextrins and finally into glucose by the enzyme. Because CGTase crystals disintegrate near the end-point of the degradation process, we suspected that an intermediate might be responsible for the stabilization of the crystals. From inspection of electron density maps we have identified three carbohydrate binding sites in which α-maltose could be modelled. The three maltose molecules make extensive hydrophobic contact from stacking of their apolar face onto the flat surfaces of aromatic rings. In addition, most of the sugar hydroxyl oxygen atoms make hydrogen bonds to the protein, either directly or through water molecules. In all three binding sites maltose is modelled in an extended conformation that allows formation of an intramolecular hydrogen bond between the C-2 hydroxyl group of the non-reducing sugar (g_2) and the C-3 hydroxyl group of the reducing sugar (g_1); in sites 1 and 3 the inferred hydrogen-bonding pattern suggests that the (g_1)C-3 hydroxyl is the proton



(a)

Fig. 6.

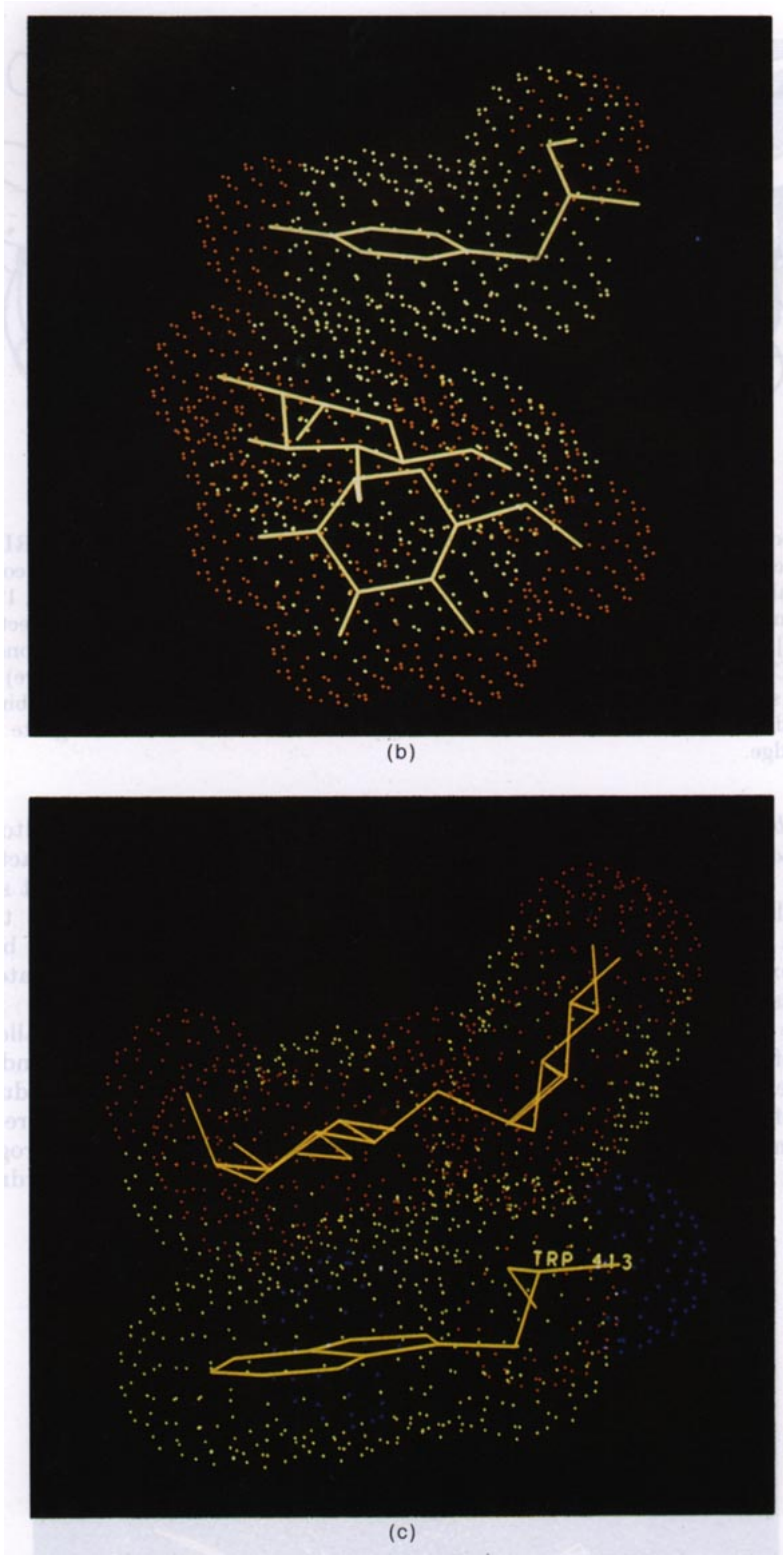
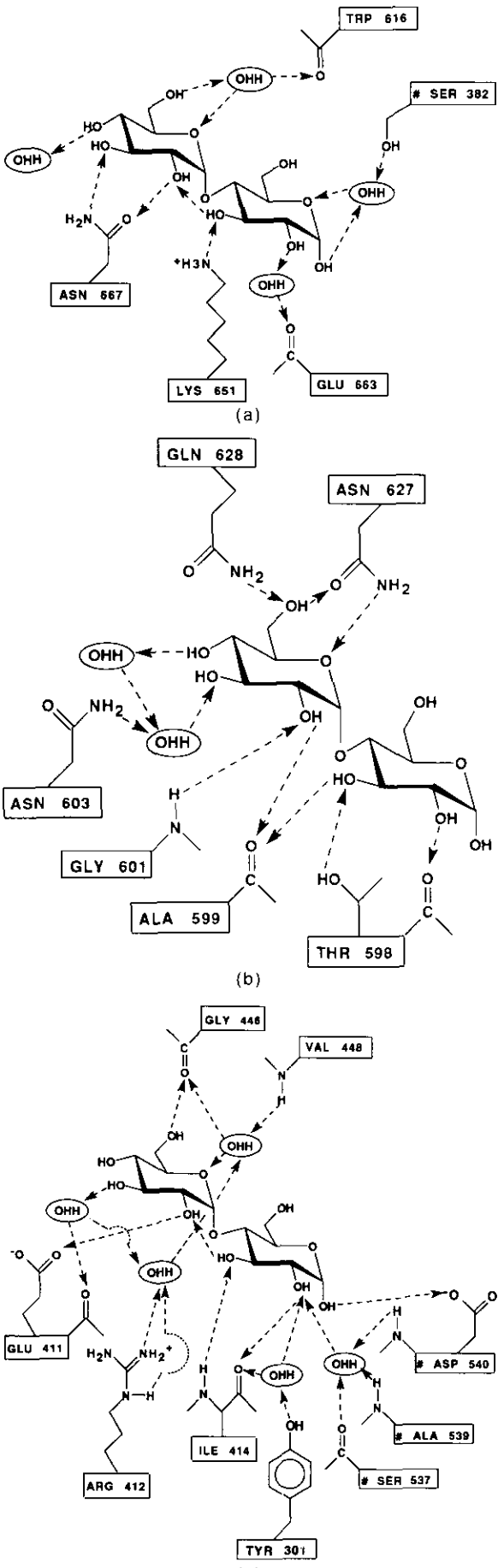


Figure 6. Interactions of the 3 maltose molecules bound to CGTase. (a) Maltose at the 1st binding site in the E domain near Trp616 and Trp662. (b) Maltose at the 2nd binding site in the E domain near Tyr633. (c) Maltose at the binding site in the C domain near Trp413.

donor. Similar bonds are observed in the structures of α-maltose (Takusagawa & Jacobson, 1978) and cyclodextrins (Saenger, 1980).

The first two maltose binding sites are located in the E domain. Binding of maltose at the first site occurs mainly through hydrophobic stacking of both glucose rings on the indole groups of Trp616 and Trp662 (Fig. 6(a)). From our model we infer the presence of three direct hydrogen bonds with the protein to the nitrogen and oxygen atoms of the



Asn667 side-chain and the Nζ-group of Lys651. Three water molecules mediate contact between protein and maltose through a hydrogen-bonded network (Fig. 7(a)). In the second site, the reducing glucose unit of maltose is stacked over the phenyl ring of Tyr633 (Fig. 6(b)). Hydrogen bonds are made with the Thr598 Oγ and peptide oxygen, Ala599 O, Gly601 N, Asn627 Oδ1 and Nδ2 and Gln628 Nε2 atoms. We observe two water molecules within hydrogen-bonding distance of maltose 2, one providing an indirect contact between maltose and Asn603 Nδ2 (Fig. 7(b)).

In the third carbohydrate binding site, located in domain C, maltose is modelled with the apolar face of the reducing sugar stacked against the aromatic side-chain of Trp413 (Fig. 6(c)). Five direct hydrogen bonds to protein are made to one carboxyl oxygen of Glu411, both the carbonyl oxygen and the NH-group of Ile414, the carbonyl oxygen of Gly446 and one carboxyl oxygen of the symmetry related Asp540 (Fig. 7(c)). Only the O-6, O-5 and O-4 atoms of the reducing end glucose (g_1) and the O-4 atom of the non-reducing end glucose (g_2) do not participate in hydrogen bonding to protein. Four water molecules are bound between protein and maltose. In pig pancreatic α-amylase a carbohydrate binding site outside of the substrate binding cleft in the active site was observed near the N terminus of helix α-8 in the A domain, facing the C domain (Buisson *et al.*, 1987). Our third maltose binding site is distinct from the α-amylase site, although the physiological function of the site might be the same as proposed for the pig pancreatic α-amylase: it could anchor CGTase to the starch granule during enzymatic activity (Buisson *et al.*, 1987). On the other hand, we cannot exclude that the third maltose binding site is a crystallographic artefact mediating contacts between symmetry related molecules.

Two lines of evidence suggest that the observed maltose binding sites in the E domain might be involved in attaching the enzyme to raw starch granules *in vivo*. First, CGTase was found to bind starch granules at a site other than the active site (Villette *et al.*, 1992). Second, CGTase belongs to a diverse group of starch degrading enzymes for which a raw-starch binding motif was proposed, based on a region of limited sequence homology (Svensson *et al.*, 1989). This region, approximately 95 residues in length, corresponds exactly to the CGTase E domain. Four of the 11 strictly conserved residues of the raw-starch binding motif (Trp616, Lys651, Trp662 and Asn667 (CGTase numbering)), are part of the first maltose-binding site, while three others (Thr598, Gly601 and Trp636) are part of the

Figure 7. Schematic diagrams of the hydrogen bonding patterns of the maltose molecules bound to CGTase. Residues labelled with # correspond to symmetry related molecules. (a) Maltose at the 1st binding site in the E domain near Trp616 and Trp662. (b) Maltose at the 2nd binding site in the E domain near Trp633. (c) Maltose bound at the C domain near Trp413.

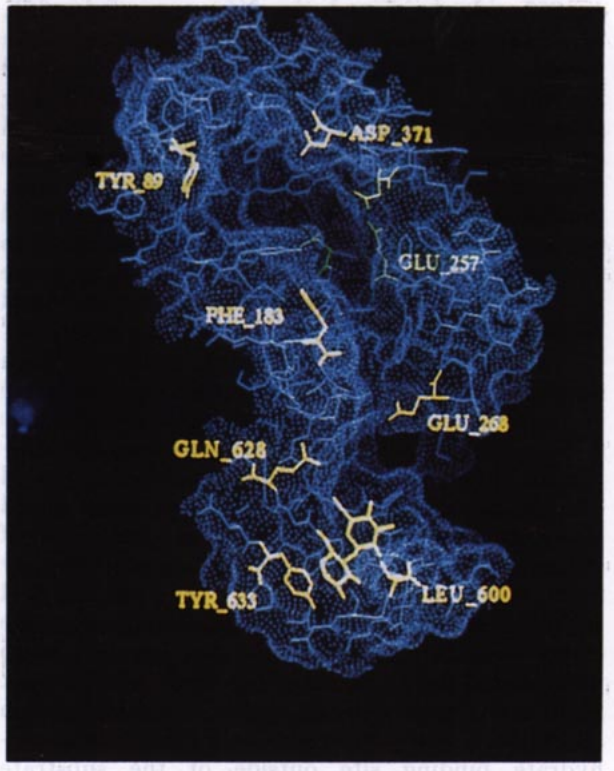


Figure 8. Putative substrate binding groove in the CGTase molecule, extending from the 2nd maltose binding site in the E domain (near Tyr633) towards the active site. The active site is indicated by Glu257.

second site. The remaining strictly conserved residues (Gly608, Leu613, Gly614 and Pro634) do not bind maltodextrin, but are probably required for structural support of the raw starch binding domain. Although we find only disaccharides at sites 1 and 2, modelling studies suggest that a longer amylose chain could be accommodated in an extended ribbon conformation interacting with both sites.

The location of two of the three maltose binding sites at intermolecular crystal contacts may explain the maltodextrin dependence of this crystal form of CGTase. In site 1, a water molecule mediates a hydrogen-bonded contact between maltose and a symmetry related molecule (Fig. 7(a)). In site 3, one direct hydrogen bond is made between maltose and another symmetry related molecule, and a water molecule mediates additional hydrogen-bonded contacts (Fig. 7(c)). It therefore appears that the requirement for maltodextrin in both crystallization and crystal stabilization is due to the critical role of maltose in forming crystal contacts.

(e) *Active site*

The active site of CGTase contains three acidic residues: Asp229, Glu257 and Asp328 (Klein & Schulz, 1991). The precise role of each of these residues is not very well understood. Mutational analysis in CGTase (Klein *et al.*, 1992; Nakamura *et al.*, 1992), as well as in α-amylases (Svensson & Sögaard, 1992; Takase *et al.*, 1992), indicates that all three residues are important in catalysis. In the electron density maps we observe a distance of 2·5 Å between Asp328 Oδ2 and Glu257 Oε1, suggesting protonation of one of the side-chains and the existence of a strong hydrogen bond between them. At present it is not known which of the two carboxylates is protonated. The active site residues are located at the bottom of a surface cavity in the A domain. Glu257 is surrounded by mostly hydrophobic residues (Phe259, Phe283, Leu281), whereas Asp328 and Asp229 are in a more polar environment. Therefore it seems most likely that Glu257 is protonated.

A long groove extends from the active site region towards the maltose binding site near Tyr633 in the E domain (Fig. 8). In contrast to the clear density for maltose in the C and E domains, we did not observe interpretable density for carbohydrate in the enzyme's active site or in this groove. Modelling studies suggest, however, that this groove, starting from the catalytic Glu257 residue, can accommodate nine to ten glucose residues, including the bound maltose of site 2 (near Tyr633). The orientation of maltose suggests that starch would bind in the active site with its reducing end towards the E domain and non-reducing end on the other side of the active site near residues Tyr89, Asp371 and Arg375 of the A domain. Major interaction sites of the groove with the starch polymer may include residues 599 to 602 (E domain) and 261 to 268 (A domain) on one side of the groove, and residues 627 to 628 (in the E domain) and 183 to 186 (B domain) on the other side. The first maltose binding site in the E domain near Trp616 and Trp662 is much further away from the active site region compared to the second site. Nevertheless, an amylose polymer of 15 to 20 glucose residues might be able to span the distance from the active site to Trp616 and Trp662. Further investigations are needed to assess the relevance of these two proposed substrate binding modes.

The elucidation of the *cgt* nucleotide sequence and the crystal structure of the *B. circulans* strain 251 CGTase has enabled us to begin rational design of mutants for detailed studies of the catalytic mechanism of CGTase, and to analyse in detail the factors that contribute to the binding of carbohydrate by the enzyme. Long term goals of our research are the improvement of the product specificity of CGTases and the reduction of product inhibition.

This work is supported by The Netherlands Programma Commissie voor Biotechnologie (PCB) of the Ministry of Economic Affairs.

References

Bergsma, J., Bruinenberg, P. M., Hokse, H. & Meiberg, J. B. M. (1988). Cyclodextrins from potato starch. Recent developments. In *Proceedings of the 4th International Symposium on Cyclodextrins* (Huber, O. & Szejtli, J., eds), pp. 41–46, Kluwer Academic Publishers, Dordrecht, the Netherlands.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer E. F., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.

Binder, F., Huber, O. & Böck, A. (1986). Cyclodextrin-glycosyltransferase from *Klebsiella pneumoniae* M5a1: cloning, nucleotide sequence and expression. *Gene*, **47**, 269–277.

Boel, E., Brady, L., Brzozowski, A. M., Derewenda, Z., Dodson, G. G., Jensen, V. J., Petersen, S. B., Swift, H., Thim, L. & Woldike, H. F. (1990). Calcium binding in α-amylases: an X-ray diffraction study at 2·1 Å resolution of two enzymes from *Aspergillus*. *Biochemistry*, **29**, 6244–6249.

Brady, R. L., Brzozowski, A. M., Derewenda, Z. S., Dodson, E. J. & Dodson, G. G. (1991). Solution of the structure of *Aspergillus niger* acid α-amylase by combined molecular replacement and multiple isomorphous replacement methods. *Acta Crystallogr. sect. B*, **47**, 527–535.

Buisson, G., Duée, E., Haser, R. & Payan, F. (1987). Three dimensional structure of porcine pancreatic α-amylase at 2·9 Å resolution. Role of calcium in structure and activity. *EMBO J.* **6**, 3909–3916.

Einstein, J. R. (1977). An improved method for combining isomorphous replacement and anomalous scattering diffraction data for macromolecular crystals. *Acta Crystallogr. sect. A*, **33**, 75–85.

Fujinaga, M., Gros, P. & van Gunsteren, W. F. (1989). Testing the method of crystallographic refinement using molecular dynamics. *J. Appl. Crystallogr.* **22**, 1–8.

Fujiwara, S., Kakihara, H., Woo, K. B., Lejeune, A., Kanemoto, M., Sakaguchi, K. & Imanaka, T. (1992). Cyclization characteristics of cyclodextrin glucanotransferase are conferred by the NH_2-terminal region of the enzyme. *Appl. Environm. Microbiol.* **58**, 4016–4025.

Hill, D. E., Aldape, R. & Rozzell, J. D. (1990). Nucleotide sequence of a cyclodextrin glucosyltransferase gene, *cgtA*, from *Bacillus licheniformis*. *Nucl. Acids Res.* **18**, 199.

Hofmann, B. E., Bender, H. & Schulz, G. E. (1989). Three-dimensional structure of cyclodextrin glycosyltransferase from *Bacillus circulans* at 3·4 Å resolution. *J. Mol. Biol.* **209**, 793–800.

Itkor, P., Tsukagoshi, N. & Udaka, S. (1990). Nucleotide sequence of the raw-starch-digesting amylase gene from *Bacillus* sp. B1018 and its strong homology to the cyclodextrin glucanotransferase genes. *Biochem. Biophys. Res. Commun.* **166**, 630–636.

Jones, T. A. (1985). Interactive computer graphics: FRODO. *Methods Enzymol.* **115**, 157–171.

Kabsch, W. (1988). Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J. Appl. Crystallogr.* **21**, 916–924.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structures: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Kaneko, T., Hamamoto, T. & Horikoshi, K. (1988). Molecular cloning and nucleotide sequence of the cyclomaltodextrin glucanotransferase from the alkalophilic *Bacillus* sp. strain no. 38-2. *J. Gen. Microbiol.* **134**, 97–105.

Kimura, K., Kataoka, S., Ishii, Y., Takano, T. & Yamane, K. (1987). Nucleotide sequence of the β-cyclodextrin glucanotransferase gene of alkalophilic *Bacillus* sp. strain 1011 and similarity of its amino acid sequence to those of α-amylases. *J. Bacteriol.* **169**, 4399–4402.

Klein, C. & Schulz, G. E. (1991). Structure of cyclodextrin glycosyltransferase refined at 2·0 Å resolution. *J. Mol. Biol.* **217**, 737–750.

Klein, C., Hollender, J., Bender, H. & Schulz, G. E. (1992). Catalytic center of cyclodextrin glycosyltransferase derived from X-ray structure analysis combined with site-directed mutagenesis. *Biochemistry*, **31**, 8740–8746.

Kubota, M., Matsuura, Y., Sakai, S. & Katsube, Y. (1990). Crystal structure of cyclodextrin glucanotransferase from *Bacillus stearothermophilus*. *Protein Eng.* **3**, 328–329.

Kubota, M., Matsuura, Y., Sakai, S. & Katsube, Y. (1991). Molecular structure of *B. stearothermophilus* cyclodextrin glucanotransferase and analysis of substrate binding site. *Denpun Kagaku*, **38**, 141–146.

Lawson, C. L., Bergsma, J., Bruinenberg, P. M., de Vries, G., Dijkhuizen, L. & Dijkstra, B. W. (1990). Maltodextrin-dependent crystallisation of cyclomaltodextrin glucanotransferase from *Bacillus circulans*. *J. Mol. Biol.* **214**, 807–809.

Lazzaroni, J. C. & Portalier, R. C. (1979). Isolation and preliminary characterization of periplasmic-leaky mutants of *Escherichia coli* K-12. *FEMS Microbiol. Letters*. **5**, 411–416.

Leslie, A. G. W. (1987). A reciprocal-space method for calculating a molecular envelope using the algorithm of B. C. Wang. *Acta Crystallogr. sect. A*, **43**, 134–136.

Luzzati, V. (1952). Traitement statistique des erreurs dans la détermination des structures cristallines. *Acta Crystallogr.* **5**, 802–810.

Matsuura, Y., Kusunoki, M., Harada, W. & Kakudo, M. (1984). Structure and possible catalytic residues of Taka-amylase. *J. Biochem.* **95**, 697–702.

Messerschmidt, A. & Pflugrath, J. W. (1987). Crystal orientation and X-ray pattern prediction routines for area detector diffractometer systems in macromolecular crystallography. *J. Appl. Crystallogr.* **20**, 306–315.

Nakamura, A., Haga, K., Ogawa, S., Kuwano, K., Kimura, K. & Yamane, K. (1992). Functional relationships between cyclodextrin glucanotransferase from an alkalophilic *Bacillus* and α-amylases. Site-directed mutagenesis of the conserved two Asp and one Glu residues. *FEBS Letters*. **296**, 37–40.

Nitschke, L., Heeger, K., Bender, H. & Schulz, G. E. (1990). Molecular cloning, nucleotide sequence and expression in *Escherichia coli* of the β-cyclodextrin glycosyltransferase gene from *Bacillus circulans* strain no. 8. *Appl. Microbiol. Biotechnol.* **33**, 542–546.

Priestle, J. P. (1988). RIBBON: a stereo cartoon drawing program for proteins. *J. Appl. Crystallogr.* **21**, 572–576.

Ramakrishnan, C. & Ramachandran, G. N. (1965). Stereochemical criteria or polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys. J.* **5**, 909–933.

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. sect. A*, **42**, 140–149.

Saenger, W. (1980). Cyclodextrin inclusion compounds in research and industry. *Angew. Chem. Int. Ed.* **19**, 344–362.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning. A Laboratory Manual*. 2nd edit.

Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Nat. Acad. Sci., U.S.A.* **74**, 5463–5467.

Schmid, G. (1989). Cyclodextrin glycosyltransferase production: yield enhancement by overexpression of cloned genes. *Tibtech.* **7**, 244–248.

Schmid, G., Englbrecht, A. & Schmid, D. (1988). Cloning and nucleotide sequence of a cyclodextrin glycosyltransferase from the alkalophilic *Bacillus* 1–1. In *Proceedings of the 4th International Symposium on Cyclodextrins* (Huber, O. & Szejtli, J., eds.), pp. 71–76, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Svensson, B. & Sögaard, M. (1992). Protein engineering of amylases. *Biochem. Soc. Trans.* **20**, 34–42.

Svensson, B., Jespersen, H. M., Sierks, M. R. & MacGregor, E. A. (1989). Sequence homology between putative raw-starch binding domains from different starch-degrading enzymes. *Biochem. J.* **264**, 309–311.

Swift, H. J., Brady, L., Derewenda, Z. S., Dodson, E. J., Dodson, G. G., Turkenburg, J. P. & Wilkinson, A. J. (1991). Structure and molecular refinement of *Aspergillus oryzae* (TAKA) α-amylase: an application of the simulated-annealing method. *Acta Crystallogr. sect. B*, **47**, 535–544.

Szejtli, J. (1982). *Cyclodextrins and Their Inclusion Complexes*. Budapest, Akademiai Kiado.

Takano, T., Fukuda, M., Monma, M., Kobayashi, S., Kainuma, K. & Yamane, K. (1986). Molecular cloning, DNA nucleotide sequencing, and expression in *Bacillus subtilis* cells of the *Bacillus macerans* cyclodextrin glucanotransferase gene. *J. Bacteriol.* **166**, 1118–1122.

Takase, K., Matsumoto, T., Mizuno, H. & Yamane, K. (1992). Site-directed mutagenesis of active site residues in *Bacillus subtilis* α-amylase. *Biochim. Biophys. Acta*, **1120**, 281–288.

Takusagawa, F. & Jacobson, R. (1978). The crystal and molecular structure of α-maltose. *Acta Crystallogr. sect. B*, **34**, 213–218.

Tronrud, D. E., Ten Eyck, L. & Matthews, B. W. (1987). An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Crystallogr. sect. A*, **43**, 489–501.

Vikmon, M. (1982). Rapid and simple spectrophotometric method for determination of microamounts of cyclodextrins. In *Proceedings of the First International Symposium on Cyclodextrins (1981)* (Szejtli, J., ed.), pp. 69–74, D. Reidel Publishing Co., Dordrecht, The Netherlands.

Villette, J. R., Krzewinski, F. S., Looten, P. J., Sicard, P. J. & Bouquelet, S. J.-L. (1992). Cyclomaltodextrin glucanotransferase from *Bacillus circulans* E 192. IV. Evidence for a raw starch-binding site and its interaction with a β-cyclodextrin copolymer. *Biotechnol. Appl. Biochem.* **16**, 57–63.

Wang, B. C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* **115**, 90–112.

Edited by R. Huber

(Received 26 July 1993; accepted 26 October 1993)