Acta Cryst. (1996). D52, 1136–1145

X-ray Structure of Cyclodextrin Glucanotransferase from Alkalophilic *Bacillus* Sp. 1011. Comparison of Two Independent Molecules at 1.8 Å Resolution

KAZUAKI HARATA,[a]* KEIKO HAGA,[b] AKIRA NAKAMURA,[b] MASANOBU AOYAGI[b] AND KUNIO YAMANE[b]

[a]*Biomolecules Department, National Institute of Bioscience and Human Technology, 1-1 Higashi, Tsukuba, Ibaraki 305, Japan, and* [b]*Institute of Biological Sciences, University of Tsukuba, Tsukuba, Ibaraki 305, Japan.*
E-mail: harata@nibh.go.jp

(*Received 4 April 1996; accepted 24 June 1996*)

Abstract

Cyclodextrin glucanotransferase (CGTase) is an enzyme which produces cyclodextrins by the degradation of starch. The enzyme from alkalophilic *Bacillus* sp. 1011, consisting of 686 amino acid residues, was crystallized from the solution containing 20% PEG 3000 and 20% 2-propanol at pH 5.6 adjusted with citrate buffer. The space group was $P1$ and the unit cell contained two molecules (V_m = 2.41 Å^3 Da^{-1}). The structure was solved by the molecular replacement method and refined to a conventional R value of 0.161 (R_{free} = 0.211) for the reflections in the resolution range 1.8–10 Å by energy minimization combined with simulated annealing. The molecule consists of five domains, designated *A*–*E*, and its backbone structure is similar to the structure of other bacterial CGTases. The molecule has two calcium binding sites where calcium ions are coordinated by seven ligands, forming a distorted pentagonal bipyramid. The two independent molecules are related by a pseudotwofold symmetry and are superimposed with an r.m.s. deviation value of 0.32 Å for equivalent C^α atoms. Comparison of these molecules indicated the relatively large mobility of domains *C* and *E* with respect to domain *A*. The active site is filled with water molecules forming a hydrogen-bond network with polar side-chain groups. Two water molecules commonly found in the active center of both molecules link to several catalytically important residues by hydrogen bonds and participate in maintaining a similar orientation of side chains in the two independent molecules.

1. Introduction

Cyclodextrin glucanotransferase (E.C. 2.4.1.19, abbreviated to CGTase) is an enzyme which degradates starch and produces cyclic oligosaccharides termed cyclodextrins. Cyclodextrins consist of six or more α-1,4-linked glucose units and form inclusion complexes with a variety of guest molecules because of the large annular cavity in the molecule (Szjitli, 1982). To form inclusion complexes, guest molecules should fit with the size of the cavity, even if only partially (Saenger, 1984; Harata, 1991). Since the formation of inclusion complexes changes the chemical and physical properties of the guest molecules, cyclodextrins have been utilized in pharmaceutical, chemical and food industries. The CGTase from several bacterial sources produces a mixture of several types of cyclodextrins and noncyclic oligosaccharides with different characteristics of the product (Szjitli, 1988). Owing to the difference in the complex-forming ability among cyclodextrins, each individual cyclodextrin should be purified prior to utilization. Therefore, it is an ultimate goal to modify the enzyme for the production of only a specific cyclodextrin.

We have been investigating the structure–function relationship of CGTase from an alkalophilic bacterium, which mainly produces β-cyclodextrin consisting of seven glucose units. The enzyme consisting of 686 amino-acid residues is stable in a markedly wide pH range (5.5–11.0) at 310 K compared with CGTases from other bacterial sources. Our recent analysis of the disproponation reaction has suggested that the transglycosidation reaction proceeds by the 'ping-pong' mechanism (Nakamura, Haga & Yamane, 1994*a*). We have applied site-directed mutagenesis for amino-acid residues which have been considered responsible for the catalytic action and sugar binding. The replacement of amino-acid residues in the active site has produced mutant enzymes having different pH dependency and also different product composition (Nakamura, Haga & Yamane, 1993; Nakamura, Haga & Yamane, 1994*b*; Penninga *et al.*, 1995). To discover the role of amino-acid residues in CGTase, which may reveal the catalytic action in alkaline conditions, the exact spacial position of the amino-acid residues in the three-dimensional enzyme structure is required. The crystal structures of CGTases from *Bacillus stearothermophilus* (Kubota, Matsuura, Sakai & Katsube, 1991), *Thermoanaerobacterium thermosulfurigenes* EM1 (Knegtel *et al.*, 1996), *Bacillus circulans* strain 8 (Klein & Schultz, 1991) and strain 251 (Lawson *et al.*, 1994) have been reported. All these are nonalkalophilic bacteria and the former are thermophilic.

As reported in a previous paper (Haga, Harata, Nakamura & Yamane, 1994), our preliminary crystallographic study indicated that the unit cell contains two independent molecules. Comparison between two independent

molecules will provide information not available from the structure of a single molecule, such as the diversity of the active-site structure and the difference in mobility of domains comprising the molecule. In this paper we report the crystal structure of CGTase from alkalophilic *Bacillus* sp. 1011 at 1.8 Å resolution and will discuss the structural difference between two independent molecules.

2. Materials and methods

2.1. *Purification, crystallization and calcium determination*

Preparation of CGTase from alkalophilic *Bacillus*. sp. 1011 was performed as described in the literature (Nakamura *et al.*, 1992). Crystals were grown in 20%(*w*/*v*) PEG 3000 and 20% (*v*/*v*) 2-propanol solution with 0.1 *M* sodium citrate buffer at pH 5.6 (Haga *et al.*, 1994). Rod-like crystals suitable for X-ray measurements were obtained by repeated macroseeding. The crystals belong to the space group *P*1 with unit-cell dimensions $a = 64.93$, $b = 74.45$, $c = 79.12$ Å, $\alpha = 85.2$, $\beta = 105.0$ and $\gamma = 101.0$. The unit cell contains two molecules ($V_m = 2.41$ Å^3 Da^{-1}).

The calcium content was determined by atomic absorption spectroscopy (AAS) with a Hitachi 180-70 spectrometer and inductively coupled plasma-atomic emission spectrometry (ICP-AES) with a Jarrell Ash ICAP 757V spectrometer. The protein sample (20 mg) in 0.2 ml of Milli Q filtered water was digested with 2 ml of concentrated nitric acid (purchased from Wako Pure Chemical Industries Ltd as analytical grade) and dried on the electric hot plate. The resultant residual was dissolved in 1 ml of 1 *N* nitric acid. The solution was diluted to 10 ml and applied to the quantitative analysis.

2.2. *Data collection*

X-ray diffraction data of CGTase crystals were collected to 2.0 Å resolution on a Nonius FAST diffractometer equipped with a Varian image intensifier and an Elliott GX21 generator (40 kV, 70 mA, focal spot size 0.3 mm). For each crystal, data were collected at three goniometer settings. At the first setting, $(\kappa,\varphi) = (0,0)$, data were collected in the 180° ω-scan range. The second and third (κ,φ) settings were $(-135,0)$ and $(-135,90)$, respectively, with a 70° ω-scan range. The data thus measured were merged to produce a set of unique reflections. Seven such data sets were obtained with an $R_{\rm merge}(I)$ value in the range 0.044–0.062. Further, two data sets with the resolution range 1.8–13.0 Å were collected on a FAST diffractometer with a DEP image intensifier and a FR571 generator (40 kV, 45 mA, focal spot size 0.2 × 2 mm). The $R_{\rm merge}(I)$ values for these data sets were 0.031 and 0.033. The merging of the nine data sets produced a final data set of 111 795 unique reflections (86.1% completeness) with the $R_{\rm merge}(I)$ value 0.080. Data statistics are given in Table 1.

Table 1. *Statistics of data collection and structure refinement*

(*a*) Data collection

Crystal	Resolution	Observed ($F_o > 0$)	Unique	$R_{\rm merge}$
1	2.0–29.6	97717	71105	0.058
2	2.0–27.6	95399	69140	0.053
3	2.1–35.0	84659	60099	0.044
4	2.1–35.0	78438	56911	0.046
5	2.0–29.6	92853	69055	0.062
6	2.0–29.6	99833	72494	0.052
7	2.0–29.6	99063	70960	0.050
8	1.8–13.0	85768	66676	0.033
9	1.8–13.0	85369	67185	0.031
Average	1.8–35.0		111795	0.080

(*b*) Structure refinement

Step	No. of residues	Resolution	No. of reflections	*R* value	$R_{\rm free}$	No. of waters
1	4–686*	2.5–10.0	43411 ($F > 3\sigma$)	0.235		
2	4–686	2.0–5.0	65212 ($F > 3\sigma$)	0.202		
3	4–686	2.0–5.0	65212 ($F > 3\sigma$)	0.188		
4	1–686	1.8–5.0	80786 ($F > 3\sigma$)	0.190		
5	1–686	1.8–10.0	85940 ($F > 3\sigma$)	0.170		735
6	1–686	1.8–10.0	85940 ($F > 3\sigma$)	0.167	0.214	776
7	1–686	1.8–10.0	101409 ($F > 2\sigma$)	0.161	0.211	805

* Coordinates of *B. stearothermophilus* CGTase.

2.3. *Structure determination and refinement*

The structure was solved by molecular replacement method using a set of coordinates of CGTase from *B. stearothermophilus* (Kubota *et al.*, 1991). A self-rotation function calculated using 4–15 Å resolution data indicated the presence of a local twofold axis. The cross-rotation search was made using Patterson vectors in the range 5–20 Å and the 129 solutions obtained were subjected to the PC refinement (Brünger, 1990), which gave a single plausible solution. The relative orientation of the second molecule was estimated from the result of the self-rotation search. To determine the relative position of the second molecule, the translation search was performed using the 4–15 Å resolution data. The rigid-body refinement for the solution with the highest correlation gave an *R* value of 0.41 for the 4–10 Å data and the successive energy minimization was converged at an *R* value of 0.28 for the 3–10 Å resolution data.

The structure was refined using the program *X-PLOR* (Brünger, Kuriyan & Karplus, 1987). Before the replacement of amino-acid residues different from those of *B. stearothermophilus* CGTase, the molecular dynamics simulation was performed at 2000 (0.5 ps, step size 0.001 ps), 1000 (0.5 ps, step size 0.001 ps) and 300 K (0.5 ps, step size 0.001 ps), followed by energy minimization (100 cycles for the coordinate and 30 cycles for the *B* factor) using the 2.5–5 Å resolution data with $|F_o| > 3\sigma(F)$. The side-chain groups were replaced and their orientation was corrected on a $2F_o - F_c$ map. Refinement using the 2.0–5 Å resolution data included

dynamic simulation at 3000, 1000 and 500 K for 0.5 ps with step size 0.0005 ps at each temperature and energy minimization with 50 cycles for coordinates and B factors. Further dynamics simulation and energy minimization were twice performed with the 1.8–5 Å resolution data. Water molecules were picked from $F_o - F_c$ and $2F_o - F_c$ maps. Electron-density peaks which are higher than 3σ in the $F_o - F_c$ map and are involved in suitable intermolecular contacts (2.5–3.3 Å) with protein atoms were taken into account, but water molecules with B values higher than 60 Å^2 were omitted during the refinement. The map calculation and structure refinement by molecular dynamics at 300 K for 1 ps with step size 0.0005 ps and energy minimization (100 cycles for coordinates and 30 cycles for B values), followed by the examination of $F_o - F_c$ and $3F_o - 2F_c$ maps, were twice repeated with the 1.8–10 Å resolution data. The refinement converged to an R value of 0.161 for 101 409 reflections with $|F_o| > 2\sigma(F)$. The final structure model included two protein molecules, four calcium ions and 805 water molecules with the full occupancy. The final $F_o - F_c$ map showed no significant peaks that could be assigned for protein atoms or water positions. Atomic coordinates have been deposited with the Protein Data Bank.*

3. Results and discussion

3.1. *The calcium ion*

It is known that the calcium ion enhances the catalytic activity of CGTase. In the CGTase structures of *B. circulans* strain 8 (Klein & Schulz, 1991) and strain 251 (Lawson *et al.*, 1994), two calcium binding sites have been defined. In the present study the calcium ion content was analytically determined by two spectroscopic methods, AAS and ICP-AES. The molar ratio of protein to calcium ions was 1:1.5 by AAS and 1:1.53 by ICP-AES. The presence of the metal ions Na^+, K^+, Mn^{2+}, Ni^{2+} and Zn^{2+} was also examined, but they were not detectable in comparison with the control experiment. The X-ray result indicates that the CGTase molecule has two calcium binding sites, but they are not fully occupied. A certain amount of calcium might be lost during the purification, because it was carried out under the calcium-free condition.

3.2. *Crystallization and data collection*

CGTase was crystallized in the presence of 20% PEG 3000 and 20% 2-propanol as precipitants. Crystals did not grow in the absence of 2-propanol, but were stable even when 2-propanol was slowly evaporated. Nine crystals were used for the collection of diffraction data. Each set of unique reflections was obtained from one crystal. Nine data sets in total were merged to produce the data set with 86% completeness. For each data set, data collections were performed at three goniometer settings to estimate local scaling factors more accurately at the data merge stage, because of the low crystallographic symmetry. However, as shown in Table 1, the relatively large R_{merge} value for the merge of nine data sets (8.0%) compared with the R_{merge} value of each data set (3.1–6.2%) indicates that the local scaling did not fully correct the intensity fluctuation resulting from absorption and/or deterioration of the crystal.

3.3. *Structure determination and refinement*

The structure was solved by the molecular replacement method using the coordinates of *B. stearothermophillus* CGTase, which has 63% sequence homology (Fig. 1) and refined to $R = 0.161$ ($R_{free} = 0.211$) for 101 409 reflections [$F_o > 2\sigma(F)$, 78% completeness] in the resolution range 1.8–10 Å. In the course of refinement, the electron density strongly indicated the 452nd residue to be Pro in both the two independent molecules, although it was thought to be Arg in the sequence data. Therefore, the 452nd residues were refined as Pro, which was also found at the corresponding position in CGTases from *B. stearothermophilus* and *B. circulans*. The r.m.s. difference of bond distances and angles from their ideal values were 0.013 Å and 2.6°, respectively. The coordinate error estimated from the Luzzati plot (Luzzati, 1952) shown in Fig. 2 was *ca* 0.2 Å. The (φ, ψ) angles fall in the normal range in the Ramachandran plot (Ramakrishnan & Ramachandran, 1965) shown in Fig. 3, except those of Ala152 and Tyr195 ($\varphi \simeq 60$, $\psi \simeq -130°$).

3.4. *Molecular structure*

The molecule consists of five domains designated *A–E*, as shown in Fig. 4. The backbone structure is very similar to those of other CGTases from *B. stearothermophillus* (Kubota *et al.*, 1991), *B. circulans* strain 8 (Klein & Schulz, 1991) and strain 251 (Lawson *et al.*, 1994), as shown in Fig. 5 Domain *A* (residues 1–406) shows a typical $(\beta\alpha)_8$ barrel structure with an additional α-helix (residues 285–292) and a long loop region containing a β-sheet (residues 80–110). The domain *B* (residues 139–203), consisting of a short α-helix, short strands and loops, is located at the wider side of the barrel, where the active site is situated. Only β-sheet structures are observed in domains *C* (residues 407–496), *D* (residues 497–584) and *E* (residues 585–686).

The asymmetric unit of the crystal contains two independent molecules. Molecule (2) is superimposed on molecule (1) by the least-squares method. The difference between equivalent C^α atoms was plotted in Fig. 6. The r.m.s. difference is 0.32 for C^α atoms. A marked

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1PAM, R1PAMSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: TS0001).

```
                          1                                                  50
B. sp1011                 APDTSVSNKQ  NFSTDVIYQI  FTDRFSDGNP  ANNPTGAAFD  GSCTNLRLYC
B. circulans strain 8     DPDTAVTNKQ  SFSTDVIYQV  FTDRFLDGNP  SNNPTGAAYD  ATCSNLKLYC
B. circulans strain 251   APDTSVSNKQ  NFSTDVIYQI  FTDRFSDGNP  ANNPTGAAFD  GTCTNLRLYC
B. stearothermophilus     ---AGNLNKV  NFTSDVVYQI  VVDRFVDGNT  SNNPSGALFS  SGCTNLRKYC
                         51                                                 100
B. sp1011                 GGDWQGIINK  INDGYLTGMG  ITAIWISQPV  ENIYSVINYS  GVNNTAYHGY
B. circulans strain 8     GGDWQGLINK  INDNYFSDLG  VTALWISQPV  ENIFATINYS  GVTNTAYHGY
B. circulans strain 251   GGDWQGIINK  INDGYLTGMG  VTAIWISQPV  ENIYSIINYS  GVNNTAYHGY
B. stearothermophilus     GGDWQGIINK  INDGYLTGMG  ITAIWISQPV  ENVFSVMN-D  ASGSASYHGY
                        101                                                 150
B. sp1011                 WARDFKKTNP  AYGTMQDFKN  LIDTAHAHNI  KVIIDFAPNH  TSPASSDDPS
B. circulans strain 8     WARDFKKTNR  YFGTMADFQN  LITTAHAKGI  KIVIDFAPNH  TSPAMETDTS
B. circulans strain 251   WARDFKKTNP  AYGTIADFQN  LIAAAHAKNI  KVIIDFAPNH  TSPASSDQPS
B. stearothermophilus     WARDFKKPNP  FFGTLSDFQR  LVDAAHAKGI  KVIIDFAPNH  TSPASETNPS
                        151                                                 200
B. sp1011                 FAENGRLYDN  GNLLGGYTND  TQNLFHHYGG  TDFSTIENGI  YKNLYDLADL
B. circulans strain 8     FSENGRLYDN  GTLVGGYTND  TNGYFHHNGG  SDFSSLENGI  YKNLYDLADF
B. circulans strain 251   FAENGRLYDN  GTLLGGYTND  TQNLFHHNGG  TDFSTTENGI  YKNLYDLADL
B. stearothermophilus     YMENGRLYDN  GTLLGGYTND  ANMYFHHNGG  TTFSSLEDGI  YRNLFDLADL
                        201                                                 250
B. sp1011                 NHNNSSVDVY  LKDAIKMWLD  LGVDGIRVDA  VKHMPFGWQK  SFMATINNYK
B. circulans strain 8     NHNNATIDKY  FKDAIKLWLD  MGVDGIRVDA  VKHMPLGWQK  SWMSSIYAHK
B. circulans strain 251   NHNNSTVDVY  LKDAIKMWLD  LGIDGIRMDA  VKHMPLGWQK  SFMAAVNNYK
B. stearothermophilus     NHQNPVIDRY  LKDAVKMWID  MGIDGIRMDA  VKHMPFGWQK  SLMDEIDNYR
                        251                                                 300
B. sp1011                 PVFTFGEWFL  GVNEISPEYH  QFANESGMSL  LDFRFAQKAR  QVFRDNTDNM
B. circulans strain 8     PVFTFGEWFL  GSAASDADNT  DFANKSGMSL  LDFRFNSAVR  NVFRDNTSNM
B. circulans strain 251   PVFTFGEWFL  GVNEVSPENH  KFANESGMSL  LDFRFAQKVR  QVFRDNTDNM
B. stearothermophilus     PVFTFGEWFL  SENEVDANNH  YFANESGMSL  LDFRFAQKLR  QVLRNNSDNW
                        301                                                 350
B. sp1011                 YGLKAMLEGS  EVDYAQVNDQ  VTFIDNHDME  RFHTSNGDRR  KLEQALAFTL
B. circulans strain 8     YALDSMINST  ATDYNQVNDQ  VTFIDNHDMD  RFKTSAVNNR  RLEQALAFTL
B. circulans strain 251   YGLKAMLEGS  AADYAQVDDQ  VTFIDNHDME  RFHASNANRR  KLEQALAFTL
B. stearothermophilus     YGFNQMIQDT  ASAYDEVLDQ  VTFIDNHDMD  RFMIDGGDPR  KVDMALAVLL
                        351                                                 400
B. sp1011                 TSRGVPAIYY  GSEQYMSGGN  DPDNRARLPS  FSTTTTAYQV  IQKLAPLRKS
B. circulans strain 8     TSRGVPAIYY  GTEQYLTGNG  DPDNRAKMPS  FSKSTTAFNV  ISKLAQLRKS
B. circulans strain 251   TSRGVPAIYY  GTEQYMSGGT  DPDNRARIPS  FSTSTTAYQV  IQKLAPLRQC
B. stearothermophilus     TSRGVPNIYY  GTEQYMTGNG  DPNNRKMMSS  FNKNTRAYQV  IQKLSSLRRN
                        401                                                 450
B. sp1011                 NPAIAYGSTE  ERWINNDVII  YERKFGNNVA  VVAINRNMNT  PASITGLVTS
B. circulans strain 8     NPAIAYGSTQ  QRWINNDVYV  YERKFGKSVA  VVAVNRNLST  SASITGLSTS
B. circulans strain 251   NPAIAYGSTQ  ERWINNDVLI  YERKFGSNVA  VVAVNRNLNA  PASISGLVTS
B. stearothermophilus     NPALAYGDTE  QRWINGDVYV  YERQFGKDVV  LVRVNRSSSS  NYSITGLFTA
                        451                                                 500
B. sp1011                 LPRGSYNDVL  GGILNGNTLT  VGAGGAASNF  TLAPGGTAVW  QYTTDATTPI
B. circulans strain 8     LPTGSYTGVL  GGVLNGNNIT  STNGSIN-NF  TLAAGATAVW  QYTTAETTPT
B. circulans strain 251   LPQGSYNDVL  GGLLNGNTLS  VGSGGAASNF  TLAAGGTAVW  QYTAATATPT
B. stearothermophilus     LPAGTYTDQL  GGLLDGNTIQ  VGSNGSVNAF  DLGPGEVGVW  AYSATESTPI
                        501                                                 550
B. sp1011                 IGNVGPMMAK  PGVTITIDGR  GFGSGKGTVY  FGTTAVTGAD  IVAWEDTQIQ
B. circulans strain 8     IGHVGPVMGK  PGNVVTIDGR  GFGSTKGTVY  FGTTAVTGAA  ITSWEDTQIK
B. circulans strain 251   IGHVGPMMAK  PGVTITIDGR  GFGSSKGTVY  FGTTAVSGAD  ITSWEDTQIK
B. stearothermophilus     IGHVGPMMGQ  VGHQVTIDGE  GFGTNTGTVK  FGTTA---AN  VVSWSNNQIV
                        551                                                 600
B. sp1011                 VKIPAVPGGI  YDIRVANAAG  AASNIYDNFE  VLTGDQVTVR  FVINNATTAL
B. circulans strain 8     VTIPSVAAGN  YAVKVA-ASG  VNSNAYNNFT  ILTGDQVTVR  FVVNNASTTL
B. circulans strain 251   VKIPAVAGGN  YNIKVANAAG  TASNVYDNFE  VLSGDQVSVR  FVVNNATTAL
B. stearothermophilus     VAVPNVSPGK  YNITVQSSSG  QTSAAYDNFE  VLTNDQVSVR  FVVNNATTNL
                        601                                                 650
B. sp1011                 GQNVFLTGNV  SELGNWDPN-N  AIGPMYNQVV  YQYPTWYYDV  SVPAGQTIEF
B. circulans strain 8     GQNLYLTGNV  AELGNWSTGST  AIGPAFNQVI  HQYPTWYYDV  SVPAGKQLEF
B. circulans strain 251   GQNVYLTGSV  SELGNWDPA-K  AIGPMYNQVV  YQYPNWYYDV  SVPAGKTIEF
B. stearothermophilus     GQNIYIVGNV  YELGNWDTS-K  AIGPMFNQVV  VSYPTWYIDV  SVPEGKTIEF
                        651                                    686
B. sp1011                 KFLKKQG-STV  TWEGGANRTF  TTPTSGTATV  NVNWQP
B. circulans strain 8     KFFKKNG-STI  TWESGSNHTF  TTPASGTATV  TVNWQ-
B. circulans strain 251   KFLKKQG-STV  TWEGGSNHTF  TAPSSGTATI  NVNWQ
B. stearothermophilus     KFIKKDSQGNV  TWESGSNHVY  TTPTNTTGKI  IVDWQ
```

Fig. 1. Amino-acid sequence of bacterial CGTases, whose structure was determined by X-ray analysis.

difference is observed in the residue range 335–338, which is a loop region linking a short β-strand (residues 320–322) and an α-helix (residues 339–350). The largest difference of 3.60 Å is observed at the Asn336 residue. Relatively large differences are also observed in the region of residues 140–200 (domain *B*), 440–500 (domain *C*) and 600–686 (domain *E*).

The *B* values of the main chain are plotted against the residue number in Fig. 7. The average *B* value in molecule 1 is 18.4 Å^2 for all atoms, 17.4 Å^2 for main-chain atoms and 19.5 Å^2 for side-chain atoms. The corresponding values in molecule 2 are 22.2, 21.4 and 23.1 Å^2. The difference of average *B* values between molecules 1 and 2 for each residue is plotted in Fig. 8. A relatively large difference is observed in the regions 440–490 (domain *C*) and 590–686 (domain *E*), which coincide with the regions of a relatively large positional difference.

Comparison of two independent molecules indicates the relatively high mobility of domains *C* and *E* with respect to domain *A*. The r.m.s. deviations for C$^\alpha$ atoms in each domain are 0.29 (domain *A*), 0.32 (domain *B*), 0.41 (domain *C*), 0.16 (domain *D*) and 0.39 Å (domain *E*). The relatively large r.m.s. deviation value of domain *A* is ascribed to the conformational difference in the residue region 335–338, as shown in Fig. 6. When these residues are excluded, the r.m.s. deviation value in domain *A* decreases to 0.17 Å, which is similar to the value of domain *D*. The larger deviation of domains *C* and *E* is ascribed to their hinge-like connection to other domains. A β-sheet consisting of three strands in domain *C* faces α-helices of domain *A* (Fig. 4). The two loops linking domain *C* to domains *A* and *D* form a hinge region. A β-sheet in domain *E* is in contact with two α-helices of domain *A*. On the other hand, domain *D* is situated between domains *C* and *E*. A β-sheet in domain *D* has contacts with loops of domain *A*. The mobility of each domain is affected by the contacts with domain *A*, because there are no short contacts among domains *C*, *D* and *E*. Domain *D* has 14 amino-acid residues having polar contacts less than 3.3 Å with domain *A* (Table 2). On the other hand, 11 and ten amino-acid residues are found in domains *C* and *E*, respectively.

Gerstein, Lesk & Chothia (1994) have classified the domain movements into two types, the hinge motion and the shear motion. In the present structure comparison of the two structures indicates a shear-like motion of domains *C* and *E*. At interfaces between these domains and domain *A*, many side-chain groups are interlocked and hydrogen bonds or salt bridges are mostly conserved in the two independent molecules, in spite of the movement of the backbone structure. The movement may be compensated at the interface by the accumulation of a small conformational change of side-chain groups.

3.5. *Intermolecular contacts between independent molecules*

The two independent molecules are related by pseudotwofold symmetry, as shown in Fig. 9. The close contacts of the two molecules are observed in the regions 301–304 (domain *A*), 411–414 (domain *C*) and 562–564 (domain *D*), as shown in Fig. 10. The carboxyl groups of Glu411 in the two molecules are related by pseudotwofold symmetry and are in contacts with the distance 2.6 Å, suggesting a strong hydrogen bond. Side-chain groups of Tyr301 and Lys304 are hydrogen-bonded to the peptide O atoms of Ile414 and Glu411, respectively. The hydroxyphenyl group of Tyr301 faces the planar side-chain group of Trp413 of the other molecule in the hydrophobic interaction. In the residue region 562–564, a salt linkage is formed between Asp562 and Arg564, and five water-mediated hydrogen-bond bridges are formed, as listed in Table 3.

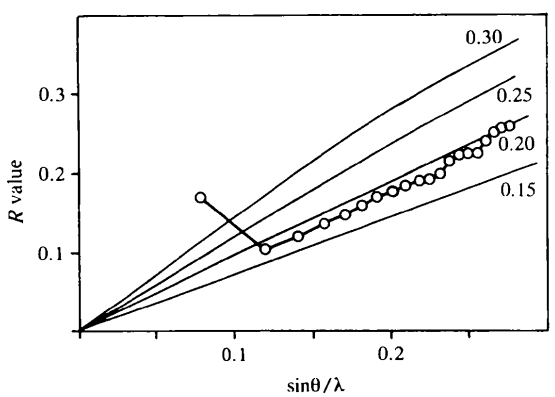

Fig. 2. Plot of the *R* value against the resolution. Solid curves are plots of the upper estimate of coordinate error according to Luzzati (1952).

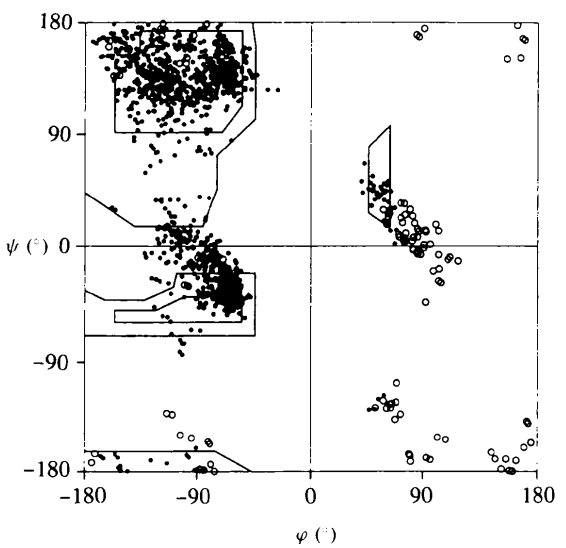

Fig. 3. Ramachandran plot of (φ, ψ) angles for two molecules. Open circles denote the glycine residues.

The formation of the dimer-type structure has not been found in crystals of other CGTases, because amino-acid residues located at the dimer interface vary among four CGTases and only the present CGTase molecule has a sequence suitable for the formation of the dimer structure. The dimer interface does not involve the active-site region or sugar binding site located at domain *E*.

3.6. *Geometry of the calcium binding sites*

The molecule contains two incompletely occupied calcium ions, as also determined by analytical methods. They are located in the loop regions of domain *A* and stabilize the conformational structure of flexible regions (Fig. 4). Especially, the coordination bond with the backbone carbonyl O atoms is effective in hindering the main-chain movement. The calcium ion near the N terminal is surrounded by two backbone carbonyl O atoms, four side-chain O atoms and one water molecule (Fig. 11*a*), and they form a distorted pentagonal bipyramid with OD1(Asn33) and OD2(Asp53) located at the tops of the bipyramid (Table 4). A similar but more distorted pentagonal bipyramid, formed by three water molecules, O(Ile90), OD1(Asn139) for the pentagon and OD1(Asp199) and O(His233) for the two tops, is observed in the coordination geometry of the second

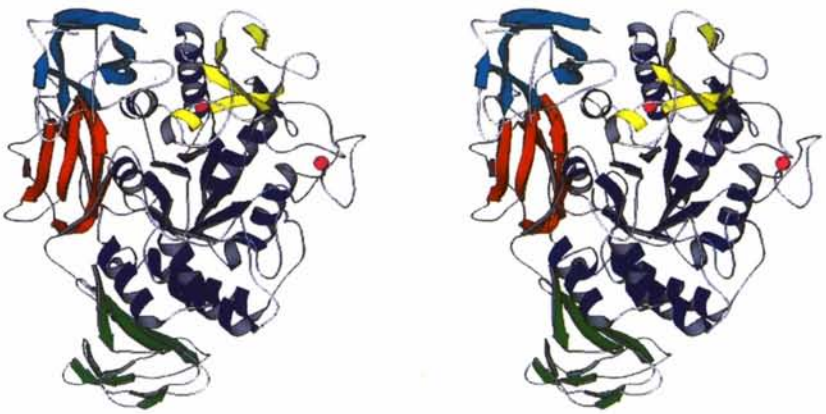

Fig. 4. The backbone structure of CGTase drawn using the program *MolScript* (Kraulis, 1991). The five domains are shown with colors, blue (domain *A*), yellow (domain *B*), green (domain *C*), red (domain *D*) and light blue (domain *E*). Calcium ions are denoted by pink-colored circles.

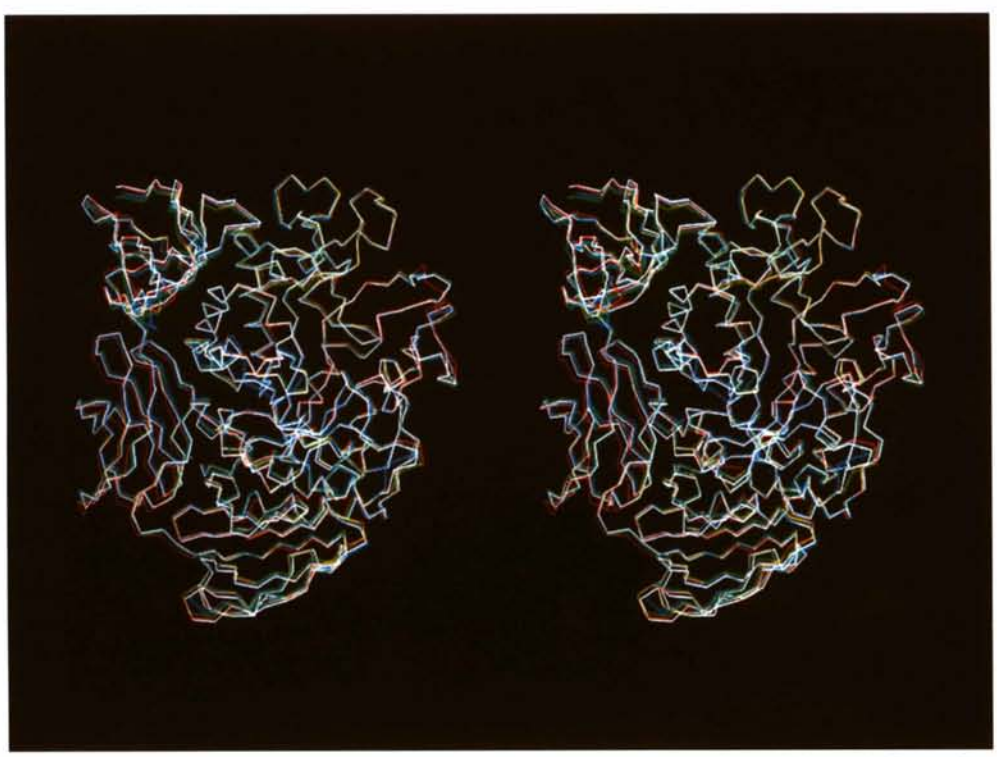

Fig. 5. Superposition of the backbone structure of CGTase from *B.* spp. 1011 (white), *B. stearothermophilus* (red), *B. circulans* strain 8 (blue) and *B. circulans* strain 251 (green).

calcium ion (Fig. 11*b*). This calcium ion is found near the active-site region and considered to be responsible for stabilizing the active-site structure. As given in Table 4, there is no significant difference in the coordination geometry between the two CGTase molecules.

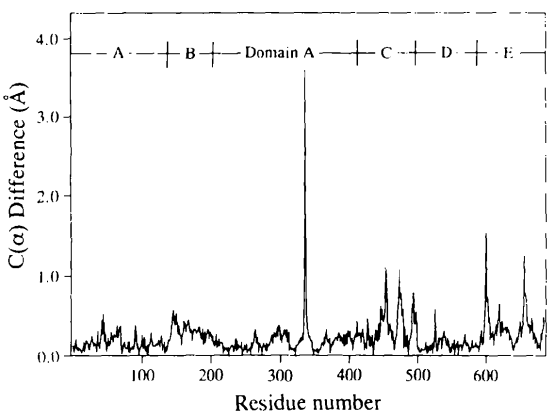

Fig. 6. Plot of the distance between equivalent C$^{\alpha}$ atoms after the least-squares superposition of two independent molecules.

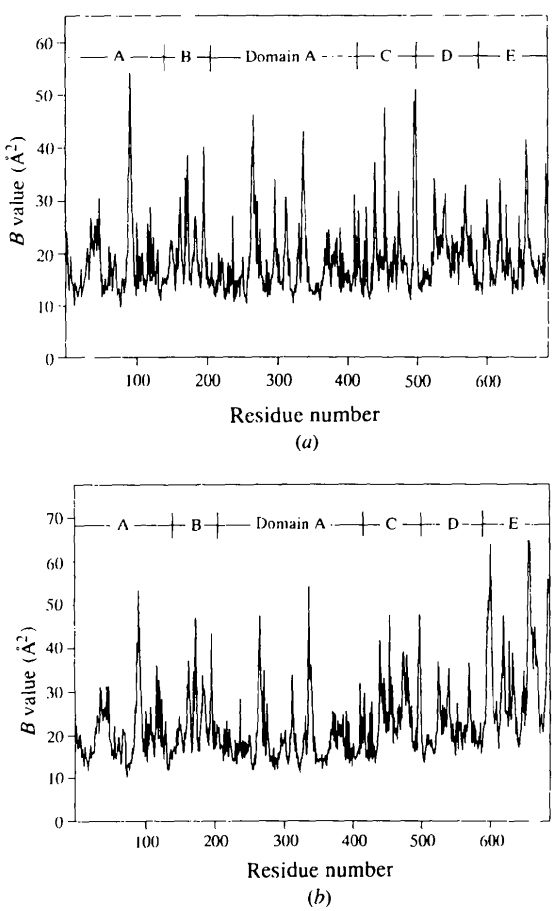

Fig. 7. Plot of the average *B* value of the main-chain peptide group (*a*) for molecule (1) and (*b*) molecule (2).

Table 2. *Contacts between domains*

In the atomic notation, G, D, E, Z and H refer to γ, δ, ε, ζ and η atom in the side chain, respectively.

(*a*) Domains *A* and *C*

Domain *A*	Domain *C*	Distance (Å) Molecule 1	Molecule 2
ND2 (Asn40)	OE1 (Gln491)	—	3.2
O (Val292)	NH1 (Arg436)	2.8	2.8
O (Asp298)	NH1 (Arg436)	2.8	2.7
NZ (Lys304)	OE1 (Glu411)	2.9	2.9
OE2 (Glu311)	N (Thr409)	—	2.9
NH1 (Arg340)	OD1 (Asn465)	2.9	2.9
NH1 (Arg340)	O (Gly485)	2.8	3.1
NE2 (Gln344)	O (Asn435)	2.8	2.8
NE2 (Agn344)	O (Gly485)	3.1	2.9
O (Thr351)	NH2 (Arg423)	2.9	2.8
OG1 (Thr351)	OH (Tyr421)	2.8	2.9

(*b*) Domains *A* and *D*

Domain *A*	Domain *D*	Distance (Å) Molecule 1	Molecule 2
N (Ala1)	OE2 (Glu544)	3.1	3.3
N (Asp3)	OG1 (Thr547)	3.0	3.1
ND2 (Asn8)	OD1 (Asp518)	3.0	2.9
NZ (Lys9)	OD2 (Asp224)	2.8	—
OE1 (Gln10)	OG1 (Thr516)	2.9	2.9
ND2 (Asn11)	OD1 (Asp518)	2.9	3.0
OG (Ser13)	OD1 (Asn503)	2.7	2.9
NZ (Lys240)	OE1 (Glu580)	3.2	3.1
ND2 (Asn247)	O (Met508)	2.9	3.2
OD1 (Asn248)	N (Lys510)	3.0	3.0
O (Asn247)	ND2 (Asn578)	3.3	3.2
NE2 (Gln316)	OD1 (Asn578)	2.9	3.1
ND2 (Asn318)	O (Gly502)	2.9	3.0
OD2 (Asp319)	ND2 (Asn503)	3.2	3.2

(*c*) Domains *A* and *E*

Domain *A*	Domain *E*	Distance (Å) Molecule 1	Molecule 2
OG1 (Thr185)	OE1 (Gln628)	3.1	—
OE1 (Glu187)	N (Glu628)	3.0	2.9
OE1 (Glu187)	N (Val629)	2.7	2.8
OE1 (Glu187)	OH (Tyr637)	—	2.6
OE2 (Glu187)	NE2 (Gln628)	2.9	3.1
OH (Tyr191)	OD1 (Asp639)	2.7	2.8
O (Asn203)	OG1 (Thr677)	2.7	2.7
OG (Ser205)	N (Thr677)	3.2	3.0
OG (Ser205)	OG1 (Thr677)	3.0	3.2
OG (Ser241)	OG1 (Thr588)	2.7	2.9

Table 3. *Contacts between the two molecules*

Molecule 1		Molecule 2	Distance (Å)	
(*a*) Direct contacts				
OH (Tyr301)		O (Ile414)	2.8	
NZ (Lys304)		O (Glu411)	3.1	
O (Glu411)		NZ (Lys304)	3.0	
OE1 (Glu411)		OE1 (Glu411)	2.6	
O (Ile414)		OH (Tyr301)	2.8	
OD2 (Asp562)		NH2 (Arg564)	3.0	
NH2 (Arg564)		OD2 (Asp562)	2.9	
(*b*) Water-mediated contacts				
NH1 (Arg412)	H_2O	OE2 (Glu308)	2.8	3.1
OH (Tyr301)	H_2O	O (Ile414)	2.7	3.0
O (Gly446)	H_2O	O (Tyr301)	2.8	2.7
OG1 (The445)	H_2O	ND2 (Asn416)	2.9	2.8
OE2 (Gly308)	H_2O	NH1 (Arg412)	2.7	2.7

3.7. *Active site*

Amino-acid residues in the active center are mostly conserved among CGTases. Three residues, Asp229, Glu257 and Asp328, have been considered as the residues responsible for the catalytic action and are located at the bottom of a surface cavity in the *A* domain. Recently, Strokopytov *et al.* (1995) reported the structure of the acarbose complex with CGTase from *B. circulans*, where these residues are hydrogen bonded to acarbose. As shown in Fig. 12, the active-site cleft is filled with water molecules which form a hydrogen-bond network linking side-chain groups of polar amino-acid residues. The carboxyl group of Glu257, which has been suggested as a proton donor, is at a distance equivalent to a hydrogen bond or salt linkage with Arg227, His327 and Asp328, and also linked to Asp229 and Tyr100 through water-mediated hydrogen-bond bridges (Fig. 13). Asp229 and Asp328 have been considered as essential residues for the catalytic action, because the replacement of these residues by Ala or Asn almost inactivated the enzyme (Klein & Schultz, 1991; Nakamura *et al.*, 1992). Furthermore, the enzyme activity was almost lost by the replacement of the His327 residue with Arg, while replacement with Asn drastically decreased the activity in the alkaline pH condition (Nakamura *et al.*, 1993). Klein, Hollender, Bender & Shultz (1992) have suggested the importance of Asp328 in the vicinity of Glu257 to increase the pK_a

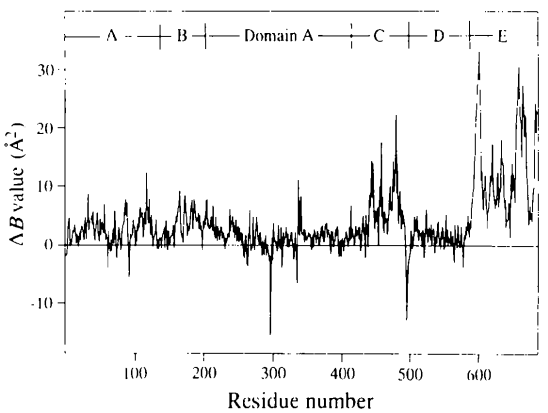

Fig. 8. Plot of the difference in *B* value between molecules (1) and (2).

Table 4. *Geometry of calcium binding*

Calcium ion	Ligands	Molecule 1 distance (Å)	Molecule 2 distance (Å)
Ca1	OD2 (Asp27)	2.4	2.4
	O (Asn29)	2.3	2.3
	OD1 (Asn32)	2.2	2.1
	OD1 (Asn33)	2.1	2.1
	O (Gly51)	2.3	2.3
	OD2 (Asp53)	2.4	2.5
	H_2O	2.3	2.2
Ca2	OD1 (Asn139)	2.3	2.3
	O (Ile190)	2.4	2.4
	OD1 (Asp199)	2.8	2.7
	O (His233)	2.4	2.3
	H_2O	2.5	2.5
	H_2O	2.3	2.2
	H_2O	2.5	2.5

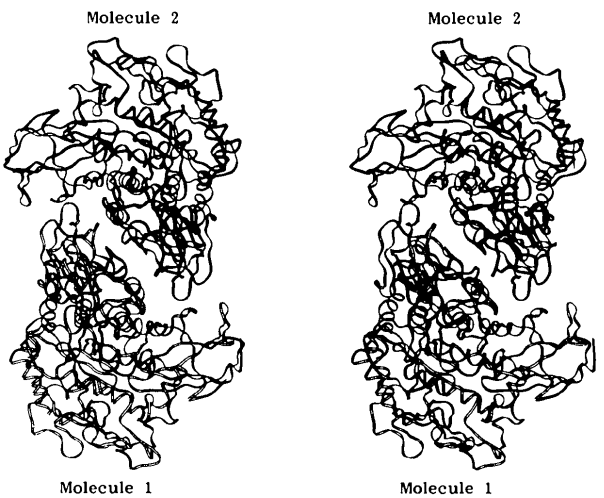

Fig. 9. Stereoview representing two independent molecules related by pseudotwofold symmetry.

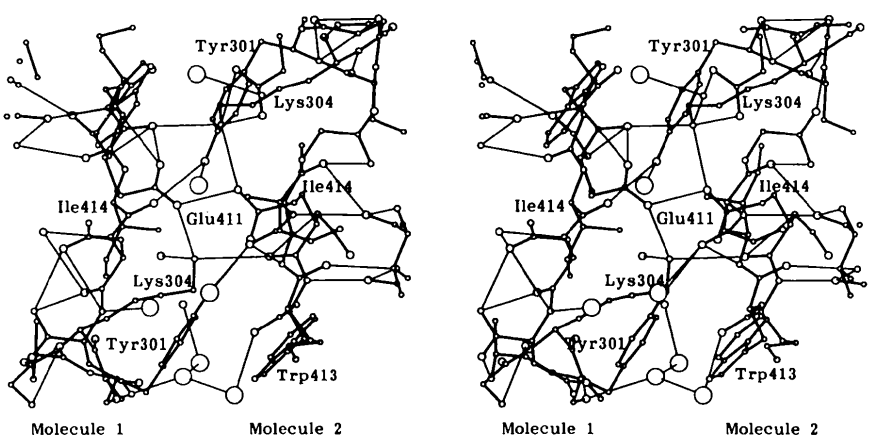

Fig. 10. Stereoview of the interface of the two molecules viewed along the pseudotwofold axis. Nonbonded polar contacts less then 3.3 Å are denoted by thin lines.

value of Glu257 as a proton donor. The imidazolyl group of His327 is within hydrogen-bonding distance to the carboxyl group of Glu257. The combination of Glu257 with His327 may also play an important role in the relay system of enzyme action at high pH conditions. There are two water molecules which link the side-chain groups of these residues. One water molecule is surrounded by Arg227, Asp229, Glu257, His327, and Tyr100 within hydrogen-bonding distance. Another water molecule links the carboxyl groups of Glu257 and Asp229. These two water molecules are commonly found in the two independent molecules and seem to be important to maintain the side-chain orientation in the active site. In spite of the distinct difference in the pH dependency of the enzyme activity, the active-site structure is very similar between the present alkalophilic CGTase and nonalkalophilic CGTases. The structure analysis of CGTase complexes with sugars, which is

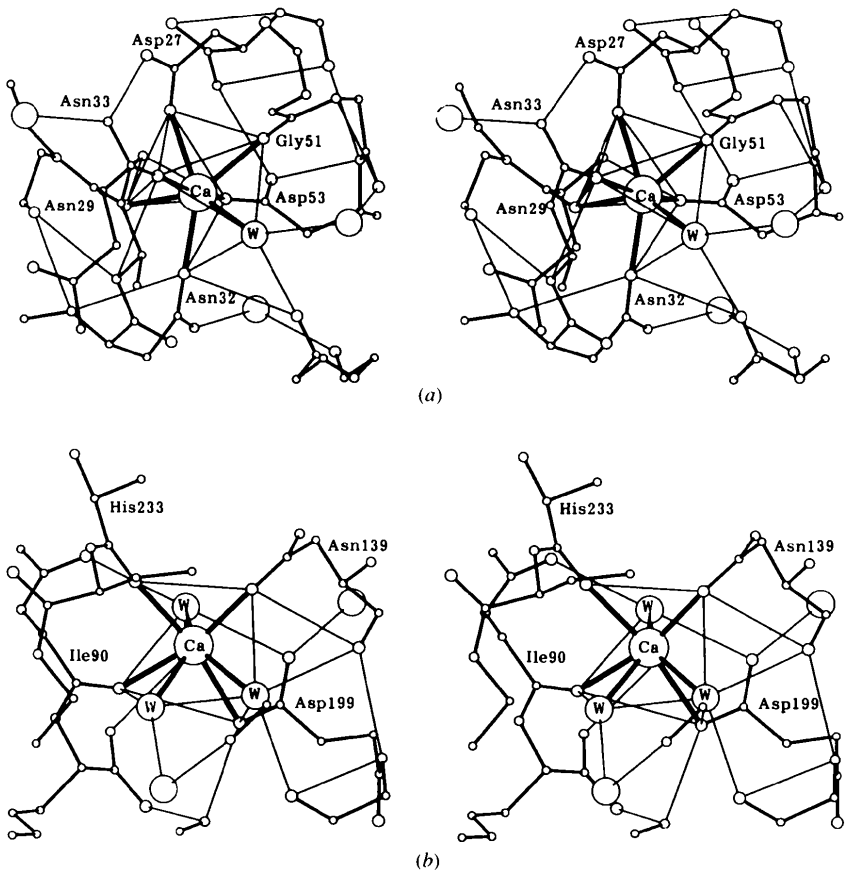

Fig. 11. Stereoviews of two calcium binding sites. Water molecules coordinated to the calcium ions are denoted by W. Nonbonded polar contacts are shown by thin lines.

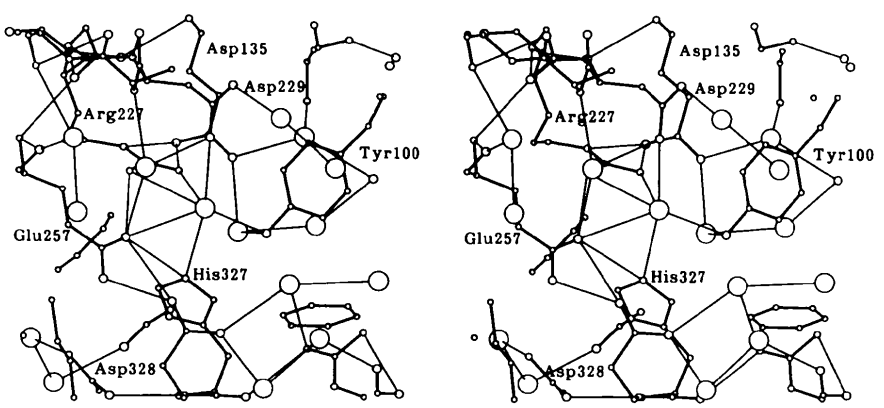

Fig. 12. Stereoview of the active center of molecule (1). Water molecules are shown by large circles. Nonbonded polar contacts are shown by thin lines.

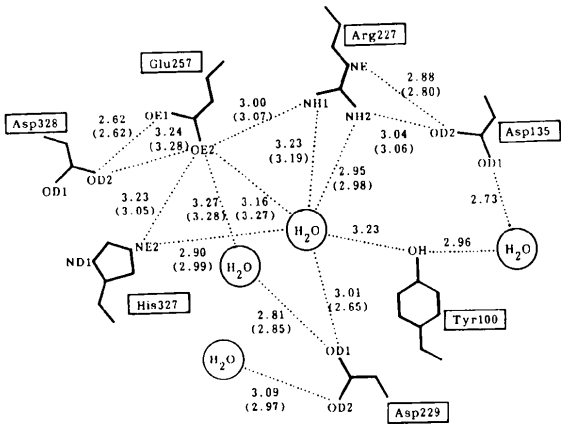

Fig. 13. Schematic drawing of the geometry of active center. Interatomic distances in molecule (2) are given in parentheses.

in progress, will provide further information on the mechanism of enzyme action, especially in alkaline conditions.

The authors wish to thank Professor Yoshiki Matsuura of Osaka University for kindly providing the atomic coordinates of *B. stearothermophilus* CGTase and Dr Peter Rehse for critical reading of the manuscript.

References

Brünger, A. T. (1990). *Acta Cryst.* A**46**, 46–57.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.

Gerstein, M., Lesk, A. M. & Chothia, C. (1994). *Biochemistry*, **33**, 6739-6749.

Haga, K., Harata, K., Nakamura, A. & Yamane, K. (1994). *J. Mol. Biol.* **237**, 163–164.

Harata, K. (1991). In *Inclusion Compounds*, edited by J. L. Atwood, J. E. D. Davies & D. D. MacNicol, Vol. 5, pp. 311–344. New York: Oxford Science Publishers.

Klein, C., Hollender, J., Bender, H. & Schultz, G. E. (1992). *Biochemistry*, **31**, 8740–8746.

Klein, C. & Schultz, G. E. (1991). *J. Mol. Biol.* **217**, 737–750.

Knegtel, R. M. A., Wind, R. D., Rozeboom, H. J., Kalk, K. H., Buitelaar, R. M., Dijkhuizen, L. & Dijkstra, B. W. (1996). *J. Mol. Biol.* **256**, 611–622.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Kubota, M., Matsuura, Y., Sakai, S. & Katsube, Y. (1991). *Denpun Kagaku*, **38**, 141–146.

Lawson, C. L., van Montfort, R., Strokopytov, B., Rozeboom, H. J., Kalk, K. H., de Vries, G. E., Penninga, D., Dijkhuizen, L. & Dijkstra, B. W. (1994). *J. Mol. Biol.* **236**, 590–600.

Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.

Nakamura, A., Haga, K., Ogawa, S., Kuwano, K., Kimura, K. & Yamane, K. (1992). *FEBS Lett.* **296**, 37–40.

Nakamura, A., Haga, K. & Yamane, K. (1993). *Biochemistry*, **32**, 6624–6631.

Nakamura, A., Haga, K. & Yamane, K. (1994*a*). *FEBS Lett.* **337**, 66–70.

Nakamura, A., Haga, K. & Yamane, K. (1994*b*). *Biochemistry*, **33**, 9929–9936.

Penninga, D., Strokopytov, B., Rozeboom, H. J., Lawson, C. L., Dijkstra, B. W., Bergsma, J. & Dijkhuizen, L. (1995). *Biochemistry*, **34**, 3368–3376.

Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.

Saenger, W. (1984). *Inclusion Compounds*, edited by J. L. Atwood, J. E. D. Davies & D. D. MacNicol, Vol. 2, pp. 231–259. London: Academic Press.

Strokopytov, B., Penninga, D., Rozeboom, H. J., Kalk, K. H., Dijkhuizen, L. & Dijkstra, B. W. (1995). *Biochemistry*, **34**, 2234–2240.

Szjitli, S. (1982). *Cyclodextrins and Their Inclusion Complexes*. Budapest: Acadmiai Kiad.

Szjitli, S. (1988). *Cyclodextrin Technology*. Dordrecht: Kluwer Academic Publishers.