A super-secondary structure predicted to be common to several α -1,4-D-glucan-cleaving enzymes

E. Ann MACGREGOR*† and Birte SVENSSON†

*Department of Chemistry, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada, and

†Department of Chemistry, Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-2500 Copenhagen Valby, Denmark

Predictions of protein secondary structure are used with amino acid sequence alignments to show that the *N*-terminal domains of cyclodextrin glucanotransferases and a yeast α -glucosidase may have the same super-secondary structure as α -amylases, i.e. an $(\alpha/\beta)_8$ -barrel fold. Sequence similarities provide evidence that glucanotransferases, and possibly the glucosidase, are, like α -amylases, Ca^{2+} -containing enzymes. The relationship between substrate specificity and the nature of the amino acid residues proposed at the active site is discussed for the transferases and α -glucosidase. A set of three programs for an Apple IIe computer to carry out the calculations described by Garnier, Osguthorpe & Robson [(1978) *J. Mol. Biol.* **120**, 97–120] and a set of four programs for an Apple IIe computer to carry out the calculations described by Levin, Robson & Garnier [(1986) *FEBS Lett.* **205**, 303–308] have been deposited as Supplementary Publication SUP 50149 (25 pages) at the British Library Document Supply Centre, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in *Biochem. J.* (1989) **257**, 5.

INTRODUCTION

α -Amylases (α -1,4-glucan-4-glucanohydrolase, EC 3.2.1.1), cyclodextrin glucanotransferases [α -1,4-glucan-4-glycosyltransferase (cyclizing), CGTase, EC 2.4.1.19] and α -glucosidases (α -1,4-glucosidase glucohydrolase, EC 3.2.1.20) all catalyse cleavage of α -1,4-glucosidic bonds in malto-oligosaccharides or polysaccharides such as amylose and amylopectin. It is therefore possible that common features exist at the active sites of the three types of enzyme. Indeed, local amino acid sequence similarities have been found in these enzymes, and their relationship to the active site has been discussed [1]. Further, considerable sequence similarities have been demonstrated between two CGTases and the α -amylase of *Aspergillus oryzae*, Taka-amylase A (TAase) [2].

α -Amylases from different sources, although varying widely in amino acid sequence, are believed to have an $(\alpha/\beta)_8$ -barrel super-secondary structure [3], i.e. a barrel of eight parallel β -strands surrounded by eight α -helices, as demonstrated by X-ray crystallography for TAase [4] and pig pancreatic α -amylase [5].

In the present paper we report results of computerized predictions of secondary structure that, along with sequence similarities, provide evidence for extensive similarities between the three-dimensional structures of the catalytic domains of CGTases, an α -glucosidase and α -amylases. Examination of sequence alignments indicates that these enzymes probably share key structural features at the active centres.

METHODS

Published amino acid sequences were used for structure predictions. These either have been determined by direct amino acid sequencing for TAase [6] and pig pancreatic

α -amylase [7,8], or have been deduced from nucleotide sequences for an 'amylase' of *Bacillus circulans* [9], for CGTases from *Bacillus macerans* [2,10], *Bacillus stearothermophilus* [2], alkalophilic *Bacillus* spp. strains 1011 [11] and 38-2 [12] and *Klebsiella pneumoniae* [13], and for an α -glucosidase from *Saccharomyces cerevisiae* [14].

Amino acid alignments were made by using the algorithm of Levin *et al.* [15], programmed for an Apple IIe computer, taking into account sequence similarities already demonstrated between α -amylases [3], CGTases [1,11,13] and α -glucosidase [1].

The method of Garnier *et al.* [16], programmed for an Apple IIe computer, was used with decision constants $\text{DC}_H = -75$ and $\text{DC}_S = -87.5$ to predict secondary structure. The predictions were improved, as described by MacGregor [3], with the aid of the method of Cid *et al.* [17].

RESULTS AND DISCUSSION

Primary, secondary and tertiary structure

Approx. 400 amino acid residues of the *N*-terminal region of α -amylases are believed to be folded to give an $(\alpha/\beta)_8$ -barrel domain [3–5], which contains the active site, and which, according to Buisson *et al.* [5], has a small separate domain inserted between one of the β -strands and a helix of the barrel. The *C*-terminal domain of α -amylases varies in length and is, at present, of unknown function.

Svensson, Kimura *et al.* and Binder *et al.* found similarities between the *N*-terminal regions of α -amylases, CGTases [1,11,13] and an α -glucosidase [1]. It was further proposed that the CGTase of *Bacillus* sp. strain 1011 is a two-domain protein, with the catalytic site in the *N*-

Abbreviations used: CGTase, cyclodextrin glucanotransferase; TAase, Taka-amylase A.

† To whom correspondence and requests for reprints should be addressed.

E1					
AO	ATPADWRSQSIYFLLTDRFARIDG	STTATCNTAD	QKYCGG		40
	7		40 47		
B1	SNKQNFSTDVYQIFIDRF	SDGNPANPETGAFFD	RLYCGG		52
	35		68 76		
BC	SNKQNFSTDVYQIVIDRFV	DGNTANNEPAGSAYD	KLYCGG		81
	7		40 47		
BM	NNKLNFTDVYQIVIDRFV	DGNSANNPTGAFFS	KLYFGG		52
	5		38 45		
KP	ETYLDFRKETIYFLIDRF	SDGOPSNNAGFNSAT	KKYTGG		50
	9				
MA	TEPKWWKEATIIYQIYPASF	K D SNN	D GWG		36
PP	QYAPQTQSGRTDIVHLF	E W			19
H1 E2 1					
AO	TWQGIIDKL D YIQGM	G FTAIWIT	P VTAQLPQDCAYG	DAYT	80
B1	DWQGIINKINDGYLTGM	G ITAIWISQ	P VENTYSVINYSGVNN	TAYH	98
BC	DWQGIMNKINDGYFTGM	G ITALWISQ	P VENTYSVINYSGVNN	TAYH	127
BM	DWQGITNKINDGYLTGM	G ITALWISQ	P VENITAVINYSGVNN	TAYH	98
KP	DLRGLINKL PYIKSL	G VTSIWITP	P IDN VNNIDAAGN	TGYH	91
MA	DLKGITSKL Q YIKDL	G VDAIWVC	P FYDSPQDD M		69
PP	FWVDIALEC ERYLGPK	G FGGVQVS	P PNENVVVTNPSR	PWWE	60
2 H2 3 E3					
AO	GYWOT DIYSIN	ENYGTADD	LKALSSALHER	GMIMDWA	120
B1	GYWAR DFKKTIN	PAYGIMQD	FNKLIDTAHAH	NIKVIIDFAP	138
BC	GYWAR DFKKTIN	PAFGSMID	FANLISAAHSR	NIKVVIDFAP	167
BM	GYWFR DFKKTIN	AAFGSFTD	FANLISAAHSR	NIKVVMDFAP	138
KP	GYWGR DYFRID	EHFGNLLD	FKELTSLMHSPDY	NMKLVLDYAP	133
MA	GYDIS NYEKW	PTYGTINED	CFELIDKTHKL	GMKFTIDLVI	109
PP	RY QP ISYKLC	TRSGNENE	FRDMVTRCANN	GVRIYDAVI	99
* H3 4					
AO	NHM GYDGAG	SSVDYSVFKPFSSQ	DYFHPFCF IQNYE	D QTQ	160
		153 170			
B1	NHT SPA	SSDDPSFAE DTQ	NLFH HYGG	T ₁ D ₁ FST	185
		182 199			
BC	NHT SPA	METNASFGE DTN	GYFH HNGG	T ₁ D ₁ FST	214
		153 170			
BM	NHT NPA	SSTDPSFAE DTA	GLFH HNGG	T ₁ D ₁ FST	185
		146 163			
KP	NHS NAN	DENE FGA ANT	GWYH HNGGVINMN	D FFQ	182
		134 149			
MA	NHC STEHEWF K	ESRSSKTNPKRWF	IPPNNWKSFFGGSATIF	D ETT	169
		126 136			
PP	NHM CGSGAAAGTIGTOSYCNPGNREF	FNDGCKKTASGGIESYN	D PYQ		156
* H3 4					
AO	VEDOWLGONTVSLP	D LDTTK DVVKNEWYDWGSLVS	NYST		200
B1	IEN GIYKNLYDLA	D LNHN SSVDVYLKDAIKMWL	DLGV		223
BC	LKN GIYKNLYDLA	D LNHN STIDTYFKNAIRLWL	DMGI		252
BM	TES GIYKNLYDLA	D INQNN NTIDSYLKESIQWL	NLGV		223
KP	VKN HNLFNLS	D LNQSN TDVYNYLLDGSKFWI	DAGV		217
MA	NEFYLRIFASRQV	D LNWNEDCRRALFESAVGFWL	DHGV		208
PP	VRDC QLVGLL	D LALEK DYVRSMIADYINKLI	DIGV		191
E4 H4 5 E5					
AO	DGLRIDTVKH	V QKDFWPG YNKAAG	V YCIGEVLDG	DPAYTC	240
		234 239			
B1	DGIRVDAVKH	M QKSEMATI NNYKP	V FTFGEWFLG	PEYHQ	271
		263 268			
BC	DGIRVDAVKH	M QKNMSSI YSYKP	V FTFGEWFLG	ANNY	300
		234 239			
BM	DGIRFDAVKH	M QKSYVSSIYSSANP	V FTFGEWFLG	QDNIN	272
		228 234			
KP	DAIRIDAIKH	M QK WTSIDIYDYSKS	F FFFGEWFGA	QNAID	275
		219 225			
MA	DGERIDTAGL	Y GLPDSP IFDKTS	I MRVGEVAHG	SDNALYTS	288
		211 228			
PP	AGFRIDASKH	MMPG DIKAVL	P FIFQEVIDL	GGEAITS	244
H5 E6 H H6					
AO	PYQNMVGVINY	PYYPLINAFKSTSGSMD	L YNNINIVK		280
B1	FANESQMSLLDFRFA	QKARQVFRD	NTD NMYG	L KAMLEGSE	311
BC	FANESQMSLLDFRFS	QKVRQVFRD	GSD TMYG	L DSMLSSTA	340
BM	FANQSGMHLDFRFA	QKIREVFRD	KSET MID	L NSVISSTG	312
KP	YANTSGSALLDFGR	DTLERVLVG	RSN TMRK	L NSYLIKRO	316
MA	AARYEVSEVFSFTHVEVGTSPFFRY	N IVPFT	L KQWKEALA		328
PP	GEMFSNGRVTEFKYG	AKLGIVVRKWSGEK	MSY	L KNWGEWNG	285

		6 E7		H7		
AO	SDCPDST	LLGTFVENHNDNRFAS	YTN	DIALAKNVAAFIIL	320	
B1	VDYAQVN	DQVTFIDNHDMEFHT	SNG	DRRKLEQALAFILT	351	
BC	ADYYSVN	DQVTFIDNHDMDRFQV	SGA	NGRKLEQALALTLT	380	
BM	SSNYNIN	NMVTIDNHDMDRFQV	AGA	STRPTEQALAVILT	352	
			342	364		
KP	TVFTSDD	WQVFMNDHDMARIGT	ALR	AQKRIDLGLVATMT	377	
MA	SNFLFTNGTDS	WATTYIENHDQARSIT	RFAD	DSPKYRKISGKLLT	373	
PP	FMPSD	RALVFVDNHDNQRGHG	AGGASILT	FWDARLYKVAVGEMLA	330	
		E8				
AO	ND	G LPITTYAGQEQHYA	GGNDPANREATWLSGYPTDSELYK		360	
B1	SR	G VPAIYYGSEQYMS	GGNDPDNR ARLPs	FSTITTAYQ	389	
BC	SR	G VPAIYYGTEQYMT	GGNDPNNR AKMS	FSTSTAYN	418	
BM	SR	G VPAIYYGTEQYMT	GGNDPNNR GMITG	FDINKTAYK	390	
			393 403			
KP	VR	G IPATYYGTEHYAA	VGSDPYNR EKMPG	FDTESEAFS	424	
MA	LLECSLT	G TLVYQGE I G QI N	FK EWP IE		402	
			341 352	373 386		
PP	HPY	G FTRVMS QDVND	WIGPPNNN GVIKE	VTIN HRWR	389	
		H8				
AO		LIASANAIRNYAISKDTGEVTKNFIKDD			390	
B1		VIQKLAPLRKSNPAIAYGS TH ERWINND			417	
BC		VISKLAPLRKSNPAIAYGT TQ QRWIND			446	
BM		VIKALAPLRKSNPAIAYGS TT QRWIND			418	
KP		IIKTLGDLRKSSPAIQNGT YT ELWIND			452	
MA		KYEDVDVKNNYEIIKKSFGKNSKE MKDF			430	
PP		QIRNMWFRNVVDGQFFANWMDGNSQVAF			419	

Fig. 1. Amino acid sequence alignment of α -amylases, CGTases and the α -glucosidase

Amino acid residues are given in the single-letter code. Sequences of α -amylases and CGTases are given without the signal peptide. Numbering of amino acid residues from the *N*-terminal end is shown, except for the *B. circulans* enzyme. The *N*-terminal residue of the mature *B. circulans* enzyme is not known, so numbering is given from the first methionine residue. The enzymes are: AO, *A. oryzae* α -amylase; B1, CGTase from *Bacillus* sp. strain 1011; BC, enzyme from *B. circulans*; BM, CGTase from *B. macerans* (sequence given by Sakai *et al.* [2]); KP, CGTase from *K. pneumoniae*; MA, α -glucosidase; PP, pig pancreatic α -amylase. Sequences of high degree of similarity and single conserved residues are boxed. $\sim$ denotes a sequence too long to be shown in the Figure. Underlining indicates the sequences used to calculate the average minimum base change/codon values. E1–E8, H1–H8 and H above the sequences indicate approximate positions of secondary-structure elements (cf. Table 1). * indicates residues implicated in Ca^{2+} binding; $\circ$ indicates residues implicated in activity of TAase and CGTases, and $\bullet$ denotes residues implicated in activity of α -amylases, CGTases and α -glucosidase.

terminal domain [11]. It was therefore decided to consider only the first 450 amino acid residues from the *N*-terminal ends of the CGTases and α -glucosidase, to examine the similarities between the *N*-terminal regions of these enzymes and the active-site domain of α -amylases.

A proposed amino acid sequence alignment is shown in Fig. 1. The sequence of the CGTase of *Bacillus* sp. strain 38-2 [12] is not given because it is almost identical with that of the enzymes from *Bacillus* sp. strain 1011 [11]. An alignment of the primary structures of the CGTases from *B. stearothermophilus* and *B. macerans* has been published recently by Sakai *et al.* [2] and is therefore not shown here. The *B. circulans* 'amylase', the CGTases and α -glucosidase are compared with TAase and pig pancreatic α -amylase, for which the three-dimensional structures have been obtained by X-ray crystallography [4,5]. The present alignment of the pig pancreatic α -amylase sequence with TAase has been altered from that reported earlier [3] for the region 270–280 of TAase, in order to take account of results on the crystal structure of pig pancreatic α -amylase [5]. The present alignment of the CGTases with the α -amylases differs extensively from that given by Binder *et al.* [13] and slightly, but significantly, from that given by Sakai *et al.* [2], particularly in areas around positions equivalent

to 70, 118, 125–200, 235–290 and 310–323 of the α -amylase of *A. oryzae*.

The boxed sequences indicate regions with a high degree of similarity between the enzymes, as already demonstrated [1,2,11,13]. Similarity between α -amylases and the α -glucosidase in region 5 is recognized in the present work. It should be pointed out that CGTases appear to be distinguished from the other enzymes in region 5 [1,3,18] by the lack of a hydrogen-bonding side chain at the position corresponding to TAase Asp-233. The sequence -(Tyr or Phe)-Ala-Pro- in region 3 also seems to be distinctive to the CGTases.

The primary structure of the CGTase of *B. macerans* deduced by Takano *et al.* [10] appears to lack region 3 [1], a region probably important for CGTase activity (see discussion below). This apparent discrepancy may have been caused by an error in DNA sequencing, since upon insertion of single bases at two positions in the DNA sequence determined for region 3 by Takano *et al.* [10] the encoded sequence becomes almost identical with that of the *Bacillus* sp. strain 1011 CGTase [11]. In the present work we therefore use the sequence reported by Sakai *et al.* [2] for the *B. macerans* CGTase.

In addition to the high degree of similarity indicated in the boxed regions of Fig. 1, some single amino acid residues may be invariant in the enzymes shown, i.e.

at positions corresponding to TAase Gly-56, Pro-64, Asp-175 and Gly-323.

There are many semiconservative substitutions of amino acid residues among the enzymes compared here, but only 17 amino acid residues appear to be invariant. Two apparent conservations indicated in Fig. 1 do not hold for the CGTase from *B. stearothermophilus*, which contains threonine and phenylalanine at the positions equivalent to TAase positions 157 and 272 respectively [2].

Both the CGTases and α -glucosidase appear to be more closely related to TAase than to the mammalian α -amylase. Calculation of the average minimum base change per codon (MBC/C) for pairs of enzymes gives values of 1.04–1.13 for the *B. circulans* 'amylase', the CGTases and α -glucosidase compared with TAase and 1.16–1.25 compared with pig pancreatic α -amylase. Values of 1.00–1.29 are typical for related proteins [3,19,20], with a value around 1.5 being expected for a random alignment [19]. Here MBC/C values were calculated over the stretches of contiguously aligned residues shown in Fig. 1.

Inspection of Fig. 1 suggests that the sequence of the *B. circulans* enzyme more closely resembles that of the *Bacillus* sp. strain 1011 CGTase than that of an α -amylase. In fact, for those sections of the sequences shown in Fig. 1, the *B. circulans* enzyme shows 77% homology with the strain 1011 CGTase, 72% homology with the *B. macerans* CGTase and only 25% and 15% homology with the α -amylases from *A. oryzae* and pig pancreas respectively. Calculation of the MBC/C for the *B. circulans* amylase compared with the CGTase of *Bacillus* sp. strain 1011 (over the contiguous sequences of Fig. 1) further emphasizes the similarity between the two. The value is 0.26, whereas the value for the MBC/C for the pair of CGTases from *Bacillus* sp. strain 1011 and *B. macerans* is 0.37, i.e. the *B. circulans* 'amylase' resembles the *Bacillus* sp. strain 1011 CGTase more closely than does another CGTase. Further, the *B. circulans* enzyme lacks the hydrogen-bonding amino acid residue equivalent to TAase position 233 and has the region 3 sequence -Phe-Ala-Pro- characteristic of CGTases (see Fig. 1).

We suggest here, therefore, that Nishizawa *et al.* [9] have determined the sequence of a cyclodextrin glucanotransferase from *B. circulans* and not an α -amylase. In the remainder of the present paper we refer to the *B. circulans* enzyme as a CGTase.

The greatest differences between the enzymes discussed here, both in the nature and number of amino acid residues present, appear to occur in regions equivalent to residues 16–40, 72–80, 126–154, 211–214, 223–225, 235–237, 252–253, 262–270, 304–307 and above 330 of TAase.

Since predictions of secondary structure based on the methods of both Garnier *et al.* [16] and Cid *et al.* [17] gave better results for α -amylases than use of the former procedure alone [3], it was decided to use a combination of the two methods, along with the sequence alignment of Fig. 1 and the known crystal structures of TAase [4] and pig pancreatic α -amylase [5]. By the Garnier procedure, each of the first 450 amino acid residues of an enzyme was assigned to one of two categories: possible regular secondary structure, consisting only of residues predicted to be in helices of at least three contiguous residues, and those predicted to be in β -structures of the

same length, or irregular structure. Examination of TAase by the Garnier method had shown that 91% of the residues found in regular secondary structure by X-ray crystallography would fall in the first category [3]. The Garnier method, however, has been demonstrated to 'overpredict', i.e. to predict an excess of secondary structure for TAase, and to be unreliable for distinguishing between helices and β -structure [3]. Thus residues in the first category were investigated further in an attempt to eliminate those residues least likely to be in regular secondary structures and to distinguish between helices and extended chains. This was carried out by using the method of Cid *et al.* [17] and careful examination of sequence similarities shown in Fig. 1. Where use of the Cid method indicated stretches of residues failing to show any of the patterns of hydrophobicity and/or hydrophilicity expected for regular secondary structure, these sequences were eliminated from further consideration. The remaining sequences were checked against the alignment of Fig. 1. Where predicted sequences coincided with regular structure in the α -amylases, such sequences were considered to have a high probability of forming helices or β -structures. The Cid method was found to be useful for distinguishing between helices, many of which appear to be amphipathic in the enzymes investigated, and β -structures, expected to be hydrophobic. The sequence similarities of Fig. 1 confirmed the assignments as helices or extended chain. The results of the predictions are given in Table 1. For a small number of residues (approx. 3% of those in Table 1) sequence similarity and the Cid method suggested possible regular structure, when the Garnier procedure did not, and these residues have been included in Table 1. In general, the error in the positions shown is likely to be ± 2 because of difficulties in locating exactly the beginnings and ends of regular structures. The uncertainty in the position of helix H8 of the α -glucosidase is greater, however, because of the low degree of sequence similarity in that area, i.e. there is a greater possibility of sequence misalignment in Fig. 1.

The CGTase from *B. macerans* is predicted to have regular structures (not given in Table 1) E1 to E4 and H1 to H4 at positions identical with those given in Table 1 for the *Bacillus* sp. strain 1011 CGTase. Because of an insertion in the *B. macerans* enzyme, the remaining structural elements are predicted to occur one residue later than in the *Bacillus* sp. strain 1011 protein, e.g. E5 is probably situated at residues 254–259 of the *B. macerans* CGTase.

As can be seen from Table 1, the CGTases and α -glucosidase appear to share the $(\alpha/\beta)_8$ -barrel super-secondary structure with the *N*-terminal domain of the α -amylases. The additional helix in the α -amylases that occurs before H6 [4,5] also seems to exist in the CGTases and α -glucosidase.

In these enzymes β -strands are predicted to alternate with helices along the polypeptide chain, and in the α -amylases from *A. oryzae* and pig pancreas adjacent β -strands and helices are found to be joined by irregular loops [4,5]. The positions of major differences in amino acid sequence between the α -amylases, CGTases and the α -glucosidase seem to coincide with the positions of such loops, particularly those linking the *C*-terminal ends of β -strands to the *N*-terminal ends of helices. The segment connecting E3 to H3 is believed to be much longer than the other loops, and has been found to contain additional

Table 1. Positions of regular secondary structures

E1–E8 are stretches of extended chain and H1–H8 are helices of the $(\alpha/\beta)_8$ structure; H is an 'extra' helix. For the *B. circulans* enzyme, amino acid residues are numbered from the *N*-terminal end of the immature protein. For the other enzymes numbering is from the end of the mature protein with signal sequence removed. Data for the α -amylases of *A. oryzae* and pig pancreas are taken from the Protein Data Bank, Brookhaven National Laboratory, Upton, NY, U.S.A., and ref. [5] respectively. The positions given for the α -amylases have been determined by X-ray crystallography. For the other enzymes, regular structures are predicted to occur at the positions shown here.

Structure	<i>A. oryzae</i> α -amylase	Pig pancreatic α -amylase	<i>Bacillus</i> sp. strain 1011 CGTase	<i>B. circulans</i> CGTase	<i>B. stearo-</i> <i>thermophilus</i> CGTase	<i>K. pneumoniae</i> CGTase	<i>S. cerevisiae</i> α -glucosidase
E1	10–15	10–16	15–21	43–49	12–18	13–19	18–23
E2	58–64	39–42	72–77	101–106	69–74	68–73	54–59
E3	112–118	92–97	131–136	160–165	127–132	126–131	102–107
E4	202–206	193–196	225–229	254–258	221–225	219–223	210–214
E5	226–231	229–231	253–258	282–287	249–254	253–258	272–275
E6	248–252	253–255	280–283	309–312	276–279	284–287	297–300
E7	289–293	292–295	320–323	349–352	316–319	325–328	341–345
E8	323–329	335–340	355–360	384–389	351–356	381–386	382–387
H1	44–53	22–31	56–61	85–90	53–58	54–59	40–49
H2	99–110	77–89	117–128	146–157	113–124	111–120	88–99
H3	182–195	173–189	209–221	238–250	205–217	210–215	191–204
H4	215–222	204–213	242–248	273–277	238–243	237–244	229–234
H5	237–242	241–246	270–274	300–303	266–270	273–278	284–290
H6	270–280	281–285	301–309	330–338	299–305	306–314	318–328
H7	308–320	321–330	342–351	371–380	338–347	368–377	363–373
H8	357–374	394–402	386–395	415–424	382–390	421–432	402–408
H	253–261	259–268	286–294	316–323	285–289	290–297	307–313

β -strands in α -amylases from *A. oryzae* [4] and pig pancreas [5]. Some of these β -strands may occur in the CGTases and α -glucosidase, but are difficult to predict by the methods used here, if, as found for the pig enzyme [5], they are very short. Thus it is possible that the CGTases and α -glucosidase may contain additional regular structural features not given in Table 1. We believe it likely, however, that the basic super-secondary structure of these proteins is $(\alpha/\beta)_8$ and that additional secondary structures are most likely to occur as regularities in loops joining adjacent β -strands to helices of the $(\alpha/\beta)_8$ -barrel.

The tertiary structures of the $(\alpha/\beta)_8$ -barrel domains of α -amylases, CGTases and the α -glucosidase seem likely, therefore, to be very similar, differing mainly in the loops linking the regular secondary structures.

α -Amylases are known to contain Ca^{2+} , suggested to interact with the side chains of residues equivalent to Asn-121, Asp-163, Asp-175 and the C=O group of residue 210 of TAase [3,5]. In some α -amylases the residue equivalent to Asp-163 is asparagine [3], and in one case may be serine [21]. The Ca^{2+} -binding residues appear to be conserved in the CGTases, suggesting that they also may contain a structurally important Ca^{2+} ion. (The Ser-188 shown in *B. macerans* CGTase at the position equivalent to Asp-163 of TAase is in some doubt; Takano *et al.* [10] found aspartate in this position.) If CGTases do indeed contain Ca^{2+} , this might explain the requirement of Ca^{2+} for CGTase synthesis [22] and stability [23], and the need for the addition of excess Ca^{2+} during purification and crystallization [24]. The Ca^{2+} ion is likely to play the same role in CGTases as in α -amylases, i.e. that of maintenance of the correct tertiary structure for activity [5].

One of the residues critical for interaction with Ca^{2+} , i.e. that equivalent to Asp-163 of TAase, appears to be Glu-171 in the α -glucosidase. Other proposed Ca^{2+} -binding groups seem to be conserved in the glucosidase [1], and thus it is tempting to propose that this enzyme also may contain a structurally important Ca^{2+} ion. Active glucosidase, however, can be prepared in the presence of EDTA [25], and so it appears less likely that Ca^{2+} is required for activity.

If it is assumed that the tertiary structures of α -amylases, CGTases and the glucosidase are similar, then it is easy to explain conservation of certain hydrophilic and hydrophobic amino acid residues in all the enzymes. Hydrophobic residues are apparently retained at positions equivalent to 57, 60, 100, 103, 112, 114, 115, 116, 176, 194, 200, 203, 205, 228, 292, 293, 314, 319, 327 and 365 of TAase. They are all situated in the interior of the TAase protein and hence are likely to occupy similar positions in the other enzymes. Conservation of hydrophilic residues seems to occur at positions equivalent to 48, 89, 96, 99, 157, 294 and 370 of TAase, and these are probably mostly on the surface of the CGTases and α -glucosidase, as they are in TAase [4]. A notable exception is the acid group at the position equivalent to 294 of TAase, which is not on the enzyme surface, but close to Arg-300. There is presumably interaction between the two side chains in the TAase, an interaction that appears to be conserved in all the enzymes discussed here (see Fig. 1), but not in α -amylases from *Bacillus* species and barley [1,3]. A second invariant side-chain interaction probably occurs in the enzymes between residues corresponding to Asp-117 and Arg-204 of TAase [1,3]. Proline, equivalent to Pro-64 of TAase, is presumably conserved for structural reasons, as are the

residues equivalent to Gly-56, -95 and -323 occurring at turns in the polypeptide chain of TAase [4]. Other residues conserved, probably because of their importance for enzymic activity, are discussed in the next subsection.

Active site

The amino acid residues at the active sites of the enzymes discussed here perform two functions: either catalysis of cleavage of an α -1,4 glucosidic bond or binding of glucose residues of the substrate.

It has been suggested that the catalytic residues of TAase are Glu-230 and Asp-297, although Asp-206 is also shown close to the glycosidic bond to be broken in the computer-generated structure of a complex between TAase and an α -glucan segment [4]. The sequence alignments given by Svensson [1], Sakai *et al.* [2], Kimura *et al.* [11], Binder *et al.* [13] and in the present paper indicate that corresponding residues are present in CGTases and the α -glucosidase. Buisson *et al.*, however, propose that the catalytic residues of pig pancreatic α -amylase are Asp-197 and -300 [5], equivalent to TAase residues 206 and 297. If this is indeed the case, then the catalytic residues of the CGTases should be Asp-229 and -328, Asp-258 and -357, Asp-229 and -329, Asp-225 and -324 and Asp-223 and -333 for the enzymes from *Bacillus* sp. strain 1011, *B. circulans*, *B. macerans*, *B. stearothermophilus* and *K. pneumoniae* respectively. For the α -glucosidase, the corresponding aspartate residues are at positions 214 and 349, as suggested by Svensson [1].

The variations in specificity among α -amylases, CGTases and the glucosidase, which all cleave α -1,4 glucosidic bonds, are likely to be reflected in differences in substrate-binding residues. Where similarities do exist, these should be greatest in the area responsible for binding rings 4 and 5 of the substrate (see Fig. 2), i.e. adjacent to the bond to be broken.

The α -glucosidase requires an α -D-glucopyranose residue at the non-reducing end of the substrate [25], which presumably binds at subsite B_4 (see Fig. 2) and is subsequently released as α -D-glucose. Binding of a glucose residue at subsite B_5 (Fig. 2) is not necessary for the activity of the glucosidase. Indeed *p*-nitrophenyl α -D-glucopyranoside is a better substrate for the enzyme than maltose [25]. Thus either a glucose residue or a *p*-nitrophenyl group can be accommodated at subsite B_5 , suggesting a greater similarity between the α -glucosidase and α -amylases at subsite B_4 than at subsite B_5 . Matsuura *et al.* [4] have suggested that His-122 and -296, Glu-230

and Asp-297 may be involved in substrate binding at subsite B_4 in TAase. These residues appear to be conserved in the α -glucosidase (see Fig. 1). Hence either His-111 or His-348 of this enzyme may represent the functionally important histidine residue found in brewer's yeast α -glucosidase by Someya *et al.* [26]. The extensive conservation found at subsite B_4 is not observed for subsite B_5 . Matsuura *et al.* [4] suggest that Lys-209, His-210, Glu-230 and a hydrophobic residue at position 231 are important in TAase at subsite B_5 . In the α -glucosidase only the last two residues seem to be retained. Thus it is likely that orientation of the substrate at the active site, as given by Matsuura *et al.* for TAase [4], holds true for the α -glucosidase. The presence of a hydrophobic acid at position 277 of the α -glucosidase (equivalent to TAase position 231) may promote binding at subsite B_5 of the aromatic *p*-nitrophenyl group.

It should be noted that the α -glucosidase discussed here is a 'maltase' i.e. catalyses hydrolysis of α -1,4-glucosidic bonds and not α -1,6-glucosidic bonds [25]. A separate enzyme of brewer's yeast catalyses α -1,6-bond hydrolysis [27], but information on its amino acid sequence or three-dimensional structure is not yet available.

CGTases and α -amylases degrade starch, whereas the α -glucosidase hydrolyses short oligosaccharides only [25,27]. In addition, both TAase and pig pancreatic α -amylase slowly hydrolyse cyclodextrins [28,29]. Thus more extensive similarities might be expected between the active sites of α -amylases and CGTases than between those of α -amylases and the glucosidase.

According to the amino acid alignment of Fig. 1, the suggestions made by Sakai *et al.* [2] concerning interaction of substrate with the CGTases of *B. macerans* and *B. stearothermophilus* seem reasonable for the binding of glucosyl rings at subsites B_3 , B_4 and B_5 (see Fig. 2). Thus a glucosyl ring at subsite B_3 may interact with residues equivalent to Trp-83 and Arg-344 of TAase and at subsite B_4 with residues equivalent to His-122 and -296 and possibly Asp-297 and Glu-230, whereas binding at subsite B_5 may involve residues equivalent to Glu-230, Lys-209, His-210 and the hydrophobic residues at positions 231 and 232. The residues suggested at subsites B_4 and B_5 are also conserved in many α -amylases [1,3], and several have been implicated by Buisson *et al.* [5] in substrate binding by pig pancreatic α -amylase.

Sakai *et al.* [2] suggest, from their alignment of CGTase sequences with TAase, that both Glu-187 and Asp-199 of the *B. macerans* CGTase interact with

Fig. 2. Schematic representation of α -glucan binding at enzyme active site

B_1 to B_6 are subsites, each interacting with one glucose residue of the substrate. Arrows indicate catalytic amino acid residues, between subsites B_4 and B_5 . The glycosidic bond between rings 4 and 5 of the substrate would be cleaved.

substrate. These residues were assumed to be equivalent to those at positions 156 and 168 of TAase. In the alignment presented here, this is not the case. We believe, however, that our alignment is more likely to be correct, since only here are Ca²⁺-binding Asp-163 and -175 of TAase [3,4] and Asp-159 and -167 of the pig amylase [5] aligned with putative Ca²⁺-binding residues in the CGTases. We suggest, therefore, that Glu-187 and Asp-199 of *B. macerans* CGTase, and corresponding residues in other CGTases, are unlikely to participate in substrate binding. In the alignment of Fig. 1, however, the CGTases have a lysine or histidine or arginine residue [2] equivalent to Asp-168 of TAase. Thus this residue may be important for substrate binding, possibly at subsite B₂.

Sakai *et al.* have further proposed that aspartate and tyrosine equivalent to residues 340 and 82 respectively of TAase are located at subsite B₂ [2]. These residues appear to be conserved in all CGTases discussed here, and the equivalent aspartate residue has been implicated in substrate binding by TAase [4]. The tyrosine residue is conserved in α -amylases [3] and also occurs in the α -glucosidase, where it is less likely to be involved in substrate binding, since subsite B₂ probably does not exist in the glucosidase. It is possible that the tyrosine residue may not be involved directly in substrate binding, but may instead be important for structural integrity of the active site.

Thus the binding of glucosyl residues at subsites B₂, B₃, B₄ and B₅ may be similar for CGTases and TAase. This should not be unexpected, since Suetsugu *et al.* concluded that TAase interacts with approximately four glucosyl residues of a cyclodextrin substrate [28].

Sakai *et al.* further proposed that Gln-55 of *B. macerans* CGTase should be able to interact with substrate ring 1 [2]. This seems unlikely, since the present alignment and structure prediction indicate that this residue should be situated just before the beginning of the first helix of the (α/β)₈ structure. For α -amylases, active-site residues are believed to be situated at and near the C-terminal ends of β -strands or in the middle of loops linking β -strands to adjacent helices [3–5]. This positioning of active-site residues on the loops allows for differences in specificity while retaining the basic (α/β)₈ skeleton of the enzymes.

The CGTases discussed here do not all have the same specificity. The enzyme from *B. macerans*, for example, produces more cyclomaltohexaose from starch than do the *B. circulans* or *B. stearothermophilus* CGTases [30]. Unlike the α -amylases, where large differences in loop length and amino acid sequence may lead to product-specificity variations [3], the *Bacillus* CGTases have similar sequences and, probably, similar loop lengths. Thus the variations in CGTase specificity are possibly due to more subtle differences between the enzymes, e.g. single amino acid substitutions, or may be due to differences in the C-terminal domain, which has not been examined here. The role of the C-terminal domain of α -amylases, CGTases and the α -glucosidase is at present unknown. It has been suggested, however, that for CGTases this domain may be important for cyclizing reactions [11], or, because of local sequence similarity with a raw starch-binding region of *Aspergillus niger* glucamylase [1], for the binding of starch granules.

Here we have shown that an α -glucosidase, which hydrolyses small oligosaccharides only, may have the same (α/β)₈-barrel super-secondary structure as

α -amylases, which hydrolyse large substrates, and CGTases, which not only degrade large substrates, but bring about cyclization, disproportionation and coupling reactions [31]. Differences in loops between the β -strands and α -helices probably account for some of the wide differences in action of α -amylases, the α -glucosidase and CGTases, yet the enzymes appear to share key functional residues [1,11] and hence active-site structural features.

We thank Dr. W. G. Baldwin of the University of Manitoba for the computer program of the procedure of Garnier *et al.* [16], and Dr. Y. Matsuura for helpful correspondence concerning the structure of *A. oryzae* α -amylase. E.A. MacG. also thanks the University of Manitoba Research Committee of Senate for financial assistance.

REFERENCES

1. Svensson, B. (1988) FEBS Lett. **230**, 72–76
2. Sakai, S., Kubota, M., Yamamoto, K., Nakada, T., Torigoe, K., Ando, O. & Sugimoto, T. (1987) Denpun Kagaku **34**, 140–147
3. MacGregor, E. A. (1988) J. Protein Chem. **7**, 399–415
4. Matsuura, Y., Kusunoki, M., Harada, W. & Kakudo, M. (1984) J. Biochem. (Tokyo) **95**, 697–702
5. Buisson, G., Duée, E., Haser, R. & Payan, F. (1987) EMBO J. **6**, 3909–3916
6. Toda, H., Kondo, K. & Narita, N. (1982) Proc. Jpn. Acad. Ser. B **58**, 208–212
7. Kluh, I. (1981) FEBS Lett. **136**, 231–234
8. Pasero, L., Mazzei-Pierron, Y., Abadie, B., Chicheportiche, Y. & Marchis-Mouren, G. (1986) Biochim. Biophys. Acta **869**, 147–157
9. Nishizawa, M., Ozawa, F. & Hishinuma, F. (1987) DNA **6**, 255–265
10. Takano, T., Fukuda, M., Monma, M., Kobayashi, S., Kainuma, K. & Yamane, K. (1986) J. Bacteriol. **166**, 1118–1122
11. Kimura, K., Kataoka, S., Ishii, Y., Takano, T. & Yamane, K. (1987) J. Bacteriol. **169**, 4399–4402
12. Hamamoto, T., Kaneko, T. & Horikoshi, K. (1987) Agric. Biol. Chem. **51**, 2019–2022
13. Binder, F., Huber, O. & Böck, A. (1986) Gene **47**, 269–277
14. Hong, S. H. & Marmur, J. (1986) Gene **41**, 75–84
15. Levin, J. M., Robson, B. & Garnier, J. (1986) FEBS Lett. **205**, 303–308
16. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) J. Mol. Biol. **120**, 97–120
17. Cid, H., Bunster, M., Arriagada, E. & Campos, M. (1982) FEBS Lett. **150**, 247–254
18. Nakajima, R., Imanaka, T. & Aiba, S. (1986) Appl. Microbiol. Biotechnol. **23**, 355–360
19. Rossmann, M. G., Moras, D. & Olsen, K. W. (1974) Nature (London) **250**, 194–199
20. Argos, P., Mahoney, W. C., Hermodson, M. A. & Hane, M. (1981) J. Biol. Chem. **256**, 4357–4361
21. Itoh, T., Yamashita, I. & Fukui, S. (1987) FEBS Lett. **219**, 339–342
22. Tilden, E. B. & Hudson, C. S. (1942) J. Bacteriol. **43**, 527–544
23. Bender, H. (1977) Arch. Microbiol. **111**, 271–282
24. Kobayashi, S., Kainuma, K. & Suzuki, S. (1978) Carbohydr. Res. **61**, 229–238
25. Needleman, R. B., Federoff, H. J., Eccleshall, T. R., Buchferer, B. & Marmur, J. (1978) Biochemistry **17**, 4657–4661

26. Someya, Y., Matsui, H. & Chiba, S. (1984) *Agric. Biol. Chem.* **48**, 873–879
27. Matsusaka, K., Chiba, S. & Shimomura, T. (1977) *Agric. Biol. Chem.* **41**, 1917–1923
28. Suetsugu, N., Koyama, S., Takeo, K. & Kuge, T. (1974) *J. Biochem. (Tokyo)* **76**, 57–63
29. Abdullah, M., French, D. & Robyt, J. F. (1966) *Arch. Biochem. Biophys.* **114**, 595–598
30. Kitahata, S. & Okada, S. (1982) *Denpun Kagaku* **29**, 13–18
31. French, D. (1957) *Adv. Carbohydr. Chem.* **12**, 189–260

Received 13 July 1988/4 October 1988; accepted 17 October 1988