Biochem. J. (1995) 305, 685–688 (Printed in Great Britain)

BIOCHEMICAL JOURNAL LETTERS

Characteristic differences in the primary structure allow discrimination of cyclodextrin glucanotransferases from α-amylases

Amylolytic and related enzymes representing 15 distinct specificities have been identified as belonging to a $(\beta/\alpha)_8$-barrel protein family, by using structure-prediction procedures in combination with known three-dimensional models [1–4]. As a result of the use of cloning techniques more than 150 primary structures are available [4,5]. Sequences of α-amylases, which are widely occurring enzymes hydrolysing starch and related oligodextrins in an endo-fashion, predominate, but cyclodextrin glucanotransferase (CGTase) sequences also occur frequently. CGTase is bacterial in origin and catalyses the cleavage of oligodextrins from non-reducing ends of starch with intramolecular glucosyl transfer to produce α-, β- or γ-cyclodextrins, although hydrolysis can also occur. Both CGTase and α-amylase catalyse starch degradation with retention of configuration at the anomeric carbon of the glycosidic bond to be cleaved, and the mechanism is likely to involve double displacement. The enzymes apply an acid catalysis mechanism [6] in which the general acid catalyst, e.g. Glu-230 in Taka-amylase A (TAA) [7], protonates the oxygen of the glucosidic bond to be cleaved with synchronous C–O bond cleavage and formation. Only in the second stage of the reaction, i.e. nucleophilic attack at the glycone C-1 by water, do the two enzymes differ. In hydrolysis, catalysed by α-amylase, the nucleophile is water, whereas for the intra- or inter-molecular glycosyl transfer brought about by the CGTase, the nucleophile is the C-4 hydroxyl group of a glucose ring [1,6]. All of the seven residues which are invariably present in the α-amylases, CGTases and related enzymes are essential in catalysis, transition-state stabilization, or maintenance of the structural integrity of the catalytic machinery [4,7–15] and their specific roles have been investigated by numerous mutational analyses [6,16–18].

The catalytic $(\beta/\alpha)_8$-domain [4,7–16] has eight β-strands forming an inner cylinder with the active site at the C-termini of the strands and extending therefrom. Three of these β-strands are easily recognized in all superfamily members. A fourth, containing the general acid catalyst, is pinpointed by considering the organization of the motifs along the polypeptide chain [1–4,18–20]. Superficial observation of this pattern, however, had led to incorrect classification of deduced sequences from *Bacillus circulans* [21], *Bacillus* sp. (strain B1018) [22] and *Clostridium thermosulfurogenes* [23] as those of α-amylase, although they all most likely belong to CGTases [1,4,24]. Indeed, Itkor et al. [22] state that the *Bacillus* species amylase sequence shows high sequence identity (87%) with a *Bacillus* species CGTase, and forms β-cyclodextrin from soluble starch [22]. In the other two cases, amylase activity was determined by measuring increased reducing power as a result of enzyme action on soluble starch, and a detailed examination of products was not carried out. Recently 37 amino acid residues from four short regions containing β-strands 3, 4, 5, and 7, i.e. the most highly conserved sequence stretches, were demonstrated to define an evolutionary tree of different amylolytic enzymes that is consistent with enzyme specificity and taxonomy [4]. Clearly, in this and in an earlier analysis of secondary-structure elements in the CGTase $(\beta/\alpha)_8$-barrel [1], sequences claimed to be α-amylases are grouped with CGTases. Since several bacteria produce both α-amylase and CGTase, it is important to classify a cloned sequence correctly. To that end we propose here a finger-print (Figure 1) that is useful in distinguishing CGTases from α-amylases. A few published sequences identified as α-amylases are proposed to represent intermediates of α-amylases and CGTases.

Highly characteristic differences between the α-amylases and CGTases are found in the sequences shown in Figure 1. Thus, at the third β-strand only CGTases have Ala-Pro, or rarely Thr-Pro, preceding Asn-His [1–4,16–20]; other amylolytic enzymes usually have Val in place of Ala. In addition, CGTases have Phe or Tyr succeeding the invariant Asp (117 in TAA). In the fourth β-strand, the CGTases typically have Ile before the invariant Arg (204 in TAA). The seventh β-strand in CGTases starts with Gln/Met.

Less easily localized in the sequence, but highly specific to CGTases, is a peptide length of eight residues between Gly and Pro, residues preceding and succeeding, respectively, the second β-strand of the $(\beta/\alpha)_8$-barrel. In all other amylolytic enzymes a heptapeptide connects the equivalent Gly and Pro [1,4,24]. At the C-terminal end of the fifth β-strand, three residues interacting with substrate succeed the general acid catalyst Glu (230 in TAA [7]). In CGTase crystal structures [7–15] and predicted structures [1] a residue (often L or G) at $n+3$ relative to the acid catalyst is unable to hydrogen-bond with substrate, in contrast with the equivalent side-chain in α-amylases [1,20]. At positions $n+1$ and $n+2$, CGTases have Trp-Phe. These residues are believed to be important for cyclizing activity [25], and the dipeptide is not found in other members of the α-amylase superfamily [17]. Recently a conserved pentapeptide stretch around a Ca^{2+}-binding Asp residue (175 in TAA [7]) in the third loop was pointed out in α-amylases [26], and CGTases have a specific sequence (Figure 1) here also.

In addition to the differences described above, full-length CGTase is usually much longer than an α-amylase. CGTases in general have five domains, whereas α-amylases have three. The fifth and C-terminal domain of a CGTase binds starch [27], and homologues are found in two bacterial exoamylases, the glucoamylases and one bacterial β-amylase, although the latter enzyme classes have a catalytic domain fold different from that of the α-amylase superfamily [2,27]. Only endo-α-amylases from *Streptomyces* have this domain [27]. Although α-amylases and related amylolytic enzymes are multidomain proteins, it is uncertain which domains in addition to the catalytic $(\beta/\alpha)_8$-barrel are needed for proper function. In α-amylases the C-terminal domain thus seems required for activity, not because it is directly involved in catalysis, but presumably for the conformational stability of the molecule [28]. The starch-binding domain of GSTases, in contrast, is dispensable [27,29].

Since crystal structures of α-amylase from pig pancreas [8,10], *Aspergillus oryzae* (TAA) [7], *Aspergillus niger* [9] and barley malt [11], and of CGTase from *Bacillus stearothermophilus* [12], *B. circulans* strain 8 [13,14] and *B. circulans* strain 251 [15] are

Enzyme	Secondary structure					
	β2	β3	Loop 3	β4	β5	β7
α-Amylases						
Bli	36-GITAVWIP-P	94-DINVYGDVVINH	198-YADID	226-DGFRLDAV	257-FTVAEYWQ	320-AVTFVDNHD
Bsu	33-GITAIQTS-P	91-GIKVIVDAVINH	144-LYDWN	171-DGFRFDAA	204-FQYGEILQ	261-LVTWVESHD
Sli	32-GYGYVQVS-P	82-GVKVVADSVINH	145-LADLD	172-DGFRIDAA	200-YWKQEAIH	260-SAVFVDNHD
Aor	56-GFTAIWIT-P	111-GMYLMVDVVANH	173-LPDLD	201-DGLRIDTV	226-YCIGEVLD	289-LGTFVENHD
Bar	34-GVTHVWLP-P	82-GVQAIADIVINH	147-APDID	175-DAWRLDFA	201-LAVAEVWD	283-AATFVDNHD
Dme	36-GYAGVQVS-P	88-GVRTYVDVVFNH	154-LRDLN	181-AGFRVDAA	219-YIVQEVID	280-SLVFVDNHD
Ppa	36-GFGGVQVS-P	90-GVRIYVDAVINH	165-LLDLA	192-AGFRLDAS	229-FIFQEVID	292-ALVFVDNHD
Intermediary						
Bfi	139-GYTAVQTS-P	200-GVAVIVDILPNH	258-LPDVD	285-DGFRIDTA	338-FVYGEVLQ	406-LVTWVESHD
Dth	179-GINTIWIS-P	226-GIRIILDFVPNH	276-MPKIN	304-SGYRMDHA	334-FYFGEIVE	396-RISFLENHD
Re-classified						
Bci	71-GVTAIWISQP	130-NIKVVIDFAPNH	198-LADLN	225-DGIRVDAV	254-FTFGEWFL	321-QVTFIDNHD
Bsp	70-GITALWISQP	129-NIKVIIDFAPNH	197-LADLN	224-DGIRMDAV	253-FTFGEWFL	320-QVTFLDNHD
Cth	70-GVTAIWISQP	128-NIKVIIDFAPNH	196-LADLN	223-DGIRLDAV	252-FTFGEWFL	319-MVTFIDNHD
CGTases						
Bci	70-GVTALWISQP	129-GIKIVIDFAPNH	197-LADFN	224-DGIRVDAV	253-FTFGEWFL	320-QVTFIDNHD
Bst	67-GVTAIWISQP	125-GIKVIIDFAPNH	193-LADLN	220-DGIRMDAV	249-FTFGEWFL	316-QVTFIDNHD
Boh	66-GITAIWISQP	122-GIKVIMDFTPNH	190-LADYD	217-DGIRVDAV	246-FTFGEWFL	313-QVTFIDNHD
Kpn	66-GVTSIWITPP	124-NMKLVLDYAPNH	191-LSDLN	218-DAIRIDAI	253-FFFGEWFG	325-QVVFMDNHD
Con	G P	D NH	l d	dg R D	E	f nHD

Figure 1 Sequence comparison of short regions of α-amylases and CGTases where specific differences can be observed

These sequences consist of the best conserved β-strands 3, 4, 5 and 7, the less-well conserved β-strand 2, and a conserved short sequence in loop 3. The enzymes are abbreviated as: α-amylases: Bli, *Bacillus licheniformis* (SwissProt accession number P06278); Bsu, *Bacillus subtilis* (P00691); Sli, *Streptomyces limosus* (P09794); Aor, *Aspergillus oryzae* (TAA, P10529); Bar, barley (P00693); Dme, *Drosophila melanogaster* (P08144); Ppa, pig pancreas (P00690); proposed intermediary enzymes: Bfi, *Butyrivibrio fibrisolvens* (P30269); Dth, *Dictyoglomus thermophilum* amy B (P14898); reclassified as CGTase: Bci, *Bacillus circulans* strain F-2 (P08137); Bsp, *Bacillus* sp. strain B1018 (P17692); Cth, *Clostridium thermosulfurogenes* (P26827); CGTases: Bci, *Bacillus circulans* strain 8 (P30920); Bst, *Bacillus stearothermophilus* (P31797); Both, *Bacillus ohbensis* (P27036); Kpn, *Klebsiella pneumoniae* (P08704). In 'Secondary structure' starting residues in the segments are numbered from the N-terminal end of the polypeptide chain. Con: consensus sequence; lower case indicates a maximum of three substitutions. The gap in aligned α-amylases in β-strand 2 is indicated by -.

available, a correlation of specificity with features of the three-dimensional structures may lead to modelling and design of amylolytic enzymes with purposely altered properties, such as higher transglycosylation capacity or hydrolytic production of oligodextrins of specific length. Mutants of CGTase in the binding residues following the acid catalyst, for example, resulted in loss of cyclizing activity [25]. Finally, insight into the evolution of enzymic starch degradation may be expanded by study of enzyme sequences of intermediate structure such as the α-amylases of [*Butyrivibrio fibrisolvens* and *Dictyoglomus thermophilum*, which have ILPNH [30] and FVPNH [31] at the third β-strand, compared with prototypes of α-amylase and CGTase, containing VVANH and FAPNH respectively.

Štefan JANEČEK, E. Ann MacGREGOR† and Birte SVENSSON‡§

* Institute of Ecobiology, Slovak Academy of Sciences, Stefanikova 3, SK-81434 Bratislava, Slovakia, †Department of Chemistry, University of Manitoba, Winnipeg R3T 2N2, Canada, and ‡Department of Chemistry, Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-2500 Copenhagen Valby, Denmark

§ To whom reprint requests and correspondence should be addressed.

1 MacGregor, E. A. and Svensson, B. (1989) Biochem. J. **259**, 145–152
2 Jespersen, H. M., MacGregor, E. A., Sierks, M. R. and Svensson, B. (1991) Biochem. J. **280**, 51–55
3 Janeček, Š. (1993) FEBS Lett. **316**, 23–26
4 Jespersen, H. M., MacGregor, E. A., Sierks, M. R. and Svensson, B. (1993) J. Protein Chem. **12**, 791–806
5 Henrissat, B. (1991) Biochem. J. **280**, 309–316
6 Svensson, B. and Søgaard, M. (1993) J. Biotechnol. **29**, 1–37
7 Matsuura, Y., Kusunoki, M., Harada, W. and Kakudo, M. (1984). J. Biochem. (Tokyo) **95**, 697–702
8 Buisson, G., Duée, E., Haser, R. and Payan, F. (1987) EMBO J. **6**, 3909–3916
9 Boel, E., Brady, L., Brzozowski, A. M., Derewenda, Z., Dodson, G. G., Jensen, V. J., Petersen, S. B., Swift, H., Thim, L. and Woeldike, H. F. (1990) Biochemistry **29**, 6244–6249
10 Qian, M., Haser, R. and Payan, F. (1993) J. Mol. Biol. **231**, 785–799
11 Kadziola, A., Abe, J., Svensson, B. and Haser, R. (1994) J. Mol. Biol. **239**, 104–121
12 Kubota, M., Matsuura, Y., Sakai, S. and Katsube, Y. (1990) Protein Eng. **3**, 328–329
13 Klein, C. and Schulz, G. E. (1991) J. Mol. Biol. **217**, 737–750
14 Klein, C., Hollender, J., Bender, H. and Schulz, G. E. (1992) Biochemistry **31**, 8740–8746
15 Lawson, C. L., van Monfort, R., Strokopytov, B., Rozeboom, H. J., Kalk, K. H., de Vries, G. E., Penninga, D., Dijkhuizen, L. and Dijkstra, B. W. (1994) J. Mol. Biol. **236**, 590–600
16 MacGregor, E. A. (1993) Starch **7**, 232–237
17 Svensson, B. (1994) Plant Mol. Biol. **25**, 141–157
18 Toda, H., Kondo, K. and Narita, K. (1982) Proc. Jpn. Acad. Ser. B **58**, 208–212
19 Nakajima, R., Imanaka, T. and Aiba, S. (1986) Appl. Microbiol. Biotechnol. **23**, 355–360
20 MacGregor, E. A. (1988) J. Protein Chem. **7**, 399–415
21 Nishizawa, M., Ozawa, F. and Hishinuma, F. (1987) DNA **6**, 255–265
22 Itkor, P., Tsukagoshoi, N. and Udaka, S. (1990) Biochem. Biophys. Res. Commun. **166**, 630–636
23 Bahl, H., Burchhardt, G., Spreinat, A., Haeckel, K., Wienecke, A., Schmidt, B. and Antranikian, G. (1991) Appl. Environ. Microbiol. **57**, 1554–1559
24 Janeček, Š. and Toth, D. (1994) Biologia **49**, 301–306
25 Fujiwara, S., Kakikora, H., Sakaguchi, K. and Imanaka, T. (1992) J. Bacteriol. **174**, 7478–7481
26 Janeček, Š. (1992) Biochem. J. **288**, 1069–1070
27 Svensson, B., Jespersen, H., Sierks, M. R. and MacGregor, E. A. (1989) Biochem. J. **264**, 309–311
28 Holm, L., Koivula, A. K., Lehtovaara, P. M., Hemminki, A. and Knowles, J. K. C. (1990) Protein Eng. **3**, 181–191
29 Bender, H. (1990) Carbohydr. Res. **206**, 257–267
30 Rumbak, E., Rawlings, D. E., Lindsey, G. G. and Woods, D. R. (1991) J. Bacteriol. **173**, 4203–4211
31 Horinouchi, S., Fukusumi, S., Ohshima, T. and Beppu, T. (1988) Eur. J. Biochem. **176**, 243–253

Received 25 August 1994