Pergamon

Prog. Biophys. molec. Biol., Vol. 67, No. 1, pp. 67–97, 1997

0079-6107/97 $32.00 + 0.00

PII: S0079-6107(97)00015-1

α-AMYLASE FAMILY: MOLECULAR BIOLOGY AND EVOLUTION

ŠTEFAN JANEČEK

Institute of Microbiology, Slovak Academy of Sciences, Štefánikova 3, SK-81434 Bratislava, Slovakia

CONTENTS

I. INTRODUCTION 67

II. α-AMYLASE ENZYME FAMILY: AN OVERVIEW 68

III. SEQUENCES AND STRUCTURES 69
3.1. *Conserved Sequence Regions* 70
3.2. $(\alpha/\beta)_8$-*Barrels* 73
3.2.1. *Active sites and mutagenesis studies* 74
3.2.2. *Calcium and chloride binding* 77
3.2.3. *Structural basis of thermostability* 78
3.3. *Domain B—Excursion in the Place of Loop 3* 78
3.4. *C-Terminal Parts* 80

IV. EVOLUTION 80
4.1. α-*Amylases* 81
4.2. *Amylases and Cyclodextrin Glycosyltransferases* 85
4.3. *Proteins Sharing Sequence Similarities and the Role of Domain B* 86
4.4. *The Entire Present-day α-Amylase Family* 89

V. CONCLUSION 91
ACKNOWLEDGEMENTS 91
REFERENCES 91

I. INTRODUCTION

When the first parallel $(\alpha/\beta)_8$-barrel of triosephosphate isomerase (TIM) was determined in 1975 (Banner *et al.*, 1975), not many people were able to imagine the scope and the impact of the new protein family that was subsequently established in the following years. The number of different enzyme specificities adopting this charming fold (Fig. 1) has risen from five in 1983 (Muirhead, 1983) through 14 in 1988 (Chothia, 1988) to 20 plus one protein without catalytic function in 1993 (Farber, 1993). α-Amylase has been known to contain this $(\alpha/\beta)_8$-barrel structure from 1980 when its first structure was determined using the α-amylase from *Aspergillus oryzae* (Matsuura *et al.*, 1980), that is often called Taka-amylase A (TAA). The most recent review on the $(\alpha/\beta)_8$-barrel family (Janeček and Bateman, 1996) described 43 different enzymes and three proteins. Remarkably, no ligase (EC 6) has been found up to now that would adopt the $(\alpha/\beta)_8$-barrel structure.

Over the years it was the question of the nature of evolutionary relationships among the $(\alpha/\beta)_8$-barrel proteins which has provoked scientists continuously. Two main evolutionary histories have been proposed: the convergent evolution to a highly symmetric and stable fold, and the divergent evolution from a single common ancestor. The eventuality that all the $(\alpha/\beta)_8$-barrels would be without whatever relatedness was reliably excluded (Brändén, 1991). The other possibility was offered by the intron–exon theory (Gilbert, 1978) that $(\alpha/\beta)_8$-barrels (most work was done with TIM) may have been formed from the combination of exons (Straus and Gilbert, 1985; Marchionni and Gilbert, 1986). There is some difference of opinion as to whether the introns are of ancient nature (Gilbert and Glynias, 1993; De Souza *et al.*, 1996) or that they do not correlate with protein structure (Kwiatowski *et al.*, 1995; Logsdon *et al.*, 1995). The arguments given in support of divergent as well as convergent evolution of $(\alpha/\beta)_8$-barrel proteins can be found in the

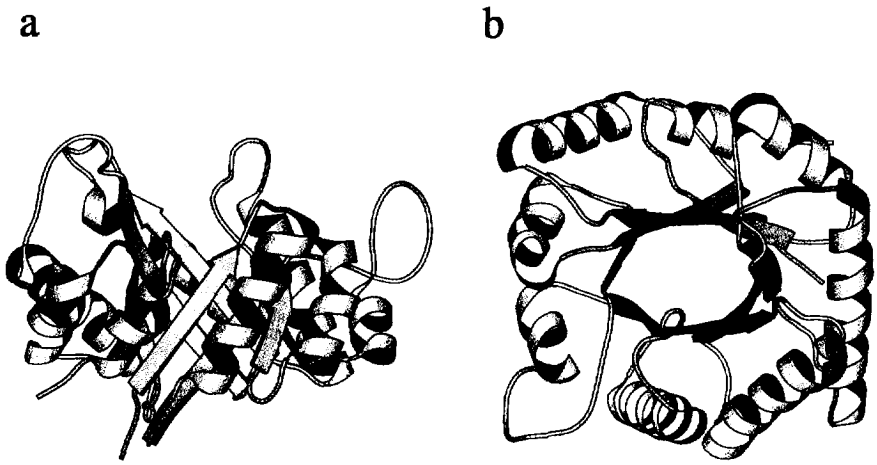

Fig. 1. The parallel $(\alpha/\beta)_8$-barrel. It is called TIM-barrel because of it first being observed in the structure of chicken muscle Triosephosphate IsoMerase (TIM) more than 20 years ago (Banner *et al.*, 1975). This folding motif is composed of eight parallel β-strands forming the inner β-barrel sheet that is surrounded by eight α-helices in such a way that there are eight repeated $(\beta\alpha)$-units in a regular $(\alpha/\beta)_8$-barrel. Therefore, this is used as the classic example to show the general arrangement of the motif. A useful classification of $(\alpha/\beta)_8$-barrel proteins can be found in the SCOP (Structural Classification Of Proteins) database that provides a detailed and comprehensive description of the structural and evolutionary relationships of the proteins of known structure (Murzin *et al.*, 1995). The division of the entire family of $(\alpha/\beta)_8$-barrels reflecting strictly their eventual divergency from a single common ancestor can be found in the excellent reviews of Farber and co-workers (Farber and Petsko, 1990; Farber, 1993; Reardon and Farber, 1995). (a) Side view. The C-terminal end of the β-barrel is toward the top of the page. (b) End view. The C-terminal end of the β-barrel is toward the reader. Figure produced using MOLSCRIPT (Kraulis, 1991). PDB (Bernstein *et al.*, 1977) file used: 1TIM.

literature (Lesk *et al.*, 1989; Farber and Petsko, 1990; Brändén, 1991; Wilmanns *et al.*, 1991; Janeček, 1993a, 1996a,b,c; Raine *et al.*, 1994; Scrutton, 1994; Henrissat *et al.*, 1995; Janeček and Baláž, 1995; Jenkins *et al.*, 1995; Reardon and Farber, 1995). It should be pointed out, however, that this fundamental evolutionary question is still far from reaching an unambiguous answer.

II. α-AMYLASE ENZYME FAMILY: AN OVERVIEW

Besides cellulose, starch ranks among the most abundant carbohydrate polymers on Earth. It is an important source of energy for animals, higher plants, and especially microorganisms. Its rather complex structure (α-1,4-linked α-D-glucose units with α-1,6-branching points) means that it is metabolized by a set of enzymes, commonly named starch hydrolases. The three amylases, α-amylase, β-amylase and glucoamylase, are the best known enzymes operating on starch. Despite their closely related functions, these three enzymes are structurally different and perhaps evolutionarily distantly related (Janeček, 1994a). The main functional difference, except for the place that the enzyme attacks the starch and discriminating the α-amylase from both β-amylase and glucoamylase, is the mechanism of glucosidic bond cleaving: the α-amylase uses the retaining mechanism (the resulting hydroxyl group retains the α-configuration) while the other two use the inverting mechanism (inversion of the anomeric configuration to β).

There are many other starch hydrolases and since these enzymes are able to hydrolyse some polysaccharides related to starch (e.g., pullulan and glycogen), they are mostly named as starch hydrolases and related enzymes. Based on the presence of the common features in their amino acid sequences, revealed by the Hydrophobic Cluster Analysis method (Gaboriaud *et al.*, 1987; Lemesle-Varloot *et al.*, 1990), the starch hydrolases and related enzymes have been classified together with the rest of the O-glycosyl hydrolases (EC 3.2.1.x) into more than fifty families (Henrissat, 1991; Henrissat and Bairoch, 1993,

1996). β-Amylase and glucoamylase form their own families: 14 and 15, respectively, while the rest of starch hydrolases are grouped into the family 13 glycosyl hydrolases. Remarkably, a few transferases (EC 2.4.1.x), the best known is cyclodextrin glycosyltransferase (CGTase), are classified together with them due to their glycosyl hydrolase/transferase activity. All these are members of the so-called α-amylase family which presently covers about 20 different enzyme specificities (Table 1). And the molecular biology and evolution of these enzymes will be the subject of the present review. There are several hundred known primary structures of the α-amylase family enzymes deposited in sequence databases and the three-dimensional structures of several enzyme specificities from various sources have already been determined. The common feature is the presence of a parallel $(\alpha/\beta)_8$-barrel domain which has in all cases inserted a long loop between the third β-strand and the third α-helix (the so-called domain B). All the similarities of both functional and sequence-structural nature reflect the very probable divergent evolution of the α-amylase family enzymes. The attractiveness of this family has recently been strengthened by theoretical demonstration of the eventual occurrence of circular permutations during the evolution of some members of the α-amylase family (MacGregor *et al.*, 1996) as well as by the fact that there are some proteins without catalytic function that may join the family based on unambiguous sequence similarities (Wells and Hediger, 1992; Janeček *et al.*, 1997). The details concerning their biochemistry can be found in several comprehensive reviews (e.g., Fogarty, 1983; Hill and MacGregor, 1988; Robyt, 1989; Vihinen and Mäntsälä, 1989; Søgaard *et al.*, 1993a; Ball *et al.*, 1996; Bauer *et al.*, 1996; Nakamura, 1996; Warren, 1996), also partly in the ENZYME database (Bairoch, 1996) and are therefore not recapitulated in this article.

III. SEQUENCES AND STRUCTURES

The α-amylase family thus consists of a large group of starch hydrolases and related enzymes (for previous reviews see e.g., MacGregor, 1993; Janeček, 1994a; Svensson, 1994)

Table 1. The members of the α-amylase family

EC*	Enzyme/Protein
3.2.1.1	α-Amylase
3.2.1.10	Oligo-1,6-glucosidase
3.2.1.60	Maltotetraohydrolase
2.4.1.19	Cyclodextrin glycosyltransferase
3.2.1.20	α-Glucosidase
3.2.1.41	Pullulanase
3.2.1.1/41	Amylopullulanase
3.2.1.54	Cyclomaltodextrinase
3.2.1.68	Isoamylase
3.2.1.70	Dextran glucosidase
3.2.1.93	Trehalose-6-phosphate hydrolase
3.2.1.98	Maltohexaohydrolase
3.2.1.116	Maltotriohydrolase
3.2.1.133	Maltogenic amylase
3.2.1.135	Neopullulanase
	Maltopentaohydrolase
	Maltooligosyltrehalose hydrolase
2.4.1.18	Glucan branching enzyme
2.4.1.25	Amylomaltase
2.4.1.25/3.2.1.33	Glucan debranching enzyme
	Maltooligosyltrehalose synthase
2.4.1.5	Glucosyltransferase
	Amino acid transport-related protein
	4F2 Heavy-chain cell surface antigen

* EC numbers are given (if known); the members are ordered in the following way: the members with crystallographically known three-dimensional structure → hydrolases → transferases → glucanotransferase that has been proposed to contain a circularly permuted version of the $(\alpha/\beta)_8$-barrel → proteins without catalytic function.

comprising about 20 different enzyme specificities (Table 1). The name "family 13 glycosyl hydrolases" (Henrissat, 1991; Henrissat and Bairoch, 1993, 1996) has also been well established. According to the recent update (March 1997) of the classification of glycosyl hydrolases (Henrissat and Bairoch, 1996), family 13 (the largest family of glycosyl hydrolases) contains 138 members, most of which are α-amylases. The real number of α-amylase family members as searched in the SWISS-PROT (Bairoch and Apweller, 1997), GenBank (Benson *et al.*, 1997) and PIR (George *et al.*, 1997) databases should be, however, even higher.

Three-dimensional structures solved by X-ray crystallography are known for: (1) α-amylases from *Bacillus licheniformis* (Machius *et al.*, 1995; Song *et al.*, 1996), *Aspergillus oryzae* (TAA; Matsuura *et al.*, 1984; Swift *et al.*, 1991), *Aspergillus niger* (Brady *et al.*, 1991), barley (Kadziola *et al.*, 1994), pig pancreas (PPA; Qian *et al.*, 1993), human pancreas (Brayer *et al.*, 1995), and human salivas (Ramasubbu *et al.*, 1996); (2) CGTases from *Bacillus circulans* strain 8 (Klein and Schulz, 1991), *Bacillus circulans* strain 251 (Lawson *et al.*, 1994), *Bacillus* sp. 1011 (Harata *et al.*, 1996), *Bacillus stearothermophilus* (Kubota *et al.*, 1991), and *Thermoanaerobacterium thermosulfurigenes* EM1 (Knegtel *et al.*, 1996); (3) oligo-1,6-glucosidase from *Bacillus cereus* (Kizaki *et al.*, 1993); and (4) maltotetraohydrolase from *Pseudomonas stutzeri* (Matsuura, 1995). The structures of five other α-amylases, those from *Alteromonas haloplanctis* (Feller *et al.*, 1994), *Bacillus licheniformis* (Declerck *et al.*, 1995), *Bacillus stearothermophilus* (Holm *et al.*, 1990; Vihinen *et al.*, 1990), *Saccharomycopsis fibuligera* (Matsui *et al.*, 1994) and *Thermoactinomyces vulgaris* (Hansen *et al.*, 1994) were modelled as well as the CGTase from *Bacillus circulans* var. *alkalophilus* (Demchuk *et al.*, 1993) and the neopullulanase from *Bacillus stearothermophilus* (Kuriki *et al.*, 1996b; Lamminmäki and Vihinen, 1996). The rest of the members of the α-amylase family, as listed in Table 1, were predicted to share a catalytic $(\alpha/\beta)_8$-barrel structure similar to that first observed in TAA (Matsuura *et al.*, 1984) based on the similarities in their amino acid sequences (MacGregor and Svensson, 1989; Raimbaud *et al.*, 1989; Jespersen *et al.*, 1991, 1993).

There are many structures solved as complexes of the enzymes with their substrates or inhibitors (Klein *et al.*, 1992; Larson *et al.*, 1994; Qian *et al.*, 1994, 1995; Knegtel *et al.*, 1995; Strokopytov *et al.*, 1995, 1996; Wiegand *et al.*, 1995; Bompard-Gilles *et al.*, 1996; Gilles *et al.*, 1996; Machius *et al.*, 1996) that have provided indispensable evidence concerning the common catalytic mechanisms of the enzymes from the α-amylase family as well as the nuance differences between the individual enzyme specificities. Also, many other structures announced in their preliminary crystallographic reports are awaited, such as the α-amylases from *Bacillus amyloliquefaciens* (Suzuki *et al.*, 1990), *Bacillus subtilis* (Chang *et al.*, 1993) and *Alteromonas haloplanctis* (Aghajari *et al.*, 1996), the neopullulanase from *Thermoactinomyces vulgaris* (Kamitori *et al.*, 1995), and mainly the amylopullulanase from the hyperthermophilic archaeon *Pyrococcus woesei* (Knapp *et al.*, 1995). It is worth mentioning that there are only two α-amylases, AmyA from the hyperthermophile *Dictyoglomus thermophilum* (Fukusumi *et al.*, 1988) and the one from the archaeon *Pyrococcus furiosus* (Laderman *et al.*, 1993), the sequences of which do not resemble the sequences of the members of the α-amylase family. These two have therefore recently been classified into the new glycosyl hydrolases family 57 (Henrissat and Bairoch, 1996) which could, however, share a common ancestor (although a very distant one) with the main family 13, as indicated by the very recently published sequence of the α-amylase from methanogenic archaeon *Methanococcus jannaschii* (Bult *et al.*, 1996), the sequence of which may contain the features characteristic of both seemingly unrelated families (Š. Janeček, unpublished results). Thus the crystal structure underway of the above-mentioned archaeal amylopullulanase (Knapp *et al.*, 1995) is expected with the special interest mainly taking into account that none of the amino acid sequence of this enzyme was released.

3.1. *Conserved Sequence Regions*

The α-amylase family enzymes exhibit in general a very low degree of sequence similarity. However, they contain several well-defined regions in their amino acid

sequences that are highly conserved (Fig. 2). This fact makes the family an attractive subject for evolutionary studies. For instance, when comparing the sequences of α-amylases originated from the entire spectrum of sources (micro-organisms, plants and animals), the overall sequence similarity is only about 10% (Nakajima *et al.*, 1986).

Toda *et al.* (1982) have pointed out three similar regions between the sequences of PPA and TAA (regions II, IV and VI in Fig. 2). These three regions including the catalytic and substrate binding sites have been demonstrated in bacterial (Friedberg, 1983) and plant (Rogers, 1985) α-amylases as well. Later, as more amino acid sequences of α-amylases have become available, the fourth conserved sequence region (region V in Fig. 2) has been determined (Nakajima *et al.*, 1986). These four conserved regions, since being more or less easily identifiable in all enzyme specificities from the α-amylase family, were then well-established as the conserved sequence regions in the frame of the entire family. They contain all three amino acid residues, Asp206, Glu230 and Asp297 (unless otherwise specified, all numbers throughout this article use the TAA numbering) that were recognized to play a role in the catalysis and in a wider sense in the active site, determined either experimentally by means of X-ray crystallography and site-directed mutagenesis (e.g., Holm *et al.*, 1990; Vihinen *et al.*, 1990; Matsuura *et al.*, 1991; Klein *et al.*, 1992; Nagashima *et al.*, 1992; Nakamura *et al.*, 1992, 1994a; Mathupala *et al.*, 1993; Podkovyrov *et al.*, 1993; Søgaard *et al.*, 1993b; Sin *et al.*, 1994; Takase, 1994; Takata *et al.*, 1994; Penninga *et al.*, 1995; Kuriki *et al.*, 1996b; McCarter and Withers, 1996; Oyama *et al.*, 1996) or theoretically by the similarity of known sequences (e.g., Svensson, 1988; Kuriki and Imanaka, 1989; MacGregor and Svensson, 1989; Watanabe *et al.*, 1990; Jespersen *et al.*, 1991, 1993; Janse *et al.*, 1993; Liu *et al.*, 1993; Oguma *et al.*, 1993; Whiting *et al.*, 1993; Kobayashi *et al.*, 1994; Lee *et al.*, 1994; Braun *et al.*, 1996; Takii *et al.*, 1996). It should be pointed out that despite the fact that region V in Fig. 2 including the catalytic glutamate (Glu230) has been established as one of the best conserved regions

---

Caption for Fig. 2 on following page.

Fig. 2. The conserved sequence regions of the members of the α-amylase family. The best conserved parts of the α-amylase-type $(\alpha/\beta)_8$-barrel comprise the strands β2, β3, β4, β5, β7 and β8. There is also a short conserved sequence located near the C-terminus of the longest loop 3. The three proposed catalytic residues (Asp, Glu and Asp in the strands β4, β5 and β7, respectively) are marked by an asterisk. The segments of following representatives are shown (for known EC numbers, see Table 1): α-amylase from *Aspergillus oryzae* (Toda *et al.*, 1982); oligo-1,6-glucosidase from *Bacillus cereus* (Watanabe *et al.*, 1990); α-glucosidase from *Saccharomyces carlsbergensis* (Hong and Marmur, 1986); pullulanase from *Klebsiella pneumoniae* (Kornacker and Pugsley, 1990); amylopullulanase from *Thermoanaerobacter ethanolicus* (Mathupala *et al.*, 1993); cyclomaltodextrinase from *Bacillus sphaericus* (Oguma *et al.*, 1993); maltotetraohydrolase from *Pseudomonas stutzeri* (Fujita *et al.*, 1989); isoamylase from *Pseudomonas amyloderamosa* (Amemura *et al.*, 1988); dextran glucosidase from *Streptococcus mutans* (Russell and Ferretti, 1990); trehalose-6-phosphate hydrolase from *Escherichia coli* (Rimmele and Boos, 1994); maltohexaohydrolase from *Bacillus* sp. strain 707 (Tsukamoto *et al.*, 1988); maltotriohydrolase from *Natronococcus* sp. strain Ah-36 (Kobayashi *et al.*, 1994); maltogenic amylase from *Bacillus stearothermophilus* (Diderichsen and Christiansen, 1988); neopullulanase from *Thermoactinomyces vulgaris* (Tonozuka *et al.*, 1993); maltopentaohydrolase from *Pseudomonas* sp. strain KO-8940 (Shida *et al.*, 1992); maltooligosyltrehalose hydrolase from *Arthrobacter* sp. strain Q36 (Maruta *et al.*, 1996a); maltooligosyltrehalose synthase from *Arthrobacter* sp. strain Q36 (Maruta *et al.*, 1996a); glucan branching enzyme from *Escherichia coli* (Baecker *et al.*, 1986); cyclodextrin glycosyltransferase from *Bacillus circulans* strain 8 (Nitschke *et al.*, 1990); amylomaltase from potato (Takaha *et al.*, 1993); glucan debranching enzyme from rice (Nakamura *et al.*, 1996); glucosyltransferase from *Streptococcus downei* (Gilmore *et al.*, 1990); human amino acid transport-related protein (Bertran *et al.*, 1993); human 4F2 heavy-chain cell surface antigen (Gottesdiener *et al.*, 1988). The regions with the question mark are either not present in the relevant members or not easily identifiable without additional structural information. The first group are hydrolases, the second group are transferases. The glucosyltransferases represented by the enzyme from *Streptococcus downei* are remarkable: they may contain a circularly permuted version of the α-amylase-type $(\alpha/\beta)_8$-barrel (see the numbering of the regions; MacGregor *et al.*, 1996). The last two protein members are proposed to be incorporated into the α-amylase family due to clear sequence similarities mainly with the oligo-1,6-glucosidase group (see Fig. 10; Wells and Hediger, 1992; Janeček *et al.*, 1997).

Enzyme/Protein	I β2	II β3	III loop3	IV β4	V β5	VI β7	VII β8
α-Amylase	56_GFTAIWIT-P	117_DVVANH	173_LPDLD	202_GLRIDTVKH	230_EVLD	292_FVE--NHD	323_GIPIIYAGQE
Oligo-1,6-glucosidase	44_GIDVIWLS-P	98_DLVVNH	167_QPDLN	195_GFRMDVINF	255_EMPG	324_YWN--NHD	360_GTPYIYQGEE
α-Glucosidase	52_GVDAIWVC-P	106-DLVINH	181_QVDLN	210_GFRIDTAGL	276_EVAH	344_YIE--NHD	381_GTLYVYQGQE
Pullulanase	458_GVTHVELL-P	590_DVVYNH	632_CSDSA	661_GRFRDLMGY	694_EGWD	817_YVS--KHD	859_GIAFDQQGSE
α-Amylase-pullulanase	435_GISVIYLN-P	487_DGVFNH	565_WADFI	593_GWRLDVANE	626_ELWG	698_LLG--SHD	745_GMPSIYYGDE
Cyclomaltodextrinase	187_GVNALYFN-P	240_DAVFNH	294_MPKLN	323_GWRLDVANE	356_EIMH	418_LLG--SHD	450_GTPCIYYGDE
Maltotetraohydrolase	50_GFSAIWMPVP	112_DVVPNH	160_DADLN	189_GFRFDFVRG	219_ELWK	289_FVD--NHD	327_GTPVVYWDHM
Isoamylase	217_GVTAVEFL-P	291_DVVYNH	341_GANFN	370_GFRFDLASV	416_EFTV	502_FID--VHD	570_GTPLMQGGDE
Dextran glucosidase	44_GVMAIWLS-P	98_DLVVNH	162_QPDLN	190_GFRMDVIDM	236_ETWG	308_FWN--NHD	344_GTPYIYQGEE
Trehalose-6-phosphate hydrolase	46_GVDAIWLT-P	100_DMVFNH	168_QADLN	196_GLRLDVVNL	251_EMSS	320_FWC--NHD	356_GTPYIYQGEE
Maltohexaohydrolase	38_GITAVWIP-P	102_DVVMNH	203_YADID	232_GFRIDAVKH	266_EFWK	328_FVD--NHD	362_GYPSVFYGDY
Maltotriohydrolase	35_GVSAIWIPQP	119_DIVLNH	172_LPSMD	200_GLRIDAAAH	232_EVWD	292_FVQ--NHD	325_GMPMLYRGGA
Maltogenic amylase	65_GVTTIWLS-P	127_DFVPNH	196_LADLS	221_GLRIDAVKH	253_EWYG	321_FID--NHD	353_VRPPIYYGTE
Neopullulanase	186_GVTALYFT-P	239_DAVFNH	293_MPKLR	321_GWRLDVANE	354_EIWH	416_LLD--SHD	448_GTPLIYYGDE
Maltopentaohydrolase	31_GFAAVQIS-P	95_DAVINH	153_LQDLN	181_GLRVDAAKH	216_EVIG	280_FVD--NHD	318_GYPALMSATA
Maltooligosyltrehalose hydrolase	147_GVDFIELL-P	202_DVVYNH	235_NLDGP	263_GLRLDAVHA	304_ESDL	395_CSQ--NHD	432_FTPMLLMGEE
Maltooligosyltrehalose synthase	32_GVDWVYLS-P	87_DIVPNH	---?	233_GLRIDHPDG	264_EKIL	478_TLS--THD	613_GVPDVYQGTE
Glucan branching enzyme	280_GFTHLELL-P	335_DWVPGH	---?	401_ALRVDAVAS	458_EEST	519_FVLPLSHD	562_GKKLLFMGNE
Cyclodextrin glycosyltransferase	70_GVTALWISQP	135_DFAPNH	197_LADFN	225_GIRVDAVKH	257_EWFL	323_FID--NHD	354_GVPAIYYGTE
Amylomaltase	106_GLLKMEEL-P	252_DVWANK	290_LYDWK	317_EFRIDHFRG	368_EDLG	416_YTG--THD	485_NIPATQFGNW
Glucan debranching enzyme	403_GLTHVHLL-P	502_DVVYNH	---?	572_GFRFDLMGH	613_EGWD	740_YVS--AHD	782_GIPFFHAGDE
Glucosyltransferase	849_GITQFEMA-P	915_DLVPNQ	---?	433_GVRVDAVDN	475_EAWS	542_FIR--AHD	614_TVTRVYYGDM
				*	*	*	
Amino acid transport protein	156_NIKTVWIT-S	210_DFIPNH	282_QPDLN	310_GFSLDAVKF	---?	---?	474_GTPITYYGEE
4F2 Heavy-chain antigen	154_KVKGLVLG-P	206_DLTPN-	---?	243_GFQVRDIEN	---?	---?	361_GTPVFSYGDE

in the frame of the entire α-amylase family, sometimes it is not very easy to recognize this region, its erroneous identifications also being found in the literature (e.g., Abe *et al.*, 1996). Some sequence nuances and recommendations on how to detect all the conserved sequence regions shown in Fig. 2 reliably were given in the work by Janeček *et al.* (1995).

The very low overall sequence similarity among the α-amylase family enzymes can be manifested by the number of invariant residues, which is only ten (Fig. 2). Seven of them are involved in the active site of these enzymes, Asp117, His122, Arg204 and His296, in addition to the catalytic residues mentioned above, Gly56 and Pro64 flanking the strand β2, and Gly323 having a structurally important role (MacGregor, 1988; Holm *et al.*, 1990; Jespersen *et al.*, 1993; Janeček, 1994a,b; Janeček and Tóth, 1994; Svensson, 1994). Of course, there are some exceptions for more distantly related enzymes (e.g., amylomaltase) as well as for the protein members (cf. Fig. 2).

It is worth mentioning that several of the conserved sequence regions can be used as markers for discrimination between the members of the family that are very closely related by both sequence–structure and function. Thus, the region comprising the strand β2 (region I in Fig. 2) is flanked by the invariant glycine and proline. In the vast majority of the enzyme specificities, the peptide flanked by the Gly and Pro is seven residues long. For the present, only CGTases and some of the maltooligosaccharide-producing members contain an octapeptide between the Gly and Pro. Furthermore, these segments of α-amylases and CGTases are in some cases extremely similar or nearly identical (cf. the stretches of TAA and *Bacillus circulans* CGTase in Fig. 2). Only the glutamine preceding the invariant Pro in CGTase (Fig. 3) can be used as a marker for the correct classification of a newly sequenced enzyme in the absence of the relevant biochemical data (Janeček, 1995a; Janeček *et al.*, 1995). The other conserved region that contains a conserved feature assignable to a given enzyme specificity comprises the segment around the Asp175 (region III in Fig. 2) involved in the binding of a calcium ion, especially in α-amylases (e.g., Boel *et al.*, 1990). This region was originally described in the sequences of α-amylases (Janeček, 1992) and later demonstrated as the so-called fifth conserved sequence region in the other members of the α-amylase family (Janeček, 1995b). Despite the fact that it is hardly traced in a few enzyme specificities (cf. Fig. 2), the lysine substitution of the calcium binding aspartate in cyclomaltodextrinases, neopullulanases and a few α-amylases is directly connected to the deletion of the middle strand of the three-stranded antiparallel β-sheet in domain B of the *Bacillus cereus* oligo-1,6-glucosidase type (Janeček *et al.*, 1997).

3.2. $(\alpha/\beta)_8$-*Barrels*

All the enzymes and proteins listed in Table 1 should contain in their structure an $(\alpha/\beta)_8$-barrel domain. This is derived from the known X-ray structures of the enzyme specificities given in the upper part of Table 1. The rest is believed to have a similar $(\alpha/\beta)_8$-barrel as is strongly supported by the presence of corresponding similarities in amino acid sequences (Fig. 2). The barrel domains found in the α-amylase family enzymes are, however, different from those found in TIM (Fig. 1) giving the first as well as the "classic" (regular) example of an eight-folded parallel $(\alpha/\beta)_8$-barrel motif. The main difference is that the α-amylase-type $(\alpha/\beta)_8$-barrel contains a large loop between strand β3 and helix α3 (Fig. 4) which constitutes, in fact, a small distinct domain, called usually domain B (e.g., Klein and Schulz, 1991; Qian *et al.*, 1993). The $(\alpha/\beta)_8$-barrel with inserted domain B is the part of the structure that is common to each enzyme from the α-amylase family (Jespersen *et al.*, 1993; Janeček, 1994a). It is usually (but not always) formed by 350–400 amino acid residues from the N-terminus of the polypeptide chain, except for some glucan-synthesizing glucosyltransferases that have been predicted recently to contain a circularly permuted α-amylase-type $(\alpha/\beta)_8$-barrel (MacGregor *et al.*, 1996). According to this work, the $(\alpha/\beta)_8$-barrel of these enzymes starts with an α-helix equivalent to helix α3 of α-amylases and their counterpart of domain B is divided to precede and succeed the barrel (see the numbering for glucosyltransferase in Fig. 2).

As already indicated by Fig. 2, the β-strands of the α-amylase-type $(\alpha/\beta)_8$-barrel are better conserved than the corresponding α-helices, which are quite variable in length and

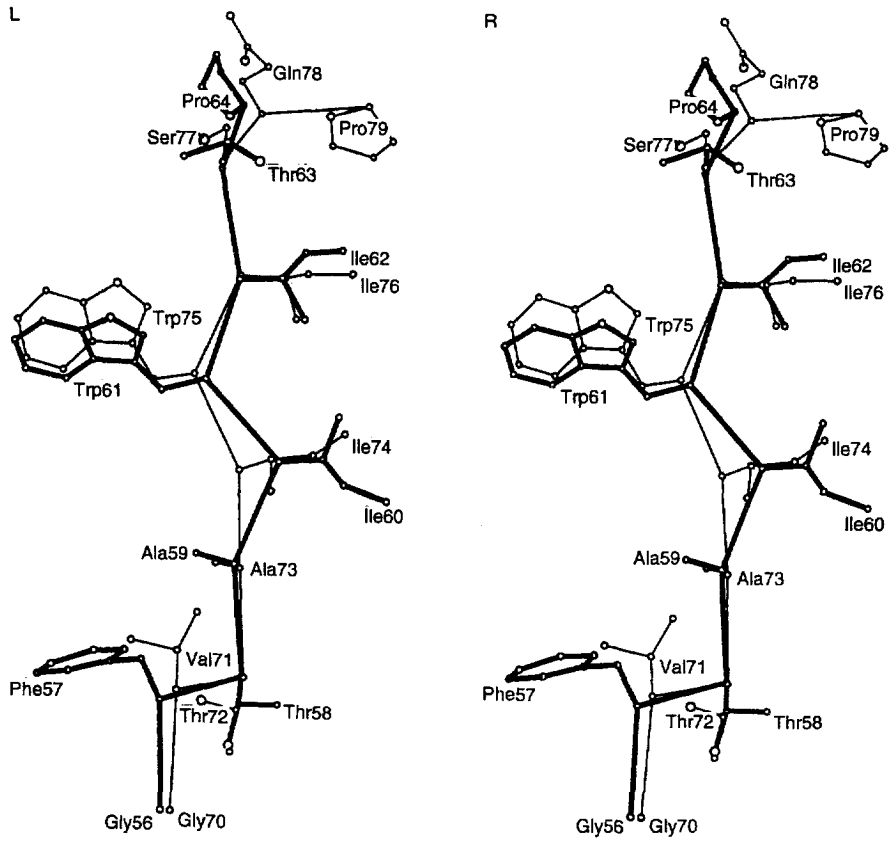

Fig. 3. The strands $\beta 2$ of α-amylase and CGTase are shown overlapped. The $\beta 2$-strand of CGTase from *Bacillus circulans* strain 251 (thin lines) is overlapped on the $\beta 2$-strand of TAA (thick lines). The PDB (Bernstein *et al.*, 1977) files used were 1CDG for CGTase (Lawson *et al.*, 1994) and 2TAA for the α-amylase (Matsuura *et al.*, 1984), respectively. Figure generated using the CCP4 software (CCP4, 1994). Taken from Janeček (1996a). The special feature of this structural alignment is the presence of Gln78 in the structure of CGTase preceding the invariant proline, common to all the α-amylase family enzymes (see Figs 2 and 9).

sequence (Jespersen *et al.*, 1993). This fact along with the differences in loops leave space for emerging various enzyme specificities in this family (cf. Fig. 3). On the other hand, the β-strands apart from their strongly conserved sequence conservation, their geometrical features and side chain packing (Lesk *et al.*, 1989; Pickett *et al.*, 1992) provide a scaffold for residues essential for substrate binding and catalysis (Jespersen *et al.*, 1993). The active site residues are located near the C-termini of the barrel β-strands (cf. Fig. 2), and, in particular, the catalytic Asp206, Glu230 and Asp297 (and their equivalents) are well superimposable (Fig. 5) not only for different α-amylases (Kadziola *et al.*, 1994; Brayer *et al.*, 1995), but also for α-amylases and CGTases (Janeček, 1994a; Machius *et al.*, 1995). Similar comparison with oligo-1,6-glucosidase (Kizaki *et al.*, 1993) as well as maltotetraohydrolase (Matsuura, 1995) has not been possible, since their X-ray co-ordinates are not presently available from the Brookhaven Protein Data Bank (PDB; Bernstein *et al.*, 1977). The location of the α-amylase family enzymes' active sites at the C-terminal side of their barrels is in agreement with that of all known $(\alpha/\beta)_8$-barrel enzymes (Farber and Petsko, 1990; Brändén, 1991).

3.2.1. *Active sites and mutagenesis studies*

During catalytic attack from the three residues, Asp206, Glu230 and Asp297 (and their equivalents), Asp206 and Glu230 are proposed to play the indispensable role of the general base and acid, respectively (Qian *et al.*, 1994; Strokopytov *et al.*, 1995; Bompard-Gilles *et al.*, 1996). The chances of Asp297 acting as a general acid are lowered by the facts

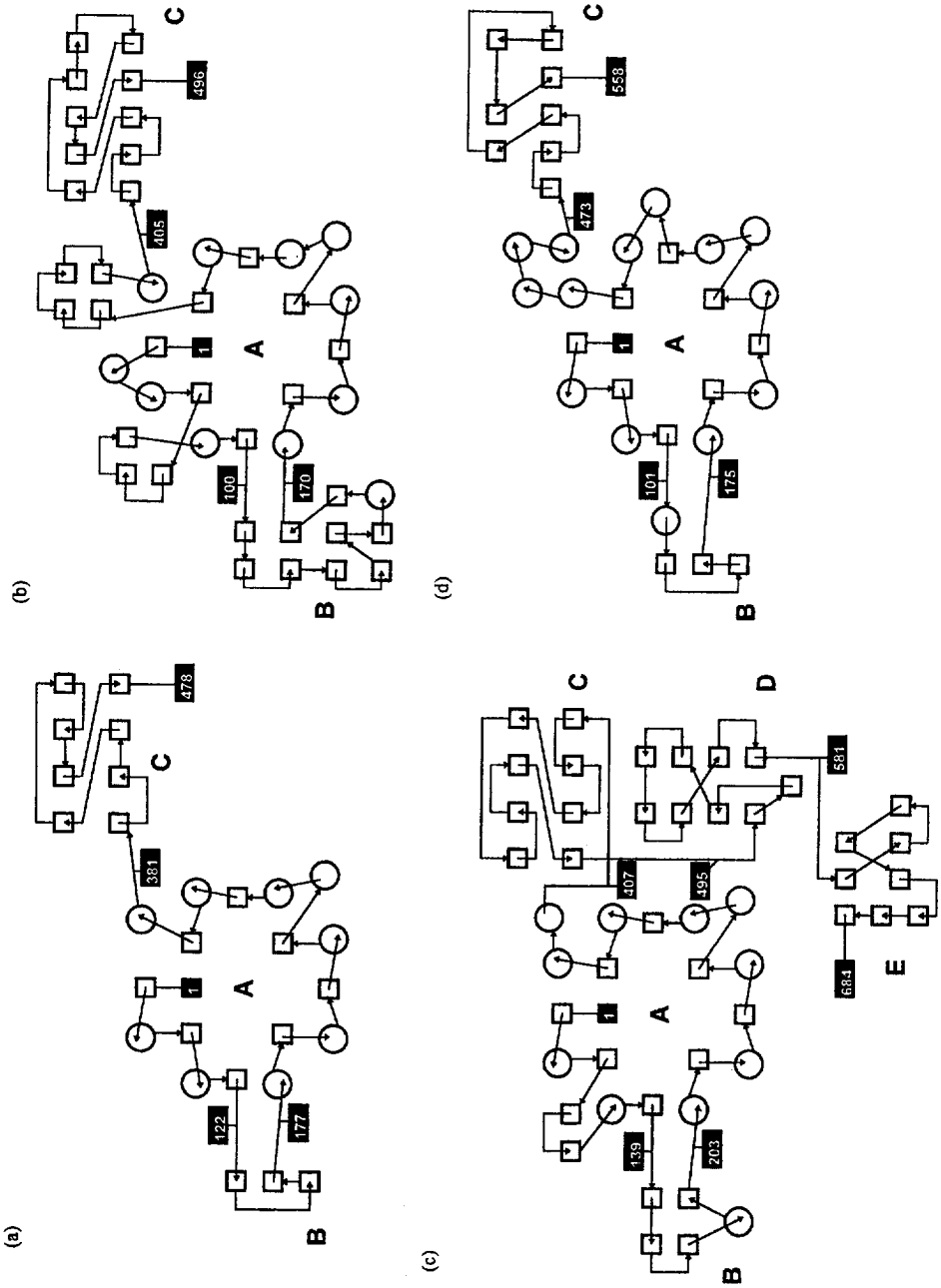

Fig. 4. A schematic representation of secondary structure elements of: (a) TAA (Matsuura *et al.*, 1984), (b) PPA (Qian *et al.*, 1993), (c) CGTase from *Bacillus circulans* strain 8 (Klein and Schulz, 1991), and (d) oligo-1,6-glucosidase from *Bacillus cereus* (Kizaki *et al.*, 1993). The squares and circles represent the strands and helices, respectively. The start of each domain is indicated by the number in the black rectangle. A, B, C, D, and E denote the corresponding domains (for details, see text).

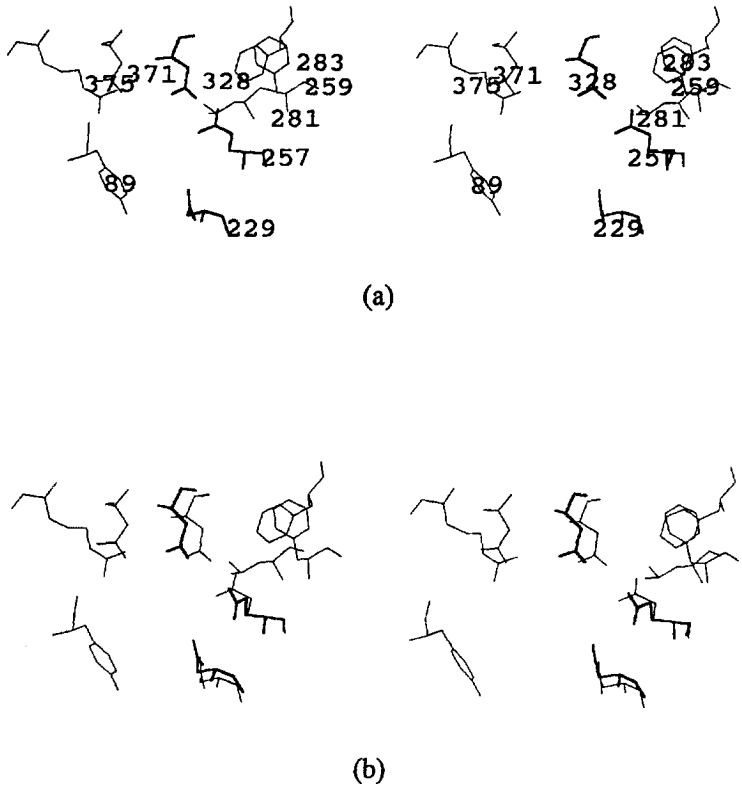

Fig. 5. The α-amylase-type active site. (a) The catalytic residues Asp229, Glu257 and Asp328 (thick lines) of CGTase from *Bacillus circulans* strain 251; PDB (Bernstein *et al.*, 1977) entry: 1CDG (Lawson *et al.*, 1994). Glu257 is surrounded by mostly hydrophobic residues (Phe259, Leu281, Phe283), whereas Asp229 and Asp328 are in a more polar environment (Tyr89, Asp371, Arg375). (b) The same picture (thin lines) with overlapped catalytic residues of TAA; PDB (Bernstein *et al.*, 1977) entry: 6TAA (Swift *et al.*, 1991): Asp206, Glu230 and Asp297 (thick lines). Figure prepared with HYPERCHEM (Autodesk, Inc.). Taken from Janeček (1994a).

that the hydrogen bond of its equivalent in the structure of PPA is longer than that of the equivalent of Glu230 and the orientation of the bond is less appropriate for donating a proton (Qian *et al.*, 1994). However, it also plays an important role, for instance, in *Bacillus circulans* strain 251, where the equivalent of this residue is involved in substrate binding and may be responsible for keeping the equivalent of the Glu230 carboxylate group protonated at higher pHs (Strokopytov *et al.*, 1995). Generally, the enzymes from the α-amylase family apply an acid catalysis mechanism in which the general acid catalyst, e.g., Glu230 in TAA, protonates the oxygen of the glucosidic bond to be cleaved, while the general base catalyst, e.g., Asp206 in TAA, attacks the carbon C-1 of the substrate glycon part (Svensson and Søgaard, 1993; Svensson, 1994). The eventual events in the attack mechanisms of several different α-amylases and CGTases are hypothesized and their discussion can be found in the literature (e.g., Holló *et al.*, 1982; Colonna *et al.*, 1988; Mazur and Nakatani, 1993; Nakamura *et al.*, 1994b; Casset *et al.*, 1995; Nakatani, 1995, 1996a,b; Alkazaz *et al.*, 1996; Suganuma *et al.*, 1996).

There are many other amino acid residues playing a role in the active site of these enzymes. Some of them are highly or invariantly conserved throughout the family, some are specific only for a given enzyme specificity (see Fig. 2). For instance, the three almost invariant histidines, His122, His210 and His296, have been shown to be involved in substrate recognition in the active site of mammalian pancreatic α-amylases (Ishikawa *et al.*, 1992, 1993) and in the transition state stabilization but not directly in catalysis of barley α-amylase (Søgaard *et al.*, 1993b). The *Bacillus stearothermophilus* α-amylase histidine residue equivalent to His210 of TAA may control the specific activity and

thermal stability (Vihinen *et al.*, 1990). A similar function has been reported for this residue in *Bacillus subtilis* α-amylase (Takase, 1994). In CGTases, the three histidines are very probably responsible for both the cycling and amylolytic activities (Mattsson *et al.*, 1995). It has been demonstrated, however, that the physicochemical properties, activity as well as the substrate specificity of these enzymes can be modified by mutation of non-essential amino acid residues adjacent to or near the catalytic residues (Takase, 1992; Inohara-Ochiai *et al.*, 1997).

In fact, most of the functionally essential or at least important residues can be found in the conserved sequence regions as they are described in Fig. 2. It has been shown that the Lys209 of TAA (preceding the very important histidine) is essential for α-amylase activity and is acting as a substrate binding residue (Nagashima *et al.*, 1992). Matsui *et al.* (1992a,b) have manifested that this lysine in the related α-amylase from *Saccharomycopsis fibuligera* controls the activity specific for substrates differing in length. The additional mutation of Tyr83 and Trp84 located in the second loop of the $(\alpha/\beta)_8$-barrel in this yeast α-amylase resulted in the increase of transglycosylation activity with a simultaneous decrease of the hydrolytic activity (Matsui *et al.*, 1994). The roles of these and several other residues, especially tyrosines and tryptophans, in various α-amylases have also been studied by chemical modification approaches (e.g., Fujimori *et al.*, 1978; Hoschke *et al.*, 1980a,b; Kochhar and Dua, 1985; Gibson and Svensson, 1986; Ishikawa and Hirata, 1989; Bealin-Kelly *et al.*, 1992; Yamashita *et al.*, 1993; Cao and Preiss, 1996).

3.2.2. *Calcium and chloride binding*

Several members of the α-amylase family are known to bind one or more calcium ions per protein molecule. Among them, the binding of a calcium has been most deeply studied in α-amylases and CGTases. It has been postulated by Buisson *et al.* (1987) that Ca^{2+} preserves the structural integrity of the active site by linking the two fragments, the catalytic $(\alpha/\beta)_8$-barrel and domain B. For instance, in the *Aspergillus niger* α-amylase, there are primary as well as secondary binding sites (Boel *et al.*, 1990). The primary site is essential in maintaining proper folding around the active site and contains a tightly bound Ca^{2+}, while the secondary site is located at the bottom of the substrate binding cleft and involves the catalytic residues (Asp206 and Glu230) which explains the inhibitory effect of calcium observed at higher concentrations (Boel *et al.*, 1990). The primary calcium binding site is well conserved and is also present in the structures of TAA (Matsuura *et al.*, 1984; Swift *et al.*, 1991), *Bacillus licheniformis* (Machius *et al.*, 1995), barley (Kadziola *et al.*, 1994) and mammalian (Qian *et al.*, 1993; Larson *et al.*, 1994; Brayer *et al.*, 1995; Ramasubbu *et al.*, 1996) α-amylases, all CGTases (Klein and Schulz, 1991; Lawson *et al.*, 1994; Matsuura and Kubota, 1995; Harata *et al.*, 1996; Knegtel *et al.*, 1996) and maltotetraohydrolase (Matsuura, 1995) as well. Generally, four residues serve as ligands (Machius *et al.*, 1995). Three of them are strictly conserved: Asn121, Asp175, His210, while the equivalents of the fourth residue which is in TAA Glu162 are Arg158 and Ile 190 in the PPA, and CGTase from *Bacillus circulans*, respectively. The Asp175 coming from the C-terminal end of domain B is the best conserved residue (cf. region III in Fig. 2) from the relevant conserved sequence region (Janeček, 1992, 1995b). Interestingly, the amylopullulanase from *Thermoanaerobacter ethanolicus* 39E which also requires calcium (Mathupala *et al.*, 1993) contains the asparagine, aspartate and glutamine residues equivalent to TAA calcium ligands while it has Glu601 instead of the invariant His210 of TAA. The same substitution can be found in the closely related *Aspergillus niger* α-amylase (Boel *et al.*, 1990). As far as the crystal structure of oligo-1,6-glucosidase is concerned (Kizaki *et al.*, 1993), the corresponding residues can be identified in all oligo-1,6-glucosidases, those two equivalent with the Asn121 and Asp175 from TAA being invariantly conserved. However, the oligo-1,6-glucosidases and other α-glucosidases unlike the α-amylase do not need calcium for their activity and stability (Y. Suzuki, personal communication).

While the binding of a calcium is considered to be a more general feature common for several different enzyme specificities from the α-amylase family, the binding of a chloride anion seems to be a feature characteristic of a few α-amylases only. Chloride activation is typical for the animal pancreatic and salivary α-amylases (Qian *et al.*, 1993; Brayer *et al.*, 1995; Ramasubbu *et al.*, 1996) and that one from *Bacillus licheniformis* (Machius *et al.*, 1995), whereas most α-amylases seem chloride independent (Vihinen and Mäntsälä, 1989). An interesting example of a bacterial α-amylase that binds a chloride is the one from *Alteromonas haloplanctis* (Feller *et al.*, 1992). This psychrophilic α-amylase has been found to be structurally closely related to the group represented by the PPA (Feller *et al.*, 1994; Janeček, 1994b), the 49% sequence identity being revealed between these two α-amylases. The ligands of the bound chloride in the animal enzymes (Qian *et al.*, 1993; Larson *et al.*, 1994; Brayer *et al.*, 1995; Ramasubbu *et al.*, 1996) are Arg195, Asn298 and Arg337 (PPA numbering), while in the structure of chloride binding site of *Bacillus licheniformis* α-amylase the Arg337 is replaced by a water molecule (Machius *et al.*, 1995). Based on the known crystal structures in conjunction with chloride titration curves Feller *et al.* (1996) proposed the role of this monovalent anion in the chloride-dependent α-amylases, whereby it allows catalytic Glu233 (PPA) to be protonated at the pH of maximal enzymatic activity. Moreover, Machius *et al.* (1995) have pointed out that the similarity to the chloride binding regions of the other enzymes from the α-amylase family is clear, so that a Cl^- could be present in their structures as well, although the authors did not report it.

3.2.3. *Structural basis of thermostability*

Here, some other work dealing with proline residues is worth mentioning. The interesting so-called proline rule for thermostabilization of proteins proposed by Suzuki and co-workers (Suzuki *et al.*, 1987; Suzuki, 1989) has been demonstrated just on the enzymes from the α-amylase family, oligo-1,6-glucosidases (Watanabe *et al.*, 1991, 1994, 1996). An increasing number of prolines (19 for *Bacillus cereus* oligo-1,6-glucosidase, 24 for *Bacillus coagulans* enzyme, and 32 for *Bacillus thermoglucosidasius* enzyme) was observed with the rise in thermostabilities of the oligo-1,6-glucosidases, the prolines at the second sites of β-turns and in the first turn of α-helices being found to contribute to the thermostability most (Watanabe *et al.*, 1996). This may have clear evolutionary implications for how the proteins have achieved their naturally enhanced thermostability.

A similar trend was observed for three bacterial α-amylases, those from *Bacillus subtilis* (Yang *et al.*, 1983), *Bacillus amyloliquefaciens* (Takkinen *et al.*, 1983) and *Bacillus stearothermophilus* (Ihara *et al.*, 1985). Their increased thermostabilities have correlated with the increased hydrophobicity of the interior, as well as the increased hydrophilicity of the exterior of their protein molecules (Janeček, 1993b). By comparing the amino acid sequences of related *Bacillus* liquefying α-amylases, it was proposed that the *Bacillus licheniformis* α-amylase (Yuuki *et al.*, 1985) could be further stabilized by mutation of His133 and Ala209 by hydrophobic amino acid residues (Declerck *et al.*, 1990; Joyet *et al.*, 1992). The constructed structural model of *Bacillus licheniformis* α-amylase has shown that the gain in protein stability can be ascribed in the case of His133 to the β-sheet-forming potential of the inserted leucine and in the case of Ala209 to the cavity-filling effect of the inserted valine (Declerck *et al.*, 1995). The mechanisms of irreversible thermoinactivation of these enzymes were also discussed (Tomazic and Klibanov, 1988a,b; Suzuki *et al.*, 1989; Brosnan *et al.*, 1992).

3.3. *Domain B—Excursion in the Place of Loop 3*

As indicated by Machius *et al.* (1995), domain B inserted between the $(\alpha/\beta)_8$-barrel elements β3 and α3, could be the least similar region of the enzymes from the α-amylase family. And indeed, this must be true, since domain B from different sources varies greatly in both length and sequence (Jespersen *et al.*, 1993; Janeček *et al.*, 1997). For instance 40 amino acid residues are enough for domain B in the *Escherichia coli* glucan

branching enzyme (Baecker *et al.*, 1986) and potato amylomaltase (Takaha *et al.*, 1993), while in the glucan debranching enzyme from rabbit muscle it takes about 250 residues (Liu *et al.*, 1993). Furthermore, this domain from two *Bacillus* α-amylases, one from the saccharifying enzyme produced by *Bacillus subtilis* (Yang *et al.*, 1983) and the other from the liquefying enzyme produced by *Bacillus licheniformis* (Yuuki *et al.*, 1985), differs in its length (51 and 101 residues, respectively; MacGregor, 1988; Machius *et al.*, 1995), the amino acid sequences of these two domains B being found to be hardly alignable (Janeček *et al.*, 1997). However, as will be shown later, this does not imply that domain B in the present-day members of the α-amylase family are the evolutionary products of more ancestors. There is one clear sequence feature near the C-terminus of domain B that can be considered to rank among the most conserved regions of these enzymes (see region III in Fig. 2). It comprises the nearly invariant aspartate residue (Asp175 in TAA) which is involved in the binding of calcium ions. This stretch was first described as a conserved sequence region in various microbial, plant and animal α-amylases (Janeček, 1992) and later demonstrated in many different enzyme specificities from the entire family (Janeček, 1995b). In spite of the fact that this aspartate or its equivalent conservative substituent seems to be absent in several (evidently more distantly related) enzyme specificities (cf. Fig. 2), this region is interestingly also present in the circularly permuted members of the family, i.e. in glucosyltransferases (B. Svensson, personal communication). Of the other functionally important residues located in domain B, Tyr195 (*Bacillus circulans* strain 251 CGTase numbering) has been shown to control the product specificity in such a way that the leucine and tryptophan mutants of several *Bacillus* CGTases (Nakamura *et al.*, 1994a; Sin *et al.*, 1994; Penninga *et al.*, 1995) exhibited a shift towards the synthesis of larger products (γ-cyclodextrins) which are relatively rare among the products of CGTases (Kato and Horikoshi, 1986). Interestingly, using the crystal structure of the Y195F mutant of the above mentioned CGTase, Strokopytov *et al.* (1996) were able to propose the mechanism of its cyclization reaction with discussion on how to tailor the product specificity of the enzyme to α-, β- or γ-cyclodextrins.

Little is known about the function of domain B in the α-amylase family enzymes. Nevertheless, some pioneering work has been done on the barley α-amylase isozymes 1 and 2 by Svensson and co-workers (Rodenburg *et al.*, 1994; Juge *et al.*, 1995). The protein sequences of these two pH isozymes have approximately 80% positional identity (Rogers and Milliman, 1983; Rogers, 1985) and differ in some physico-chemical properties. The constructed isozyme hybrids (Juge *et al.*, 1993) revealed that domain B determines several functional and stability properties, such as stability at low pH characteristic of AMY1 and sensitivity to barley amylase/subtilisin inhibitor specific to AMY2 (Rodenburg *et al.*, 1994; Juge *et al.*, 1995). In a similarly oriented study, Terashima *et al.* (1996) have shown by constructing the chimeric enzyme engineered from two rice α-amylase isozymes, that protein modules of homologous proteins are interchangeable, and that chimeric enzymes with new reaction properties can be tailored by the combination of the suitable protein modules. They prepared a chimeric rice α-amylase containing the N-terminal part from isozyme Amy1A (first part of the $(\alpha/\beta)_8$-barrel with domain B) and the C-terminal part from isozyme Amy3D (the rest of the structure following the end of domain B), the two isozymes being found to exhibit 70% similarity (O'Neil *et al.*, 1990). Due to the lack of an N-linked carbohydrate chain which is present in the isozyme Amy1A only and is crucial for its enhanced thermostability (Terashima *et al.*, 1994), the chimeric α-amylase exhibited also lowered thermostability. On the other hand, the nearly 85% similarity of the N-terminal part of the $(\alpha/\beta)_8$-barrel and domain B with those parts of the AMY2 of barley α-amylase (Kadziola *et al.*, 1994) including the perfectly conserved Ca^{2+} binding residues in domain B, may indicate that this region could be responsible for maintaining the proper structure required for high efficiency in soluble-starch hydrolysis (Terashima *et al.*, 1996). The same functions, i.e., the correct folding and enzymatic activity, were ascribed to the first part of the $(\alpha/\beta)_8$-barrel involving domain B for *Bacillus* α-amylases (Conrad *et al.*, 1995).

3.4. *C-Terminal Parts*

As far as the C-terminal parts of the structure of the α-amylase family members are concerned, it seems that all of them contain a domain C that succeeds the $(\alpha/\beta)_8$-barrel (Janeček, 1994a). The prediction study by Jespersen *et al.* (1991) revealed that this domain may differ slightly in structure with different enzyme specificity. The basic motif adopted by the domain C is the Greek-key motif (Richardson, 1977; Blundell *et al.*, 1981) that should be present, in fact, in all α-amylases, CGTases and oligo-1,6-glucosidases (see Fig. 4). In the crystal structure of *Pseudomonas stutzeri* maltotetraohydrolase (Matsuura, 1995), this domain is composed of only one β-sheet of five antiparallel β-strands, one of which is in TAA replaced by a four-stranded antiparallel β-sheet (cf. Fig. 4A). The function of the C-domain is unknown, although inactivation caused by mutations in (Holm *et al.*, 1990) and truncations of (Vihinen *et al.*, 1994) domain C of *Bacillus stearothermophilus* α-amylase suggests it is necessary for enzyme activity. This is supported by the fact that both isolated fragments comprising either the $(\alpha/\beta)_8$-barrel with domain B or domain C prepared by limited proteolysis of PPA were inactive (Desseaux *et al.*, 1991).

The following domain D which is characteristic of CGTases only (see Fig. 4) adopts an immunoglobulin folding pattern (Bork *et al.*, 1994). Based on the sequence similarity between CGTases and the maltogenic amylase from *Bacillus stearothermophilus* (Jespersen *et al.*, 1991; Janeček, 1995a), it can be suggested that the maltogenic amylase (Diderichsen and Christiansen, 1988) will contain a domain highly similar to domain D of the CGTases. The exact function of domain D is unknown (Jespersen *et al.*, 1991), however, it was speculated that it might be involved in carbohydrate binding (Bork *et al.*, 1994).

More is known about the least C-terminal domain of CGTases, domain E (Klein and Schulz, 1991; Lawson *et al.*, 1994; Matsuura and Kubota, 1995; Harata *et al.*, 1996; Knegtel *et al.*, 1996). This domain was predicted in a number of different starch-degrading enzymes (Svensson *et al.*, 1989). It is, in fact, the only joining feature of the α-amylase family enzymes, β-amylases and glucoamylases consisting of several β-strands (Fig. 6). The function of binding granular starch hypothesized for this domain previously (Svensson *et al.*, 1989) was recently clearly confirmed (Dalmia *et al.*, 1995; Penninga *et al.*, 1996; Sorimachi *et al.*, 1996). This domain has been shown in the *Bacillus circulans* strain 8 CGTase, to function in the binding of raw starch and in guiding linear starch chains into the active site (Penninga *et al.*, 1996). Furthermore, it behaves so independently that it can be fused to the other enzymes (e.g., β-galactosidase), its unique enzymatic properties being saved (Dalmia *et al.*, 1995). It should be pointed out that this domain is highly characteristic of CGTases but is rare for the other α-amylase family enzymes. It is present, for instance, in a few *Streptomyces* α-amylases (Long *et al.*, 1987; Virolle *et al.*, 1988; Vigal *et al.*, 1991; Petricek *et al.*, 1992) and recently it was found in the yeast α-amylase from *Cryptococcus* sp. S-2 (Iefuji *et al.*, 1996).

There are also N-terminal domains in some members of the α-amylase family, such as pullulanase, isoamylase, neopullulanase and glucan branching enzyme. The analysis and discussion of these segments was given by Jespersen *et al.* (1991).

IV. EVOLUTION

The evolution of the α-amylase family may look like a "never-ending story". Despite the fact that its basic frames are known and well understood (dictated by the similarities of the conserved sequence regions shown in Fig. 2), the great variety of sources (bacteria, eucarya and archea) as well as enzyme specificities (Table 1) have been offering new, often fascinating observations that justify the interest in their evolutionary relationships. As an example, the α-amylase from the Antarctic psychrophile *Alteromonas haloplanctis* can be used (Feller *et al.*, 1992), the amino acid sequence of which was found to resemble sequences of animal (i.e. also human) α-amylases, thus indicating the close evolutionary relations among this psychrophilic and all known animal α-amylases (Feller *et al.*, 1994; Janeček, 1994b). There are many other examples of more close or more distant relatedness

```
                ----β1----          ---β2--              -β3-          --β4--
Aspka_AMY   516 TTLPITFEEL-VTTTYGEEVYLSGSISQLGEWHT-SDAVKLSADDYTSSNPEWSVTVS
Crysp_AMY   544 SAGTVVFDVY-VQTQYGQSVVIAGNIPQLGNWSP-ANGLNLNANQYTASSPKWTGTIT
Strli_AMY   441 GQTSASFHVN-ATTAWGENIYVTGDQAALGNWDP-ARALKLDP----AAYPVWKLDVP
Bci_8_CGT   583 DQVTVRFVVNNASTTLGQNLYLTGNVAELGNWSTGSTAIGPAFNQVIHQYPTWYYDVS
Psest_M4H   428 ALVSVSFRCDNGATQMGDSVYAVGNVSQLGNWSP-AAALRLTD---TSGYPTWKGSIA
Cloth_BMY   420 GTIPVTFTINNATTYYGQNVYIVGSTSDLGNWNT-TYARGPAS---CPNYPTWTITLN
Aspni_GMY   513 TAVAVTFDLT-ATTTYGENIYLVGSISQLGDWET-SDGIALSADKYTSSDPLWYVTVT
                    *      *  *      *    ** *                  * *
consensus                  T  G      G    LG W                  P W

                ---- ----β5----          -β6    --β7-        ---β8---
Aspka_AMY       -LPVGTTFEYKFIKVDEGG---SVTWESDPNREYTVPECGSGSGETVVDTWR- 619
Crysp_AMY       GVAPGTTFQWKPIVVTNG----NDNWYPGNNQQATTGSACSSPAADIEFTWSS 646
Strli_AMY       -LAAGTPFQYKYLRKDAAG---KAVWESGANRTATVGTTG---ALTLNDTWRG 538
Bci_8_CGT       -VPAGKQLEFKFFKKNGS----TITWESGSNHTFTTPASG---TATVTVNWQ- 684
Psest_M4H       -LPAGQNEEWKCLIRNEANATQVRQWQGGAN-NSLTPSEG---ATTVGRL--- 526
Cloth_BMY       -LLPGEQIQFKAVKIDSSG---NVTWEGGSNHTYTVPTSG---TGSVTITWQN 519
Aspni_GMY       -LPAGESFEYKFIRIESDD---SVEWESDPNREYTVPQACGTSTATVTDTWR- 616
                    *     *              *     *
consensus               K                W     N
```

Fig. 6. Starch binding domain of starch-hydrolases. Sources of enzymes: Aspka_AMY, α-amylase from *Aspergillus kawachii* (Kaneko *et al.*, 1996); Crysp_AMY, α-amylase from *Cryptococcus* sp. S-2 (Iefuji *et al.*, 1996); Strli_AMY, α-amylase from *Streptomyces limosus* (Long *et al.*, 1987); Bci_8_CGT, CGTase from *Bacillus circulans* strain 8 (Nitschke *et al.*, 1990); Psest_M4H, maltotetraohydrolase from *Pseudomonas stutzeri* (Fujita *et al.*, 1989); Cloth_BMY, β-amylase from *Clostridium thermosulfurogenes* (Kitamoto *et al.*, 1988); and Aspni_GMY, glucoamylase from *Aspergillus niger* (Boel *et al.*, 1984). Gaps are indicated by dashes. Invariant residues are signified by asterisks. The residues identified as consensus in the first report on the motif (Svensson *et al.*, 1989) are provided. The β-strands elements as identified in the CGTase (Klein and Schulz, 1991) are indicated above the alignment. The Strli_AMY represents the related α-amylases from *Streptomyces violaceus* (Virolle *et al.*, 1988), *Streptomyces griseus* (Vigal *et al.*, 1991) and *Thermomonospora curvata* (Petricek *et al.*, 1992). The alignment was produced by CLUSTAL W (Thompson *et al.*, 1994).

either in the frame of one enzyme specificity or among the members belonging to different enzyme specificities. Some members of the α-amylase family exhibit the so-called "intermediary" character (Janeček, 1995a; Janeček *et al.*, 1995), i.e., they contain sequence features characteristic of the enzyme specificity other than that which they belong to. The amino acid transport-related proteins recognized as sequentially similar to α-glucosidases (Wells and Hediger, 1992) can be added to the entire α-amylase family, being very probably recruited from the bacterial oligo-1,6-glucosidases (Janeček *et al.*, 1997). Thus, the evolution of the α-amylase family can really be an attractive subject.

4.1. *α-Amylases*

The evolution of α-amylases was presented in the original study by Janeček (1994b). For 37 different microbial, plant and animal α-amylases an unrooted evolutionary tree was made which indicated three main evolutionarily related groups: (1) fungi and yeasts; (2) plants; and (3) *Alteromonas haloplanctis*, streptomycetes, *Thermomonospora curvata*, insects and mammals. The other microbial α-amylases were scattered in the tree by grouping the known related α-amylases, such as the intracellular enzymes from *Escherichia coli* and *Salmonella typhimurium* (Raha *et al.*, 1992), liquefying α-amylases (Yuuki *et al.*, 1985), and those from *Aeromonas hydrophila* and *Xanthomonas campestris* (Gobius and Pemberton, 1988; Hu *et al.*, 1992). The grouping together of the α-amylases from *Butyrivibrio fibrisolvens* (with the unusually large polypeptide chain, 943 amino acid residues of the mature enzyme (Rumbak *et al.*, 1991)) and *Bacillus subtilis* (Yang *et al.*, 1983) was an interesting observation reflecting their more than 40% sequence identity (Fig. 7). The three bacterial α-amylases, those from *Bacillus megaterium* (Metz *et al.*, 1988), *Dictyoglomus thermophilum* AmyB (Horinouchi *et al.*, 1988) and *Micrococcus* sp.

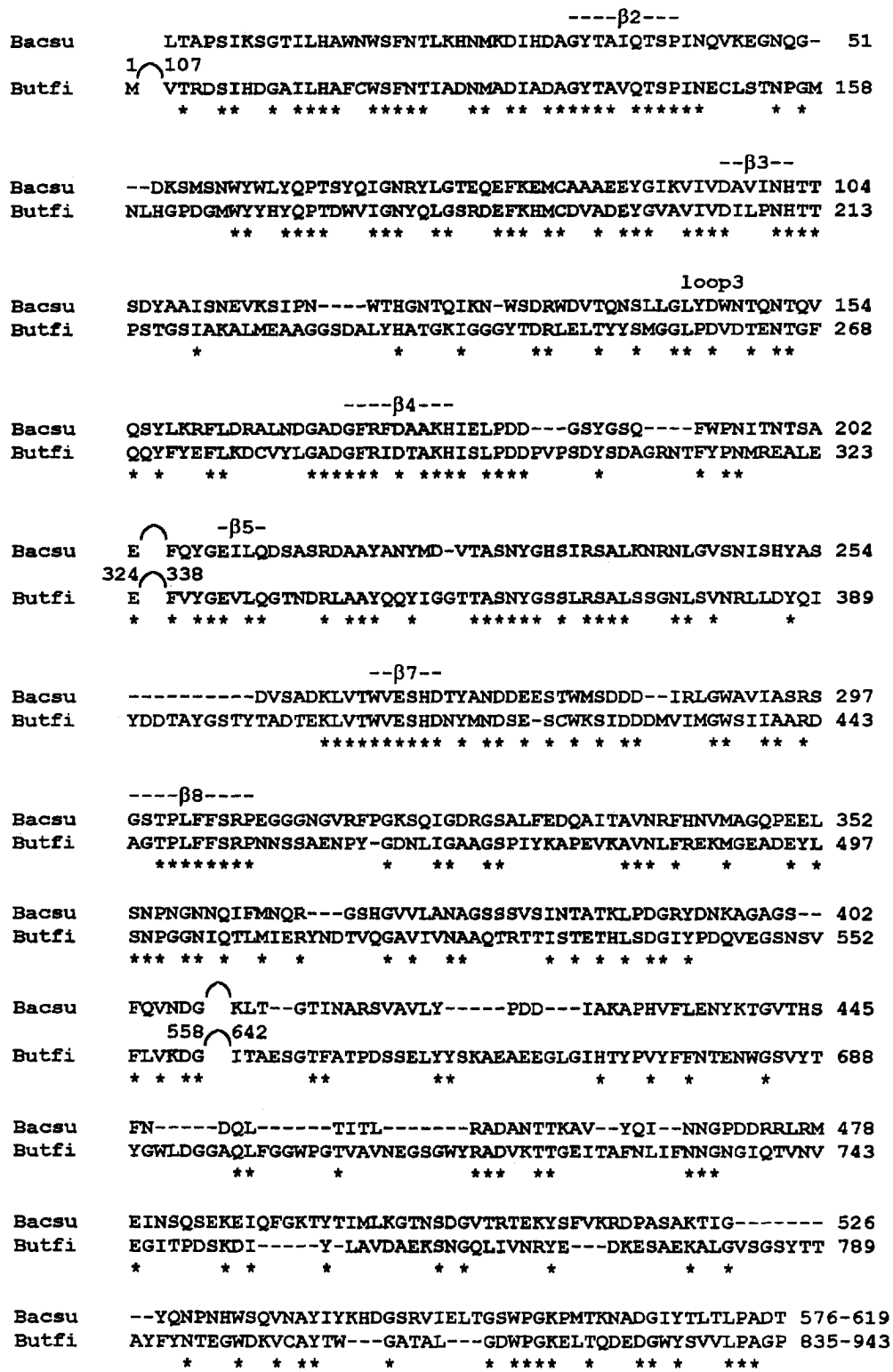

Fig. 7. Sequence alignment of the α-amylase from *Bacillus subtilis* (Bacsu; Yang *et al.*, 1983) with the α-amylase from *Butyrivibrio fibrisolvens* (Butfi; Gobius and Pemberton, 1988) indicating their evolutionary relatedness (Janeček, 1994b) in spite of the quite different lengths of their polypeptide chains: Bacsu: 619, Butfi: 943 amino acid residues. The identities (40.8%) are marked by the asterisks. The conserved regions characteristic of the entire α-amylase family (Fig. 2) are given above the alignment. Remarkably, the short conserved sequence near the C-terminus of domain B of Bacsu enzyme (identified by "loop3": 144_LYDWN) is highly similar to the corresponding stretch of potato amylomaltase (Takaha *et al.*, 1993; 290_LYDWK; cf. Fig. 2) indicating the existence of "intermediaries" in the sequences of more distantly related members of the α-amylase family (Janeček, 1995b). The alignment was made using CLUSTAL V (Higgins *et al.*, 1992).

207 (Kimura and Horikoshi, 1990) were on the longest branches of that tree, indicating their more distant relatedness to the rest of the α-amylases studied (Janeček, 1994b).

Since that time, however, about 15 new α-amylase sequences from different sources covering micro-organisms, plants and animals have been reported. Many of them can be added to the previously recognized groups. Thus, for instance, the intracellular α-amylase from *Streptococcus bovis* (Whitehead and Cotta, 1995) goes well with the liquefying enzymes from the genus *Bacillus* (Yuuki *et al.*, 1985). The periplasmic α-amylase from *Xanthomonas campestris* K-11151 that was reported to have the mixed substrate specificity of α-amylase, cyclodextrinase and neopullulanase (Abe *et al.*, 1996) is closely related to the *Bacillus megaterium* α-amylase (43.2% sequence identity) which can also act on starch-related materials, pullulan and cyclodextrins (Brumm *et al.*, 1991). Based on the unambiguous sequence similarity with the *Bacillus cereus* oligo-1,6-glucosidase (Watanabe *et al.*, 1990) in the domain B region (Fig. 8), from the evolutionary point of view, the α-amylase from *Xanthomonas campestris* K11151 ranks among the oligo-1,6-glucosidase group with the structure in domain B formed by one α-helix and a three-stranded antiparallel β-sheet (Janeček *et al.*, 1997). This could be the same case for the lipoprotein α-amylase from the hyperthermophilic bacterium *Thermotoga maritima* (Liebl *et al.*, 1997; cf. Fig. 8).

Two animal α-amylases (still not published, just released from the GenBank), those from chicken (Benkel *et al.*, 1996) and mollusc *Pecten maximus* (Le Moine *et al.*, 1996), are closely related to the pancreatic and salivary α-amylase isozymes of mammals (Janeček, 1994b). The chicken enzyme is more closely related to mammalian ones. It exhibits 80.2% sequence identity to PPA, whereas the value for the α-amylase from *Pecten maximus* is 58.5%. These evolutionary relationships are in accordance with the proposed disulphide bridges based on the alignment of cysteine residues: 5 (the same as in the PPA; Pasero *et al.*, 1986) in the chicken α-amylase and 4 (the one corresponding to 70–115 in PPA is missing) in the *Pecten maximus* enzyme. The same observations as reported here for the *Pecten maximus* α-amylase are applicable for the shrimp α-amylase from *Penaeus vannamei* (Van Wormhoudt and Sellos, 1996). Similar close relationships to this mammalian group can be awaited for the ostrich pancreatic α-amylase (Oosthuizen *et al.*, 1994) and the α-amylases from the digestive tube of two sea urchins, *Strongylocentrotus nudas* (Nakatani and Kobayashi, 1996) and *Anthocidaris crassispina* (Nakatani *et al.*, 1996).

As far as the insect α-amylases are concerned, a lot of work was done using the different

```
                --II--      --Bα1-              -Bβ1-                  -Bβ2
                *** ***** * **       **     *  * *                *  * **
Xanca_AMY   102 DLVINHTSDQHPWFKAALDPKD--AHRSWYTWAGPGTN---LKAVSAVGGPAWHAN
Bacce_OGL    98 DLVVNHTSDEHNWFIESRKSKD-NKYRDYYIWRPGKEGKEPNNWGAAFSGSAWQYD
Thema_AMY   141 DLVINHTSDEHPWFKDAVENTTSSPYWDYYIMS--------LEDHSGQDHWHWKIN
                *** ******* **           * ****                      *


                   -Bβ3-      -III-                          ----IV---
                     **  *     **** *   **         **** ** **** *
Xanca_AMY          GK-Q--HYLGDFTGAMPDLNYDEPAVRREMIAVGKFWLDKGADGFRLDAARH 199
Bacce_OGL          EMTD-EYYLHLFSKKQPDLNWDNEKVRQDVYEMMKFWLEKGIDGFRMDVINF 203
Thema_AMY          SKGQKVWYFGLFGYNMPDLNHDSQKVREEVKKIVDFWISKGVDGFRIDAAKH 240
                     *  **     **** *  ***  *      **  ** **** *
```

Fig. 8. Sequence alignment of the domain B region of the oligo-1,6-glucosidase from *Bacillus cereus* (Bacce_OGL; Watanabe *et al.*, 1990) with the corresponding regions of the α-amylases from *Xanthomonas campestris* (Xanca_AMY; Abe *et al.*, 1996) and *Thermotoga maritima* (Thema_AMY; Liebl *et al.*, 1997). The conserved sequence regions as described in Fig. 2 are indicated by "II", "III", and "IV", respectively. The secondary structure elements forming the Bacce_OGLU domain B (Watanabe *et al.*, 1994) are also given. The sequence identities between Xanca_AMY and Bacce_OGLU as well as those between Thema_AMY and Bacce_OGLU are signified by asterisks above and below the alignment, respectively. The alignment was made using CLUSTAL W (Thompson *et al.*, 1994).

species of *Drosophila* (e.g., Boer and Hickey, 1986; Hickey *et al.*, 1987; Brown *et al.*, 1990; Magoulas *et al.*, 1993; Da Lage *et al.*, 1996). The studies were oriented mostly to providing the evidence for the non-independent (concerted) evolution of members of the insect amylase gene families (Hickey *et al.*, 1991; Popadic and Anderson, 1995; Popadic *et al.*, 1996; Shibata and Yamazaki, 1995; Inomata and Yamazaki, 1996). The interesting results have been observed by Da Lage *et al.* (1996) who have found that most of the drosophilid species have one intron of ancestral nature corresponding with one of the introns of mammalian α-amylases. This suggests that the vertebrates would have gained several additional introns during their evolution (Gumucio *et al.*, 1988) or, on the other hand, insects would have lost most of the primordial introns, leading in some species (e.g., in *Drosophila melanogaster*; Boer and Hickey, 1986; Shibata and Yamazaki, 1995) to a monoexonic structure (Da Lage *et al.*, 1996).

The group of plant α-amylases defined by Janeček (1994b) can be extended by the enzyme from potato (Gausing and Kreiberg, 1991), that seems to be on its own, the longest branch connecting the two previously described taxonomic subgroups: Leguminosae (*Vigna mungo*; Yamauchi and Minamikawa, 1990) and Gramineae [barley (Rogers and Milliman, 1983), maize (Young *et al.*, 1993), rice (Huang *et al.*, 1990) and wheat (Baulcombe *et al.*, 1987)].

Two remarkable α-amylase sequences belonging to the group of fungi and yeasts were reported recently, the acid-stable α-amylase from *Aspergillus kawachii* (Kaneko *et al.*, 1996) and the raw-starch-digesting and thermostable yeast α-amylase from *Cryptococcus* sp. S-2 (Iefuji *et al.*, 1996). They can be added to their own taxonomic fungal and yeast subgroups along with the other new α-amylase from *Lipomyces kononenkoae* (Steyn *et al.*, 1995). The remarkability of the two former enzymes is justified by the fact that they both contain the starch-binding domain (see Fig. 6) which is rather rare among α-amylases (Svensson *et al.*, 1989; Janeček *et al.*, 1995). Moreover, Iefuji *et al.* (1996) showed that a mutant of their *Cryptococcus* sp. S-2 α-amylase that lacks the C-terminal (i.e. the starch-binding) domain lost not only its ability to bind or digest raw starch, but also its thermostability, thus indicating the new, additional role for the structural motif known as the starch-binding domain described in Fig. 6.

Hofemeister *et al.* (1994) have reported the primary structure of the bacterial α-amylase from *Thermoactinomyces vulgaris* which revealed similarity to the other fungal α-amylases but interestingly 74% sequence identity and thus evolutionary relatedness to *Bacillus polymyxa* α-amylase (Hansen *et al.*, 1994) that is produced as a fusion protein of both β- and α-amylase (Uozumi *et al.*, 1989). The amylase from *Alicyclobacillus acidocaldarius* has an unusually large polypeptide chain of about 1300 residues (Koivula *et al.*, 1993; Schwermann *et al.*, 1994). It exhibits sequence similarities with amylopullulanase and neopullulanase but due to its main reaction products it is designated as α-amylase (Schwermann *et al.*, 1994). Another very long α-amylase from *Streptomyces lividans* (Tsao *et al.*, 1993) is interesting not only by its length but also by the fact that it exhibits just less than 30% identity with other related α-amylases from the genus *Streptomyces*. The other very recently released α-amylase from the same organism (Yin *et al.*, 1997) is of comparable length, however, the recognized $(\alpha/\beta)_8$-barrel elements are shifted to its C-terminus by about 200 amino acid residues in comparison with the former enzyme of Tsao *et al.* (1993), the sequence identity of the corresponding $(\alpha/\beta)_8$-barrel parts of the two *Streptomyces lividans* α-amylase sequences being found to be 27.5%.

It will also be of interest to know the sequences of further α-amylases, such as those from psychrotrophic *Vibrio* (Hamamoto and Horikoshi, 1991), *Bacillus caldovelox* (Bealin-Kelly *et al.*, 1990) or from vine shoot inter-nodes (Berbezy *et al.*, 1996). The four archaeal α-amylases from *Sulfolobus acidocaldarius* (Kobayashi *et al.*, 1996), *Sulfolobus solfataricus* (Kobayashi *et al.*, 1996), *Pyrococcus* sp. KOD1 (Tachibana *et al.*, 1996) and *Thermococcus profundus* (Lee *et al.*, 1996) rank clearly among the α-amylase family based on the presence of conserved sequence regions covering the β-strands of the catalytic $(\alpha/\beta)_8$-barrel in their sequences. This is an important observation mainly if the fact is taken into account that there is another, sequentially unrelated archeal α-amylase from

Pyrococcus furiosus (Laderman *et al.*, 1993) that served as a basis for the new α-amylase family, family 57 glycosyl hydrolases (Henrissat and Bairoch, 1996). However, the detailed analysis of the amino acid sequence of *Methanococcus jannaschii* α-amylase, an archaebacterial amylase reported to be sequentially similar to the family 57 *Pyrococcus furiosus* α-amylase (Bult *et al.*, 1996), may indicate the existence of an ancestor (although a very distant one) for both seemingly unrelated families (13 and 57) since the relics of the conserved sequence regions characteristic of the family of "true" α-amylases (family 13; see Fig. 2) can be detected in the sequence of *Methanococcus jannaschii* α-amylase (Š. Janeček, unpublished results).

4.2. *Amylases and Cyclodextrin Glycosyltransferases*

α-Amylases and CGTases are perhaps the most closely related members of the α-amylase family. Despite the fact that the CGTases are longer and contain in general two more domains at the C-terminus (designated as D and E; cf. Fig. 4), the parts of their structures that are common to both enzyme specificities, including the Ca^{2+} binding site, are indeed very similar (MacGregor and Svensson, 1989). The fact of their mutual sequence similarity has been known for at least 10 years (Kimura *et al.*, 1987). Moreover, the functions of these two enzymes are also very similar. They differ only in the second stage of the catalytic reaction, i.e. nucleophilic attack at the glycone C-1 by water: in hydrolysis, catalysed by α-amylase, the nucleophile is water, whereas for intra- or intermolecular glycosyl transfer brought about by the CGTase, the nucleophile is the C-4 hydroxyl group of a glucose ring (Svensson and Søgaard, 1993). These functional as well as sequence similarities had led in some cases to the incorrect classification of deduced nucleotide sequences of CGTases as those of α-amylase. These were the enzymes from *Bacillus circulans* (Nishizawa *et al.*, 1987), *Bacillus* sp. strain B1018 (Itkor *et al.*, 1990) and *Clostridium thermosulfurogenes* EM1 (Bahl *et al.*, 1991). The proposals for the re-classification of these three "α-amylases" have appeared in the theoretical papers (MacGregor, 1988; Jespersen *et al.*, 1993; Janeček, 1994b; Janeček and Tóth, 1994). The results from deep analysis of the α-amylase and CGTase sequences were summarized by Janeček *et al.* (1995) giving the sequence "fingerprints" that allow discrimination of CGTases from α-amylases (Fig. 9). Following the Itkor *et al.* (1990) statement that their enzyme forms β-cyclodextrin from soluble starch, Wind *et al.* (1995) provided clear biochemical evidence that the "α-amylase" from *Clostridium thermosulfurogenes* (i.e. *Thermoanaerobacterium thermosulfurigenes*) EM1 produces α-, β- and γ-cyclodextrins. This definitively confirmed the former theoretical proposals (MacGregor, 1988; Jespersen *et al.*, 1993; Janeček, 1994b; Janeček and Tóth, 1994; Janeček *et al.*, 1995). Finally, the evolutionarily oriented work by Janeček (1995a) concluded these uncertainties by the evolutionary tree of α-amylases and CGTases in which the three re-classified CGTases were unambiguously grouped in the CGTase cluster.

The further interesting findings of that work are the "intermediary" sequence features between α-amylases and CGTases also including the maltooligosaccharide-producing amylases (Janeček, 1995a). They are the features characteristic of CGTases present in the sequences of α-amylases. Their importance is that they allow the derivation of the evolutionary relationships of these two enzyme specificities in detail. These are: (i) the tryptophan in the β2-strand region (Trp75 in the sequence of *Bacillus circulans* strain 8 CGTase); (ii) the proline preceding the invariant AsnHis in the β3-strand region (Pro138); (iii) the isoleucine (leucine) in contrast to phenylalanine preceding the invariant arginine in the β4-strand region (Ile226); (iv) the glycine preceding the catalytic glutamate in the β5-strand region (Gly256); and (v) the isoleucine or leucine (6-C amino acid) succeeding the invariant phenylalanine in the β7-strand region (Ile324). Remarkably, these features were found in the fungal and yeast α-amylases and the two enzymes from *Dictyoglomus termophilum* AmyB and AmyC (Horinouchi *et al.*, 1988) as well as, to a lesser extent, in the liquefying α-amylases of the genus *Bacillus* (Yuuki *et al.*, 1985). As far as the maltooligosaccharide-producing amylases are concerned, the maltogenic amylase from *Bacillus stearothermophilus* (Diderichsen and Christiansen, 1988) and maltotriose-

	β2	β3	loop3	β4	β5	β7
α-Amylases						
Bacli	36_GITAVWIP-P	94_DINVYGDVVINH	198_YADID	226_DGFRLDAV	257_FTVAEYWQ	320_AVTFVDNHD
Bacsu	33_GITAIQTS-P	91_GIKVIVDAVINH	144_LYDWN	171_DGFRFDAA	204_FQYGEILQ	261_LVTWVESHD
Strli	32_GYGYVQVS-P	82_GVKVVADSVINH	145_LADLD	172_DGFRIDAA	200_YWKQEAIH	260_SAVFVDNHD
Aspor	56_GFTAIWIT-P	111_GMYLMVDVVANH	173_LPDLD	201_DGLRIDTV	226_YCIGEVLD	289_LGTFVENHD
Barle	34_GVTHVWLP-P	82_GVQAIADIVINH	147_APDID	175_DAWRLDFA	201_LAVAEVWD	283_AATFVDNHD
Drome	36_GYAGVQVS-P	88_GVRTYVDVVFNH	154_LRDLN	181_AGFRVDAA	219_YIVQEVID	280_SLVFVDNHD
Pigpa	36_GFGGVQVS-P	90_GVRIYVDAVINH	165_LLDLA	192_AGFRLDAS	229_FIFQEVID	292_ALVFVDNHD
Re-classified	! !	!!!	!	!	! !	! !
Bacci	71_GVTAIWISQP	130_NIKVVIDFAPNH	198_LADLN	225_DGIRVDAV	254_FTFGEWFL	321_QVTFIDNHD
Bacsp	70_GITALWISQP	129_NIKVIIDFAPNH	197_LADLN	224_DGIRMDAV	253_FTFGEWFL	320_QVTFLDNHD
Cloth	70_GVTAIWISQP	128_NIKVIIDFAPNH	196_LADLN	223_DGIRLDAV	252_FTFGEWFL	319_MVTFIDNHD
CGTases	! !	!!!	!	!	! !	! !
Bacci	70_GVTALWISQP	129_GIKIVIDFAPNH	197_LADFN	224_DGIRVDAV	253_FTFGEWFL	320_QVTFIDNHD
Bacst	67_GVTAIWISQP	125_GIKVIIDFAPNH	193_LADLN	220_DGIRMDAV	249_FTFGEWFL	316_QVTFIDNHD
Bacoh	66_GITAIWISQP	122_GIKVIMDFTPNH	190_LADYD	217_DGIRVDAV	246_FTFGEWFL	313_QVTFIDNHD
Klepn	66_GVTSIWITPP	124_NMKLVLDYAPNH	191_LSDLN	218_DAIRIDAI	253_FTFGEWFG	325_QVVFMDNHD

Fig. 9. Sequence fingerprints of α-amylases and CGTases. The sequence features highly characteristic of CGTases are signified by exclamation mark. The enzymes are abbreviated as: α-amylases: Bacli, *Bacillus licheniformis* (Yuuki *et al.*, 1985); Bacsu, *Bacillus subtilis* (Yang *et al.*, 1983); Strli, *Streptomyces limosus* (Long *et al.*, 1987); Aspor, *Aspergillus oryzae* (Toda *et al.*, 1982); Barle, barley (Rogers and Milliman, 1983); Drome, *Drosophila melanogaster* (Boer and Hickey, 1986); Pigpa, pig pancreas (Pasero *et al.*, 1986); re-classified as CGTase: Bacci, *Bacillus circulans* strain F-2 (Nishizawa *et al.*, 1987); Bacsp, *Bacillus* sp. strain B1018 (Itkor *et al.*, 1990); Cloth, *Clostridium thermosulfurogenes* (*Thermoanaerobacterium thermosulfurigenes*; Bahl *et al.*, 1991); CGTases: Bacci, *Bacillus circulans* strain 8 (Nitschke *et al.*, 1990); Bacst, *Bacillus stearothermophilus* (Fujiwara *et al.*, 1992); Bacoh, *Bacillus ohbensis* (Sin *et al.*, 1991); Klepn, *Klebsiella pneumoniae* (Binder *et al.*, 1986). Adapted from Janeček *et al.* (1995).

producing amylase from *Natronococcus* sp. (Kobayashi *et al.*, 1994) possess most features of CGTases and appeared next to them in the distance tree (Janeček, 1995a). On the other hand, the maltopentaose-forming amylase from *Pseudomonas* sp. (Shida *et al.*, 1992) was clustered with the α-amylases from the animal group, which should also be the case for the N-terminal part of the enzyme from an alkaliphilic Gram-positive bacterium (Candussio *et al.*, 1990). This amylase has been recently found to resemble the maltotetraose-forming alkaline amylase from *Bacillus* strain GM8901 (Shin and Byun, 1996). The maltohexaose-producing amylase from *Bacillus* sp. has already been reported to be sequentially similar to the liquefying type α-amylases (Tsukamoto *et al.*, 1988). Interestingly, maltotetraohydrolase from *Pseudomonas saccharophila* (Zhou *et al.*, 1989) contained no clear features of either α-amylases or CGTases and therefore it occupied the position that is in the tree approximately halfway between the animal α-amylases and the CGTases just near the plant and *Bacillus* liquefying α-amylases (Janeček, 1995a). The same position should be occupied by the closely related amylase from *Pseudomonas stutzeri* (Fujita *et al.*, 1989), as supported also by the tree described by Jespersen *et al.* (1993). All these observations may indicate that the maltooligosaccharide-producing amylases could play, perhaps, an important role during the evolution of "true" α-amylases and "true" CGTases (Janeček, 1995a).

In general, the amino acid sequences of CGTases exhibit quite high identity (mostly 50% and higher) which very probably reflects the fact that they are all but one of *Bacillus* origin. The only one CGTase from the other source is that from *Klebsiella pneumoniae* (Binder *et al.*, 1986) which, indeed, revealed the lower sequence identity (about 30%; see also Fig. 9) with the *Bacillus* CGTases (Nitschke *et al.*, 1990). An interesting observation has been reported by Nógrády *et al.* (1995) that, in *Bacillus macerans*, CGTase is the only starch-degrading enzyme and hence it may be fully responsible for degradation of starch in this bacterium.

4.3. *Proteins Sharing Sequence Similarities and the Role of Domain B*

Several of the $(\alpha/\beta)_8$-barrel enzymes have evolutionary protein counterparts that have no known catalytic function (Janeček and Bateman, 1996). These are, for instance, the

chitinases (Perrakis *et al.*, 1994; Coulson, 1994; Tews *et al.*, 1996) with narbonin (Hennig *et al.*, 1992), concanavalin B (Hennig *et al.*, 1995) and nodulin (Perlick *et al.*, 1996) or the enzymes from the aldo-keto reductase family with a protein regulated by fibroblast growth factor-1 (Wilson *et al.*, 1995). The α-amylase family is not in these terms an exception (Reardon and Farber, 1995). In 1992, two sequencing papers of special importance appeared that have revealed the amino acid sequences of rat (Wells and Hediger, 1992) and rabbit (Bertran *et al.*, 1992) kidney proteins stimulating dibasic and neutral amino acid transport. These animal proteins have had sequence similarity to the enzymes from the α-amylase family. A year later, the sequence of this protein from human kidney was also reported (Bertran *et al.*, 1993).

These three 85% identical amino acid transport-related proteins were found to clearly possess characteristic features of the α-amylase family, including β-strands β2, β3, β4 and β8 of the $(\alpha/\beta)_8$-barrel, as well as the conserved stretch near the C-terminus of domain B (see Fig. 2). Deep analysis of domain B sequences has revealed that the amino acid transport-related proteins may have recruited from the enzymes of the oligo-1,6-glucosidase group (Janeček *et al.*, 1997). It is of interest that there are some other animal proteins, the so-called 4F2 heavy-chain cell surface antigens (Quackenbush *et al.*, 1987; Gottesdiener *et al.*, 1988) which are highly similar to the transport-related proteins (Wells and Hediger, 1992). The 4F2 antigen also exhibits sequence similarities with the α-amylase family enzymes, especially with those from the oligo-1,6-glucosidase group, but it lacks domain B, or at least its major part (Fig. 10) that might have been lost during the evolution from the transport-related proteins (Janeček *et al.*, 1997). The 4F2 antigen contains a part of domain B only that has a close sequence resemblance to the middle β-strand of the three-stranded antiparallel β-sheet of domain B from the oligo-1,6-glucosidase from *Bacillus cereus* (Kizaki *et al.*, 1993; Watanabe *et al.*, 1994). In contrast, some catalytically active members of the α-amylase family (e.g., neopullulanase and cyclomaltodextrinase) that contain a proper domain B related to that of oligo-1,6-glucosidase, miss just this middle β-strand segment (Janeček *et al.*, 1997).

Of the three catalytic residues in the α-amylase family (i.e., Asp206, Glu230 and Asp297) only aspartate corresponding to the Asp206 of TAA in the β4-strand region can be traced unambiguously in the amino acid transport-related proteins and tentatively in the 4F2 heavy-chain antigens (Fig. 10). The sequence in the β3-strand region of the transport-related proteins resembling that of CGTases (see Fig. 2) supports the view of "intermediary" sequence features present among the α-amylase family members (Janeček, 1995a,b).

Caption for Fig. 10 on following page.

Fig. 10. Sequence alignment of the enzyme and protein members from the α-amylase family covering their $(\alpha/\beta)_8$-barrels involving domain B. Sources: Bacce_OGL, oligo-1,6-glucosidase from *Bacillus cereus* (Watanabe *et al.*, 1990); Bacsp_CMD, cyclomaltodextrinase from *Bacillus sphaericus* (Oguma *et al.*, 1993); Human_AAT, human amino acid transport-related protein (Bertran *et al.*, 1993); Human_4F2, human 4F2 heavy-chain cell surface antigen (Gottesdiener *et al.*, 1988). The secondary structure elements of $(\alpha/\beta)_8$-barrel (starting with "A") and domain B (starting with "B") of Bacce_OGL (Watanabe *et al.*, 1994) are given above the alignment. The identities among all four sequences and among three whichever ones are signified by asterisks and black points, respectively. The lack of the middle β-strand in domain B ("Bβ2") in Bacsp_CMD is strongly supported by the Hydrophobic Cluster Analysis (Janeček *et al.*, 1997). This lack seems to be remarkably joined with the Asp→Lys substitution in the short conserved sequence near the C-terminus of domain B (denoted by "loop3"), as supported by the sequences of cyclomaltodextrinase from *Thermoanaerobacter ethanolicus* (Podkovyrov and Zeikus, 1992), neopullulanases from *Bacillus* sp. strain KSM-1876 (Igarashi *et al.*, 1992) and *Thermoactinomyces vulgaris* (Tonozuka *et al.*, 1993) and the α-amylase AmyB from *Dictyoglomus thermophilum* (Horinouchi *et al.*, 1988). The segments around the strands β5 and β7 of the $(\alpha/\beta)_8$-barrel domain with the catalytic Glu255 and Asp329, respectively, of Bacce_OGL (Kizaki *et al.*, 1993) belong to the weak parts of the alignment and the residues of the protein members corresponding to the catalytic Glu and Asp can be hardly identified. However, the entire alignment is justified by the unambiguous correspondence comprising the following strand β8 of the barrel. The alignment was produced using CLUSTAL W (Thompson *et al.*, 1994).

```
                --Aβ1--                                      ------Aα1-----
Bacce_OGL   WWKESVVYQIYPRSFM--DSNGD--------GI--------GDLRGIISKLDYLKELGID  46
Bacsp_CMD   WVKEAIFYQIFPERFANGDPSNDPEGVQEWGGTPSAGNFFGGDLQGVIDHLDYLSDLGVN 189
Human_AAT   WWQEGPMYQIYPRSFK--DSNKD--------GN--------GDLKGIQDKLDYITALNIK 158
Human_4F2   WWHTGALYRIG--DLQ--AFQGH--------GA--------GNLAGLKGRLDYLSSLKVK 156
            ** *   *** * *   *   *         *          *** *    ****  *

            Aβ2-                                -----Aα2-----  ---Aβ3--
Bacce_OGL   VIWLSPVYESPNDDNGYDISDYCKIMNEFGTMEDWDELLHEMHERNMKLMMDLVVNHTSD 106
Bacsp_CMD   ALYFNPLFAATTNHK-YDTADYMKIDPQFGTNEKLKELVDACHARGMRVLLDAVFNHCG- 247
Human_AAT   TVWITSFYKSSLKDFRYGVEDFREVDPIFGTMEDFENLVAAIHDKGLKLIIDFIPNHTSD 218
Human_4F2   GLVLGPIHKNQKDD--VAQTDLLQIDPNFGSKEDFDSLLQSAKKKSIRVILDLTPN---- 210
                 *       *  *   *   *** *** **   *    *         *   **

              --Bα1-            -Bβ1-                          -Bβ2    -Bβ3-
Bacce_OGL   EHNWFIESRKSKDNKYRDYYIWRPGKE--GK--EPNNWGAAFSGSAWQYDEMTDEYYLHL 162
Bacsp_CMD   -HTFPPFVDVLNNGLNSRYADWFHVRE------WPLRVVDGIPT-----------YDTFA 289
Human_AAT   KHIWFQLSR-TRTGKYTDYYIWHDCTHENGKTIPPNNWLSVYGNSSWHFDEVRNQCYFHQ 277
Human_4F2   ---------------YR-----------------------GENSWFFTQV-------- 222
             *              *  *  *           *          *

                loop3   -------Aα3------      -Aβ4
Bacce_OGL   FSKKQPDLNWDNEKVRQDVYEMMKFWLEK-GIDGFRMDVINFISKEE----GLPTVETEE 217
Bacsp_CMD   FEPIMPKLNTGNEEVKAYLLNVGRYWLEEMGLDGWRLDVANEVDHQF-----WREFRSEI 344
Human_AAT   FMKEQPDLNFRNPDVQEEIKEILRFWLTK-GVDGFSLDAVKFLLEAKHLRDEIQVNKTQI 336
Human_4F2   ------------DTVATKVKDALEFWLQA-GVDGFQVRDIENLKDAS----SFLAEWQNI 265
            *    * **  *  *          ***   * ***  *                      *

                              ----Aα4----       --Aβ5-      --Aα5--
Bacce_OGL   EGYVSG-HKHFMNGPNIHKYLHEMNEEVLSHYDIMTVGEMPGVTTEEAKLYTGEERKELQ 276
Bacsp_CMD   KRINPS--------------------------AYILGEIMHDSMPWLQGDQFDAVMNYP 377
Human_AAT   PDTVTQYSELYHDFTTTQVGMHDIVRSFRQTMDQYSTEPGRYRFMGTEAYAESIDRTVMY 396
Human_4F2   TKGFSE-----------------------DRLLIAGTNSSDLQQILSLLESNKDLLLTS 301

            -Aβ6 -Aα6'                 ------Aα6-----      -Aβ7-         Aα7'
Bacce_OGL   MVFQFEHMDLDSGEGGKWDVKPCSLLTLKENLTKWQKALEHTGWNSLYWNNHDQPRVVSR 336
Bacsp_CMD   FTNILLNFFARRLTNAAE-----FAQAIGTQLAGYPQQVTEVSFN--LLGSHDTTRLLTL 430
Human_AAT   YGLPFIQEADFPFNNYLSMLDTVSGNSVYEVITSWMENMPEGKWPNWMIGGPDSSRLTSR 456
Human_4F2   --------------SYLS-DSGSTGEHTKSLVTQYLNATG-NRWCSWSLS---QARLLTS 342
                                          *          *         *  **

            -       -----Aα7-----     -Aβ8-                        ----Aα8'---
Bacce_OGL   FGNDGMYRIESAKMLATVLHMMKGTPYIYQGEEIGMTNVRFESIDEYRDIETLNMYKEKV 396
Bacsp_CMD   CSGN----VERMKLATLFQLTYQGTPCIYYGDEIGMDGEYDP----------LNRK-CME 475
Human_AAT   LGN------QYVNVMNMLLFTLPGTPITYYGEEIGMGNIVAANLNESYDINTLRSKSPMQ 510
Human_4F2   FLP-----AQLLRLYQLMLFTLPGTPVFSYGDEIGLDAAALP-------GQPMEAP-VML 399
                          * *  ***  *** ****                 *     *

            -     ---Aα8''--                                        Aα8'
Bacce_OGL   MERGEDIEKVMQSIYIKGRDNARTPMQWDDQNHAG---FTTGEPWITVNPNYKEINVKQA 453
Bacsp_CMD   WDKSK-------------QNTELLAF------------FRSMISLRKAHPALR-GSGLRF 509
Human_AAT   WDNSSNAGFSEASNTWLPTNSDYHTVNVDVQKTQPRSALKLYQDLSLLHANELLLNRGWF 570
Human_4F2   WDESS--------FPDIPGAVSANMTVKGQSEDPGS--LLSLFRRLSDQRSKERSLLHGDF 450
            ** *                                        *              *

            ''      ------Aα8-----
Bacce_OGL   IQNK-DSIFYYYKKLIELRKNNEIVVYGSYDLILEN  488-558
Bacsp_CMD   LPVLEHPQLLVYERWDD-NER-FLIMLNNEDAPVNV  543-591
Human_AAT   CHLRNDSHYVVYTRELDGIDRIFIVVLNFGESTLLN  606-685
Human_4F2   HAFSAGPELFSYIRHWDQNER-FLVVLNFGDVGLSA  485-539
                     * *  *   * * ****  *  *
```

4.4. *The Entire Present-day α-Amylase Family*

The evolutionary relationships of the entire α-amylase family were described a few years ago in the works by Jespersen *et al.* (1993) and Janeček (1994a). Here six new enzyme specificities are considered in addition to those works: (1) trehalose-6-phosphate hydrolase (Rimmele and Boos, 1994; Helfert *et al.*, 1995); (2) maltotriohydrolase (Kobayashi *et al.*, 1994); (3) maltooligosyltrehalose hydrolase (Maruta *et al.*, 1996a,b); (4) amylomaltase (Takaha *et al.*, 1993); (5) maltooligosyltrehalose synthase (Maruta *et al.*, 1996a,b); and (6) glucosyltransferase (Ferretti *et al.*, 1987; Gilmore *et al.*, 1990; Monchois *et al.*, 1996). The two proteins without catalytic function, the amino acid transport-related proteins (Wells and Hediger, 1992; Bertran *et al.*, 1992, 1993) and the 4F2 heavy-chain cell surface antigens (Quackenbush *et al.*, 1987; Gottesdiener *et al.*, 1988), are also included for the first time.

The relationships can be deduced from the evolutionary tree shown in Fig. 11. The tree is based on the alignment of conserved sequence regions presented in Fig. 2. For the sake of simplicity only one representative for each enzyme specificity (or protein group for the two proteins) was considered. This may result in some oversimplifying since, for instance, the members belonging to the same EC do not exhibit in their conserved sequence regions 100% identity, and therefore those members with the so-called "intermediary" sequence features are not demonstrated in the tree. It has been suggested, however, that all the members of the α-amylase family are mutually closely

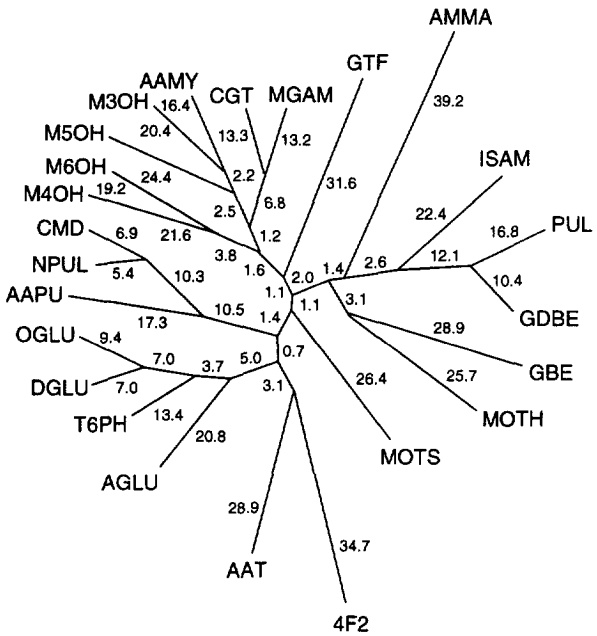

Fig. 11. Evolutionary tree of the representatives of the entire α-amylase family. The α-amylase family members are those from Fig. 2 using the following abbreviations: AAMY, α-amylase; OGLU, oligo-1,6-glucosidase; AGLU, α-glucosidase; PUL, pullulanase; AAPU, amylopullulanase; CMD, cyclomaltodextrinase; M4OH, maltotetraohydrolase; ISAM, isoamylase; DGLU, dextran glucosidase; T6PH, trehalose-6-phosphate hydrolase; M6OH, maltohexaohydrolase; M3OH, maltotriohydrolase; MGAM, maltogenic amylase; NPUL, neopullulanase; M5OH, maltopentaohydrolase; MOTH, maltooligosyltrehalose hydrolase; GBE, glucan branching enzyme; CGT, cyclodextrin glycosyltransferase; AMMA, amylomaltase; GDBE, glucan debranching enzyme; MOTS, maltooligosyltrehalose synthase; GTF, glucosyltransferase; AAT, amino acid transport-related protein; 4F2, 4F2 heavy-chain cell surface antigen. The tree is based on the alignment of conserved sequence regions shown in Fig. 2. It was constructed by the neighbor-joining method (Saitou and Nei, 1987) implemented in the CLUSTAL W package (Thompson *et al.*, 1994). The branch lengths are proportional to the divergency of the amino acid sequences of the individual members of the α-amylase family, the sum of the lengths of the branches linking any members being a measure of the evolutionary distance between them.

related (some of them are only more closely related and the others are only less closely related), and thus the strict evolutionary borders separating the individual enzyme specificities can hardly be defined (Janeček, 1995b). In general, the tree presented in Fig. 11 reveals and, in fact, confirms that there are two basic groups in the α-amylase family containing very closely related enzymes: (1) enzymes with the α-1,4-glucosidic bond specificity comprising the α-amylases, CGTases and maltooligosaccharide-producing amylases; and (2) enzymes with the α-1,6-glucosidic bond specificity including the oligo-1,6-glucosidases, dextran glucosidases, trehalose-6-phosphate hydrolases and α-glucosidases. The enzymes having dual bond specificity (cyclomaltodextrinases, neopullulanases and amylopullulanases) are clustered between the two above mentioned basic groups, indicating, however, a closer relatedness to the α-1,6-specificity enzymes. The position of the two animal members without catalytic function (amino acid transport-related proteins and the 4F2 heavy-chain cell surface antigens) next to the cluster of this α-1,6-specificity group is in accordance with the previously revealed mutual sequence similarities (Wells and Hediger, 1992) as well as confirming the proposal that the amino acid transport-related proteins may have evolved by recruitment from this group of enzymes from the α-amylase family (Janeček *et al.*, 1997). The very long branch of the 4F2 antigens agrees with the less evident similarities with the α-amylase family enzymes (see Fig. 2) as well as possibly reflecting the absence (i.e., the loss during evolution) of substantial parts of domain B in these antigens (Janeček *et al.*, 1997; see Fig. 10).

With regard to the glucosyltransferases that have been recently indicated to contain a circularly permuted version of the α-amylase-type $(\alpha/\beta)_8$-barrel (MacGregor *et al.*, 1996), their branch (though a quite long one) is, in fact, in the cluster of amylases and CGTases (the eventuality of circular permutation is not reflected in the tree). These interesting enzymes (for a review see Robyt, 1996) very probably contain the three β-strand segments (β4, β5, β7) comprising the catalytic triad of amino acid residues (Mooser *et al.*, 1991; MacGregor *et al.*, 1996; see Fig. 2). The suggestion of the occurrence of circular permutations in these enzymes, however seemingly attractive and justified by the analysis of their amino acid sequences (MacGregor *et al.*, 1996), has to be confirmed by the crystal structure of a representative of the glucosyltransferases.

The grouping of the isoamylase and pullulanase together with the glucan branching and debranching enzymes was demonstrated previously (Jespersen *et al.*, 1993; Janeček, 1994a) and can also be seen in the present tree, the amylomaltase and maltooligosyltrehalose hydrolase being added (Fig. 11). The discussion concerning the similarities of branching and debranching enzymes from various sources together with pullulanases and isoamylases can be found in the recent review by Nakamura (1996). As far as the position of the maltooligosyltrehalose synthase is concerned, its sequences from *Arthrobacter* sp. Q36 (Maruta *et al.*, 1996a) as well as *Rhizobium* sp. M-11 (Maruta *et al.*, 1996b) were determined only recently. A more detailed analysis of their amino acid sequences as well as of the other enzyme specificities from this cluster will be needed to elucidate their relationships correctly.

The amylomaltase seems to be the most distantly related member in the frame of the entire α-amylase family (it lies on the longest branch in the tree). This may reflect the fact that it does not contain an important histidine in the β3-strand region which is, in fact, an invariant residue throughout the enzymatic members of the family. Interestingly, its region comprising the short conserved sequence near the C-terminus of domain B (290_LYDWK; designated as "loop3" in Fig. 2) is highly similar to that of *Bacillus subtilis* α-amylase (144_LYDWN; Fig. 7). This indicates that often very subtle changes in the conserved sequence regions and/or perhaps sometimes outside them (depending on the concrete case) may have resulted in appearing with a new enzyme specificity, keeping the basic features of catalytic mechanism and three-dimensional structure. The two protein members (transport-related proteins and the 4F2 antigens) evince the possibility that the above-mentioned changes might also have been responsible for the lost of catalytic activity.

V. CONCLUSION

The aim of the present article has been to describe and discuss the molecular biology and evolution of the α-amylase family. This broad protein family covers about 20 different enzymes from hydrolases and transferases as well as two proteins with no catalytic function. It very probably represents a compact cluster of divergently related members in the frame of the entire $(\alpha/\beta)_8$-barrel protein family. On the other hand, the functionally closely related β-amylase that also adopts an $(\alpha/\beta)_8$-barrel fold (Mikami *et al.*, 1993; Cheong *et al.*, 1995) seems to be "solitary" from the evolutionary point of view (Reardon and Farber, 1995; Pujadas *et al.*, 1996).

The eventuality of occurrence of circular permutations during the evolution of some members of the α-amylase family, in glucosyltransferases (MacGregor *et al.*, 1996), has opened further attractive trends for evolutionary studies despite the fact that these permutations have not been experimentally evidenced yet. Similar permutations have been already confirmed in the other group of $(\alpha/\beta)_8$-barrel enzymes, in the class I aldolase family (Jia *et al.*, 1996).

Another feature of more general interest could be provided by the enzymes having both α-amylase as well as pullulanase specificities. These enzymes can be expressed as a single amylopullulanase with one active site responsible for both specificities (Mathupala *et al.*, 1993) or as a "biheaded" enzyme with two active sites localized at two $(\alpha/\beta)_8$-barrel domains in one polypeptide chain (Hatada *et al.*, 1996).

As far as the other $(\alpha/\beta)_8$-barrels seemingly unrelated to the α-amylase family enzymes are concerned, three of them have their $(\alpha/\beta)_8$-barrel domains discontinued by a long loop 3, called in the α-amylase family domain B. These are pyruvate kinase (Muirhead *et al.*, 1986; Larsen *et al.*, 1994; Mattevi *et al.*, 1995), old yellow enzyme (Fox and Karplus, 1994) and tRNA-guanine transglycosylase (Romier *et al.*, 1996). While the former two enzymes have been indicated to have some relationship to the enzymes from the α-amylase family (Farber and Petsko, 1990; Janeček, 1996c), the relatedness between the tRNA-guanine transglycosylase and the α-amylase family (if any) is still waiting to be discovered.

In conclusion, it seems acceptable that despite the fact that each new observation makes our knowledge wider and deeper, it simultaneously opens further questions for our mind. It is hoped that this article may help to narrow this discrepancy.

ACKNOWLEDGEMENTS

My research was supported by grant No. 2/3013/96 from the Slovak Grant Agency for Science. The part concerning the study of domain B was supported by the FEBS Short-Term Fellowship. And my sincere thanks go to Drs (in alphabetical order) Alex Bateman (Cambridge, England), Greg Farber (University Park, Pennsylvania), Bernard Henrissat (Grenoble, France), Yuzuru Suzuki (Shimogamo, Japan), Birte Svensson (Valby, Denmark), and Jozef Ševčík (Bratislava, Slovakia).

REFERENCES

Abe, J.-I., Shibata, Y., Fujisue, M. and Hizukuri, S. (1996) *Microbiology* **142,** 1505–1512.

Aghajari, N., Feller, G., Gerday, C. and Haser, R. (1996) *Prot. Sci.* **5,** 2128–2129.

Alkazaz, M., Desseaux, V., Marchis-Mouren, G., Payan, F., Forest, E. and Santimone, M. (1996) *Eur. J. Biochem.* **241,** 787–796.

Amemura, A., Chakraborty, R., Fujita, M., Noumi, T. and Futai, M. (1988) *J. biol. Chem.* **263,** 9271–9275.

Baecker, P. A., Greenberg, E. and Preiss, J. (1986) *J. biol. Chem.* **261,** 8738–8743.

Bahl, H., Burchhardt, G., Speinat, A., Haeckel, K., Wienecke, A., Schmidt, B. and Antranikian, G. (1991) *Appl. environ. Microbiol.* **57,** 1554–1559.

Bairoch, A. (1996) *Nucleic Acids Res.* **24,** 221–222.

Bairoch, A. and Apweller, R. (1997) *Nucleic Acids Res.* **25,** 31–36.

Ball, S., Guan, H.-P., James, M., Myers, A., Keeling, P., Mouille, G., Buléon, A., Colonna, P. and Preiss, J. (1996) *Cell* **86,** 349–352.

Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D. and Waley, S. G. (1975) *Nature* **255,** 609–914.

Bauer, M. W., Halio, S. B., Kelly, R. M. (1996) *Adv. Prot. Chem.* **48,** 271–310.

Baulcombe, D. C., Huttly, A. K., Martienssen, R. A., Barker, R. F. and Jarvis, M. G. (1987) *Molec. gen. Genet.* **209,** 33–40.

Bealin-Kelly, F. J., Kelly, C. T. and Fogarty, W. M. (1990) *Biochem. Soc. Trans.* **18,** 310–311.
Bealin-Kelly, F. J. D., Kelly, C. T. and Fogarty, W. M. (1992) *J. indust. Microbiol.* **9,** 63–68.
Benkel, B. F., Nguyen, T., Ahluwalia, N., Benkel, I. and Hickey, D. A. (1996) *GenBank*: U63411 (unpublished).
Benson, D. A., Boguski, M. S., Lipman, D. J. and Ostell, J. (1997) *Nucleic Acids Res.* **25,** 1–6.
Berbezy, P., Legendre, L. and Maujean, A. (1996) *Plant Physiol. Biochem.* **34,** 353–361.
Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer Jr., E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) *J. molec. Biol.* **122,** 535–542.
Bertran, J., Werner, A., Moore, M. L., Stange, G., Markovich, D., Biber, J., Testar, X., Zorzano, A., Palacin, M. and Murer, H. (1992) *Proc. natn. Acad. Sci. U.S.A.* **89,** 5601–5605.
Bertran, J., Werner, A., Chillaron, J., Nunes, V., Biber, J., Testar, X., Zorzano, A., Estivill, X., Murer, H. and Palacin, M. (1993) *J. biol. Chem.* **268,** 14842–14849.
Binder, F., Huber, O. and Böck, A. (1986) *Gene* **47,** 269–277.
Blundell, T., Lindley, P., Miller, L., Moss, D., Slinsly, C., Tickle, I., Turwell, B. and Wistow, G. (1981) *Nature* **289,** 771–777.
Boel, E., Hjort, I., Svensson, B., Norris, F., Norris, K. E. and Fiil, N. P. (1984) *Eur. molec. Biol. Org. J.* **3,** 1097–1102.
Boel, E., Brady, L., Brzozowski, A. M., Derewenda, Z., Dodson, G. G., Jensen, V. J., Petersen, S. B., Swift, H., Thim, L. and Woldike, H. F. (1990) *Biochemistry* **29,** 6244–6249.
Boer, P. H. and Hickey, D. A. (1986) *Nucleic Acids Res.* **14,** 8399–8411.
Bompard-Gilles, C., Rousseau, P., Rougé, P. and Payan, F. (1996) *Structure* **4,** 1441–1452.
Bork, P., Holm, L. and Sander, C. (1994) *J. molec. Biol.* **242,** 309–320.
Brady, R. L., Brzozowski, A. M., Derewenda, Z. S., Dodson, E. J. and Dodson, G. G. (1991) *Acta Cryst.* **B47,** 527–535.
Brändén, C.-I. (1991) *Curr. Opin. struct. Biol.* **1,** 978–983.
Braun, C., Lindhorst, T., Madsen, N. B. and Withers, S. G. (1996) *Biochemistry* **35,** 5458–5463.
Brayer, G. D., Luo, Y. and Withers, S. G. (1995) *Prot. Sci.* **4,** 1730–1742.
Brosnan, M. P., Kelly, C. M. and Fogarty, W. M. (1992) *Eur. J. Biochem.* **203,** 225–231.
Brown, C. J., Aquadro, C. F. and Anderson, W. W. (1990) *Genetics* **126,** 131–138.
Brumm, P. J., Hebeda, R. E. and Teague, W. M. (1991) *Starch* **43,** 315–319.
Buisson, G., Duée, E., Haser, R. and Payan, F. (1987) *Eur. molec. Biol. Org. J.* **6,** 3909–3916.
Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J.-F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., Weidman, J. F., Fuhrmann, J. L., Presley, E. A., Nguyen, D., Utterback, T. R., Kelley, J. M., Peterson, J. D., Sadow, P. W., Hanna, M. C., Cotton, M. D., Hurst, M. A., Roberts, K. M., Kaine, B. P., Borodovsky, M., Klenk, H.-P., Fraser, C. M., Smith, H. O., Woese, C. R. and Venter, J. C. (1996) *Science* **273,** 1058–1073.
Candussio, A., Schmid, G. and Böck, A. (1990) *Eur. J. Biochem.* **191,** 177–185.
Cao, H. and Preiss, J. (1996) *J. Prot. Chem.* **15,** 291–304.
Casset, F., Imberty, A., Haser, R., Payan, F. and Perez, S. (1995) *Eur. J. Biochem.* **232,** 284–293.
CCP4 (1994) Collaborative Computational Project Number 4. *Acta Cryst.* **D50,** 760–763.
Chang, C., Kim, K. K., Hwang, K. Y., Choi, M.-U. and Suh, S. W. (1993) *J. molec. Biol.* **229,** 235–238.
Cheong, C. G., Eom, S. H., Chang, C., Shin, D. H., Song, H. K., Min, K., Moon, J. H., Kim, K. K., Hwang, K. Y. and Suh, S. W. (1995) *Proteins: Struct. Funct. Genet.* **21,** 105–117.
Chothia, C. (1988) *Nature* **333,** 598–599.
Colonna, P., Buléon, A. and Lemarié, F. (1988) *Biotechnol. Bioeng.* **31,** 895–904.
Conrad, B., Hoang, V., Polley, A. and Hofemeister, J. (1995) *Eur. J. Biochem.* **230,** 481–490.
Coulson, A. F. W. (1994) *Fedn Eur. biochem. Socs Lett.* **354,** 41–44.
Da Lage, J.-L., Wegnez, M. and Cariou, M.-L. (1996) *J. molec. Evol.* **43,** 334–347.
Dalmia, B. K., Schütte, K. and Nikolov, Z. L. (1995) *Biotechnol. Bioeng.* **47,** 575–584.
Declerck, N., Joyet, P., Gaillardin, C. and Masson, J.-M. (1990) *J. biol. Chem.* **265,** 15481–15488.
Declerck, N., Joyet, P., Trosset, J.-Y., Garnier, J. and Gaillardin, C. (1995) *Prot. Eng.* **8,** 1029–1037.
Demchuk, E., Vihinen, M., Wade, R. and Korpela, T. (1993) In *Proceedings of an International Symposium on Stability and Stabilizations of Enzymes*, pp. 291–298 (eds. W. J. J. Van der Tweel, A. Harder and R. M. Buitelaar), Elsevier, Amsterdam.
De Souza, S. J., Long, M., Schoenbach, L., Roy, S. W. and Gilbert, W. (1996) *Proc. natn. Acad. Sci. U.S.A.* **93,** 14632–14636.
Desseaux, V., Payan, F., Ajandouz, E. H., Svensson, B., Haser, R. and Marchis-Mouren, G. (1991) *Biochim. biophys. Acta* **1080,** 237–244.
Diderichsen, B. and Christiansen, L. (1988) *Fedn Eur. microbiol. Socs Microbiol. Lett.* **56,** 53–60.
Farber, G. K. (1993) *Curr. Opin. struct. Biol.* **3,** 409–412.
Farber, G. K. and Petsko, G. A. (1990) *Trends Biochem. Sci.* **15,** 228–234.
Feller, G., Lonhienne, T., Deroanne, C., Libioulle, C., Van Beeumen, J., Gerday, C. (1992) *J. biol. Chem.* **267,** 5217–5221.
Feller, G., Payan, F., Theys, F., Qian, M., Haser, R. and Gerday, C. (1994) *Eur. J. Biochem.* **222,** 441–447.
Feller, G., le Bussy, O., Houssier, C. and Gerday, C. (1996) *J. biol. Chem.* **271,** 23836–23841.
Ferretti, J. J., Gilpin, M. L. and Russell, R. R. B. (1987) *J. Bacteriol.* **169,** 4271–4278.
Fogarty, W. M. (1983) In *Microbial Enzymes and Biotechnology*, pp. 1–92, Elsevier Applied Science Publishers, Barking, U.K.
Fox, K. M. and Karplus, P. A. (1994) *Structure* **2,** 1089–1105.
Friedberg, F. (1983) *Fedn. Eur. biochem. Socs Lett.* **152,** 139–140.
Fujimori, H., Ohnishi, M. and Hiromi, K. (1978) *J. Biochem.* **83,** 1503–1510.

Fujita, M., Torigoe, K., Nakada, T., Tsusaki, K., Kubota, M., Sakai, S. and Tsujisaka, Y. (1989) *J. Bacteriol.* **171,** 1333–1339.
Fujiwara, S., Kanemoto, M., Kim, B., Lejeune, A., Sakaguchi, K. and Imanaka, T. (1992) *Appl. environ. Microbiol.* **58,** 4016–4025.
Fukusumi, S., Kamizono, A., Horinouchi, S. and Beppu, T. (1998) *Eur. J. Biochem.* **174,** 15–21.
Gaboriaud, C., Bissery, V., Benchetrit, T. and Mornon, J. P. (1987) *Fedn Eur. biochem. Socs Lett.* **224,** 149–155.
Gausing, K. and Kreiberg, T. D. (1991) *GenBank*: M79328 (unpublished).
George, D. G., Dodson, R. J., Garavelli, J. S., Haft, D. H., Hunt, L. T., Marzec, C. R., Orcutt, B. C., Sidman, K. E., Srinivasarao, G. Y., Yeh, L. S. L., Arminski, L. M., Ledley, R. S., Tsugita, A. and Barker, W. C. (1997) *Nucleic Acids Res.* **25,** 24–27.
Gibson, R. M. and Svensson, B. (1986) *Carlsberg Res. Commun.* **51,** 295–308.
Gilbert, W. (1978) *Nature* **271,** 501.
Gilbert, W. and Glynias, M. (1993) *Gene* **135,** 137–144.
Gilles, C., Astier, J.-P., Marchis-Mourem, G., Cambillau, C. and Payan, F. (1996) *Eur. J. Biochem.* **238,** 561–569.
Gilmore, K. S., Russell, R. R. B. and Ferretti, J. J. (1990) *Infect. Immun.* **58,** 2452–2458.
Gobius, K. S. and Pemberton, J. M. (1988) *J. Bacteriol.* **170,** 1325–1332.
Gottesdiener, K. M., Karpinski, B. A., Lindsten, T., Strominger, J. L., Jones, N. H., Thompson, C. B. and Leiden, J. M. (1988) *Molec. cell. Biol.* **8,** 3809–3819.
Gumucio, D. L., Wiebauer, K., Caldwell, R. M., Samuelson, L. and Meisler, M. H. (1988) *Molec. cell. Biol.* **8,** 1197–1205.
Hamamoto, T. and Horikoshi, K. (1991) *Fedn Eur. microbiol. Socs Microbiol. Lett.* **84,** 79–84.
Hansen, G., Heese, O., Höhne, W. E. and Hofemeister, B. (1994) *Int. J. Peptide Prot. Res.* **44,** 245–252.
Harata, K., Haga, K., Nakamura, A., Aoyagi, M., Yamane, K. (1996) *Acta Cryst.* **D52,** 1136–1145.
Hatada, Y., Igarashi, K., Ozaki, K., Ara, K., Hitomi, J., Kobayashi, T., Kawai, S., Watabe, T. and Ito, S. (1996) *J. biol. Chem.* **271,** 24075–24083.
Helfert, C., Gotsche, S. and Dahl, M. K. (1995) *Molec. Microbiol.* **16,** 111–120.
Hennig, M., Schlesier, B., Dauter, Z., Pfeffer, S., Betzel, C., Höhne, W. and Wilson, K. S. (1992) *Fedn Eur. biochem. Socs Lett.* **306,** 80–84.
Hennig, M., Jansonius, J. N., Terwisscha van Scheltinga, A., Dijkstra, B. W. and Schlesier, B. (1995) *J. molec. Biol.* **254,** 237–246.
Henrissat, B. and Bairoch, A. (1996) *Biochem. J.* **316,** 695–696 (A permanently updated version of the glycosyl hydrolases classification is available at the URL: http://expasy. hcuge.ch/cgi-bin/lists?glycosid.txt).
Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P. and Davies, G. (1995) *Proc. natn. Acad. Sci. U.S.A.* **92,** 7090–7094.
Henrissat, B. (1991) *Biochem. J.* **280,** 309–316.
Henrissat, B. and Bairoch, A. (1993) *Biochem. J.* **293,** 781–788.
Hickey, D. A., Benkel, B. F., Boer, P. H., Genest, Y., Abukashawa, S. and Ben-David, G. (1987) *J. molec. Evol.* **26,** 252–256.
Hickey, D. A., Bally-Cuif, L., Abukashawa, S., Payant, V. and Benkel, B. F. (1991) *Proc. natn. Acad. Sci. U.S.A.* **88,** 1611–1615.
Higgins, D. G., Bleasby, A. J. and Fuchs, R. (1992) *Comput. Applic. Biosci.* **8,** 189–191.
Hill, R. D. and MacGregor, A. W. (1988) *Adv. Cereal Sci. Technol.* **9,** 217–261.
Hofemeister, B., König, S., Hoang, V., Engel, J., Mayer, G., Hansen, G. and Hofemeister, J. (1994) *Appl. environ. Microbiol.* **60,** 3381–3389.
Holló, J., László, E., Hoschke, Á., El Hawary, F. and Bánky, B. (1982) *Starch* **34,** 304–308.
Holm, L., Koivula, A. K., Lehtovaara, P. M., Hemminki, A. and Knowles, J. K. C. (1990) *Prot. Eng.* **3,** 181–191.
Hong, S. H. and Marmur, J. (1986) *Gene* **41,** 75–84.
Horinouchi, S., Fukusumi, S., Ohshima, T. and Beppu, T. (1988) *Eur. J. Biochem.* **176,** 243–253.
Hoschke, Á., László, E. and Holló, J. (1980a) *Carbohydr. Res.* **81,** 145–156.
Hoschke, Á., László, E. and Holló, J. (1980b) *Carbohydr. Res.* **81,** 157–166.
Hu, N.-T., Hung, M.-N., Huang, A.-M., Tsai, H.-F., Yang, B.-Y., Chow, T.-Y. and Tseng, Y.-H. (1992) *J. gen. Microbiol.* **138,** 1647–1655.
Huang, N., Sutliff, T. D., Litts, J. C. and Rodriguez, R. L. (1990) *Plant molec. Biol.* **14,** 655–668.
Iefuji, H., Chino, M., Kato, M. and Iimura, Y. (1996) *Biochem. J.* **318,** 989–996.
Igarashi, K., Ara, K., Saeki, K., Ozaki, K., Kawai, S. and Ito, S. (1992) *Biosci. Biotech. Biochem.* **56,** 514–516.
Ihara, H., Sasaki, T., Tsuboi, A., Yamagata, H., Tsukagoshi, N. and Udaka, S. (1985) *J. Biochem.* **98,** 95–103.
Inohara-Ochiai, M., Nakayama, T., Goto, R., Nakao, M., Ueda, T. and Shibano, Y. (1997) *J. biol. Chem.* **272,** 1601–1607.
Inomata, N. and Yamazaki, T. (1996) *J. Genet.* **75,** 125–137.
Ishikawa, K. and Hirata, H. (1989) *Arch. Biochem. Biophys.* **272,** 356–363.
Ishikawa, K., Matsui, I., Honda, K. and Nakatani, H. (1992) *Biochem. Biophys. Res. Commun.* **183,** 286–291.
Ishikawa, K., Matsui, I., Kobayashi, S., Nakatani, H. and Honda, K. (1993) *Biochemistry* **32,** 6259–6265.
Itkor, P., Tsukagoshi, N. and Udaka, S. (1990) *Biochem. Biophys. Res. Commun.* **166,** 630–636.
Janeček, Š. (1992) *Biochem. J.* **288,** 1069–1070.
Janeček, Š. (1993a) *Fedn Eur. biochem. Socs Lett.* **316,** 23–26.
Janeček, Š. (1993b) *Int. J. Biol. Macromol.* **15,** 317–318.
Janeček, Š. (1994a) *Fedn Eur. biochem. Socs Lett.* **353,** 119–123.
Janeček, Š. (1994b) *Eur. J. Biochem.* **224,** 519–524.
Janeček, Š. (1995a) *Biologia* **50,** 515–522.
Janeček, Š. (1995b) *Fedn Eur. biochem. Socs Lett.* **377,** 6–8.
Janeček, Š. (1996a) *Prot. Sci.* **5,** 1136–1143.
Janeček, Š. (1996b) *Biochem. J.* **319,** 1005–1006.
Janeček, Š. (1996c) *J. evol. Biol.* **9,** 1017–1021.

Janeček, Š. and Baláž, Š. (1995) *Prot. Eng.* **8,** 809–813.
Janeček, Š. and Bateman, A. (1996) *Biologia* **51,** 613–628.
Janeček, Š., MacGregor, E. A. and Svensson, B. (1995) *Biochem. J.* **305,** 685–686.
Janeček, Š., Svensson, B. and Henrissat, B. (1997) *J. molec. Evol.* (in press).
Janeček, Š. and Tóth, D. (1994) *Biologia* **49,** 301–306.
Janse, B. J. H., Steyn, A. J. C. and Pretorius, I. S. (1993) *Curr. Genet.* **24,** 400–407.
Jenkins, J., Lo Leggio, L., Harris, G. and Pickersgill, R. (1995) *Fedn Eur. biochem. Socs Lett.* **362,** 281–285.
Jespersen, H. M., MacGregor, E. A., Sierks, M. R. and Svensson, B. (1991) *Biochem. J.* **280,** 51–55.
Jespersen, H. M., MacGregor, E. A., Henrissat, B., Sierks, M. R. and Svensson, B. (1993) *J. Prot. Chem.* **12,** 791–805.
Jia, J., Huang, W. J., Schorken, U., Sahm, H., Sprenger, G. A., Lindqvist, Y. and Schneider, G. (1996) *Structure* **4,** 715–724.
Joyet, P., Declerck, N. and Gaillardin, C. (1992) *Bio/Technology* **10,** 1579–1583.
Juge, N., Søgaard, M., Chaix, J.-C., Martin-Eauclaire, M. F., Svensson, B., Marchis-Mouren, G. and Guo, X. (1993) *Gene* **130,** 159–166.
Juge, N., Rodenburg, K. W., Guo, X.-J., Chaix, J.-C. and Svensson, B. (1995) *Fedn Eur. biochem. Socs Lett.* **363,** 299–303.
Kadziola, A., Abe, J., Svensson, B. and Haser, R. (1994) *J. molec. Biol.* **239,** 104–121.
Kamitori, S., Satou, T., Tonozuka, T., Sakano, Y., Matsuzawa, H. and Okuyama, K. (1995) *J. struct. Biol.* **114,** 229–231.
Kaneko, A., Sudo, S., Takayasu-Sakamoto, Y., Tamura, G., Ishikawa, T. and Oba, T. (1996) *J. Ferment. Bioeng.* **81,** 292–298.
Kato, T. and Horikoshi, K. (1986) *J. Jpn Soc. Starch Sci.* **33,** 137–143.
Kimura, K., Kataoka, S., Ishii, Y., Takano, T. and Yamane, K. (1987) *J. Bacteriol.* **169,** 4399–4402.
Kimura, T. and Horikoshi, K. (1990) *Fedn Eur. microbiol. Socs Microbiol. Lett.* **71,** 35–42.
Kitamoto, N., Yamagata, H., Kato, T., Tsukagoshi, N. and Udaka, S. (1988) *J. Bacteriol.* **170,** 5848–5854.
Kizaki, H., Hata, Y., Watanabe, K., Katsube, Y. and Suzuki, Y. (1993) *J. Biochem.* **113,** 646–649.
Klein, C. and Schulz, G. E. (1991) *J. molec. Biol.* **217,** 737–750.
Klein, C., Hollender, J., Bender, H. and Schulz, G. E. (1992) *Biochemistry* **31,** 8740–8746.
Knapp, S., Rüdiger, A., Antranikian, G., Jorgensen, P. L. and Ladenstein, R. (1995) *Proteins: Struct. Funct. Genet.* **23,** 595–597.
Knegtel, R. M. A., Strokopytov, B., Penninga, D., Faber, O. G., Rozeboom, H. J., Kalk, K. H., Dijkhuizen, L. and Dijkstra, B. W. (1995) *J. biol. Chem.* **270,** 29256–29264.
Knegtel, R. M. A., Wind, R. D., Rozeboom, H. J., Kalk, K. H., Buitelaar, R. M., Dijkhuizen, L. and Dijkstra, B. W. (1996) *J. molec. Biol.* **256,** 611–622.
Kobayashi, T., Kanai, H., Aono, R., Horokoshi, K. and Kudo, T. (1994) *J. Bacteriol.* **176,** 5131–5134.
Kobayashi, K., Kato, M., Miura, Y., Kettoku, M., Komeda, T. and Iwamatsu, A. (1996) *Biosci. Biotech. Biochem.* **60,** 1720–1723.
Kochhar, S. and Dua, R. D. (1985) *Arch. Biochem. Biophys.* **240,** 757–767.
Koivula, T. T., Hemilä, H., Pakkanen, R., Sibakov, M. and Palva, I. (1993) *J. gen. Microbiol.* **139,** 2399–2407.
Kornacker, M. G. and Pugsley, A. P. (1990) *Molec. Microbiol.* **4,** 73–85.
Kraulis, P. J. (1991) *J. appl. Cryst.* **24,** 946–950.
Kubota, M., Matsuura, Y., Sakai, S. and Katsube, Y. (1991) *Denpun Kagaku* **38,** 141–146.
Kuriki, T. and Imanaka, T. (1989) *J. gen. Microbiol.* **135,** 1521–1528.
Kuriki, T., Guan, H., Sivak, M. and Preiss, J. (1996a) *J. Prot. Chem.* **15,** 305–313.
Kuriki, T., Kaneko, H., Yanase, M., Takata, H., Shimada, J., Handa, S., Takada, T., Umeyama, H. and Okada, S. (1996b) *J. biol. Chem.* **271,** 17321–17329.
Kwiatowski, J., Krawczyk, M., Kornacki, M., Bailey, K. and Ayala, F. J. (1995) *Proc. natn. Acad. Sci. U.S.A.* **92,** 8503–8506.
Laderman, K. A., Asada, K., Uemori, T., Mukai, H., Taguchi, Y., Kato, I. and Anfinsen, C. B. (1993) *J. biol. Chem.* **268,** 24402–24407.
Lamminmäki, U. and Vihinen, M. (1996) *Biochim. biophys. Acta* **1295,** 195–200.
Larsen, T. D., Laughlin, L. T., Holden, H. M., Rayment, I. and Reed, G. H. (1994) *Biochemistry* **33,** 6301–6309.
Larson, S. B., Greenwood, A., Cascio, D., Day, J. and McPherson, A. (1994) *J. molec. Biol.* **235,** 1560–1584.
Lawson, C. L., Van Montfort, R., Strokopytov, B., Rozeboom, H. J., Kalk, K. H., De Vries, G. E., Penninga, D., Dijkhuizen, L. and Dijkstra, B. W. (1994) *J. molec. Biol.* **236,** 590–600.
Lee, J. T., Kanai, H., Kobayashi, T., Akiba, T. and Kudo, T. (1996) *J. Ferment. Bioeng.* **82,** 432–438.
Lee, S.-P., Morikawa, M., Takagi, M. and Imanaka, T. (1994) *Appl. environ. Microbiol.* **60,** 3764–3773.
Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A. and Mornon, J. P. (1990) *Biochimie* **72,** 555–574.
Le Moine, S., Sellos, D., Moal, D., Daniel, J. Y. and San Juan Serrano, F. (1996) *GenBank*: X99729 (unpublished).
Lesk, A. M., Brändén, C.-I. and Chothia, C. (1989) *Proteins: Struct. Funct. Genet.* **5,** 139–148.
Liebl, W., Stemplinger, I. and Ruile, P. (1997) *J. Bacteriol.* **179,** 941–948.
Liu, W., de Castro, M. L., Takrama, J., Bilous, P. T., Vinayagamoorthy, T., Madsen, N. B. and Bleackley, R. C. (1993) *Arch. Biochem. Biophys.* **306,** 232–239.
Logsdon, J. M.Jr, Tyshenko, M. G., Dixon, C.-D., Jafari, J., Walker, V. K. amd Palmer, J. D. (1995) *Proc. natn. Acad. Sci. U.S.A.* **92,** 8507–8511.
Long, C. M., Virolle, M.-J., Chang, S.-Y., Chang, S. and Bibb, M. J. (1987) *J. Bacteriol.* **169,** 5745–5754.
MacGregor, E. A. (1988) *J. Prot. Chem.* **7,** 399–415.
MacGregor, E. A. (1993) *Starch* **7,** 232–237.
MacGregor, E. A., Jespersen, H. M. and Svensson, B. (1996) *Fedn Eur. biochem. Socs Lett.* **378,** 263–266.
MacGregor, E. A. and Svensson, B. (1989) *Biochem. J.* **259,** 145–152.

Machius, M., Wiegand, G. and Huber, R. (1995) *J. molec. Biol.* **246,** 545–559.
Machius, M., Vértesy, L., Huber, R. and Wiegand, G. (1996) *J. molec. Biol.* **260,** 409–421.
Magoulas, C., Loverre-Chyurlia, A., Abukashawa, S., Bally-Cuif, L. and Hickey, D. A. (1993) *J. molec. Evol.* **36,** 234–242.
Marchionni, M. and Gilbert, W. (1986) *Cell* **46,** 133–141.
Maruta, K., Hattori, K., Nakada, T., Kubota, M., Sugimoto, T. and Kurimoto, M. (1996a) *Biochim. biophys. Acta* **1289,** 10–13.
Maruta, K., Hattori, K., Nakada, T., Kubota, M., Sugimoto, T. and Kurimoto, M. (1996b) *Biosci. Biotech. Biochem.* **60,** 717–720.
Mathupala, S. P., Lowe, S. E., Podkovyrov, S. M. and Zeikus, J. G. (1993) *J. biol. Chem.* **268,** 16332–16344.
Matsui, I., Ishikawa, K., Miyairi, S., Fukui, S. and Honda, K. (1992a) *Biochemistry* **31,** 5232–5236.
Matsui, I., Ishikawa, K., Miyairi, S., Fukui, S. and Honda, K. (1992b) *Fedn Eur. biochem. Socs Lett.* **310,** 216–218.
Matsui, I., Yoneda, S., Ishikawa, K., Miyairi, S., Fukui, S., Umeyama, H. and Honda, K. (1994) *Biochemistry* **33,** 451–458.
Matsuura, Y., Kusunoki, M., Harada, W., Tanaka, N., Iga, Y., Yasuoka, N., Toda, H., Narita, K. and Kakudo, M. (1980) *J. Biochem.* **87,** 1555–1558.
Matsuura, Y., Kusunoki, M., Harada, W. and Kakudo, M. (1984) *J. Biochem.* **95,** 697–702.
Matsuura, Y., Kusunoki, M. and Kakudo, M. (1991) *Denpun Kagaku* **38,** 137–139.
Matsuura, Y. (1995) In *Enzyme Chemistry and Molecular Biology of Amylases and Related Enzymes*, pp. 137–145 (ed. T. Yamamoto), CRC, Ann Arbor.
Matsuura, Y. and Kubota, M. (1995) In *Enzyme Chemistry and Molecular Biology of Amylases and Related Enzymes*, pp. 153–162 (ed. T. Yamamoto), CRC, Ann Arbor.
Mattevi, A., Valentini, G., Rizzi, M., Speranza, M. L., Bolognesi, M. and Coda, A. (1995) *Structure* **3,** 729–741.
Mattsson, P., Battchikova, N., Sippola, K. and Korpela, T. (1995) *Biochim. biophys. Acta* **1247,** 97–103.
Mazur, A. K. and Nakatani, H. (1993) *Arch. Biochem. Biophys.* **306,** 29–38.
McCarter, J. D. and Withers, S. G. (1996) *J. biol. Chem.* **271,** 6889–6894.
Metz, R. J., Allen, L. N., Cao, T. M. and Zeman, N. W. (1988) *Nucleic Acids Res.* **16,** 5203.
Mikami, B., Hehre, E. J., Sato, M., Katsube, Y., Hirose, M., Morita, Y. and Sacchettini, J. C. (1993) *Biochemistry* **32,** 6836–6845.
Monchois, V., Willemot, R. M., Remaud-Simeon, M., Croux, C. and Monsan, P. (1996) *Gene* **182,** 23–32.
Mooser, G., Hefta, S. A., Payton, R. J., Shively, J. E. and Lee, T. D. (1991) *J. biol. Chem.* **266,** 8916–8922.
Muirhead, H. (1983) *Trends Biochem. Sci.* **8,** 326–330.
Muirhead, H., Clayden, D. A., Barford, D., Lorimer, C. G., Fothergill-Gilmore, L. A., Schiltz, E. and Schmitt, W. (1986) *Eur. molec. Biol. Org. J.* **5,** 475–481.
Murzin, A. G., Brenner, S. E., Hubbard, T. and Chothia, C. (1995) *J. molec. Biol.* **247,** 536–540 (The SCOP database is freely accessible at the URL: http://scop.mrc-lmb.cam.ac.uk/scop/).
Nagashima, T., Tada, S., Kitamoto, K., Gomi, K., Kumagai, C. and Toda, H. (1992) *Biosci. Biotech. Biochem.* **56,** 207–210.
Nakajima, R., Imanaka, T. and Aiba, S. (1986) *Appl. Microbiol. Biotechnol.* **23,** 355–360.
Nakamura, A., Haga, K., Ogawa, S., Kuwano, K., Kimura, K. and Yamane, K. (1992) *Fedn Eur. biochem. Socs Lett.* **296,** 37–40.
Nakamura, A., Haga, K. and Yamane, K. (1994a) *Biochemistry* **33,** 9929–9936.
Nakamura, A., Haga, K. and Yamane, K. (1994b) *Fedn Eur. biochem. Socs Lett.* **337,** 66–70.
Nakamura, Y. (1996) *Plant Sci.* **121,** 1–18.
Nakamura, Y., Umemoto, T., Ogata, N., Kuboki, Y., Yano, M. and Sasaki, T. (1996) *Planta* **199,** 209–218.
Nakatani, H. (1995) *J. theoret. Biol.* **176,** 195–197.
Nakatani, H. (1996a) *J. theoret. Biol.* **178,** 423–424.
Nakatani, H. (1996b) *Biopolymers* **39,** 665–669.
Nakatani, H. and Kobayashi, I. (1996) *Comp. Biochem. Physiol.* **113B,** 383–386.
Nakatani, H., Kobayashi, I. and Miyauchi, T. (1996) *Comp. Biochem. Physiol.* **115B,** 389–392.
Nishizawa, M., Ozawa, F. and Hishinuma, F. (1987) *DNA* **6,** 255–265.
Nitschke, L., Heeger, K., Bender, H. and Schulz, G. E. (1990) *Appl. Microbiol. Biotechnol.* **33,** 542–546.
Nógrády, N., Pócsi, I. and Szentirmai, A. (1995) *Biotechnol. appl. Biochem.* **21,** 233–243.
Oguma, T., Matsuyama, A., Kikuchi, M. and Nakano, E. (1993) *Appl. Microbiol. Biotechnol.* **39,** 197–203.
O'Neil, S. D., Kumagai, M. H., Majumdar, A., Huang, N., Sutliff, T. D. and Rodriguez, R. L. (1990) *Molec. gen. Genet.* **221,** 235–244.
Oosthuizen, V., Naudé, R. J., Oelofsen, W., Muramoto, K. and Kamiya, H. (1994) *Int. J. Biochem.* **26,** 1313–1321.
Oyama, T., Myojin, M., Nitta, Y., Toda, H., Nagashima, T. and Kitamoto, K. (1996) *Biosci. Biotech. Biochem.* **60,** 1351–1352.
Pasero, L., Mazzei-Pierron, Y., Abadie, B., Chicheportiche, Y. and Marchis-Mouren, G. (1986) *Biochim. biophys. Acta* **869,** 147–157.
Penninga, D., Strokopytov, B., Rozeboom, H. J., Lawson, C. L., Dijkstra, B. W., Bergsma, J. and Dijkhuizen, L. (1995) *Biochemistry* **34,** 3368–3376.
Penninga, D., van der Veen, B. A., Knegtel, R. M. A., van Hijum, S. A. F. T., Rozeboom, H. J., Kalk, K. H., Dijkstra, B. W. and Dijkhuizen, L. (1996) *J. biol. Chem.* **271,** 32777–32784.
Perlick, A. M., Frühling, M., Schröder, G., Frosch, S. C. and Pühler, A. (1996) *Plant Physiol.* **110,** 147–154.
Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A. B., Chet, I., Wilson, K. S. and Vorgias, C. E. (1994) *Structure* **2,** 1169–1180.
Petricek, M., Tichý, P. and Kuncová, M. (1992) *Gene* **112,** 77–83.
Pickett, S. D., Saqi, M. A. S. and Sternberg, M. J. E. (1992) *J. molec. Biol.* **228,** 170–187.
Podkovyrov, S. M. and Zeikus, J. G. (1992) *J. Bacteriol.* **174,** 5400–5405.

Podkovyrov, S. M., Burdette, D. and Zeikus, J. G. (1993) *Fedn Eur. biochem. Socs Lett.* **317,** 259–262.
Popadic, A. and Anderson, W. W. (1995) *Molec. Biol. Evol.* **12,** 564–572.
Popadic, A., Norman, R. A., Doane, W. W. and Anderson, W. W. (1996) *Molec. Biol. Evol.* **13,** 883–888.
Pujadas, G., Ramírez, F. M., Valero, R. and Palau, J. (1996) *Proteins: Struct. Funct. Genet.* **25,** 456–472.
Qian, M., Haser, R. and Payan, F. (1993) *J. molec. Biol.* **231,** 785–799.
Qian, M., Haser, R., Buisson, G., Duée, E. and Payan, F. (1994) *Biochemistry* **33,** 6284–6294.
Qian, M., Haser, R. and Payan, F. (1995) *Prot. Sci.* **4,** 747–755.
Quackenbush, E., Clabby, M., Gottesdiener, K. M., Barbosa, J., Jones, N. H., Strominger, J. L., Speck, S. and Leiden, J. L. (1987) *Proc. natn. Acad. Sci. U.S.A.* **84,** 6526–6530.
Raha, M. A., Kawagishi, I., Mueller, V., Kihara, M. and Macnab, R. M. (1992) *J. Bacteriol.* **174,** 6644–6652.
Raimbaud, E., Buleon, A., Perez, S. and Henrissat, B. (1989) *Int. J. Biol. Macromol.* **11,** 217–225.
Raine, A. R. C., Scrutton, N. S. and Mathews, F. S. (1994) *Prot. Sci.* **3,** 1889–1892.
Ramasubbu, N., Paloth, V., Luo, Y., Brayer, G. D. and Levine, M. J. (1996) *Acta Cryst.* **D52,** 435–446.
Reardon, D. and Farber, G. K. (1995) *Fedn Am. Socs exp. Biol. J.* **9,** 497–503.
Richardson, J. S. (1977) *Nature* **268,** 495–500.
Rimmele, M. and Boos, W. (1994) *J. Bacteriol.* **176,** 5654–5664.
Robyt, J. F. (1989) *Denpun Kagaku* **36,** 287–301.
Robyt, J. F. (1996) In *Enzymes for Carbohydrate Engineering*, pp. 1–22 (eds. K. H. Park, J. F. Robyt and Y.-D. Choi), Elsevier, Amsterdam.
Rodenburg, K. W., Juge, N., Guo, X.-J., Søgaard, M., Chaix, J.-C. and Svensson, B. (1994) *Eur. J. Biochem.* **221,** 277–284.
Rogers, J. C. (1985) *J. biol. Chem.* **260,** 3731–3738.
Rogers, J. C. and Milliman, C. (1983) *J. biol. Chem.* **258,** 8169–8174.
Romier, C., Reuter, K., Suck, D. and Ficner, R. (1996) *Eur. molec. Biol. Org. J.* **15,** 2850–2857.
Rumbak, E., Rawlings, D. E., Lindsey, G. G. and Woods, D. R. (1991) *J. Bacteriol.* **173,** 4203–4211.
Russell, R. R. B. and Ferretti, J. J. (1990) *J. gen. Microbiol.* **136,** 803–810.
Saitou, N. and Nei, M. (1987) *Molec. Biol. Evol.* **4,** 406–425.
Schwermann, B., Pfau, K., Liliensiek, B., Schleyer, M., Fischer, T. and Bakker, E. P. (1994) *Eur. J. Biochem.* **226,** 981–991.
Scrutton, N. S. (1994) *BioEssays* **16,** 115–122.
Shibata, H. and Yamazaki, T. (1995) *Genetics* **141,** 223–236.
Shida, O., Takano, T., Takagi, H., Kadowaki, K. and Kobayashi, S. (1992) *Biosci. Biotech. Biochem.* **56,** 76–80.
Shin, Y. C. and Byun, S. M. (1996) In *Enzymes for Carbohydrate Engineering*, pp. 61–82 (eds. K. H. Park, J. F. Robyt and Y.-D. Choi), Elsevier, Amsterdam.
Sin, K. A., Nakamura, A., Kobayashi, K., Masaki, H. and Uozumi, T. (1991) *Appl. Microbiol. Biotechnol.* **35,** 600–605.
Sin, K.-A., Nakamura, A., Masaki, H., Matsuura, Y. and Uozumi, T. (1994) *J. Biotechnol.* **32,** 283–288.
Song, S. W., Hwang, K. Y., Chang, C. and Suh, S. W. (1996) In *Enzymes for Carbohydrate Engineering*, pp. 167–170 (eds. K. H. Park, J. F. Robyt and Y.-D. Choi), Elsevier, Amsterdam.
Sorimachi, K., Jacks, A. J., Le Gal-Coëffet, M.-F., Williamson, G., Archer, D. B. and Williamson, M. P. (1996) *J. molec. Biol.* **259,** 970–987.
Søgaard, M., Abe, J.-I., Martin-Eauclaire, M. F. and Svensson, B. (1993a) *Carbohydr. Polymers* **21,** 137–146.
Søgaard, M., Kadziola, A., Haser, R. and Svensson, B. (1993b) *J. biol. Chem.* **268,** 22480–22484.
Steyn, A. J. C., Marmur, J. and Pretorius, I. S. (1995) *Gene* **166,** 65–71.
Straus, D. and Gilbert, W. (1985) *Molec. cell. Biol.* **5,** 3497–3506.
Strokopytov, B., Penninga, D., Rozeboom, H. J., Kalk, K. H., Dijkhuizen, L. and Dijkstra, B. W. (1995) *Biochemistry* **34,** 2234–2240.
Strokopytov, B., Knegtel, R. M. A., Penninga, D., Rozeboom, H. J., Kalk, K. H., Dijkhuizen, L. and Dijkstra, B. W. (1996) *Biochemistry* **35,** 4241–4249.
Suganuma, T., Ohnishi, M., Hiromi, K. and Nagahama, T. (1996) *Carbohydr. Res.* **282,** 171–180.
Suzuki, A., Yamane, T., Ito, Y., Nishio, T., Fujiwara, H. and Ashida, T. (1990) *J. Biochem.* **108,** 379–381.
Suzuki, Y. (1989) *Proc. Jpn Acad.* **65B,** 146–148.
Suzuki, Y., Ito, N., Yuuki, T., Yamagata, H. and Udaka, S. (1989) *J. biol. Chem.* **264,** 18933–18938.
Suzuki, Y., Oishi, K., Nakano, H. and Nagayama, T. (1987) *Appl. Microbiol. Biotechnol.* **26,** 546–551.
Svensson, B. (1988) *Fedn Eur. biochem Socs Lett.* **230,** 72–76.
Svensson, B. (1994) *Plant molec. Biol.* **25,** 141–157.
Svensson, B., Jespersen, H., Sierks, M. and MacGregor, E. A. (1989) *Biochem. J.* **264,** 309–311.
Svensson, B. and Søgaard, M. (1993) *J. Biotechnol.* **29,** 1–37.
Swift, H. J., Brady, L., Derewenda, Z. S., Dodson, E. J., Dodson, G. G., Turkenburg, J. P. and Wilkinson, A. J. (1991) *Acta Cryst.* **B47,** 535–544.
Tachibana, Y., Mendez Leclere, M., Fujiwara, S., Takagi, M. and Imanaka, T. (1996) *J. Ferment. Bioeng.* **82,** 224–232.
Takaha, T., Yanase, M., Okada, S. and Smith, S. M. (1993) *J. biol. Chem.* **268,** 1391–1396.
Takase, K. (1992) *Eur. J. Biochem.* **211,** 899–902.
Takase, K. (1994) *Biochemistry* **33,** 7925–7930.
Takata, H., Takaha, T., Kuriki, T., Okada, S., Takagi, M. and Imanaka, T. (1994) *Appl. environ. Microbiol.* **60,** 3096–3104.
Takii, Y., Takahashi, K., Yamamoto, K., Sogabe, Y. and Suzuki, Y. (1996) *Appl. Microbiol. Biotechnol.* **44,** 629–634.
Takkinen, K., Pettersson, R. F., Kalkkinen, N., Palva, I., Söderlund, H. and Kääriäinen, L. (1983) *J. biol. Chem.* **258,** 1007–1013.
Terashima, M., Kubo, A., Suzawa, M., Itoh, Y. and Katoh, S. (1994) *Eur. J. Biochem.* **226,** 249–254.

Terashima, M., Kawai, M., Kumagai, M. H., Rodriguez, R. L. and Katoh, S. (1996) *Appl. Microbiol. Biotechnol.* **45,** 607–611.
Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K. S. and Vorgias, C. E. (1996) *Nature struct. Biol.* **3,** 638–648.
Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680.
Toda, H., Kondo, K. and Narita, K. (1982) *Proc. Jpn Acad.* **B58,** 208–212.
Tomazic, S. J. and Klibanov, A. M. (1988a) *J. biol. Chem.* **263,** 3086–3091.
Tomazic, S. J. and Klibanov, A. M. (1988b) *J. biol. Chem.* **263,** 3092–3096.
Tonozuka, T., Ohtsuka, M., Mogi, S.-I., Sakai, H., Ohta, T. and Sakano, Y. (1993) *Biosci. Biotech. Biochem.* **57,** 395–401.
Tsao, L.-S., Lin, L.-L., Chen, J.-C., Chen, J.-H. and Hsu, W.-H. (1993) *Biochim. biophys. Acta* **1171,** 255–262.
Tsukamoto, A., Kimura, K., Ishi, Y., Takano, T. and Yamane, K. (1988) *Biochem. Biophys. Res. Commun.* **151,** 25–31.
Uozumi, N., Sakurai, K., Sasaki, T., Takekawa, S., Yamagata, H., Tsukagoshi, N. and Udaka, S. (1989) *J. Bacteriol.* **171,** 375–382.
Van Wormhoudt, A. and Sellos, D. (1996) *J. molec. Evol.* **42,** 543–551.
Vigal, T., Gil, J. A., Daza, A., Garcia-Gonzalez, M. D. and Martin, J. F. (1991) *Molec. gen. Genet.* **225,** 278–288.
Vihinen, M. and Mäntsälä, P. (1989) *Crit. Rev. Biochem. molec. Biol.* **24,** 329–418.
Vihinen, M., Ollikka, P., Niskanen, J., Meyer, P., Suominen, I., Karp, M., Holm, L., Knowles, J. and Mäntsälä, P. (1990) *J. Biochem.* **107,** 267–272.
Vihinen, M., Peltonen, T., Iitiä, A., Suominen, I. and Mäntsälä, P. (1994) *Prot. Eng.* **7,** 1255–1259.
Virolle, M.-J., Long, C. M., Chang, S. and Bibb, M. J. (1988) *Gene* **74,** 321–334.
Warren, R. A. J. (1996) *A. Rev. Microbiol.* **50,** 183–212.
Watanabe, K., Chishiro, K., Kitamura, K. and Suzuki, Y. (1991) *J. biol. Chem.* **266,** 24287–24294.
Watanabe, K., Kitamura, K., Iha, H. and Suzuki, Y. (1990) *Eur. J. Biochem.* **192,** 609–620.
Watanabe, K., Kitamura, K. and Suzuki, Y. (1996) *Appl. environ. Microbiol.* **62,** 2066–2073.
Watanabe, K., Masuda, T., Ohashi, H., Mihara, H. and Suzuki, Y. (1994) *Eur. J. Biochem.* **226,** 277–283.
Wells, R. G. and Hediger, M. A. (1992) *Proc. natn. Acad. Sci. U.S.A.* **89,** 5596–5600.
Whitehead, T. R. and Cotta, M. A. (1995) *Curr. Microbiol.* **30,** 143–148.
Whiting, G. C., Sutcliffe, I. C. and Russell, R. R. B. (1993) *J. gen. Microbiol.* **139,** 2019–2026.
Wiegand, G., Epp, O. and Huber, R. (1995) *J. molec. Biol.* **247,** 99–110.
Wilmanns, M., Hyde, C. C., Davies, D. R., Kirschner, K. and Jansonius, J. N. (1991) *Biochemistry* **30,** 9161–9169.
Wilson, D. K., Nakano, T., Petrash, J. M. and Quiocho, F. A. (1995) *Biochemistry* **34,** 14323–14330.
Wind, R. D., Liebl, W., Buitelaar, R. M., Penninga, D., Spreinat, A., Dijkhuizen, L. and Bahl, H. (1995) *Appl. environ. Microbiol.* **61,** 1257–1265.
Yamashita, H., Nakatani, H. and Tonomura, B. (1993) *Biochim. biophys. Acta* **1202,** 129–134.
Yamauchi, D. and Minamikawa, T. (1990) *Nucleic Acids Res.* **18,** 4250.
Yang, M., Galizzi, A. and Henner, D. (1983) *Nucleic Acids Res.* **11,** 237–249.
Yin, X. H., Francou, F. X., Gerbaud, C., Guerineau, M. and Virolle, M. J. (1997) *GenBank*: Z86113 (unpublished).
Young, T. E., Demason, D. A. and Close, T. J. (1994) *Plant Physiol.* **105,** 759–760.
Yuuki, T., Nomura, T., Tezuka, H., Tsuboi, A., Yamagata, H., Tsukagoshi, N. and Udaka, S. (1985) *J. Biochem.* **98,** 1147–1156.
Zhou, J., Baba, T., Takano, T., Kobayashi, S. and Arai, Y. (1989) *Fedn Eur. biochem Socs Lett.* **255,** 37–41.