CMLS, Cell. Mol. Life Sci. 59 (2002) 1945–1959
1420-682X/02/111945-15

CMLS Cellular and Molecular Life Sciences

Research Article

Oligo-1,6-glucosidase and neopullulanase enzyme subfamilies from the α-amylase family defined by the fifth conserved sequence region

A. Oslancová and Š. Janeček*

Institute of Molecular Biology, Slovak Academy of Sciences, Dúbravská cesta 21, 84251 Bratislava (Slovakia), Fax + 421 2 5930 7416, e-mail: stefan.janecek@savba.sk

Received 12 July 2002; received after revision 28 August 2002; accepted 24 September 2002

Abstract. The α-amylase enzyme family is the largest family of glycoside hydrolases. It contains almost 30 different enzyme specificities covering hydrolases, transferases and isomerases. Some of the enzyme specificities from the family are closely related, others less so. This study, based on the analysis of 79 amino acid sequences, postulates two subfamilies in the framework of the α-amylase family: the oligo-1,6-glucosidase subfamily and the neopullulanase subfamily. The specific sequence in the fifth conserved sequence region of the family served as the basis for defining the subfamilies: QpDln for the oligo-1,6-glucosidase subfamily and MPKln for the neopullulanase subfamily. This conserved sequence region is proposed to be the selection marker that enables one to distinguish between the two subfamilies. The 'intermediary' sequence MPDLN can be characteristic of the so-called intermediary group with a mixed enzyme specificity of α-amylase, cyclomaltodextrinase and neopullulanase. The evolutionary trees clearly supported the proposed definition of the two subfamilies.

Key words. Alpha-amylase enzyme family; oligo-1,6-glucosidase; neopullulanase; conserved sequence region; evolutionary relatedness; protein bioinformatics.

The α-amylase enzyme family contains almost 30 different enzyme specificities covering hydrolases, transferases and isomerases [1]. In the sequence-based classification system of glycoside hydrolases it forms a clan GH-H grouping the families 13, 70 and 77 [2]. These enzymes can cleave and/or synthesise the α-1,4-, α-1,6- and less commonly α-1,2- and α-1,3-glucosidic linkages, as well as act on sucrose (α-1,5-bonds) and trehalose (α-1,1-bonds) [1]. The catalytic domain of all of these enzymes should adopt the structure of a parallel $(\beta/\alpha)_8$ barrel [1–3] first recognised for amylolytic enzymes in the structure of Taka-amylase A [4] which is an α-amylase from *Aspergillus oryzae*. From sequence and evolutionary points of view, some mammalian transport proteins and antigens have been suggested to be related to the enzymatically active members of the family [5].

Although the amino acid sequences of α-amylase family members have diverged so efficiently that numerous specificities have emerged, they possess several well-conserved sequence stretches known as conserved sequence regions [1, 3, 6–13]. These contain most of the residues that play important functional roles. Four of them, the regions I, II, III and IV, were definitively established in 1986 [9]. They cover the strands β3, β4, β5 and β7 of the $(\beta/\alpha)_8$ barrel with the three catalytic residues: Asp206, Glu230 and Asp297 (Taka-amylase A numbering) positioned near the C termini of β strands 4, 5 and 7, respectively. Further, three conserved sequence regions, regions V, VI and VII, were identified at the beginning of the 1990s [10–13]. These cover the strands β2 and β8 (regions VI and VII)

* Corresponding author.

and a short stretch (region V) near the C terminus of domain B protruding from the catalytic barrel between strand β3 and helix α3. The conserved strands β2 and β8 were shown to exhibit features characteristic of certain specificities or taxonomic groups [14, 15]. The fifth conserved sequence region, region V (173_LPDLD in Taka-amylase A), was originally described for α-amylases [10] and later also identified in the other enzyme specificities from the family [13]. The aspartate from the middle of the region (Asp175 in Taka-amylase A) is well conserved and may be involved in binding of a calcium ion in the enzymes from this family [16].

There have been several reports on evolutionary relationships among the individual enzyme specificities as well as in the frame of a given specificity from the α-amylase family (e.g. [11, 12, 15, 17–25]). In the past few years, however, the number of available sequences has substantially increased (almost 1000 according to the CAZY server accessed on 3 May 2002; [2]) and new enzyme specificities have also been recognised to belong to the family [1]. There may be groups of enzymes in the α-amylase family that are more closely related to each other [1]. In cases when the degree of sequence similarity is very high and the data concerning the characterisation of biochemical activity of a given sequence are lacking or not unambiguous, this may result in incorrect classification. For example, a few sequences of cyclodextrin glucanotransferases were classified incorrectly as α-amylases [26, 27] due to a superficial observation of four conserved sequence regions without deeper analysis [14]. It is thus of special importance to know as exactly as possible the sequence features that are required or at least highly characteristic of each enzyme specificity or enzyme group.

The aim of the present work was to define two enzyme subfamilies within the framework of the α-amylase family: the oligo-1,6-glucosidase subfamily and the neopullulanase subfamily. Both comprise several enzyme specificities and can be distinguished from each other by the specific sequence motif of the fifth conserved sequence region. This region can thus be used as an identification and selection marker. Several amylases that possess a motif containing the features of both the oligo-1,6-glucosidase and neopullulanase subfamilies are proposed to constitute an intermediary group.

Materials and methods

All amino acid sequences as well as enzyme specificities used in the present study are listed in table 1. The listing

Table 1. The α-amylase family enzymes used in the present study.

Enzyme	EC number	Organism	Abbreviation	GenPept	Reference
Oligo-1,6-glucosidase	3.2.1.10	*Bacillus cereus*	Bacce.ogl	CAA37583.1	[28]
		Bacillus coagulans	Bacco.ogl	BAA11354.1	[29]
		Bacillus flavocaldarius	Bacfl.ogl	BAB18518.1	[30]
		Bacillus sp. F5	Bac-F5.ogl	BAA00534.1	[31]
		Bacillus subtilis	Bacsu.ogl	AAG23399.1	GenBank Acc. No. AY008307
		Bacillus thermoglucosidasius	Bacth.ogl	BAA01368.1	[32]
		Erwinia rhapontici	Erwrh.ogl	AAK28737.1	[33]
α-Glucosidase	3.2.1.20	*Aedes aegypti*	Aedae.agl	AAA29352.1	[34]
		Anopheles gambiae (Agm1)	Anoga-1.agl	CAA60857.1	[35]
		Anopheles gambiae (Agm2)	Anoga-2.agl	CAA60858.1	[35]
		Apis meliffera	Apime.agl	BAA11466.1	[36]
		Bacillus sp. DG0303	Bac-DG.agl	AAF71997.1	[37]
		Bacillus sp. SAM1606	Bac-SAM.agl	CAA54266.1	[38]
		Bacillus stearothermophilus	Bacst.agl	BAA12704.1	[39]
		Bifidobacterium adolescentis (AglA)	Bifad-1.agl	AAK27723.1	GenBank Acc. No. AF411186
		Bifidobacterium adolescentis (AglB)	Bifad-2.agl	AAL05573.1	GenBank Acc. No. AF411186
		Brevibacterium fuscum	Brefu.agl	BAB60692.1	GenBank Acc. No. AB025195
		Candida albicans	Canal.agl	AAA34350.2	[40]
		Drosophila virilis (Mav1)	Drovi-1.agl	AAB82327.1	[41]
		Drosophila virilis (Mav2)	Drovi-2.agl	AAB82328.1	[41]
		Erwinia rhapontici	Erwrh.agl	AAK28739.1	[33]
		Kluyveromyces lactis	Klula.agl	CAB46746.1	GenBank Acc. No. AJ007636
		Pediococcus pentosaceus	Pedpe.agl	CAA83671.1	GenBank Acc. No. L32093
		Penicillium minioluteum	Penmi.agl	CAC09327.1	[42]
		Pichia angusta	Pican.agl	AAF69018.1	[43]
		Saccharomyces cerevisiae CB11	Sacce-CB.agl	AAA34757.1	[44]

Table 1 (continued)

Enzyme	EC number	Organism	Abbreviation	GenPept	Reference
		Saccharomyces cerevisiae FSP2	Sacce-FSP.agl	BAA07818.1	GenBank Acc. No. DA3761
		Sinorhizobium meliloti	Sinme.agl	AAD12047.1	[45]
		Staphylococcus xylosus	Staxy.agl	CAA55409.1	[46]
		Streptomyces lividans	Strli.agl	AAC46450.1	[47]
		Thermomonospora curvata	Thscu.agl	AAA57313.1	[48]
		Thermus caldophilus	Theca.agl	AAD50603.1	[49]
Dextran glucosidase	3.2.1.70	*Aspergillus parasiticus*	Asppa.dgl	AAF26276.1	[50]
		Streptococcus equisimilis	Stceq.dgl	CAA51348.1	[51]
		Streptococcus mutans	Stcmu.dgl	AAA26939.1	[52]
		Streptococcus suis	Stcsu.dgl	AAB65079.1	[53]
Trehalose-6-P-hydrolase	3.2.1.93	*Bacillus subtilis*	Bacsu.t6p	CAA91015.1	[54]
		Escherichia coli	Ecoli.t6p	AAC43382.1	[55]
		Pseudomonas fluorescens	Psefl.t6p	AAG31032.1	[56]
Amylosucrase	2.4.1.4	*Neisseria polysaccharea*	Neipo.asu	CAA09772.1	[22]
Sucrose phosphorylase	2.4.1.7	*Agrobacterium vitis*	Agrvi.sph	CAA80424.1	[57]
		Leuconostoc mesenteroides	Leume.sph	BAA14344.1	GenBank Acc. No. D90314
		Pseudomonas saccharophila	Psesa.sph	AAD40317.1	GenBank Acc. No. AF158367
		Streptococcus mutans	Stcmu.sph	AAA26937.1	[58]
Isomaltulose synthase	5.4.99.11	*Erwinia rhapontici*	Erwrh.isy	AAK28735.1	[33]
		Klebsiella sp. LX3	Kle-LX3.isy	AAK82938.1	[59]
Trehalose synthase	5.4.99.16	*Pimelobacter* sp. R48	Pim-R48.tsy	BAA11303.1	[60]
		Streptomyces coelicolor	Strco.tsy	CAA04601.2	[61]
		Thermus aquaticus	Theaq.tsy	BAA19934.1	[62]
Cyclomalto-dextrinase	3.2.1.54	Alicyclobacillus acidocaldarius	Aliac.cmd	CAB40078.1	[24]
		Bacillus sp. A2-5a	Bac-A2.cmd	BAA31576.1	GenBank Acc. No. AB015670
		Bacillus sp. I-5	Bac-I-5.cmd	AAA92925.1	[63]
		Bacillus sphaericus	Bacsf.cmd	CAA44454.1	[64]
		Thermoanaerobacter thermohydrosulfuricus	Thbth.cmd	AAA23219.1	[65]
		Thermococcus sp. B1001	Thc-B1001.cmd	BAB18100.1	[66]
		Thermotoga maritima	Thtma.cmd	AAD36898.1	[67]
		Thermotoga neapolitana	Thtne.cmd	CAA08867.1	GenBank Acc. No. AJ009832
Maltogenic amylase	3.2.1.133	*Bacillus acidopullulyticus*	Bacac.mam	CAA80246.1	GenBank Acc. No. 222520
		Bacillus stearothermophilus	Bacst.mam	AAC46346.1	[68]
		Bacillus subtilis	Bacsu.mam	AAF23874.1	[69]
		Thermus sp. IM6501	The-IM.mam	AAC15072.1	[70]
Neopullulanase	3.2.1.135	*Bacillus* sp. KCTC8848P	Bac-KCT.npu	AAL07400.1	[71]
		Bacillus sp. KSM-1876	Bac-KSM.npu	BAA02521.1	[72]
		Bacillus polymyxa	Bacpo.npu	AAD05199.1	[73]
		Bacillus stearothermophilus IMA6503	Bacst-IMA.npu	AAK15003.1	[74]
		Bacillus stearothermophilus TRS40	Bacst-TRS.npu	AAA22622.1	[75]
		Bacteroides thetaiotaomicron	Batth.npu	AAC44670.1	[76]
		Thermoactinomyces vulgaris (TVAI)	Thevu-1.npu	BAA02471.1	[77]
		Thermoactinomyces vulgaris (TVAII)	Thevu-2.npu	BAA02473.1	[78]
Intermediary group (MpDln)	3.2.1.1	*Bacillus megaterium*	Bacme.amy	CAA30247.1	[79]
	3.2.1.1	*Clostridium acetobutylicum* ATCC824	Cloac-ATCC.amy	AAD47072.1	[80]
	3.2.1.1	*Dictyoglomus thermophilum* (AmyC)	Dicth-C.amy	CAA34072.1	[81]
	3.2.1.1	*Mycoplasma pulmonis*	Mycpu.amy	CAC13805.1	[82]
	3.2.1.1	*Thermotoga maritima*	Thtma.amy	AAD36902.1	[83]
	3.2.1.1	*Xanthomonas campestris* (periplasmic)	Xanca-P.amy	BAA07401.1	[84]
Neopullulanase-like (MpKln)	3.2.1.41	*Bacillus flavocaldarius*	Bacfl.pul	BAB18516.1	[85]
	3.2.1.1	*Dictyoglomus thermophilum* (AmyB)	Dicth-B.amy	CAA31586.1	[81]
	not det.	*Klebsiella oxytoca* (CymH)	Kleox.nd	CAA60007.1	[86]
	3.2.1.41	*Thermococcus aggregans*	Thcag.pul	CAB94218.1	[87]

for the clan GH-H provided by the CAZy web-server (3 May 2002) served as a database [2]. The sequences were retrieved from GenPept on the ENTREZ system [88, 89]. The final set consisting of 79 sequences was obtained using the strategy described as follows:

(1) First, the search focused on the α-amylase family specificities indicated previously [13, 20] as those with a characteristic sequence motif in their fifth conserved sequence region. These specificities were oligo-1,6-glucosidase and neopullulanase with sequences 167_QPDLN for the *Bacillus cereus* oligo-1,6-glucosidase [28] and 295_MPKLN for the *B. stearothermophilus* neopullulanase [75]. As specificities possessing similar motifs, α-glucosidase, dextran glucosidase, trehalose-6-phosphate hydrolase, amylosucrase, sucrose phosphorylase, isomaltulose synthase and trehalose synthase in addition to oligo-1,6-glucosidase, and cyclomaltodextrinase, maltogenic amylase in addition to neopullulanase were recognised. The C-terminal parts of two exceptionally long sequences (*Thermus aquaticus* trehalose synthase [62] residues 565–963, and *Clostridium acetobutylicum* ATCC-824 α-amylase [80] residues 541–760) were disregarded.

(2) In the second step, attention was given to those sequences that were experimentally proven. In other words, all sequences of putative and hypothetical proteins from various complete-genome sequencing projects were disregarded despite their apparently convincing sequence similarity to experimentally confirmed members of the α-amylase family.

(3) In the next step, BLAST tools [90] were used to find sequences (especially from the α-amylase family) that might be closely related to oligo-1,6-glucosidase and neopullulanase via the fifth conserved sequence region with the previously identified intermediary motif MPDLN [91]. Thus seven sequences were added, all designated as α-amylases. However, since the sequence of the 'α-amylase' from *C. acetobutylicum* DSM 792 [92] contains a histidine in the position corresponding to catalytic β4 strand Asp206 from Taka-amylase A, it was not included in the final set. The remaining six sequences were marked as an 'intermediary group'.

(4) BLAST was also applied for finding the sequences (especially from the α-amylase family) that would have the sequence motif characteristic of either oligo-1,6-glucosidase (QpDln) or neopullulanase (MpKln). This search yielded four sequences with a motif similar to that of neopullulanase (two pullulan hydrolases, one α-amylase and one sequence without assigned specificity). These four sequences were referred to as 'neopullulanase-like'.

(5) Finally, the odd sequence of trehalose synthase from *Pseudomonas stutzeri* [GenBank Acc. No. AF113617] was eliminated from further analysis. It contains many longer insertions in comparison with other trehalose synthases, a strange motif even in the fifth conserved sequence region (259_QPSLN), as well as trehalose-synthase-unlike features in other conserved sequence regions.

All sequence alignments were performed using the program CLUSTAL W [93] and subsequently manually tuned where applicable. The neighbour-joining method was used for building the evolutionary trees [94]. The Phylip format tree output was applied using the bootstrapping procedure [95]; the number of bootstrap trials used was 1000. The trees were drawn with the program TreeView [96].

The three-dimensional structures of oligo-1,6-glucosidase from *B. cereus*, neopullulanase TVAII from *Thermoactinomyces vulgaris* and the high-pI α-amylase isozyme from barley were retrieved from the Protein Data Bank [97] under the PDB entry codes: 1UOK [98], 1BVZ [99] and 1AMY [100], respectively. The protein structures were displayed using the program WebLab-ViewerLite (Molecular Simulations).

Results and discussion

The fifth conserved sequence region as a definition marker

In this study, 79 amino acid sequences of several related enzyme specificites (table 1) from the α-amylase family were compared. The main goal of the present work was to use the fifth conserved sequence region (173_LPDLD in Taka-amylase A) for defining the two closely related subfamilies, i.e. the oligo-1,6-glucosidase subfamily and the neopullulanase subfamily. As can be seen from fig. 1, the sequence of the fifth conserved region is highly specific for both subfamilies with QpDln for the oligo-1,6-glucosidase subfamily and MPKln for the neopullulanase subfamily. These short sequence stretches are proposed as the identification and selection markers or sequence fingerprints for each of the subfamilies.

With regard to enzyme specificities, the oligo-1,6-glucosidase subfamily consists at present of oligo-1,6-glucosidase (for EC numbers, see table 1), α-glucosidase, dextran glucosidase, trehalose-6-phosphate hydrolase, amylosucrase, sucrose phosphorylase, isomaltulose synthase and trehalose synthase. All these enzymes contain in the fifth conserved region the sequence that corresponds to the template sequence QpDln. Worth mentioning is that amylosucrase and sucrose phosphorylases have a hydrophobic residue (tryptophan or isoleucine) in place of the well-conserved proline that is only in five cases re-

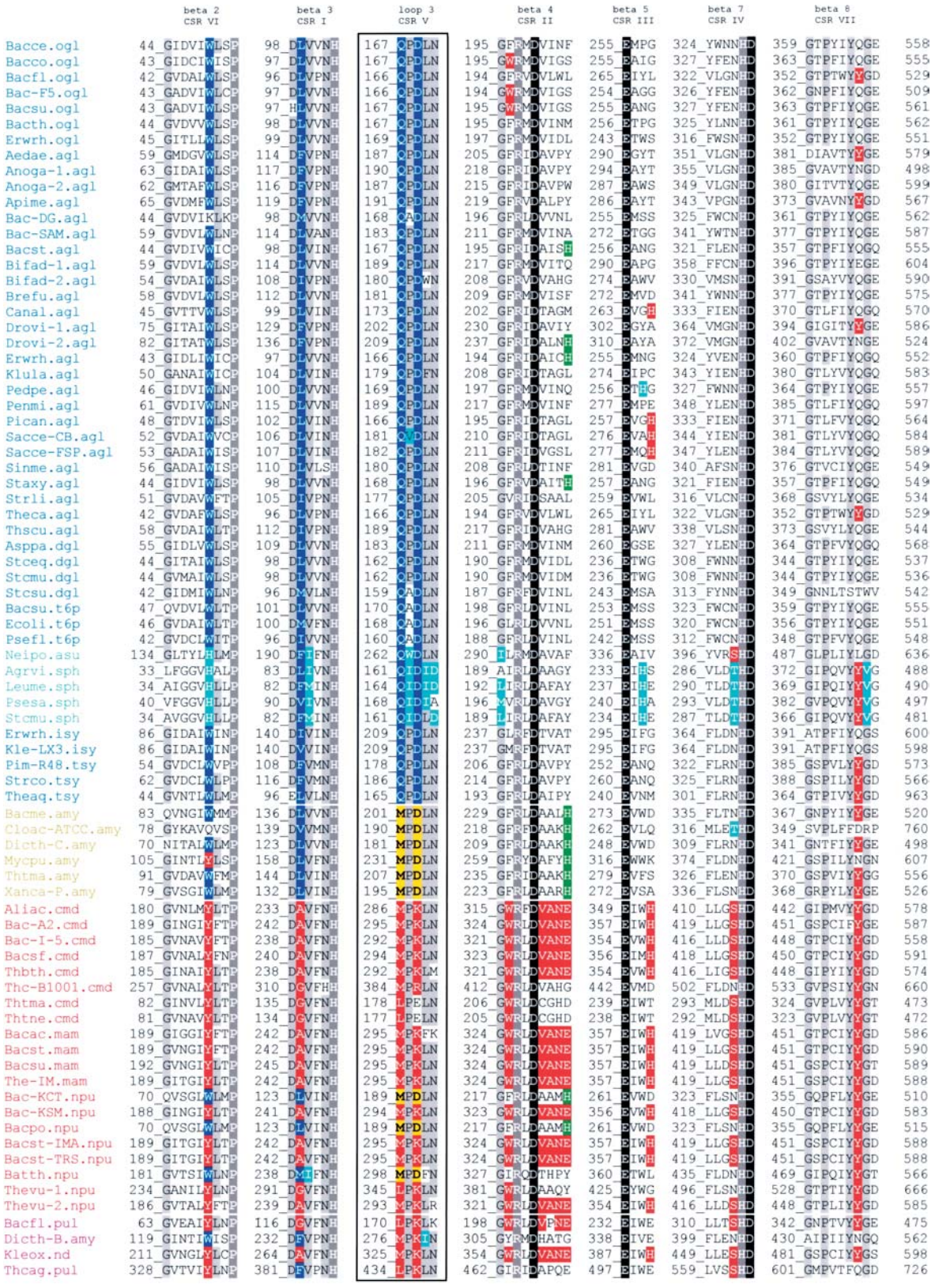

Figure 1. The conserved sequence regions (CSRs) of oligo-1,6-glucosidase and neopullulanase subfamilies from the α-amylase family. The abbreviations of enzyme sources are given in table 1. The best conserved parts of the sequence of an α-amylase family member comprise the strands β2, β3, β4, β5, β7 and β8 as well as the short stretch located near the C terminus of loop 3 connecting strand β3 and helix α3. The latter fifth conserved sequence region is central to the present study and is therefore boxed. The colour code for the enzymes is as follows: oligo-1,6-glucosidase subfamily – blue; neopullulanase subfamily – red; indermediary group – dark yellow; neopullulanase-like members – pink, transferases – turquoise. The sequence features are highlighted as follows: oligo-1,6-glucosidase subfamily specific – blue; neopullulanase subfamily specific – red; transferase specific – turquoise; intermediary group specific – yellow (in CSR V) and green (in CSR II); three catalytic residues – black and white inversion; invariant residues – dark-grey and white inversion; conserved residues (present in at least 75 % of the sources) – light grey.

placed by alanine (mostly in trehalose-6-phophate hydrolases). A hydrophobic residue (valine) in this position was furthermore found only (fig. 1) in the α-glucosidase from *Saccharomyces cerevisiae* strain CB11 [44].

As far as the enzyme specificities from the neopullulanase subfamily are concerned, at present this subfamily is formed by three specificities: cyclomaltodextrinase, maltogenic amylase and neopullulanase (for EC numbers, see table 1). The sequence of these enzymes in the fifth conserved region corresponds satisfactorily with the template MPKln. Although in a few cases, the methionine can be substituted with leucine and the lysine can be replaced by arginine (fig. 1), both changes preserve the chemical nature of the amino acid residue side chain.

The cyclomaltodextrinases from *Thermotoga maritima* [67] and *Thermotoga neapolitana* [GenBank Acc. No. AJ009832] contain a glutamate (LPELN; fig. 1) instead of lysine, a fact that could make these two enzymes more similar to the oligo-1,6-glucosidase subfamily with a dominating aspartate. The cyclomaltodextrinase from *T. maritima* was biochemically characterised as a novel type of this enzyme specificity [101]. Its sequence [67] shares more than 88% sequence identity with that of the *T. neapolitana* enzyme [GenBank Acc. No. AJ009832]. In most of the remaining conserved sequence regions, however, these two *Thermotoga* cyclomaltodextrinases exhibit the features of the neopullulanase subfamily (fig. 1) and also contain the neopullulanase-like shorter domain B (for details, see below). Since they lack the N-terminal domain (characteristic of the true members of the neopullulanase subfamily), both these enzymes could represent a new type of cyclomaltodextrinase and thus should not be classified into the neopullulanase subfamily.

Remarkably, three enzymes designated as neopullulanases exhibit an intermediary sequence in the fifth conserved region (MPDlN): *Bacillus* sp. KCTC8848P [71], *B. polymyxa* [73] and *Bacteroides thetaiotaomicron* [76]. The neopullulanase from *B. polymyxa* was reported as a 'novel' neopullulanase [102] that produced only panose as a final product from pullulan hydrolysis unlike the typical neopullulanase described originally by Imanaka and Kuriki [103]. From the sequence point of view also, it was found [73] to be most similar to α-amylases from *Bacillus megaterium* [79] and *Dictyoglomus thermophilum* AmyC [81] both of which are proposed in this work to constitute an 'intermediary group' (see below). This should also be the case of *Bacillus* sp. KCTC8848P neopullulanase with 92% sequence identity with the *B. polymyxa* enzyme [71]. The neopullulanase activity of the enzyme from *B. thetaiotaomicron* [76] was assigned based on the production of panose from pullulan but the activity of this enzyme against pullulan was approximately equal to its activity against amylose [104].

The sequence MPDLN mentioned above is exclusively characteristic of several α-amylases constituting the 'intermediary group'. The intermediary character of the sequence MPDLN is in the replacement of oligo-1,6-glucosidase-like glutamine by the neopullulanase-like methionine, whereas the aspartate is conserved as in the oligo-1,6-glucosidase subfamily, i.e. not substituted by the neopullulanase-like lysine (fig. 1). This is consistent with experimental findings that some of these 'α-amylases' appear to have the mixed substrate specificity of α-amylase, cyclomaltodextrinase and neopullulanase [84, 105]. The α-amylase from *T. maritima* was found, however, to be active against soluble starch and pullulan in the ratio 100 : 4 [83], indicating that its neopululanase activity is very low.

In this respect, the two α-amylases from *D. thermophilum*, designated AmyC and AmyB [81], containing MPDLN and MPKIN, respectively (fig. 1), should be of great interest. Unfortunately, one cannot say unambiguously whether or not these sequence features correlate with the enzyme specificities, due to the lack of specificity analysis data, since both *Dictyoglomus* α-amylases were tested on soluble starch only [81]. Nevertheless, the MPKIN sequence of AmyB indicates that this *Dictyoglomus* 'α-amylase' could rank among the neopullulanase-like members together with the CymH protein from *Klebsiella oxytoca* with as yet not determined specificity [86] having MPKLN (fig. 1). Finally, two interesting, extremely thermostable pullulanases, those from *Bacillus flavocaldarius* [85] and *Thermococcus aggregans* [87], with LPKLK and LPKLN, respectively, are also included in the group of the so-called 'neopullulanase-like' enzymes (fig. 1).

All these proposals concerning the definition of the oligo-1,6-glucosidase subfamily, the neopullulanase subfamily and the intermediary group, based on the sequence fingerprint of the fifth conserved region, can be supported by the following analysis of the remaining parts of the amino acid sequences.

Subfamily-associated features in the remaining parts of the amino acid sequences

In this section, we will focus on sequence features characteristic of the individual subfamilies and groups that are present in the other conserved sequence regions (fig. 1) as well as in the remaining segments of the complete sequence alignment (not shown).

Thus, in the conserved region VI (strand β2) there is a tryptophan (Trp49 in *B. cereus* oligo-1,6-glucosidase) characteristic for the oligo-1,6-glucosidase subfamily in the *i-3* position with respect to the conserved C-terminal proline (fig. 1). The neopullulanase subfamily has a tyrosine (Tyr191 in *T. vulgaris* neopullulanase TVAII) in that position (fig. 1). Importantly, the three neopullulanases with the intermediary sequence in the fifth conserved region (Bac-KCT.npu, Bacpo.npu and Batth.npu) do not contain the neopullulanase-like tyrosine but an oligo-1,6-

glucosidase-like tryptophan similar to most of the enzymes from the intermediary group (fig. 1). A glutamine residue present in this position in the α-amylase from *C. acetobutylicum* ATCC824 is a conserved feature of the so-called animal group of α-amylases [11]. Of the four enzymes from the neopullulanase-like group, only the 'α-amylase' AmyB from *D. thermophilum* has in this position tyrosine replaced by the tryptophan (fig. 1). Of interest is that in the frame of the oligo-1,6-glucosidase subfamily, a group of transferases (EC 2: amylosucrase and sucrose phosphorylases; see table 1) exhibits its own sequence features. All these enzymes have only histidine in the Trp/Tyr position (fig. 1).

A sequence feature discriminating the two subfamilies can also be found in the conserved region I (strand β3). There is a hydrophobic residue (Leu99 in *B. cereus* oligo-1,6-glucosidase) versus alanine (glycine) in the oligo-1,6-glucosidase versus neopullulanase subfamilies, respectively, in the position succeeding the almost invariant N-terminal aspartate (fig. 1). Again, the three neopullulanases (Bac-KCT.npu, Bacpo.npu and Batth.npu) with leucine and methionine fulfil the criteria of the intermediary group or, in a wider sense, the entire oligo-1,6-glucosidase subfamily. The fact that the four neopullulanase-like enzymes contain hydrophobic phenylalanine (Dicth-B.amy and Thcag.pul) as well as the alanine and glycine (Bacfl.pul and Kleox.nd) supports the contention that they could also have an intermediary character. As far as the transferases are concerned, in the succeeding position, that is throughout hydrophobic, they have either isoleucine or methionine in comparison with an almost invariantly conserved valine (fig. 1). Note that the isoleucine is also found in the 'odd' neopullulanase from *B. thetaiotaomicron* (fig. 1).

In conserved sequence region II (strand β4), there are a few features characteristic of the neopullulanase subfamily. These are the tryptophan (Trp322 in *T. vulgaris* neopullulanase TVAII) and the stretch VANE at the C terminus of this region (fig. 1). Both these features are not consistently present in all neopullulanases. However, the mutant TVAI enzyme from *T. vulgaris* whose AAQY stretch is substituted by VANE (as in the TVAII enzyme; cf. fig. 1), in contrast to expectations, did not exhibit the anticipated neopullulanase-like properties [106]. Note that some amylopullulanases also contain at the C-terminal end of this conserved region the stretch VANE or VENE but they have a completely different sequence in their fifth conserved region [S. Janecek, unpublished results]. What is however more important is that the three neopullulanases discussed above (Bac-KCT.npu, Bacpo.npu and Batth.npu) lack both the tryptophan and the VANE stretch completely (fig. 1). Moreover, two of the three (Bac-KCT.npu, Bacpo.npu) contain at the C termini of this region, a histidine that is characteristic for many α-amylases and several other specificities from the α-amylase family [1] and which is conserved also in the intermediary group shown in fig. 1. This histidine is found in a few α-glucosidases (Bacst.agl, Drovi-2.agl, Erwrh.agl and Staxy.agl) as is the neopullulanase-like tryptophan in oligo-1,6-glucosidases from *Bacillus coagulans*, *Bacillus* sp. strain F5 and *Bacillus subtilis* (fig. 1). The group of transferases may be distinguished from the rest of the enzymes again by the hydrophobic residue replacing the N-terminal conserved glycine (except for the sucrose phosphorylase from *Agrobacterium vitis*; fig. 1).

Concerning the conserved sequence region III (strand β5), a C-terminal histidine (*i+3* position from the catalytic glutamate; Glu354 in *T. vulgaris* neopullulanase TVAII) can perhaps be considered to be the feature of the neopullulanase subfamily, but its presence in neopullulanases as well as in the neopullulanase-like group is not convincing (fig. 1). As could be expected from the previous analysis, the three – from the sequence point of view – not unambiguous neopullulanases (Bac-KCT.npu, Bacpo.npu and Batth.npu) do not have this histidine. It is, on the other hand, found in four yeast α-glucosidases (Canal.agl, Pican.agl, Sacce-CB.agl and Sacce-FSP.agl; fig. 1). Of the five transferases, all four sucrose phosphorylases (i.e. except for the amylosucrase from *Neisseria polysaccharea*) contain histidine in the *i+2* position from the catalytic glutamate; Glu255 in *B. cereus* oligo-1,6-glucosidase). A corresponding histidine is found only in the α-glucosidase from *Pediococcus pentosaceus*.

A similar situation is found for the last two conserved sequence regions, region IV (strand β7) and region VII (strand β8). There is a serine preceding the invariant C-terminal dipeptide HD (His420-Asp421 in *T. vulgaris* neopullulanase TVAII) in the region of strand β7 and a tyrosine (Tyr454 in *T. vulgaris* neopullulanase TVAII) in the region of strand β8 (fig. 1) that could be ascribed to the neopullulanase subfamily. These two residues, like the C-terminal histidine from the previously conserved sequence region of strand β5, are not exclusively present in the neopullulanase subfamily members (fig. 1) and thus cannot be used as specificity markers. Interestingly, while in the conserved region VI (strand β2), the intermediary group contains rather the feature of the oligo-1,6-glucosidase subfamily, in conserved region VII (strand β8), this group behaves like the members of the neopullulanase subfamily (fig. 1). The four sucrose phosphorylases again have their own features in these two regions: a threonine in β7 and a valine in β8 (Thr289 and Val379 in *A. vitis* sucrose phosphorylase). Of these, only the β7 strand threonine can be found in one of the remaining enzymes, namely in the α-amylase from *C. acetobutylicum* ATCC824 (fig. 1).

With regard to the five transferases (one amylosucrase and four sucrose phosphorylases; see table 1), their specific features can also be traced in the fifth conserved sequence region that is used as a definition marker in this

study. Although they seem to belong to the oligo-1,6-glucosidase subfamily satisfying the sequence criterion QxDln, they have a hydrophobic residue at position 'x' (amylosucrase – tryptophan; sucrose phosphorylases – isoleucine) and the dipeptide 'ln' has been changed to 'id' (fig. 1). Note that from the entire set of 79 sequences studied here, there are only two that are partly similar in this respect: the α-glucosidase from *S. cerevisiae* CB11 contains a valine in position 'x' whereas the α-amylase AmyB from *D. thermophilum* has an isoleucine in the position of 'l' (fig. 1).

As far as the remaining parts of the amino acid sequence alignment are concerned (not shown), several differences can be found that discriminate the two subfamilies from each other. Thus, at the N-terminal part of the alignment, the neopullulanase subfamily enzymes contain an alignable segment that has no equivalent among the oligo-1,6-glucosidase subfamily enzymes. This segment corresponds to the N-terminal domain, preceding the catalytic $(\beta/\alpha)_8$ barrel, found in three-dimensional structures of neopullulanase TVAII from *T. vulgaris* [99] and *Thermus* maltogenic amylase [107]. In agreement with observations mentioned above, the 'neopullulanases' from *B. polymyxa* [73] and *Bacillus* sp. KCTC8848P [71], lacking the N-terminal domain, behave like the members of the oligo-1,6-glucosidase subfamily or, at least, the enzymes from the intermediary group. The same is true for the *T. maritima* and *T. neapolitana* cyclomaltodextrinases [67; GenBank Acc. No. AJ009832] that were also discussed above as exhibiting the neopullulanase-unlike sequence feature in the fifth conserved region (LPELN; fig. 1). The 'odd' neopullulanase from *B. thetaiotaomicron* [76] seems to contain the N-terminal segment; however, its sequence is evidently different from those of true neopullulanase subfamily enzymes. Of the four neopullulanase-like group enzymes (table 1), the pullulanase from *B. flavocaldarius* [85] lacks the N-terminal segment, while the other three have some N-terminal sequence. Of these, only the N-terminal segment of the *K. oxytoca* CymH protein [86] agrees well with the equivalent parts of the enzymes from the neopullulanase subfamily.

The other significant sequence feature distinguishing the two subfamilies can also be seen in domain B. The enzymes belonging to the oligo-1,6-glucosidase subfamily should share the structure of domain B from *B. cereus* oligo-1,6-glucosidase [98], i.e. one α helix and a three-stranded antiparallel β sheet, whereas the enzymes belonging to the neopullulanase subfamily seem to lack the second β strand (153_WQYD in *B. cereus* oligo-1,6-glucosidase) from the antiparallel β sheet, as indicated previously [20]. There are, however, a few enzymes originally designated as cyclomaltodextrinase and neopullulanase that do contain the domain B strand β2. These are again the three 'odd' neopullulanases (Bac-KCT.npu, Bacpo.npu and Batth.npu) and very probably the cyclomaltodextrinase from *Thermococcus* sp. B1001 [66]. Remarkably, the CymH protein from *K. oxytoca*, that looks in the other parts of its sequence like a neopullulanase subfamily enzyme, shares unambiguously the structure of domain B from an oligo-glucosidase, i.e. it contains the second β strand.

The transferases (EC 2; one amylosucrase and four sucrose phosphorylases, table 1) seem to constitute an independent group in the frame of the oligo-1,6-glucosidase subfamily. This was shown in the above discussion concerning the similarities and differences in conserved sequence regions (fig. 1) and is also clear when comparing the entire sequences. For example, there is a strongly conserved tyrosine position (Tyr63 in *B. cereus* oligo-1,6-glucosidase), which is also conserved in the neopullulanase subfamily, but which only in the four sucrose phosphorylases is substituted by a phenylalanine (not shown). In domain B, structurally at least, the amylosucrase should share the structure of the *B. cereus* domain B with the three-stranded antiparallel β sheet [108]. The amylosucrase from *N. polysaccharea* [108], moreover, contains a domain B′ inserted between the seventh β strand and seventh α helix. Interestingly, all four sucrose phosphorylases have a segment of comparable length inserted in that part of the sequence, but these insertions are not sequentially similar to domain B′ of amylosucrase (not shown). The difference between the amylosucrase and sucrose phosphorylases seems to be located at the N-terminal end where the amylosucrase has an N-terminal domain [108] while the sucrose phosphorylases seem to start directly with the catalytic $(\beta/\alpha)_8$ barrel domain (alignment not shown).

Tertiary structure comparison

The sequence changes in the fifth conserved region analysed above (fig. 1) are reflected in the tertiary structures of these enzymes. Figure 2 shows the situation concerning the calcium-binding aspartate from the fifth conserved sequence region (Asp175 in Taka-amylase A) and its equivalents in the counterpart enzymes from the oligo-1,6-glucosidase and neopullulanase subfamilies. In the oligo-1,6-glucosidase subfamily, based on the three-dimensional structure of *B. cereus* oligo-1,6-glucosidase [98], the aspartate is conserved (fig. 1) but the Ca^{2+} ion, present in α-amylases (fig. 2A), is absent. It is replaced with a presumably protonated Nζ atom of lysine (Lys206 in *B. cereus* oligo-1,6-glucosidase; fig. 2B). A similar architecture was proposed in amylosucrase [108]. On the other hand, in the neopullulanase subfamily, based on the three-dimensional structure of *T. vulgaris* neopullulanase [99], the aspartate is replaced by lysine (fig. 1) and the Ca^{2+} ion is absent as expected. In this case, the Nζ atom of the lysine substituting the aspartate (Lys295 in *Thermoactinomyces vulgaris* neopullulanase) directly occupies the Ca^{2+} position (fig. 2C).

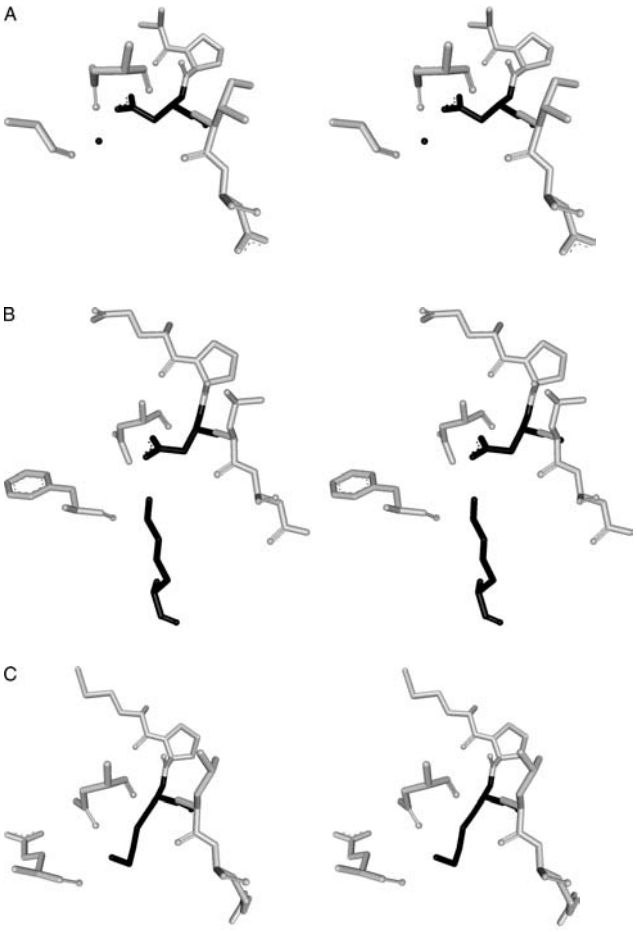

Figure 2. The arrangement of the selected residues involved in binding of a calcium ion in barley α-amylase (*A*) and the corresponding residues in *Bacillus cereus* oligo-1,6-glucosidase (*B*) and *Thermoactinomyces vulgaris* TVAII neopullulanase (*C*). The calcium ion in *A* is shown as a black sphere. For illustration, the entire fifth conserved sequence region with conserved aspartate (*A, B*) (coloured black) and substituted lysine (*C*) (coloured black) is depicted. While in α-amylase, the aspartate is involved in Ca^{2+} binding, there is no calcium in either oligo-1,6-glucosidase and neopullulanase. Since in the oligo-1,6-glucosidase (subfamily), the aspartate is conserved, the calcium position is occupied by a binding atom of another residue; Nζ atom of Lys206 in the particular case shown in *B* (coloured black). In neopullulanase (subfamily), the aspartate is substituted by a residue with the longer side chain of Lys or Arg, and the calcium position is occupied directly by a binding atom of that residue, the Nζ atom of Lys295 in the particular case shown in *C*.

Evolutionary relationships

The evolutionary relationships among all 79 enzymes studied in this work are shown in figure 3. Both trees clearly reflect the existence of the two subfamilies in the frame of the α-amylase family that was postulated above. Regardless of whether the tree was based on the alignment of complete amino acid sequences (fig. 3 A) or conserved sequence regions only (fig. 3 B), the basic arrangement of both trees is very similar. The two subfamilies form their own large clusters with a separated intermediary group and the group of transferases.

In the neopullulanase subfamily part of the tree, the individual enzyme specificities belonging to this subfamily are more or less indistinguishable from each other [1, 24, 109–111] because there are no special branches leading separately to cyclomaltodextrinases, maltogenic amylases and neopullulanases.

Of the four neopullulanase-like enzymes (table 1), three (Kleox.nd, Bacfl.pul and Thcag.pul) can be found among the neopullulanase subfamily members. The most convincing example is the CymH protein from *K. oxytoca* [86], while the less convincing one seems to be the unique pullulanase from *T. aggregans* [87]. Indeed, the latter, due to its unique action on pullulan, has been named as pullulan hydrolase type III [87]. Its position on a quite long separate branch in both trees (fig. 3) indicates that, although this enzyme may be closely related to the neopullulanase subfamily, it nevertheless retains its own uniqueness. The enzyme from *B. flavocaldarius* designated as a pullulanase [85, 112], placed in fig. 3 A on a branch adjacent to the cyclomaltodextrinase from *Alicyclobacillus acidocaldarius* [24], is, in the tree based on the conserved sequence regions (fig. 3 B), on a separate branch. The two *Thermotoga* cyclomaltodextrinases, which simultaneously have the fifth conserved sequence region LPELN, lack the N-terminal domain, contain a shorter domain B and exhibit the neopullulanase-like sequence features (fig. 1), are positioned near the border of the neopullulanase subfamily cluster in both trees (fig. 3). Thus they very probably do not belong to the true neopullulanase subfamily. The 'intermediary' positions of the *Bacillus flavocaldarius* pullulanase (fig. 3) reflect also both the presence of the neopullulanase-like sequence features (fig. 1) and the lack of the N-terminal domain characteristic for true neopullulanases (data not shown) [85, 109]. The fourth neopullulanase-like enzyme, the 'α-amylase' AmyB from *D. thermophilum* [81], is positioned in both trees outside the neopullulanase-subfamily part (fig. 3). Its position among the intermediary group (fig. 3 A) seems to be more convincing since it reflects the similarities and differences over the entire amino acid sequence.

The positions in both trees of the three interesting 'neopullulanases' from *Bacillus* sp. KCTC8848P [71], *B. polymyxa* [73] and *B. thetaiotaomicron* [76] deserve special interest. All of these are unambiguously placed outside the neopullulanase subfamily part of the trees, a fact that is in agreement with the analysis of their amino acid sequences given in the previous section. While the enzyme from *B. thetaiotaomicron* occupies slightly different locations in the trees (compare fig. 3 A and B), the two *Bacillus* 'neopullulanases' (Bac-KCT.npu and Bacpo.npu) go well together with each other and with the 'α-amylases' from the intermediary group (fig. 3).

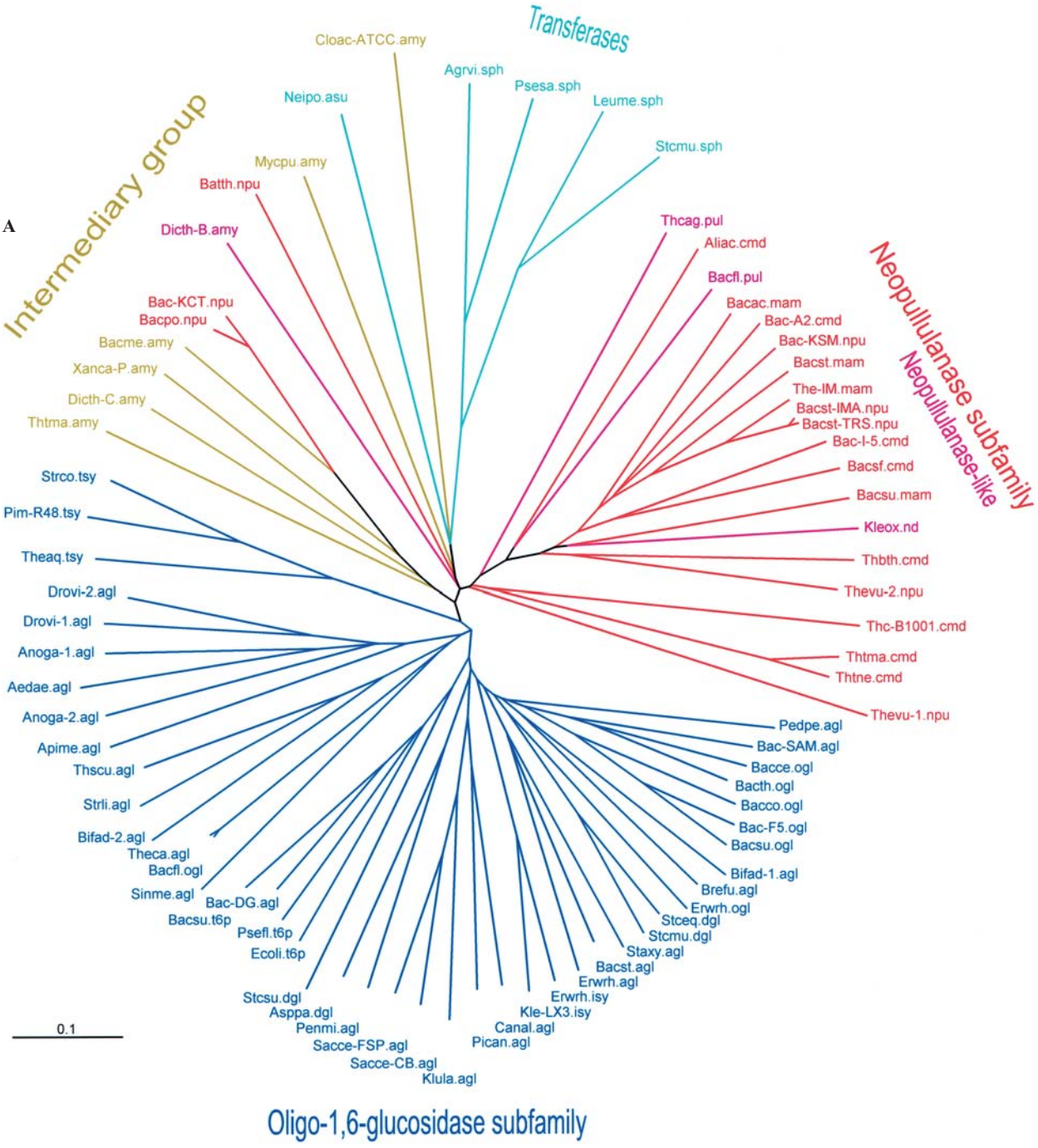

Figure 3. Evolutionary trees of oligo-1,6-glucosidase and neopullulanase subfamilies from the α-amylase family. The abbreviations of enzyme sources are given in table 1. The trees are based on the alignment of complete sequences (*A*) and conserved sequence regions shown in figure 1 (*B*). The branch lengths are proportional to the sequence divergence. The colour code for the enzymes is as follows: oligo-1,6-glucosidase subfamily – blue; neopullulanase subfamily – red; intermediary group – dark yellow; neopullulanase-like members – pink, transferases – turquoise.

B

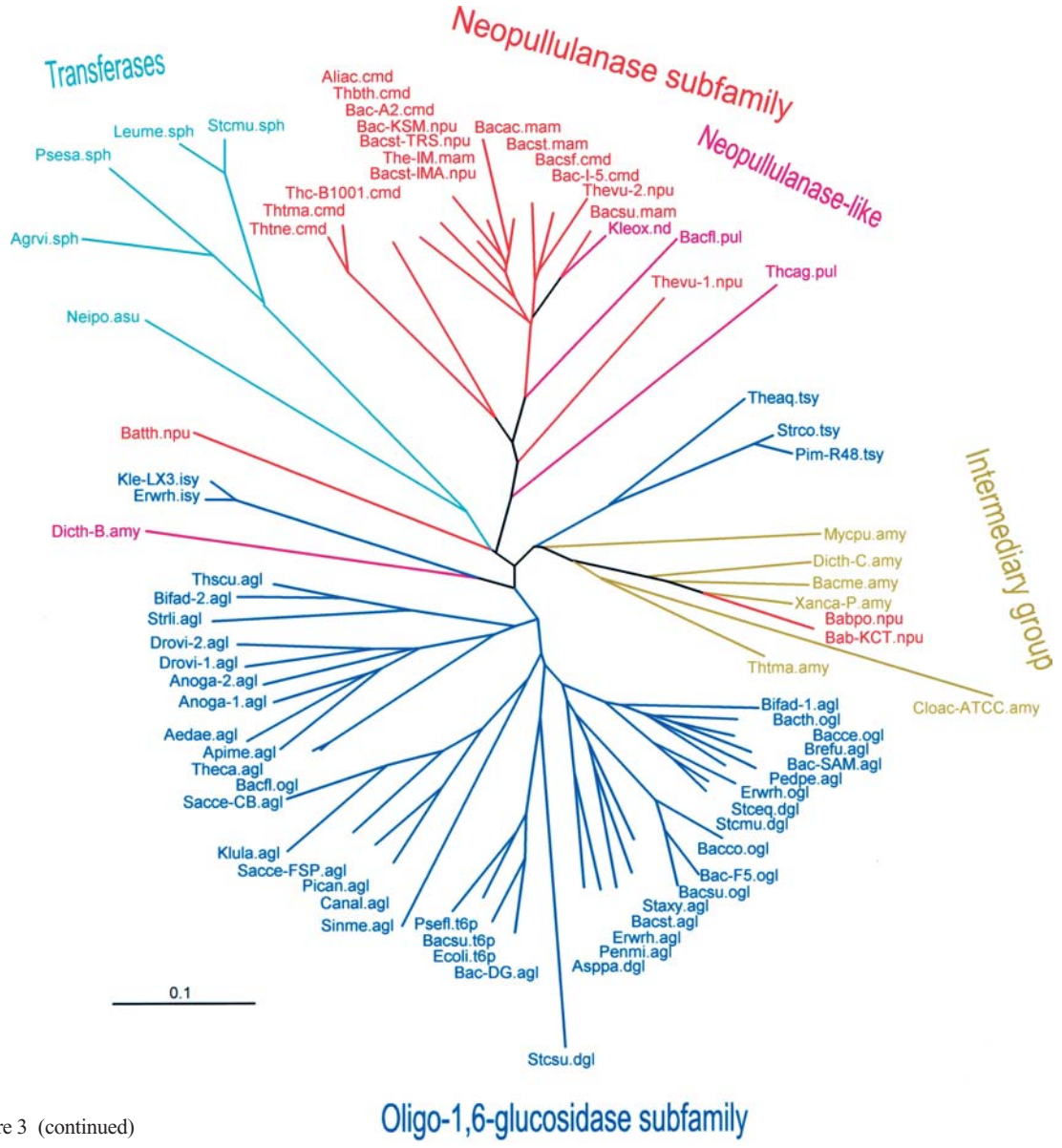

Figure 3 (continued)

The oligo-1,6-glucosidase subfamily parts of both evolutionary trees are compact, similar to the neopullulanase subfamily parts. The detailed subtle differences in clustering of the individual enzyme specificities are not so important, and the entire subfamily behaves as one larger cluster. Nevertheless, a few groups can be found as conserved in both trees, e.g. the insect and yeast α-glucosidases (fig. 3). The compactness of the oligo-1,6-glucosidase subfamily is especially evident in the tree based on the alignment of complete sequences (fig. 3A). In the tree based on the alignment of conserved sequence regions (fig. 3B), the intermediary group was inserted between the trehalose synthases and the rest of the enzymes from the oligo-1,6-glucosidase subfamily. This may indicate that the similarities observed in the framework of isolated, although well-conserved and functionally important sequence stretches might not be extendable into a generalisation for the remaining parts of the amino acid sequences.

The group of transferases (EC 2) is also worth mentioning. Although these enzymes are proposed to belong to the oligo-1,6-glucosidase subfamily, their specific sequence features (fig. 1) discriminating them from the rest of the enzyme specificities from the subfamily were discussed in the previous section. This is clearly reflected in both evolutionary trees, where transferases form, in fact, their own clusters (fig. 3), the sucrose phosphorylase and amylosucrase specificities being clustered separately from each other.

Conclusions

To summarise, this work describes the definition of two subfamilies in the framework of the α-amylase family: the oligo-1,6-glucosidase subfamily and the neopullulanase subfamily. They are defined based on the sequence of the fifth conserved sequence region, i.e. identification marker QpDln for the oligo-1,6-glucosidase subfamily and MPKln for the neopullulanase subfamily. The region can simultaneously be used as a marker distinguishing the two subfamilies from each other, i.e. as a selection marker. The sequence MPDLN is proposed as characteristic of the so-called intermediary group with mixed enzyme specificity. The subfamily-associated sequence features are also found in the other conserved sequence regions. The evolutionary trees support the proposed existence of the two subfamilies.

Acknowledgements. This work was financially supported in part by VEGA grant No. 2/2057/22 from the Slovak Grant Agency for Science.

1 MacGregor E. A., Janeček Š. and Svensson B. (2001) Relationship of sequence and structure to specificity in the α-amylase family of enzymes. Biochim. Biophys. Acta **1546:** 1–20
2 Coutinho P. M. and Henrissat B. (1999) Carbohydrate-active enzymes, http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html
3 Kuriki T. and Imanaka T. (1999) The concept of the α-amylase family: structural similarity and common catalytic mechanism. J. Biosci. Bioeng. **87:** 557–565
4 Matsuura Y., Kusunoki M., Harada W. and Kakudo M. (1984) Structure and possible catalytic residues of Taka-amylase A. J. Biochem. **95:** 697–702
5 Janeček Š. (2000) Proteins without enzymatic function with sequence relatedness to the α-amylase family. Trends Glycosci. Glycotechnol. **12:** 363–371
6 Toda H., Kondo K. and Narita K. (1982) The complete amino acid sequence of Taka-amylase A. Proc. Jpn Acad. **B58:** 208–212
7 Friedberg F. (1983) On the primary structure of amylases. FEBS Lett. **152:** 139–140
8 Rogers J. C. (1985) Conserved amino acid sequence domains in α-amylases from plants, mammals, and bacteria. Biochem. Biophys. Res. Commun. **128:** 470–476
9 Nakajima R., Imanaka T. and Aiba S. (1986) Comparison of amino acid sequences of eleven different α-amylases. Appl. Microbiol. Biotechnol. **23:** 355–360
10 Janeček Š. (1992) New conserved amino acid region of α-amylases in the third loop of their $(\beta/\alpha)_8$-barrel domains. Biochem. J. **288:** 1069–1070
11 Janeček Š. (1994) Sequence similarities and evolutionary relationships of microbial, plant and animal α-amylases. Eur. J. Biochem. **224:** 519–524
12 Janeček Š. (1994) Parallel β/α-barrels of α-amylase, cyclodextrin glycosyltransferase and oligo-1,6-glucosidase versus the barrel of β-amylase: evolutionary distance is a reflection of unrelated sequences. FEBS Lett. **353:** 119–123
13 Janeček Š. (1995) Close evolutionary relatedness among functionally distantly related members of the $(\alpha/\beta)_8$-barrel glycosyl hydrolases suggested by the similarity of their fifth conserved sequence region. FEBS Lett. **377:** 6–8
14 Janeček Š., MacGregor E. A. and Svensson B. (1995) Characteristic differences in the primary structure allow discrimination of cyclodextrin glucanotransferases from α-amylases. Biochem. J. **305:** 685–686
15 Janeček Š., Leveque E., Belarbi A. and Haye B. (1999) Close evolutionary relatedness of α-amylases from archaea and plants. J. Mol. Evol. **48:** 421–426
16 Boel E., Brady L., Brzozowski A. M., Derewenda Z., Dodson G. G., Jensen V. J. et al. (1990) Calcium binding in α-amylases: an X-ray diffraction study at 2.1-Å resolution of two enzymes from *Aspergillus*. Biochemistry **29:** 6244–6249
17 Jespersen H. M., MacGregor E. A., Henrissat B., Sierks M. R. and Svensson B. (1993) Starch- and glycogen-debranching and branching enzymes: prediction of structural features of the catalytic $(\beta/\alpha)_8$-barrel domain and evolutionary relationship to other amylolytic enzymes. J. Protein Chem. **12:** 791–805
18 MacGregor E. A., Jespersen H. M. and Svensson B. (1996) A circularly permuted α-amylase-type α/β-barrel structure in glucan-synthesizing glucosyltransferases. FEBS Lett. **378:** 263–266
19 del-Rio G., Morett E. and Soberon X. (1997) Did cyclodextrin glycosyltransferases evolve from α-amylases? FEBS Lett. **416:** 221–224
20 Janeček Š., Svensson B. and Henrissat B. (1997) Domain evolution in the α-amylase family. J. Mol. Evol. **45:** 322–331
21 Da Lage J. L., Renard E., Chartois F., Lemeunier F. and Cariou M. L. (1998) *Amyrel*, a paralogous gene of the amy-lase gene family in *Drosophila melanogaster* and the *Sophophora* subgenus. Proc. Natl. Acad. Sci. USA **95:** 6848–6853
22 De Montalk G. P., Remaud-Simeon M., Willemot R. M., Planchot V. and Monsan P. (1999) Sequence analysis of the gene encoding amylosucrase from *Neisseria polysaccharea* and characterization of the recombinant enzyme. J. Bacteriol. **181:** 375–381
23 Jones R. A., Jermiin L. S., Easteal S., Patel B. K and Beacham I. R. (1999) Amylase and 16S rRNA genes from a hyperthermophilic archaebacterium. J. Appl. Microbiol. **86:** 93–107
24 Matzke J., Herrmann A., Schneider E. and Bakker E. P. (2000) Gene cloning, nucleotide sequence and biochemical properties of a cytoplasmic cyclomaltodextrinase (neopullulanase) from *Alicyclobacillus acidocaldarius*, reclassification of a group of enzymes. FEMS Microbiol. Lett. **183:** 55–61
25 Pujadas G. and Palau J. (2001) Evolution of α-amylases: architectural features and key residues in the stabilization of the $(\beta/\alpha)_8$ scaffold. Mol. Biol. Evol. **18:** 38–54
26 Nishizawa M., Ozawa F. and Hishinuma F. (1987) Molecular cloning of an amylase gene of *Bacillus circulans*. DNA **6:** 255–265
27 Bahl H., Burchhardt G., Spreinat A., Haeckel K., Wienecke A., Schmidt B. et al. (1991) α-Amylase of *Clostridium thermosulfurogenes* EM1: nucleotide sequence of the gene, processing of the enzyme, and comparison of other α-amylases. Appl. Environ. Microbiol. **57:** 1554–1559
28 Watanabe K., Kitamura K., Iha H. and Suzuki Y. (1990) Primary structure of the oligo-1,6-glucosidase of *Bacillus cereus* ATCC7064 deduced from the nucleotide sequence of the cloned gene. Eur. J. Biochem. **192:** 609–620
29 Watanabe K., Kitamura K. and Suzuki Y. (1996) Analysis of the critical sites for protein thermostabilization by proline substitution in oligo-1,6-glucosidase from *Bacillus coagulans* ATCC 7050 and the evolutionary consideration of proline residues. Appl. Environ. Microbiol. **62:** 2066–2073
30 Kashiwabara S., Matsuki Y., Kishimoto T. and Suzuki Y. (1998) Clustered proline residues around the active-site cleft in thermostable oligo-1,6-glucosidase of *Bacillus flavocaldarius* KP1228. Biosci. Biotechnol. Biochem. **62:** 1093– 1102
31 Yamamoto M. and Horikoshi K. (1990) Nucleotide sequence of alkalophilic *Bacillus* oligo-1,6-glucosidase gene and the properties of the gene product by *Escherichia coli* HB101. Denpun Kagaku **37:** 137–144

32 Watanabe K., Chishiro K., Kitamura K. and Suzuki Y. (1991) Proline residues responsible for thermostability occur with high frequency in the loop regions of an extremely thermostable oligo-1,6-glucosidase from *Bacillus thermoglucosidasius* KP1006. J. Biol. Chem. **266:** 24287–24294

33 Börnke F., Hajirezaei M. and Sonnewald U. (2001) Cloning and characterization of the gene cluster for palatinose metabolism from the phytopathogenic bacterium *Erwinia rhapontici*. J. Bacteriol. **183:** 2425–2430

34 James A. A., Blackmer K. and Racioppi J. V. (1989) A salivary gland-specific, maltase-like gene of the vector mosquito, *Aedes aegypti*. Gene **75:** 73–83

35 Zheng L., Whang L.H., Kumar V. and Kafatos F. C. (1995) Two genes encoding midgut-specific maltase-like polypeptides from *Anopheles gambiae*. Exp. Parasitol. **81:** 272– 283

36 Ohashi K., Sawata M., Takeuchi H., Natori S. and Kubo T. (1996) Molecular cloning of cDNA and analysis of expression of the gene for α-glucosidase from the hypopharyngeal gland of the honeybee *Apis mellifera* L. Biochem. Biophys. Res. Commun. **221:** 380–385

37 Lee Y. E. (2000) Cloning and characterization of α-glucosidase gene from thermophilic *Bacillus* sp. DG0303. J. Microbiol. Biotechnol. **10:** 244–250

38 Nakao M., Nakayama T., Kakudo A., Inohara M., Harada M., Omura F. et al. (1994) Structure and expression of a gene coding for thermostable α-glucosidase with a broad substrate specificity from *Bacillus* sp. SAM1606. Eur. J. Biochem. **220:** 293–300

39 Takii Y., Takahashi K., Yamamoto K., Sogabe Y. and Suzuki Y. (1996) *Bacillus stearothermophilus* ATCC12016 α-glucosidase specific for α-1,4 bonds of maltosaccharides and α-glucans shows high amino acid sequence similarities to seven α-D-glucohydrolases with different substrate specificity. Appl. Microbiol. Biotechnol. **44:** 629–634

40 Geber A., Williamson P. R., Rex J. H., Sweeney E. C. and Bennett J. E. (1992) Cloning and characterization of a *Candida albicans* maltase gene involoved in sucrose utilization. J. Bacteriol. **174:** 6992–6996

41 Vieira C. P., Vieira J. and Hartl D. L. (1997) The evolution of small gene clusters: evidence for an independent origin of the maltase gene cluster in *Drosophila virilis* and *Drosophila melanogaster*. Mol. Biol. Evol. **14:** 985–993

42 Garcia B., Castellanos A., Menendez J. and Pons T. (2001) Molecular cloning of an α-glucosidase-like gene from *Penicillium minioluteum* and structure prediction of its gene product. Biochem. Biophys. Res. Commun. **281:** 151–158

43 Liiv L., Parn P. and Alamae T. (2001) Cloning of maltase gene from a methylotrophic yeast, *Hansenula polymorpha*. Gene **265:** 77–85

44 Hong S. H. and Marmur J. (1986) Primary structure of the maltase gene of MAL6 locus of *Saccharomyces carlsbergensis*. Gene **41:** 75–84

45 Willis L. B. and Walker G. C. (1999) A novel *Sinorhizobium meliloti* operon encodes an α-glucosidase and a periplasmic-binding-protein-dependent transport system for α-glucosides. J. Bacteriol. **181:** 4176–4184

46 Egeter O. and Bruckner R. (1995) Characterization of a genetic locus essential for maltose-maltotriose utilization in *Staphylococcus xylosus*. J. Bacteriol. **177:** 2408–2415

47 Piendl W., Eichenseer C., Viel P., Altenbuchner J. and Cullum J. (1994) Analysis of putative DNA amplification genes in the element AUD1 of *Streptomyces lividans* 66. Mol. Gen. Genet. **244:** 439–443

48 Janda L., Damborsky J., Petricek M., Spizek J. and Tichy P. (2000) Molecular characterization of the *Thermomonospora curvata* aglA gene encoding a thermotolerant α-1,4-glucosidase. J. Appl. Microbiol. **88:** 773–783

49 Nashiru O., Koh S., Lee S. Y. and Lee D. S. (2001) Novel α-glucosidase from extreme thermophile *Thermus caldophilus* GK24. J. Biochem. Mol. Biol. **34:** 347–354

50 Yu J., Chang P., Bhatnagar D. and Cleveland T. E. (2000) Cloning of a sugar utilization gene cluster in *Aspergillus parasiticus*. Biochim. Biophys. Acta **1493:** 211–214

51 Mechold U., Steiner K., Vettermann S. and Malke H. (1993) Genetic organization of the streptokinase region of the *Streptococcus equisimilis* H46A chromosome. Mol. Gen. Genet. **241:** 129–140

52 Russell R. R. and Ferretti J. J. (1990) Nucleotide sequence of the dextran glucosidase (dexB) gene of *Streptococcus mutans*. J. Gen. Microbiol. **136:** 803–810

53 Serhir B., Dugourd D., Jacques M., Higgins R. and Harel J. (1997) Cloning and characterization of a dextranase gene (dexS) from *Streptococcus suis*. Gene **190:** 257–261

54 Helfert C., Gotsche S. and Dahl M. K. (1995) Cleavage of trehalose-phosphate in *Bacillus subtilis* is catalysed by a phospho-α-(1-1)-glucosidase encoded by the *treA* gene. Mol. Microbiol. **16:** 111–120

55 Boos W., Ehmann U., Forkl H., Klein W., Rimmele M. and Postma P. (1990) Trehalose transport and metabolism in *Escherichia coli*. J. Bacteriol. **172:** 3450–3461

56 Matthijs S. L. C., Koedam N., Cornelis P. and De Greve H. (2000) The trehalose operon of *Pseudomonas fluorescens* ATCC 17400. Res. Microbiol. **151:** 845–851

57 Fournier P., Ruffray P. de and Otten L. (1994) Natural instability of *Agrobacterium vitis* Ti plasmid due to unusual duplication of a 2.3-kb DNA fragment. Mol. Plant Microbe Interact. **7:** 164–172

58 Ferretti J. J., Huang T. T. and Russell R. R. (1988) Sequence analysis of the glucosyltransferase A gene (*gtfA*) from *Streptococcus mutans* Ingbritt. Infect. Immun. **56:** 1585–1588

59 Zhang D., Li X. and Zhang L. H. (2002) Isomaltulose synthase from *Klebsiella* sp. strain LX3: gene cloning and characterization and engineering of thermostability. Appl. Environ. Microbiol. **68:** 2676–2682

60 Tsusaki K., Nishimoto T., Nakada T., Kubota M., Chaen H., Sugimoto T. et al. (1996) Cloning and sequencing of trehalose synthase gene from *Pimelobacter* sp. R48. Biochim. Biophys. Acta **1290:** 1–3

61 Schneider D., Bruton C. J. and Chater K. F. (2000) Duplicated gene clusters suggest an interplay of glycogen and trehalose metabolism during sequential stages of aerial mycelium development in *Streptomyces coelicolor* A3(2). Mol. Gen. Genet. **263:** 543–553

62 Tsusaki K., Nishimoto T., Nakada T., Kubota M., Chaen H., Fukuda S. et al. (1997) Cloning and sequencing of trehalose synthase gene from *Thermus aquaticus* ATCC33923. Biochim. Biophys. Acta **1334:** 28–32

63 Kim T. J., Shin J. H., Oh J. H., Kim M. J., Lee S. B., Ryu S. et al. (1998) Analysis of the gene encoding cyclomaltodextrinase from alkalophilic *Bacillus* sp. I-5 and characterization of enzymatic properties. Arch. Biochem. Biophys. **353:** 221–227

64 Oguma T., Matsuyama A., Kikuchi M. and Nakano E. (1993) Cloning and sequence analysis of the cyclomaltodextrinase gene from *Bacillus sphaericus* and expression in *Escherichia coli* cells. Appl. Microbiol. Biotechnol. **39:** 197–203

65 Podkovyrov S. M. and Zeikus J. G. (1992) Structure of the gene encoding cyclomaltodextrinase from *Clostridium thermohydrosulfuricum* 39E and characterization of the enzyme purified from *Escherichia coli*. J. Bacteriol. **174:** 5400–5405

66 Hashimoto Y., Yamamoto T., Fujiwara S., Takagi M. and Imanaka T. (2001) Extracellular synthesis, specific recognition, and intracellular degradation of cyclomaltodextrins by the hyperthermophilic archaeon *Thermococcus* sp. strain B1001. J. Bacteriol. **183:** 5050–5057

67 Nelson K. E., Clayton R. A., Gill S. R., Gwinn M. L., Dodson R. J., Haft D. H. et al. (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature **399:** 323–329

68 Cha H. J., Yoon H. G., Kim Y. W., Lee H. S., Kim J. W., Kweon K. S. et al. (1998) Molecular and enzymatic characterization of a maltogenic amylase that hydrolyzes and transglycosylates acarbose. Eur. J. Biochem. **253:** 251–262

69 Cho H. Y., Kim Y. W., Kim T. J., Lee H. S., Kim D. Y., Kim J. W. et al. (2000) Molecular characterization of a dimeric intracellular maltogenic amylase of *Bacillus subtilis* SUH4-2. Biochim. Biophys. Acta **1478:** 333–340

70 Kim T. J., Kim M. J., Kim B. C., Kim J. C., Cheong T. K., Kim J. W. et al. (1999) Modes of action of acarbose hydrolysis and transglycosylation catalyzed by a thermostable maltogenic amylase, the gene for which was cloned from a *Thermus* strain IM6501. Appl. Environ. Microbiol. **65:** 1644–1651

71 Kim H. J., Park J. N., Kim H. O., Shin D. J., Chin J. E., Lee H. B. et al. (2002) Cloning and expression of a *Paenibacillus* sp. neopullulanase gene in *Saccharomyces cerevisiae* producing *Schwanniomyces occidentalis* glucoamylase. J. Microbiol. Biotechnol. **12:** 340–344

72 Igarashi K., Ara K., Saeki K., Ozaki K., Kawai S. and Ito S. (1992) Nucleotide sequence of the gene that encodes a neopullulanase from an alkalophilic *Bacillus*. Biosci. Biotechnol. Biochem. **56:** 514–516

73 Yebra M. J., Blasco A. and Sanz P. (1999) Expression and secretion of *Bacillus polymyxa* neopullulanase in *Saccharomyces cerevisiae*. FEMS Microbiol. Lett. **170:** 41–49

74 Cheong K. A., Kim T. J., Yoon J. W., Park C. S., Lee T. S., Kim Y. B. et al. (2002) Catalytic activities of intracellular dimeric neopullulanase on cyclodextrin, acarbose and maltose. Biotechnol. Appl. Biochem. **35:** 27–34

75 Kuriki T. and Imanaka T. (1989) Nucleotide sequence of the neopullulanase gene from *Bacillus stearothermophilus*. J. Gen. Microbiol. **135:** 1521–1528

76 D'Elia J. N. and Salyers A. A. (1996) Contribution of a neopullulanase, a pullulanase, and an α-glucosidase to growth of *Bacteroides thetaiotaomicron* on starch. J. Bacteriol. **178:** 7173–7179

77 Tonozuka T., Mogi S., Shimura Y., Ibuka A., Sakai H., Matsuzawa H. et al. (1995) Comparison of primary structures and substrate specificities of two pullulan-hydrolyzing α-amylases, TVA I and TVA II, from *Thermoactinomyces vulgaris* R-47. Biochim. Biophys. Acta **1252:** 35–42

78 Tonozuka T., Ohtsuka M., Mogi S., Sakai H., Ohta T. and Sakano Y. (1993) A neopullulanase-type α-amylase gene from *Thermoactinomyces vulgaris* R-47. Biosci. Biotechnol. Biochem. **57:** 395–401

79 Metz R. J., Allen L. N., Cao T. M. and Zeman N. W. (1988) Nucleotide sequence of an amylase gene from *Bacillus megaterium*. Nucleic Acids Res. **16:** 5203

80 Sabathe F., Cornillot E., Croux C. and Soucaille P. (2002) Molecular characterization of amyP, a pSOL1 located gene coding the major α-amylase of *Clostridium acetobutylicum* ATCC824, and its use as a reporter system for strain degeneration. FEMS Microbiol. Lett. **210:** 93–98

81 Horinouchi S., Fukusumi S., Ohshima T. and Beppu T. (1988) Cloning and expression in *Escherichia coli* of two additional amylase genes of a strictly anaerobic thermophile, *Dictyoglomus thermophilum*, and their nucleotide sequences with extremely low guanine-plus-cytosine contents. Eur. J. Biochem. **176:** 243–253

82 Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F. et al. (2001) The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulmonis*. Nucleic Acids Res. **29:** 2145–2153

83 Liebl W., Stemplinger I. and Ruile P. (1997) Properties and gene structure of the *Thermotoga maritima* α-amylase AmyA, a putative lipoprotein of a hyperthermophilic bacterium. J. Bacteriol. **179:** 941–948

84 Abe J., Shibata Y., Fujisue M. and Hizukuri S. (1996) Expression of periplasmic α-amylase of *Xanthomonas campestris* K-11151 in *Escherichia coli* and its action on maltose. Microbiology **142:** 1505–1512

85 Kashiwabara S., Ogawa S., Miyoshi N., Oda M. and Suzuki Y. (1999) Three domains comprised in thermostable molecular weight 54,000 pullulanase of type I from *Bacillus flavocaldarius* KP1228. Biosci. Biotechnol. Biochem. **63:** 1736– 1748

86 Fiedler G., Pajatsch M. and Bock A. (1996) Genetics of a novel starch utilisation pathway present in *Klebsiella oxytoca*. J. Mol. Biol. **256:** 279–291

87 Niehaus F., Peters A., Groudieva T. and Antranikian G. (2000) Cloning, expression and biochemical characterisation of a unique thermostable pullulan-hydrolysing enzyme from the hyperthermophilic archaeon *Thermococcus aggregans*. FEMS Microbiol. Lett. **190:** 223–229

88 Benson D. A., Karsch-Mizrachi I., Lipman D. J., Ostell J., Rapp B. A. and Wheeler D. L. (2002) GenBank. Nucleic Acids Res. **30:** 17–20

89 Schuler G. D., Epstein J. A., Ohkawa H. and Kans J. A. (1996) Entrez: molecular biology database and retrieval system. Methods Enzymol. **266:** 141–162

90 Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhang Z., Miller W. et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402

91 Janeček Š. (1997) α-Amylase family: molecular biology and evolution. Prog. Biophys. Mol. Biol. **67:** 67–97

92 Gerischer U. and Durre P. (1990) Cloning, sequencing, and molecular analysis of the acetoacetate decarboxylase gene region from *Clostridium acetobutylicum*. J. Bacteriol. **172:** 6907–6918

93 Thompson J. D., Higgins D. G. and Gibson T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment trough sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680

94 Saitou N. and Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425

95 Felsenstein J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791

96 Page R. D. (1996) TreeView: an application to display phylogenetic trees on personal computers. Comput. Applic. Biosci. **12:** 357–358

97 Berman H. M., Battistuz T., Bhat T. N., Bluhm W. F., Bourne P. E., Burkhardt K. et al. (2002) The protein data bank. Acta Crystallogr. **D58:** 899–907

98 Watanabe K., Hata Y., Kizaki H., Katsube Y. and Suzuki Y. (1997) The refined crystal structure of *Bacillus cereus* oligo-1,6-glucosidase at 2.0 Å resolution: structural characterization of proline-substitution sites for protein thermostabilization. J. Mol. Biol. **269:** 142–153

99 Kamitori S., Kondo S., Okuyama K., Yokota T., Shimura Y., Tonozuka T. et al. (1999) Crystal structure of *Thermoactinonyces vulgaris* R-47 α-amylase II (TVAII) hydrolyzing cyclodextrins and pullulan at 2.6 Å resolution. J. Mol. Biol. **287:** 907–921

100 Kadziola A., Abe J. I., Svensson B. and Haser R. (1994) Crystal and molecular structure of barley α-amylase. J. Mol. Biol. **239:** 104–121

101 Lee M. H., Kim Y. W., Kim T. J., Park C. S., Kim J. W., Moon T. W. et al. (2002) A novel amylolytic enzyme from *Thermotoga maritima*, resembling cyclodextrinase and α-glucosidase, that liberates glucose from the reducing end of the substrates. Biochem. Biophys. Res. Commun. **295:** 818–825

102 Yebra M. J., Arroyo J., Sanz P. and Prieto J. A. (1997) Characterization of novel neopullulanase from *Bacillus polymyxa*. Appl. Biochem. Biotechnol. **68:** 113–120
103 Imanaka T. and Kuriki T. (1989) Pattern of action of *Bacillus stearothermophilus* neopullulanase on pullulan. J. Bacteriol. **171:** 369–374
104 Smith K. A. and Salyers A. A. (1991) Characterization of a neopullulanase and an α-glucosidase from *Bacteroides thetaiotaomicron* 95-1. J. Bacteriol. **173:** 2962–2968
105 Brumm P. J., Hebeda R. E. and Teague W. M. (1991) Purification and characterization of the commercialized, cloned *Bacillus megaterium* α-amylase. I. Purification and hydrolytic properties. Starch **43:** 315–319
106 Ibuka A., Tonozuka T., Matsuzawa H. and Sakai H. (1998) Conversion of neopullulanase-α-amylase from *Thermoactinomyces vulgaris* R-47 into an amylopullulanase-type enzyme. J. Biochem. **123:** 275–282
107 Kim J. S., Cha S. S., Kim H. J., Kim T. J., Ha N. C., Oh S. T. et al. (1999) Crystal structure of a maltogenic amylase provides insights into a catalytic versatility. J. Biol. Chem. **274:** 26279–26286
108 Skov L. K., Mirza O., Henriksen A., De Montalk G. P., Remaud-Simeon M., Sarcabal P. et al. (2001) Amylosucrase, a glucan-synthesizing enzyme from the α-amylase family. J. Biol. Chem. **276:** 25273–25278
109 Park K. H., Kim T. J., Cheong T. K., Kim J. W., Oh B. H. and Svensson B. (2000) Structure, specificity and function of cyclomaltodextrinase, a multispecific enzyme of the α-amylase family. Biochim. Biophys. Acta **1478:** 165–185
110 Horváthová V., Janeček Š. and Šturdík E. (2001) Amylolytic enzymes: molecular aspects of their properties. Gen. Physiol. Biophys. **20:** 7–32
111 Lee H. S., Kim M. S., Cho H. S., Kim J. I., Kim T. J., Choi J. H. et al. (2002) Cyclomaltodextrinase, neopullulanase, and maltogenic amylase are nearly indistinguishable from each other. J. Biol. Chem. **277:** 21891–21897
112 Suzuki Y., Hatagaki K. and Oda H. (1991) A hyperthermostable pullulanase produced by an extreme thermophile, *Bacillus flavocaldarius* KP 1228, and evidence for the proline theory of increasing protein thermostability. Appl. Microbiol. Biotechnol. **34:** 707–714